The Eastern Magician's Apprentice in Philosophising about the Mind

TAKASHIMA, ELEENA

The Eastern Magician's Apprentice in Philosophising about the Mind

by Eleena Takashima

A thesis submitted to the University of Durham for the degree of DOCTOR OF PHILOSOPHY.

Philosophy Department

University of Durham

April 2025

Abstract

Physicalism, loosely defined, is the view that everything which exists is 'physical'. According to physicalism, consciousness does not exist over and above the physical, and all mental activity reduces to scientific 'physical' concepts (Crane & Mellor 1990). To put it another way, a physicalist attitude rejects supernaturalism, religious/spiritual entities, 'spooky' ethereal concepts, and anything that seems too 'Magical' – in the Harry Potter sense. This concept of Magic, however, can have a dual meaning. There is Harry Potter Magic which Dennett (2003) has referred to as 'Real Magic'; and there is the card trick kind of 'Stage Magic'. Ironically, 'Real Magic' refers to a supernatural concept which is not actually real, whereas 'Stage Magic' is the kind of Magic that is real and uses methods accessible to Muggles (i.e., non-magical folk; Rowling 1997) like us. Philosophical Illusionism is a physicalist view which claims that consciousness is like Stage Magic. Consciousness is an illusion (Dennett 2016; Frankish 2016).

This thesis proposes an anti-physicalist version of a similar idea, which will be called 'Magicalism', the view that the fundamental essence of consciousness is Magic (but in the sense of a card trick, not the Harry Potter sense). This thesis defends a variation of Eastern spiritualism and rejects a physicalist ontology. In the same way that in a card trick, the cards themselves are real but what the magician does with them creates an illusion around it, this thesis will argue that consciousness itself is also real, but it is what philosophers do with the concept which can create a sense of illusion around it. Western academia itself can be like a magic show. In this decolonisation project, Stage Magic is used as a bridge between an Eastern and Western approach to philosophy of mind to try and take the best of both worlds.

Table of Contents

Preface

There was a recent debate which took place between physicist Sean Carroll and Buddhist scholar Alan Wallace, discussing the nature of reality and how consciousness fits into this picture (Carroll & Wallace 2022). As can be observed in this discourse, the 'scientific' perspective and Eastern Buddhist perspective take a very different stance in the way that they deal with the topic of consciousness. As Wallace criticises, scientific discussions tend to claim objectivity and yet rarely take any Eastern accounts into considerations. He argues "The age of colonialism should be over by now" (*ibid*, p.54) and Eastern perspectives are in greater demand if we truly wish to get to the bottom of the issue of consciousness. There is a "blindness", he claims, "to the possibility of profound insights into the fundamental nature of reality from any of the great civilizations of Asia over the past five thousand years. In all those stories, where do we ever hear of any references to discoveries made in the traditional cultures of China, Japan, India, southeast Asia, Tibet, or central Asia?" (*ibid,* p.43). In response to these considerations, this thesis is written in the spirit of decolonisation and takes relevant Eastern views into account.

A Buddhist perspective may connotate something mystical or spiritual, whereas the dominant Western view in philosophy of mind – physicalism – stands in stark opposition. Physicalism is the view that everything which exists is 'physical', or in other words, are the sorts of things that belong in the domain of science. It rejects all senses of supernaturalism or esoterism, including psychic powers, souls, fundamental mentality, and other 'spooky' concepts that have a more ethereal nuance. That said, some physicalists such as Dennett (2016) and Frankish (2016) have argued for a version of physicalism that alludes to the concepts of magic and illusion, albeit in the sense of stage magic rather than supernatural 'Real Magic' that we might observe in fantasy worlds such as a Harry Potter novel. There is an interesting overlap here between the mystical Buddhist perspective which values storytelling to capture the essence of reality, and the rigorous and logical but sometimes poetic illusionist attitude.

This thesis will argue for an anti-physicalist view which attempts to take the best of both worlds. To do so, there are some preliminary matters that need to be acknowledged here before any of our discussions begin. Firstly, this thesis will essentially argue that certain methods of Western philosophy create a blind spot in the way we philosophise about the mind. In the same way that when a magician performs a card trick, the cards themselves are real but what the magician does with them will create an illusion around it, so too is consciousness itself real but what Western philosophers do with it can create a deceptive illusion of physicalism around the topic. It is argued, therefore, that it is not consciousness itself which is illusory, but rather, it is Western academia which is like a magic show, creating a deceptive illusion around the concept of consciousness. To phrase this another way, it is acknowledged that IF we are limited to the typical methods of analytic philosophy, THEN physicalism probably is the best fit answer. A central purpose of this thesis is to overcome such blind spots, to reject physicalism, and find an even better answer.

This thesis, technically and formally speaking, is a philosophy thesis. However, different disciplines are not always cleanly divided in the natural world, and this current piece of work falls in the ambiguous area occupied by philosophy, magic, psychology, and Eastern spirituality. There is a common ground between both magicians and philosophers – where there is an awe about mystery that drives us forwards, even when the mystery is sometimes at odds with our expectations of what a rational, logical, or analytic appraisal of reality looks like.

> "Ultimately, whatever we call ourselves, we are to take on board the one thing the modern rational mind doesn't much know what to do with: *mystery.* […] Once upon a time in the West, mystery found a cosy and well-furnished home in our oppressively exotic image of the East." (Brown 2023, p.476)

Magic is not about deception and fakery, but (especially in Eastern magic), what is most essential is the storytelling (Fujiyama 2010; see also Brown 2023). This thesis has strived to capture as much of this Eastern storytelling methodology as possible while still fitting into the structure as expected and required in a Western PhD. Early chapters are written more in the conventional structure of a PhD, and in later chapters, deviation in methodology may become noticeable. These deviations were methodologically necessary for the purpose of conveying Eastern nuances without compromising important connotations behind the more Buddhist ideas.

The author of this thesis is half-Japanese and half-American, a native speaker of both Japanese and English, and has lived in both the East (Japan) and the West (UK, US, Canada). This 'positionality' of the author is important to acknowledge in this decolonisation project, which not only introduces Eastern thought, but also is grounded in an autoethnographically informed perspective. Furthermore, it will be helpful if readers can acknowledge the author's background in the sciences, having studied pre-medicine (covering the basics of physics, chemistry, biology, physiology, neuroscience, psychiatry, and psychology) in North America before switching to philosophy. The author comes from a scientifically informed background, and the goal of this thesis is to find a better middle-ground between the two extremes of science and spirituality – as opposed to rejecting science and favouring spirituality, or vice versa. It is the hope that this decolonisation project will contribute not only to the knowledgebase for philosophy of mind, but also in spreading a more accurate understanding of an Eastern approach, combining the best of the two worlds in a harmonious and mutually constructive relationship, with Stage Magic as the bridge between the different views of ontological reality.

Acknowledgements

I express my deepest gratitude to my supervisor, Professor Stephen Mumford, for your extensive support, critiques, feedback, and an openness for allowing me to explore my philosophical interests. The thesis has changed, evolved, and developed drastically over the years which I could not have done without your guidance. The PhD journey itself has been like a philosophy apprenticeship, and it has not only helped me to hone my philosophical skillset but has greatly improved and enhanced the heart of the theories themselves as well.

Next, to my partner, Mars, for your endless sympathy and patience of philosophising alongside me. I would never have come anywhere near this far without your unwavering dedication to help me reach for my potential. The faith you have put in me all these years has been deeply, deeply appreciated. I could not have overcome my fears of expressing my theories without your persistent encouragement. I hope to make you proud by flourishing in my career.

I am thankful to my other supervisors, Philip Goff and Benedict Smith for your help as well. Thank you for the inspiration, feedback, and support throughout my years at Durham, beginning with undergraduate and watching me grow up over many years. My sincere thanks to all of the *sensei* at Durham University for all of the opportunities that this higher-education journey has created for me. Finally, I would like to extend my gratitude also to Derren Brown, for the extensive inspiration and guidance that your books and stage shows have given to me.

Chapter 0. Introduction

0.1 Philosophy of Mind: What is the Question?

When philosophers ask what consciousness is, what it is made of, or how our phenomenal experiences fit into a scientific picture, these questions are open to different interpretations. "Consciousness studies are frequently criticized for failing to define precisely what consciousness is." (Frith & Rees 2017, p.12) which leads to ambiguity in such discussions. This can be further exacerbated when key terminology is loaded with nuances that create a gap between a layperson's pre-theoretical impressions, and the way in which the words are used in more technical, academic contexts. It is easy to talk about 'the mind-body problem' or 'the problem of consciousness' as though this refers to a single, unified question, but philosophising about the mind can take on many different forms. It is helpful, therefore, to clarify the question(s) before attempting to search for any answers.

One of the most discussed problems is what Chalmers (2017) has referred to as the hard problem of consciousness, which asks why brain processes give rise to a rich inner life. He contrasts this with the easy (or easier) problem of consciousness, which is instead concerned with how brain processes give rise to cognitive mechanisms. The easy problem is considered to be 'easy' (relatively speaking) because it is a question that is addressable using scientific methods, which are well-suited for addressing questions about analysing structures and mechanisms. The hard problem, however, is not as straightforward of a question. One issue is that it would be question-begging to presuppose that science can give us answers, since that is part of what we are asking when we ask whether consciousness is fundamentally 'physical'. Another issue is that we also cannot presuppose the answer will be about functionalities.

Physicalism, summarised briefly, is the doctrine which claims that everything which exists is fundamentally physical (Stoljar 2024). Dualism, in contrast, is the doctrine that both mental and physical substances exist fundamentally (Robinson 2023). Under a layperson's pre-

theoretical impressions about physical and mental matter, it would be easy to act as though physicalism and dualism exhaust all the options. A naïve way of posing the question would be to ask whether souls (or other spiritual entities, like angels, ghosts, God, etc.) exist. If the answer is yes to any of these seemingly supernatural entities, it would suggest a dualist world. If the answer is no, and only physical matter exists, then this would point towards physicalism.

But again, even 'the hard problem of consciousness' is not a single, unified question where all philosophers are agreed on its interpretation. Firstly, a dualist need not claim that mental substances are supernatural or antiscientific, which makes it somewhat less clear how to interpret the concept of fundamental mentality (Lowe 2008a). But perhaps more pressingly, the key notion of 'physicality' is also ambiguous. Hempel's Dilemma, for instance, shows that while physicalism seems to indicate something to do with the physical sciences, it is unclear whether this refers to current physics, with all its flaws and incompleteness, or to future completed physics, which we currently know nothing about (Hempel 1980; 1966). If the former, then a physicalism that is loyal to an incomplete and still-flawed science must surely be an incorrect doctrine. Maybe electrons will turn out to be just as make-believe as ghosts. If the latter, then this leads to a vacuous version of physicalism that tells us nothing about what this doctrine entails. Maybe future parapsychologists will find undeniable evidence for the existence of ghosts (Would this mean that a non-physical entity was finally discovered, or would it mean that ghosts were 'physical' after all?) These arguments have led some philosophers to conclude that physicalism is either clearly wrong or completely vacuous, and thus is not a doctrine which can be taken seriously in the first place (Crane & Mellor 1990).

The ambiguity of the question (if there is any sense of 'THE' question to be salvaged) does not end with Hempel's Dilemma. Philosophers also disagree upon how other key terminology is to be defined, such as the words 'phenomenal consciousness' or 'qualia' (Crane 2012). The general topic of discussion seems to begin with the idea that there are certain *what*

it's like to X type of experiences. There is a certain *redness* kind of experience when we see the colour red, which we can imagine a person who was born colour-blind might simply have no impression about. We know *what it is like* to be in pain, but we do not really know *what it is like* to be a bat (Nagel 1980). Nonetheless, there is a certain phenomenal experience – the *what it's like* experience – which appears to have some qualitative *feels* to it, which we might refer to as 'qualia' (though again, Crane 2012 explains the various ways this term has been used, and we should exercise caution as different people may use this word differently).

There is a sense in which these *feels* seem to not fit in very neatly with the ontological picture that the sciences seem to paint. Very loosely explained, physics can tell us about the complex building blocks of matter, the mathematical or functional relation between different forces or fields, and it tells us how different macro-entities emerge from the micro-structures underlying them. The fact that salt dissolves can be explained by the chemical bonds of its building blocks of how sodium and chlorine molecules bind together (Hendry 2019), and this requires no 'feely' explanations about how the sodium and chlorine molecules feel about the situation. Similarly, the way in which objects move might be describable by analysing the forces which are involved, and it is not as though there is a 'soul' that possesses a toy car that slides down a ramp during a physics demonstration as gravity takes hold. The toy car does not 'decide' to go down the ramp no more than it experiences *what it's like* to be a toy car. There appears to be *something* that is categorically different about a living being's consciousness.

Chalmers (2003) has set up a classification system which broadly distinguishes the types of responses that people can have to bridging this gap between the 'feely' and the science. Very loosely, Type A-C Physicalism refers to philosophers who believe that consciousness belongs fully to the domain of the physical sciences in one way or another. Type D-E Dualists believe that there is a separate kind of 'mental' substance that is distinct from the type of substances

which science tells us about. Finally Type F Monists believe that only a single type of substance exists, but it is a more neutral substance rather than being physical or mental.

However, in the absence of a proper definition of 'physicality', even this classification system is difficult to make sense of. If future parapsychologists discover that ghosts do indeed haunt graveyards at night which causes trees to sway without any wind or other physical influences, does this entail that Dualism is true (because ghosts are real!) or does it entail that Physicalism is true (because ghosts are physically moving the trees, so the ghosts themselves must also be physical!) Also consider an original Buddhist Reincarnation thought experiment. Imagine that future science discovers a new kind of gaseous substance which is emitted at death. This gas, sometimes travelling across the globe, will eventually find a pregnant woman's womb to enter, which in turn alters the brain chemistry of the foetus to resemble that of the recently deceased person where the gas came from. This appears to be empirical evidence to support Buddhist reincarnation, thus raising the question as to whether such gaseous 'souls' ought to be considered 'physical' substances or not. If it truly is just a gas and nothing more, then one may argue that the substance is very obviously 'physical'. But then – if proving the existence of Buddhist soul reincarnation is *still* not enough to disprove physicalism – what would it take? Is physicalism unfalsifiable even in theory? Is physicalism nothing more than just a tautology?

These are the types of worries which will be investigated more deeply in chapters one to three. We investigate the concern that physicalism may be at worst vacuous, and at best a tautology which is necessarily true but only for logical and semantic reasons, thereby making it unfalsifiable (but only because of the sneaky way that definitions have been set up). For now, what is important is to see that the word 'physical' contains enough ambiguity to render 'the question' also ambiguous, making it difficult to even get the discussion off the ground in the first place. Our task here would be simple and straightforward if only we could ask the very simple question of "Is consciousness reducible to the physical?" Unfortunately, even this small

initial step is made difficult unless we can be more precise about what is meant by 'physicality', and we still find ourselves at the stage of trying to define a question. It also does not help our situation that 'feely' consciousness (qualia?) also seems to be rather ill-defined.

The concept of physicality will be investigated properly in chapters one to three, but because this requires a great deal of argumentation, we will need to get the discussion started somewhere, even in the absence of clearcut or unified definitions of important key terminology. At this point, it may be helpful to take a step back and return to the pre-theoretical starting point taken by a layperson. While it may be tempting to assume that naïve, pre-theoretical intuitions discussed by laypersons are merely less-refined versions of the rigorous, technical discussions in the academic literature, this assumption should be approached with caution. Consider the following examples where layperson and technical concerns might diverge:

(A) A person who is grieving the loss of a loved one and wonders whether it is conceivable to be reunited with their loved one in the afterlife may be asking a valid question, and yet, telling them that philosophical zombies are inconceivable seems to miss the point of this discussion.

(B) A person who reads in philosophy books that consciousness doesn't exist might misinterpret illusionist philosophers and wrongly infer that if people aren't conscious anyway, we might as well go around hurting people. This would be misinterpreting the technical discussions, yet there is a separate interesting discussion here about ethical implications that might arise if consciousness was just as 'illusory' as ghosts.

(C) A believer of psychic powers goes to a faith healer to cure their cancer, and experiences temporary improvement. However, they later discover it was just an illusion. How (if at all) does this link to technical discussion of physicalists claiming consciousness itself is an illusion?

In situation (A), the layperson is grappling with an emotional 'story' about how to process grief, and the issue in question here is not a technical one. While a technical argument that supports (or disproves) the existence of souls that persist to the afterlife may have relevance for how this person tells a story to themselves, the technical discussion and the emotional journey of overcoming grief are two separate matters. In situation (B), the layperson is contemplating an ethical 'story' about how to treat other people. Again, while a technical argument which (this person interprets) trivialises phenomenal experiences might lead the individual to start treating their friends more poorly and the two topics could be interlinked, ultimately, the moral attitudes of an individual in their personal life is a separate concern from the technical ontological discussions. Finally, in situation (C), the layperson is trying to figure out what is or is not a 'scam' in the sense of being deceived out of their money, or even made to do something which damages their health. While a technical argument in support of dualism might incline a person to take 'mental causation' of faith healing powers more seriously, the 'illusory' nature of a scam may contain a different connotation from the 'illusory' nature of consciousness as discussed in the technical physicalist literature (e.g., Frankish 2016).

The layperson's interests, as we can see from the above examples, may pertain to the relevance of philosophically related topics in our personal lives, where the debates will have some sense of personal impact. (A) How do I process grief, or a fear of my own mortality? (B) How do I treat other people, and what ethical treatments are justified on the basis that my friends are conscious, experiencing beings? (C) How do I distinguish between something which is fake news, a scam, or an illusion, so that I can protect my finances and my health?

The technical discussions, in contrast, provide a more detached platform for discussions to take place in a way that does not become biased by these personal concerns. A person may wish to argue for physicalism despite being a vegan in their personal life, believing in the importance of animal rights in virtue of animals being consciousness, while simultaneously

arguing that ‘qualia’ or ‘phenomenal experiences’ are pseudo-concepts for ontological discussion. This does not suggest hypocrisy, again, because academic philosophy provides a platform where technical, narrow-scoped arguments can be separated from personal messages.

That being said, we can still acknowledge the duality between ‘storytelling’ of the layperson in their personal lives (or even of the professional philosopher in their own personal life) and the ‘technical argumentation’ that is often a central focus in analytic philosophical discussion that are more readily accepted for publication. This duality in some ways mimic the seeming duality between the ‘mental’ and ‘physical’ dichotomy investigated in philosophy of mind, and therefore, it will be important to tease apart these distinct layers. There is a sense in which the ‘stories’ that we tell ourselves have a fundamentally mental quality. It is a cognitive activity we engage in which can affect our emotions, our motivations, and our outlooks on life. In contrast, the ‘technical arguments’ we engage in do have a more fundamentally ‘scientific’ kind of rigorous quality. It strives for objectivity and to be unbiased where possible, so that the logical, mathematical, or functionalist structures of the sciences can sit in harmony with it.

Depending on the story we tell ourselves about what consciousness *is*, this may shape our image of what it means to be a human being – a thinking, feeling creature with a rich inner life and the capacity to feel both love and pain. If we were to focus exclusively on the technical argument, then the significance that the ‘feely’ nature has on our lives may become lost. For example, let us say that physicalism ends up winning the technical debate once and for all, and future philosophers can agree that the ‘physical’ stuff as described by the physical sciences is all that exists. Consciousness is no different from tables, chairs, clocks, calculators, and other machinery. In which case, this would raise two further questions: (1) Does this suggest that there is no distinction between a ‘Tamagotchi’ (i.e., a pixilated ‘pet’ contained in an egg-shaped videogame popularised in the 1990’s) and a real pet cat? (2) Is the philosopher just a middleman who merely tells us who it is we should be listening to (i.e., scientists) but nothing of content?

O'Rourke (1998), for instance, has written about the anxiety sparked by such virtual pets in the early days of machine-related sources of affection. There is a sense of rapport and empathy that we can establish with beings that we take to be phenomenally conscious. Yet, if it turns out that living conscious beings are no different than pixel art carried inside a beeping toy egg, then this might make us wonder whether there is still something about the 'specialness' of consciousness that we neglected to explain. Surely there is something about the 'feely' nature of living things that justifies a deeper explanation than just opening a physics book?

We can take one of two responses to this question. Firstly, we could reply that metaphysics is concerned not about human nature and other social concepts, but it is concerned with how philosophy relates to physics. In which case, the philosopher could argue that being over-attached to a pixel entity is concerning, maybe from a psychological perspective, but this is a separate issue from metaphysics. The *metaphysical* discussion, this opponent would say, has been concluded once future philosophers have agreed that the stuff of science is all that exists. This is technically a valid response, but it renders the metaphysician a mere 'middleman', making it slightly confusing what role (if any) they played in this discussion. Could we not simply do away with the philosopher, and just listen to the physicists and social psychologists?

The other response is to accept that philosophy of mind – the 'love of wisdom' of mind – will inevitably have overlaps with things that relate to human matters, such as emotions, passions, motivations, affection, love, and so on. To truly explain the fundamental essence of consciousness would require not only explaining the relation between mind and physics, but also begs for something to be said on 'feely' elements that make us conscious human beings. In this case, 'the question' will have to be multifaceted. It is not limited to the hard problem, but will also need to ask a wider question: *What is the fundamental essence of consciousness?*

In this thesis, the first three chapters will focus on the conventional types of discussion had in the consciousness literature, including topics such as Hempel's Dilemma, philosophical

zombies, and the relation between the mind and the physical sciences. Then, chapter four will advocate a novel reply to the mind-body problem called Magicalism:

> **Magicalism** = the neutral monist theory which claims that the fundamental essence of consciousness is Magic (in the sense of a card trick, not the Harry Potter sense).

This, of course, requires us to investigate the next question about what it is that we mean by Magic in the card trick or stage show sense. Stage magic is not as simple as just deceiving the audience using sleight of hand, and a more in-depth explanation is needed on what Magic *is.*

0.2 What is Magic?

Daniel Dennett, a physicalist who argues for the eliminativist view that consciousness does not really exist (Dennett 1991), has written a couple papers on the relation between magical thinking and consciousness (2016; 2003). However, not only does his usage of the word 'physicality' suffer from the types of ambiguities highlighted in this thesis, the way he uses the word magic is also pre-theoretical, such as his saying that the "magic" of consciousness refers to "the supposed unexplainability" (2003, p.7). While it is often the case that magical stage shows tend to involve performances that are seemingly inexplicable, this is not always the case in a magic show, nor is it a necessary requirement. In Japanese traditional magic known as 'tezuma', for example, magic shows centre around poetic storytelling and may contain very little deception. The Butterfly Act performed by Fujiyama Shinatro and his apprentices is an example of an entirely non-deceptive (yet still Magical) performance. The act uses scrap pieces of paper to create an illusion of butterflies dancing on stage, but the performance is not designed to make the 'illusion' misleading in any sense. It is obviously clear to the audience that the 'butterflies' are just pieces of paper, but the performance relies on the audience's willingness

to imagine that the paper comes to life, thereby completing the 'illusion'. From there, the Magic of the performance is in the story being told by the butterflies: A story of the beauty in impermanence of life, as the butterflies find a mate, have their offsprings, and live out their short but meaningful lives in virtue of leaving behind their mortal legacy. There is nothing fake or deceptive about it, and the act it alludes to the idea that *life itself* is Magical (Fujiyama 2013).

If a philosopher is going to make any claims about consciousness being analogous to illusions or anything pertaining to stage magic, it is helpful if we also have an understanding of what stage magic is, and how to distinguish between deception and wonder found in magical illusions. While Magicalism advocated in this thesis is not a form of physicalism (largely due to the rejection of the word 'physicality'), it may still be understood as a close cousin to philosophical illusionism in that it makes claims about consciousness being like a magical illusion. However, Illusionism and Magicalism use the term 'illusion' in different ways. Dennett, for instance, while analysing Frankish's illusionism, claimed:

> "An illusionist is an expert in sleight-of-hand and the other devious methods of stage magic. We philosophical illusionists are also illusionists in the everyday sense — or should be. That is, our burden is to figure out and explain how the 'magic' is done." (Dennett 2016, p.65)

After stating this, he goes on to quote Frankish:

> "Illusionism replaces the hard problem with the illusion problem — the problem of explaining how the illusion of phenomenality arises and why it is so powerful. This problem is not easy but not impossibly hard either. The method is to form hypotheses about the underlying cognitive mechanisms and their bases in neurophysiology and neuroanatomy, drawing on evidence from across the cognitive sciences." (Frankish 2016, p. 37)

What can be observed in both quotes is that the concept of illusion is heavily emphasising the idea that it is the *method* which is of the essence in a magical performance. Among the literature written by magicians, however, we find discussions claiming that the effect, showmanship, and storytelling in a magic trick are just as essential to it (if not more important) than the method or mechanism used to achieve that effect (see Brown 2023). The *method*, in this case, is understood as the implementation of the sleight of hand, and the *effect* refers to the impact on the spectator that creates a sensation of Magicalness (Lamont & Wiseman 2005, p.1). When this dichotomy is defined as such, we can begin to tease the ideas apart. An amateur magician, we can imagine, might become proficient at figuring out, explaining, analysing, and performing the sleight of hand, yet their performance may lack the Magic if they neglect building rapport with the audience, and merely do tricks 'at' them rather than 'with' them (Brown 2023; 1999).

An analogy exists here between the magic show and the problem of consciousness, as Frankish's above quote indicates. We can imagine that an apprentice approaches a master magician and asks them what Magic really *is*. The apprentice might try to insist that there must be an easy problem of Magic (i.e., how the methods or mechanisms work) and the hard problem of Magic (i.e., what the Magic *is*, or why Magic could possibly exist in our otherwise 'Muggle' world). The Frankish-equivalent of a magician could reply to this that the two questions are basically, one in the same. There is no hard problem of magic, strictly speaking, because nothing non-Muggle is really happening. However, it is still possible to explain how it is that the powerful illusion arises. Perhaps we could even say that there is an easy problem of Magic which pertains strictly to sleight of hand, and a harder problem of Magic that is about why human beings experience illusion only under certain conditions of good showmanship. However, again, strictly speaking, the 'hard problem' of Magic does not truly exist. There is a sense in which showmanship, rapport-building, and storytelling are still part of the 'method'.

This reply, however, would still be incomplete. A philosophical illusionist – who is a reductive physicalist – would go a step further and say that showmanship and rapport-building are all reducible to neurophysiology and neuroanatomy. The magic show fundamentally reduces to the atoms and molecules that compose the playing cards, the neural responses in the audience's brains, and the vocal, neural, and muscle activity of the magician as they talk to the audience and move the cards around in a certain manner. A magic show, therefore, is just as 'physical' as any other show, including a Shakespeare play, or a psychic faith healing show.

At this point, the problem of pseudo-psychics will start to become an issue. The term 'pseudo-psychics' refers to people who claim to be using supernatural powers even though in actuality, they are using the same methods used by other magicians (Lamont & Wiseman 2005, p.102). Faith healers, in particular, refer to those who claim that physical ailments can be cured by using psychical energies – which appears to have some heavy dualist connotations if we wish to draw an analogy with the consciousness debate. For our current purposes, we will use the term 'pseudo-psychic' strictly in the sense of referring to people who advertise their shows not as a magic show, but as a demonstration of some kind of psychic or non-physical power. These shows are often dishonest or fraudulent, they tend to be more expensive (since they are advertised as genuine demonstrations of the supernatural), and audience members will be composed primarily of 'believers' of psychic powers. In contrast, a magic show is advertised as a form of entertainment, and even if patter involves claims about psychic abilities, audiences tend to expect witnessing illusions rather than supernaturalism (*ibid,* Ch4).

What is noteworthy for our purposes is to acknowledge that there is a difference between an apprentice asking a magician "Where does the Magic come from?" and an audience member at the pseudo-psychic's performance who asks the same question. The former is asking a slightly more metaphorical question, and the latter is asking a supernatural question. In this context, the overly reductive reply that it is all just neurophysiology adds no information to the

former's question ("I *know* it's an illusion, but what is it that makes illusions feel special?") but does add new information to the latter's question ("Wait, it's not supernatural after all?").

The magician's apprentice already knows that one does not need to have attended Hogwarts to become a professional magician, and that mastering Muggle-skills are enough to become proficient in the craft. Therefore, telling the magician's apprentice that "Hey, it's just stage magic!" (Dennett 2016, p.66) does not add any new information. Instead, what the apprentice here is asking relates more to matters of human nature. Where does that Magical feeling come from? How do certain performers manage to create that wonderous sensation, whereas other amateur performers who are proficient in sleight of hand still 'ruin' the Magic?

That being said, while Muggle-skills may be enough to successfully vanish a card, there is still something to be said about honing the Magic of being a human being – a performer – while on stage. A robotic and dull performer would likely receive a poor reception, whereas a relatable or charismatic figure might be able to create a greater sense of wonder. Again, as we saw in the Japanese butterfly act, playing up the idea of the Magic that is inherent in life itself can be brought onto the stage to create a sense of Real Magic. Not supernatural, yet still Real.

The audience members at the pseudo-psychic's performance, in contrast, are not asking a metaphorical question about the wonders and Magic of life, but instead, are asking more of a causal type of functional question. It is a *how* question about what makes the trick work. Is it some kind of mind-power, like a wizard literally moving objects with his mind, or is it more like a stage magician who sneakily hides an object under their sleeve? This functional question is more similar to the hard problem of consciousness as discussed in the academic literature. It asks a structural question about what substances exist in the world, and the causal impact of mental powers if they turn out to exist. In this context, telling the audience that "Hey, it's just stage magic!" may be a shocking reveal for the audience who believed they had been witnessing something Harry Potter-esque, and the topic of the Magic of Life is left untouched.

Oddly enough, in the Harry Potter universe where a dualist ontology is quite explicitly accepted, the philosophical problem of consciousness might still prevail anyway. In the Harry Potter universe, ideas such as souls, ghosts, afterlife, and spellcasting as a form of mental causation are readily accepted (Rowling 2007; 1999). Yet, if we lived in the Harry Potter universe, would we be any closer to solving the question about *what* consciousness *is*? Yes, we would know for a fact that 'magical' concepts exist independently from the Muggle world which physics and technology are unable to explain. But does that mean that dualism is true?

One option would be to develop Magic as a science in its own right, and begin analysing it in terms of substances, mechanisms, and causal powers (see, for example, 'The Science of Harry Potter: How Magic Really Works' by Highfield 2003). Let us refer to this as Hogwarts Physics (call it h-physics), which focuses on quantifying and reducing the magical types of entities to causal mechanisms that are inaccessible to Muggle physicists. However, this only delays the issue and creates a new kind of physicalism (h-physicalism). Atoms, molecules, fields, forces, spellpower, souls, ghosts, etc., are all entities discussed by h-physicists. But how does consciousness fit in? Is it reducible to a 'soul' (an h-physical entity), or does consciousness exist over and above the h-physical? If Voldemort splits his soul into seven horcruxes, does each horcrux have an independent what-it's-like experience, or are they more like philosophical zombies containing a fragment of Voldemort's functionalities? What about the paintings that move and speak. Do they feel pain if attacked, or do they just say they do? The h-physicalists could still ask if it is conceivable to have soulless h-zombies. But how could a mere structural object like a 'soul' ever give rise to something 'feely' and experiential?

Despite proving the existence of the supernatural, we would still be back to where we started. The philosophical problem of what consciousness *is* does not seem to immediately go away, and the reason for this, once again, is that the type of questions some laypersons wish to ask is different from the type of questions that technical discussions attempt to answer. If a

Tamagotchi pixelated pet starves due to neglect, then does this cause just as much suffering as if a living cat starved while the family was away on holiday? What if we simply drew a cartoon of a starving pet – does it 'feel' any sensation of suffering? Common sense would tell us that these are silly questions to ask. *Of course* only conscious entities feel things like suffering. If we become too attached to a pixelated or cartoon pet and succumb to the idea that we are causing someone harm, this would merely be an illusion. No 'feely' suffering exists.

We could analyse to infinity the mechanisms of Voldemort splitting his soul in seven pieces, how a Horcrux (an inanimate object containing a part of a soul) relates to Muggle physics, and what Descartes may have said on Hogwartsian Interactionism or on the divisibility of a soul, but these would all be separate considerations from addressing the significance of the 'feely'. If our partner professes love to us, this *means* something because we know that they are conscious, and are, indeed, feeling something warm and fuzzy towards us. It would be a mistake to tell them that their declaration of love is 'nothing but an illusion' because 'feely' things just are neurophysiological activity that are fundamentally made of indifferent particles.

What we wish to do now is to start developing a theory about the 'feely' in a way that makes it more refined and thorough than pre-theoretical common sense. Perhaps common-sense intuitions already tell us that our partner's love or a cat's suffering are very much real (i.e., non-illusory) in contrast to how a faith healing ritual's power to cure cancer is *not* real. Having a technical discussion on how to capture the essence of the 'feely' can be a useful addition to our academic knowledge base. This is where Magicalism comes in.

Magicalism can be understood as a close cousin to Illusionism as both create a link between consciousness and stage magic. However, the two doctrines understand the concept of an illusion in slightly different ways. Notice that 'illusion' can have two separate meanings:

Illusion$_{(A)}$ = A surprising or seemingly impossible story that brings a feeling of wonder.

Illusion$_{(B)}$ = A deception or a falsity. Based in misinformation, or something that is not real.

Magicalism emphasises the similarity between consciousness and stage magic in the sense of Illusion$_{(A)}$ where the purpose of discussion is focused upon explaining the wonderous effect of the 'feely' stuff; whereas Illusionism tends to emphasise the similarity between consciousness and stage magic in the sense of Illusion$_{(B)}$ where a purpose is to debunk certain ways of talking about the 'feely'. As in the case of Japanese stage magic, we saw that stage magic is not strictly speaking required to include elements of deception. To say the least, for an Eastern magician, it is Illusion$_{(A)}$ that plays the important role in stage magic performances (Fujiyama 2013).

In this light, when the Eastern magician's apprentice approaches the master magician and asks, "What is Magic?", it would clearly be a mistake to give an overly reductive response about neurophysiological mechanisms. Instead, what is being asked is not a structural, or even a methodological question, but it is asking about the human 'feely' experience. What is this surprising thing that means our partner's love is Real, but is at the same time, Magical?

The response to such a question would have to be given, not in structural language (e.g., The soul is a divisible entity that is distinct from Muggle physics, etc.) but in the form of *storytelling*. The type of psychic powers professed by the pseudo-psychic are mere Illusions$_{(B)}$ in the sense that the faith healing performance is not what it claims to be. The cancer is not actually healed due to psychical abilities, and if the recipient temporarily experiences a sense of healing, this is an Illusions$_{(B)}$, or a deception. In contrast, if the apprentice asks the magician if magic is real, the magician can distinguish between Illusions$_{(A)}$ and Illusions$_{(B)}$. That *feeling* of wonder (or love or pain or suffering) is Real, even when the *method* might involve deception.

The Magic is not just in the *method* but is in the *effect*. The method can be Illusory$_{(B)}$ but the experience itself is very much real, just as much as the things of physics. The difficulty

with the conventional technical discussions in philosophy of mind is that there is not much room (or interest) for storytelling about human nature within the field of physics. Therefore, trying to answer questions about consciousness and its relation to physics will inevitably leave out the discussions about how to tell the right stories about the 'feely' in a way that makes it more rigorous than common sense, yet does not limit itself to the language of mere mechanisms.

To reiterate, Magicalism in this thesis will be defined as follows:

> **Magicalism** = the neutral monist theory which claims that the fundamental essence of consciousness is Magic (in the sense of a card trick, not the Harry Potter sense).

Furthermore, 'Magic' can be understood in the following way:

> **Magic** = the way that a (philosophical) story makes us *feel* (the *effect* of wonder).

Effectively, the argument will be that to understand the fundamental essence of consciousness, the answer cannot be given in a solely structural or functionalist sort of way, and storytelling will be an essential feature to fully address questions about the 'feely'. Equally, we cannot merely explain away or reject the 'feely' without sounding a little sociopathic – such as if we were to trivialise the love of a partner, or if we were to become indifferent to causing suffering to other conscious beings because we believe it is 'just an $\text{Illusion}_{(B)}$'. There is something here still to be explained on the 'feely' which deserves a more technical analysis than common sense, and yet, it cannot be captured exclusively under a structural or functional argument. Thus, the hard problem of consciousness is replaced not with the problem of illusions, but with the problem of *aversions* for analytic philosophers engaging with matters of the passions.

0.3 The Eastern Apprentice

Some philosophical topics (e.g., existentialism) can require engagement with passions at least to some extent, to capture the essence of the discussion. For instance, Benatar argues:

> "Continental philosophers might agree that existential questions cannot be answered by analysis and technical argument. On this view, momentous questions of ultimate significance cannot be addressed satisfactorily in the coolly dispassionate manner of analytic philosophy. Instead, such questions must be engaged passionately, stirring the heart rather than sharpening the mind." (Benatar 2010, p.2)

There is great overlap between Japanese philosophy and continental philosophy, (Heidegger in particular, see Ishihara 2019; Krummel 2018; Rigsby 2010) where existential problems are taken very seriously. Although continental philosophy exceeds the scope of this current thesis, we will discuss some important parts of Japanese thinking on issues in philosophy of mind.

More specifically, in the Japanese language, the word 'kokoro' is used to mean both 'mind' and 'heart', and this linguistic point is reflected in the way in which philosophy of mind is analysed by Japanese philosophers (Kasulis, Yasuo & Yuasa 1987). When the word for 'mind' is identical to the word for 'heart', then philosophising about the mind is not completed without also philosophising matters of the heart and passions as well.

> "Passions and clarity need not be antagonistic. Although the passions often cloud the mind, questions about which one cares deeply can be ones about which one thinks clearly. […] Indeed, these questions are best answered clearly and rigorously." (Benatar 2010, p.2)

Benatar further discusses the "very rigorous methodology characteristic of the philosophy that happens to be done in English-speaking countries. Indeed, philosophers of this kind are often

highly critical of the lack of clarity, precision, coherence, and logical structure of much existential and other continental philosophy. Much existential thought is more like evocative literature than like technical philosophy. […] if generalizations are to be made, English-speaking philosophers, sometimes called "analytic" philosophers, have tended to be more concerned with precision and clarity (including drawing distinctions and avoiding ambiguities) and technical rigor (such as logical structure) than their European counterparts." (*ibid,* p.1-2).

In so far as we are addressing the hard problem of consciousness, especially if one is arguing in favour of physicalism, then it is possible (maybe even preferable) to write in a technical and logical structure that almost resembles language used in scientific journals. Even if it is not quite as technical as the writing style of physics journals, evocative language can risk distracting from the technical and logical arguments. In contrast, when it is our goal to analyse the 'feely' nature of consciousness in a way that is distinct from structural or functional arguments, then focusing exclusively on logical structure may start to become an impediment in its own right. Avoiding ambiguities in definitions or drawing distinctions between different categories is useful up to a point, but falls short for expressing the attributes of 'feely' human nature that is more rigorous than common sense, yet not quite as 'dry' as in the hard sciences.

Robert Nozick (1981), when discussing the topic on 'meaning of life', expresses this conflict of interest using a joke about a person who seeks philosophical wisdom from an Eastern holy man living in a cave. The story begins with a person who travels to the Himalayas, and upon encountering the wise old man living in an isolated cave, asks him what the meaning of life is. The wise man returns a seemingly nonsensical answer ("Life is a fountain") and when he is pressed for clarity, the supposed 'sage' turns out not to have any analytic answer to give. The joke is that the questioner had hoped for an intellectual formula or method to achieve a life with meaning, and yet is unhappy with any analytic answer that could have been provided.

> "What did he expect to hear from this meditating man in a cave high in the mountains? "Go back to the posh suburb and continue your present life, but shift to a less pressured job and be more accessible to your children?" […] Could *any* formula answer the question satisfactorily?" (Nozick 1981, p. 572)

The questioner in this scenario has turned to the Eastern wise man precisely in order to find something that may have escaped the notice of his analytic mindset, and yet, when presented with a variety of possible answers, he is quick to reject them as it does not satisfy his own expectations. In a way, this questioner is like a magician's apprentice, trying to figure out the *method* for achieving the *effect* of making life feel more magical (fulfilling, wonderous, etc.).

However, the Eastern pursuit of wisdom does not work this way. When the question itself is a fundamentally spiritual one (in the sense of 'being about the spirit/passions' rather than 'being about substance dualism'), then the solution too will require something more nuanced. This thesis, as the title would suggest, is written from the perspective of an Eastern apprentice who is philosophising about the mind. Unlike in Nozick's caricatured version, however, the Eastern apprentice does not presume that there is a hidden 'sage' somewhere out there who has all the answers. Rather than being about seeking 'THE' factual answer, instead, this can be understood as a journey to overcome the 'problem of aversions', when a philosophical topic is not easily reducible to something that is technical, dry, detached, or passionless.

Furthermore, before proceeding, it should be noted that in this thesis, the discussion on 'Eastern Philosophy' will be primarily focused on Japanese Buddhism supplemented by Japanese Shinto nuances, and some claims made in this thesis may contradict other forms of 'Eastern Philosophy' (e.g., Hindu traditions, etc.). This is done for a few reasons. Firstly, because 'Eastern Philosophy' is itself covers such an enormously large area, it is impossible to focus on the entirety of Eastern traditions all at once, particularly when this thesis already covers many other large topics such as philosophy of mind and magic, and therefore, some sort

of narrower focus needs to be defined. Secondly, because this thesis is mainly interested in the overlap between Eastern traditions and the claims about the storytelling of stage magic, it makes sense to focus on Japanese philosophy which also places a very strong emphasis on storytelling. Once again, it should be noted that this thesis strives to introduce some Eastern attitudes to philosophy of mind in an attempt at decolonisation, and further expanding the contents of this thesis to other branches of Eastern philosophy should be partaken in future projects, as it unfortunately exceeds the scope of this current thesis.

0.4 Aims, Methods, and Conclusions of this Project

In broad terms, the primary aim of this thesis can be understood as a defence of Magicalism. The argument requires three separate subgoals that are addressed separately:

1. Rejecting Physicalism [Ch1-3]
2. Advocating Magicalism [Ch4] and defending why it does not turn into a supernatural or 'unscientific' kind of belief system [Ch5-7]
3. Advocating a Magician's Research Programme to which Magicalism is a part of [Ch8]

Each of these tasks can be broken down further into the following methods, each bullet point roughly (though not exactly) corresponding to each chapter of this thesis:

1. Structural Concerns [Chapters 1-3]
    a. Discuss the ambiguity in the word 'physicality' and show why it is not definable in a way that allows for the construction of a coherent version of physicalism.

b. With various interpretations of physicalism rejected, the remaining possible interpretation of 'physicality' is to see it as an 'anti-magic' type of attitude (i.e., a scepticism of seemingly 'magical' explanations of ontological phenomenon).
c. Misconceptions in physics are addressed where they may be causing philosophical confusion.

2. Magic and Pseudo-Theories [Chapters 4-7]
    a. Magicalism is explained as a conjunction of an (i) Eastern (ii) Magician's (iii) Apprentice's perspective, where each attribute is described in turn.
    b. Problematic features of 'unscientific' belief systems are discussed, including astrology and psychic faith healing. This involves discussing three kinds of problems with these belief systems (i) The Empirical Problem, (ii) The Theoretical Problem, and (iii) The Pragmatic Problem.
    c. Similar problems are observed in the scientific field of medical psychiatry, showing that appealing to 'science' does not always address these issues.
    d. Finally, a logical argument is given to summarise the common error in deductive arguments, where analytic premises can mistakenly be accepted as empirically justified premises. A deductive conclusion is only as sound as the premises which it accepts, and a fundamental flaw with analytic philosophy is explained.

3. The Magician's Research Programme [Chapter 8]
    a. Earlier in chapter five, the concept of a Lakatosian research programme (1978) is introduced as a set of hypotheses, methods, and definitional commitments that lead to a 'progressive' framework that contribute novel predictions, or a 'degenerative' framework that only make *ad hoc* claims to save its own system.

b. A novel research programme based in Magic is advocated as a way of addressing consciousness studies in the future, and Magicalism is understood as the ontological reply that exists within the overarching research programme.

0.5 Chapter-by-Chapter Summary

Chapter 1 will begin with proposing a novel Xombie thought experiment to replace the classic zombie thought experiment (Kirk 1999). Due to the ambiguity in the word 'physical', the Xombie thought experiment is constructed to avoid the usage of this word while still preserving the intuitions behind the original thought experiment. To do this, Xombies are defined as 'third-person identical' but 'first-person distinct' creatures, and we explore how this novel version links with the original versions, as well as discussing its further implications towards the mind-spirit problem explored in the second half of this thesis.

Chapter 2 then works on classifying different versions of physicalism into two broad categories: Structural and Scientific Physicalism. Again, by exploiting the ambiguity in the word 'physicality', we begin to see that physicalism can either be defined in a structural sense which tells us *a priori* what properties or qualities a physical entity has or lacks (e.g. the No Fundamental Mentality Principle), or in the scientific sense of following whatever scientists tell us. This chapter argues that the two are not always compatible, and that physicalists must pick a camp. Both versions of physicalism each struggle, though for different reasons.

Chapter 3 builds on the previous argument and discusses misconceptions about the field of physics, particularly in relation to field theory which does not reduce to an atomist model. After seeing that turning to physics is still not particularly helpful for defining 'physical' in a way that is useful for *meta*-physical discussions, we argue against the possibility of Hybrid Physicalism – which a conjunction of Structural and Scientific Physicalism that tries to have its cake and eat it too. This is untenable as 'THE' question is just too broad and also too vague.

Chapter 4 then discusses Magicalism as an answer to the question, "What is the fundamental essence of consciousness?" The chapter begins by advocating neutral monism as a default, and then proceeds to discussing the characteristic features of taking an (a) Eastern (b) Magician's (c) Apprentice Approach to philosophising about the mind. The chapter illustrates the importance of three features: (i) a sceptical attitude as the default, (ii) a need for empirical 'contact points' with reality when making ontological claims, and (iii) the importance of poetic storytelling in order to satisfy our philosophical curiosities in a *meta*-physical debate, where without the storytelling, consciousness is explained away just like a 'ruined' magic trick where the magician gives away the spoilers about how a card trick is performed.

Chapter 5 looks into some 'unscientific' belief systems such as astrology or faith healing, explaining what it is that makes these frameworks problematic. After a quick overview of some key figures in philosophy of science – Popper, Kuhn, and Lakatos – we discuss what has gone wrong in belief systems such as astrology or psychic faith healing. Importantly, the issue is not just a conceptual vice of accepting entities that don't really exist, but there is (i) an empirical issue where they tend not to take a genuine interest in whether their system does its own heavy lifting, (ii) a theoretical issue where it tends to set up tautological arguments that cannot be falsified, and (iii) a pragmatic issue that it can be damaging for the believers if they are fooled into making real life decisions based on a framework that turns out to lack credibility.

Chapter 6 then turns to psychiatry as an example of a supposedly scientific framework that still suffers from similar kinds of issues. The context of psychiatry is compared against a discussion about hypnosis magic shows, where the common feature in both is that they explore situations where human agency works in unexpected ways. In the same way that some cynical sceptics might accuse a person with depression for just 'faking it' or being 'weak-willed' a similar misconception can occur in a hypnosis show if a participant spontaneously collapses

on stage after being given the command to 'sleep'. While fakery or an appeal to weakness of will may be simplistic solutions, a deeper investigation of agency is explored in this chapter.

Chapter 7 returns the topic to Magicalism, further elaborating on the analogy between magic and consciousness. After investigating some problematic characteristics about deductive arguments which can create an illusion of logical rigour in a belief system that is logically valid but unsound, it is argued that Magicalism advocates a worldview where consciousness essentially just is 'the warm fuzzies', in a somewhat metaphorical sense. The essence of consciousness can only be captured in storytelling that allows us to discuss 'love' and other warm, fuzzy concepts that exceed the scope of physics and other hard sciences, which tend to rely heavily on statistical testing or quantifiable methods. There is something qualitative about 'love' that makes it irreducible to matters of 'indifferents'.

Finally, Chapter 8 argues for a novel research programme that is specifically designed for consciousness studies. There are three central elements to the 'hard core' of the research programme: (i) that it is a normative endeavour, (ii) that it is a teleological endeavour, and (iii) it only applies to studying conscious beings. In this final chapter, the 'warm fuzzies' are elaborated on, and we end up with an ontological picture that includes Buddhist concepts such as 'Karmic Fortune' and 'Buddhist souls'. These, however, are metaphorical claims rather than supernatural claims. It is a reminder that human storytelling plays an important role in any metaphysical discussion because philosophy itself is about abstract *stories* rather than 'hard facts' or raw data. It is important to remind ourselves that to understand consciousness, we may be required to step outside of familiar methodology in favour of the *teleological* task of having academic disciplines that improve our quality of life. We do not wish to leave these tasks to self-help gurus and social media influencers, nor do we need to believe that academia must be confined to using statistical types of 'scientific' methods. Instead, we can allow for a more Magical approach to deepening our understanding of consciousness: That method and

storytelling are *both* important in understanding the wonderous elements of the conscious human mind. This thesis is written under the spirit of both Buddhism, as well as the spirit of a magician's mindset, with a focus on discussions about Magic by the psychological illusionist / magician, Derren Brown. In this thesis, The Eastern (unofficial) Apprentice will attempt to illustrate the relevance of Magical craftsmanship in the task of philosophising about the mind.

Chapter 1. The Xombie Argument

In the typical thought experiment involving philosophical zombies, we are asked to imagine a being which has a 'molecule-for-molecule identity' as a normal conscious person yet is lacking in phenomenal experiences. Zombies are physically identical to us, yet there is 'nothing it is like' to be a zombie, as they are beings which are lacking any first-person experiences. If a zombie is stabbed, they will have the pain behaviours of screaming or shouting "I am in pain"; and if a zombie's brain is scanned, all the same brain activity will be recorded as we would observe in a conscious human being. Yet, the zombie is understood as a creature that lacks any first-person perspective and has no 'what it's like' inner experience. We are asked to imagine creatures where 'no one is home' despite them being 'physically' identical to us humans. We then explore whether this thought experiment is conceivable and/or possible (Kirk 2023, 2008, 2005, 1999; Fischer & Sytsma 2021; Márton 2019; Papineau 2013, 2011, 2002, 1998; Marcus 2004; Chalmers 1996).

In recent years, there has been increasing controversy on how 'physical' should be understood (Spiegal 2023; Firt, Hemmo, & Shenker, 2021; Dove & Elpidorou, 2021; Dove 2018; Witmer 2018; Vicente 2011; Stoljar 2009; Judisch 2008; Ney 2008a, 2008b; Montero 2009, 2006; Montero & Papineau 2005; Melnyk 1997; Crane & Mellor 1990). This is relevant for the zombie argument in so far as ambiguity in the notion of 'physicality' can create sources of confusion regarding what it is that we are meant to be imagining in the first place when discussing these 'molecule-for-molecule' identical (i.e., 'physically' identical) zombie creatures.

There is a general sense in which we tend to think that 'physical' things are the types of entities which belong to the domain of physics (Ney, 2008a; Poland 2003; Lower 2001; Melnyk 1997; Pettit 1993). There is another sense in which 'physicality' entails a commitment towards

a No Fundamental Mentality principle, in which case 'physical' would simply be defined as 'non-mental' (Wilson 2006; Worley 2006; Montero 1999; Montero & Papineau 2005; Crook & Gillett 2001; Gillett & Witmer 2001; Spurrett & Papineau 1999). When discussing the zombie argument, a lack of a clear understanding in 'physical' identity can lead to exploitation of this ambiguity.

The problems begin to occur due to the two central components that physicalism tends to involve: the Pro-Science attitude, and the No Fundamental Mentality principle. These two attitudes do not necessarily go hand-in-hand, as it is possible that future science will postulate fundamentally mental entities. The issue with 'physicality' is that the notion may seem entirely straightforward and easy to grasp under a pre-theoretical and intuitive understanding of 'physicality' yet can become elusive once more complex considerations are taken into account.

Later in chapter 2, we will engage in a deeper discussion about various problems with the notion of 'physicality'. However, the purpose of the current chapter is to propose the novel Xombies – defined as 'third-person identical' but 'first-person distinct' creatures – are suggested as a replacement for classic zombies, which allows for a discussion on this thought experiment in a way that avoids using the word 'physical', thus preserving the essence of the thought experiment while side-stepping the issues of ambiguity. The first section begins by discussing the ambiguity of 'physicality' which can create problems in the original zombie experiment. In section two, the construction of the Xombie experiment is explained. Section three will discuss the conceivability of Xombies, and section four discusses the possibility of Xombies with an emphasis on addressing the phenomenal concept strategy. Finally, section five gives a brief note on how Xombies relate to the recently proposed meta-problem of consciousness as brought up by Chalmers (2018).

1.1 'Molecule-For-Molecule' Identity

In the classic zombie experiment, we are asked to begin by imagining two 'molecule-for-molecule' identical worlds, as though we have copied and pasted all 'physical' factors. Then, we explore the conceivability and possibility of this imagined scenario, questioning at a deeper level whether it makes sense for two physically identical worlds to turn out being mentally distinct. It is typically thought to be the case that if two 'physically' identical worlds turn out to be phenomenologically distinct (i.e., experientially distinct), this suggests that there is 'something more' to mental experiences which cannot simply be accounted for by all the physical facts, thus undermining the physicalist claim that only physical facts exist in our world (Chalmers 1996). But if we are uncertain what is being copied and pasted to begin with, the thought experiment cannot take off the ground in the first place. Are we supposed to imagine two worlds that appear identical from the perspective of modern physics, i.e., the limited kind of stuff that our current science talks about? Or we could imagine two worlds with identical chemical compositions, but what happens when physics tells us there is more to science than just molecular compositions?

The simplest way of understanding the concept of 'molecule-for-molecule' identity would be to imagine molecules as being Lego-like building blocks. If we were to imagine that all macro-entities in the world are like brick houses and are composed by microscopic Lego-like building blocks, then we could straightforwardly imagine two worlds that are identical in their Lego composition. Then, we could move on to asking if the two worlds are experientially distinct. If this is conceivable and possible, it might suggest that there is more to the world than just Legos.

The above construal is problematic, however, since not only is this Lego-world view not very accurate from the perspective of contemporary physics particularly when taking into

account field theory (see Lancaster 2019), it is also too specific for metaphysical purposes and does not allow any leeway or nuance for understanding the scientific picture of fundamental reality. Rather than limiting ourselves to a Lego-esque model, the idea of 'molecule-for-molecule' identity might instead be understood as a placeholder term for 'whatever physical composing parts underlie the macro-entity'. In the same way that 'c-fibres firing' is often used as a placeholder term for whatever physical brain processes turn out to underlie pain responses (Kind 2018), 'molecule-for-molecule' identity can simply be used to mean 'identical in all the fundamental ways that science tells us about'. Thus, a 'molecule-for-molecule' identical zombie would be identical to us conscious people in every brain sort of way, but if souls turn out to exist, we might wish to say that zombies lack souls, as this is not typically considered a scientific entity.

This description is still rather vague, however, and does not allow for a distinction to be made between 'physical' and 'non-physical/neutral' composing parts, as neutral monism advocates (Stubenberg & Wishon 2023). To illustrate this point, we can consider a variation of a zombie argument using the Kripkean 'Water = H2O' analogy (Kripke 1980). We begin by imagining two worlds which are 'molecule-for-molecule' identical. There is water on both Worlds A and B, and because 'Water = H2O', we can also say that there is H2O on both worlds. However, let us now imagine that in World A panpsychism is true, and the water molecules are conscious. There is a *what it's like* to be water kind of experience that is being enjoyed by the H2O molecules. In contrast, we can imagine that in World B, panpsychism is false, and H2O molecules are just as inanimate as we might expect them to be. In this case, there is ambiguity in terms of what it means for the worlds to be 'molecule-for-molecule' identical. Technically, both worlds are identical in terms of the amount of H2O molecules that exist, as well as the location where these molecules are placed. However, even though the molecular chemical composition is identical in the two worlds, the properties of the molecules themselves

is very drastically different. Ultimately, this would just be a delayed-zombie argument, but now the thought experiment has not even taken off the ground because we are stuck on the initial set-up phase of working out what it is that we were allowed to copy and paste onto the two worlds at the start of the experiment.

This is an idea that we will return to later when discussing the phenomenal concept strategy. For now, we can start to see that when we are told to imagine two 'physically' identical worlds that are 'molecule-for-molecule' identical, we need to have a consensual and consistent understanding of what it is that we are imagining in the first place. Are we supposed to be imagining two worlds where (a) only chemical composition is identical, or (b) all the molecules have fundamentally identical properties? Presumably, it is the latter. As was mentioned earlier, the word 'molecule' here is used as a placeholder term for 'whatever physical composing parts underlie that entity at the most fundamental level'. Therefore, molecular composition is only of limited value. What we really want to know is the composition at the most fundamental level.

However, this still just delays the question. What if future scientists discover a fundamental particle, like a boson or a graviton, but we will call it a 'phenomon' particle, which is the 'seed' required for any macro-entity to enjoy a conscious experience? Some may argue that this is like a 'seed' of a soul whereas others might argue that because it is an entity postulated by physicists, we should understand this as being a physical kind of entity. But what if future physics derives the conclusion of the existence of 'phenomon' particles for philosophical reasons (thought experiments, etc.,) rather than empirical testing? If the concept gets published in a peer-reviewed physics journal, does this entail that it is a 'physical' entity – even if the article was written by a dualist philosopher hired by a team of physicists? And again, we are back to a confusion about what exactly we mean when we talk about two worlds which are 'physically' identical.

This leads to what has become known as Hempel's Dilemma. The dilemma is this: the word 'physical' can be defined either in terms of its loyalty to current physics (which is an incomplete discipline as of today) or its loyalty towards a future completed physics (which we currently know nothing about in terms of its content). Physicalism which abides with current physics will end up being a philosophical theory that is based in an incomplete science that is sure to be proven false in the future in many respects – thus, physicalism of this type will clearly be false. However, physicalism which references future completed physics says nothing about what entities future physics will tell us about, leading to a completely vacuous version of physicalism. Therefore, both horns of Hempel's Dilemma are problematic, but for different reasons (Ney 2008; Crane & Mellor 1990; Hempel 1980, 1966).

According to currentist physicalism, panpsychist laws or 'phenomon' particles would be 'non-physical', as contemporary physics does not discuss these concepts. However, according to futurist physicalism, these entities could very well turn out to be 'physical' depending on how future physics happens to develop, which is something that we currently know nothing about. More pressingly, Dove and Elpidorou (2021) have recently argued for a new version of Hempel's Dilemma that shows the same dilemma occurs when defining the notion of 'mental' as well. That is, our current intuitive understanding of what 'fundamental mentality' entails may differ drastically from what future philosophers may conclude about the nature or essence of fundamental mentality. Thus, even if we define 'physical' as meaning 'non-mental', a new dilemma emerges: What do we mean by fundamental 'mentality'?

Defining 'physical' as 'non-mental' just delays the issue and is an insufficient reply. Judisch (2008) further discusses the ambiguity of the causal closure principle, arguing that there is a vagueness about whether the world is truly 'physically' causally closed, or if it is causally closed in a more neutral monist sort of way. But again, for the classic zombie argument, it is important to start the thought experiment by having some sense of the two worlds being

either ‘physically’ identical, or ‘molecule-for-molecule’ identical. The thought experiment cannot even take off from the ground until we know what it is that we are asked to copy and paste onto the two worlds as part of the construction of the thought experiment, and which parts need to be derived as part of the argument or conclusion. This clarity is needed to avoid tautologies and question-begging.

Consider, for example, the issue of tautologies which would emerge if we were to define future completed physics’ as being synonymous with the phrase ‘future completed theory of everything’. In this case, physicalism would amount to the claim that:

1. Physicalism = The claim that everything which exists is physical
2. Physical = Things which are covered by future completed physics
3. Future completed physics = Future completed theory of everything which exists

Therefore,

4. Physicalism = The claim that everything which exists are the things which are covered by a future completed theory of everything which exists.

Which implies that,

5. Anti-Physicalism = The claim that some things which exists are things that do not actually exist (i.e., things which are not covered by a future completed theory of *everything* that exists).

This is a tautological version of physicalism that is necessarily true, but only because it presupposes an all-powerful definition of future completed physics. If we are going to presuppose that future physics is a theory of *everything* which exists, then anti-physicalism

would amount to the belief that some things which exist are things that do not actually exist, which is obviously a logical contradiction. Furthermore, in the context of a zombie argument, if we were to start the thought experiment by copying and pasting *everything* which exists onto both worlds (including souls and whatnot), then obviously, the idea that one world would be different from the other is inconceivable – but only in virtue of the way we constructed the thought experiment in the first place. While this kind of physicalism is necessarily true, it is also meaningless due to its vacuity.

To debate the conceivability and possibility of zombies in a way that is philosophically interesting and does not devolve into an argument about trivial definitions, it is important to begin the thought experiment by imagining a meaningful scenario in the first place, before any technical aspects can come into play. When we say that zombies are 'physically' identical or 'molecule-for-molecule' identical, we need to know which parts of the composition we can take as being 'physical' parts of its compositions, and what parts (if any) would be 'non-physical'.

Again, we will explore the more detailed considerations on the definition of 'physicality' in the next chapter. In this chapter, our goal will be about preserving the intuitions behind the zombie argument while sidestepping the problem of 'physicality'. To do this, we are going to reconstruct the original zombie experiment in such a way that avoids this confusing terminology altogether, and this novel variation of the zombie argument will be called The Xombie Argument.

1.2 'Xombies' as a Replacement for Zombies

At the most intuitive level, when we attempted to imaging classic zombies, we likely begin by imagining creatures that look and act like a normal human being from the outside, but

there is a fact of the matter that no phenomenal experiences are being had by these entities. In other words, we are imagining a being that *appears* identical from a third-person point of view, but it turns out that they are not the same as normal human beings when viewed from a first-person point of view. Considering this, we begin by defining 'Xombies' as creatures that are 'third-person identical but first-person distinct'. The next step is to further clarify what it is that is identical between the two worlds, namely, that we are going to imagine the two worlds are entirely identical from the perspective of an external observer. For instance, any brain scans, dissections, verbal reports, or behavioural observations would be identical between the two worlds. At the same time, if we were a resident in one of the worlds, either we would have some kind of experience (e.g., the ability to feel pain) or it would be like a persistent blackout state, void of any inner experiences.

Crucially, we will allow that there can exist laws of nature that only impact the first-person perspective. It should be noted that the terms 'first-person' and 'third-person' are used in this thesis as referents to methodology used for inferring about facts. Therefore, while physicalists may try to insist that there is no distinction between first-person and third person facts, this reply would miss the point of the claims being made here. Once again, we are focused on the epistemic means which are used to derive information, and currently making no claims that two distinct types (first-person and third-person) of ontological facts exist. This distinction is important so that the current discussion does not merely revert back to the physical/non-physical ontological dichotomy that we were trying to avoid in the first place.

For instance, in the Xombie experiment, we will permit that one world can contain an invisible Consciousness Field that gives rise to consciousness, just as the gravity field might 'give rise' to mass having the disposition or power of pulling other masses towards itself. Again, what is central to the Xombie experiment is that the two worlds cannot be distinguished from the third-person standpoint, but we will allow that the two worlds are permitted to contain

different laws of nature. This construction may be unpalatable, particularly for some physicalists, but this point is addressed further in section 1.3. For now, we will focus on clarifying what a Xombie argument entails.

By creating this distinction, we can side-step Hempel's Dilemma and other worries that come from the ambiguity in the word 'physical'. We are, in essence, imagining two worlds where there is a difference between them (i.e., they are not identical in *every* way possible) but at the same time, strictly third-person methodology of the sciences is unable to detect any difference. In other words, to understand the Xombie thought experiment, we must acknowledge that 'science' is not synonymous with 'all-powerful methodology' (which would lead to the tautological version of physicalism mentioned earlier). Therefore, we side-step the issue of tautologies by teasing apart the concept of 'all-powerful methodology' and the idea of 'third-person observable phenomenon'.

With these clarifications in place, we can move on to imagining a World A with conscious people, and a World B with only Xombie inhabitants. We have said that the two worlds are identical in terms of brain scans, brain dissections, verbal reports, and behaviours, and that there is no way of telling the two worlds apart form a third-person standpoint. At the same time, we have allowed that the worlds can differ in terms of the laws of nature governing the worlds, so long as the difference in laws of nature do not impact anything which is third-person observable.

We can, of course, still introspect to gain knowledge about which world we are in. If we ourselves have any kind of sensory or emotional experiences (what most philosophers refer to as 'phenomenal experiences', though Frankish may wish to call this 'quasi-phenomenal properties'; see Frankish 2016, p.15), then we can safely conclude that we are inhabitants of the 'feely' world. Because we cannot presuppose at the start of the thought experiment if

'phenomenal properties' are just illusions or not (since this is precisely what we are trying to derive from the thought experiment), we are currently trying to find common ground in various philosophical positions. Presumably, most (if not all) philosophers would be agreed that there is a quality-esque component to our lives which is different from a blackout state. Whatever this quality-esque thing is, whether it turns out to be an illusion or not, is the stuff that humans have and Xombies lack.

Interestingly, while introspection can shed light on whether I myself have experiences or not, and I can use this information to make an educated guess that others in my world are probably also conscious (since the same laws of nature that affect my brain affect that of others in my world), this information is difficult to document in, say, a psychology study. This is due to the problem of other minds (Overgard 2006). When we observe another human being and listen to their verbal testimony that they are *not* in a blackout state, we have no way of really knowing if this is true. A character in a videogame could claim that they are fully conscious, yet people who play these videogames tend not to have issues with causing pain or harm to videogame characters, even though these players might never harm anyone in real life. There is a gap here which emerges between verbal reports which we think we have good reason to take seriously, and those which we do not. Problematically, while we have direct access to our own introspection (e.g., I know *I* am not in a blackout state), we cannot tell whether our friends too are indeed conscious or not.

We could, of course, try to scan our own brains and see if other people have the same type of activity. We can extrapolate from the premise:

(i) I am conscious

Combine this with the assumption that:

(ii) The way my own brain works is likely to be the same way that other brains work.

And derive the conclusion that:

(iii) Other people are also likely conscious.

However, this logic cannot be applied to beings in another world, as we have allowed that the other world could have different laws of nature. Just because I am conscious in *my world* does not mean that creatures in a different world (with potentially different laws of nature) will be conscious in the same way. While they might superficially exhibit the same behaviours, verbal reports, and other brain activity that is measurable through scientific apparatuses, nothing guarantees that the brain regions associated with consciousness in *our world* gives rise to the same consciousness phenomena in the other world.

This is the explanatory gap (Chalmers 2006). It may be that in the other world, these brain regions are causally inert and play no role at all, even though they will fire exactly like ours do. Perhaps it is merely a byproduct of (different) evolutionary mutations in *their* world so that the Xombies have superficially identical brain activity as us, but crucially, maybe their brains are missing 'phenomon' particles – particles which might turn out to be inferable only through philosophical reasoning but not actually 'observable' through typical scientific (third-person) methodology. The bottom line is that we are imagining two worlds where there is a fact of the matter that they differ in terms of composition (one world contains phenomon particles and the other one does not) but science is not the right tool for measuring this. This leads to a situation where third-person methods cannot distinguish the two worlds, but introspection (i.e., first-person methodology) *does* still allow us to make inferences.

What this also means is that the two worlds are similar only in very superficial ways. While this entails accepting cross-world similarity relations (Lewis 1986), the laws of nature

seem to be drastically different in these two worlds. In fact, some may argue that the two worlds are outright 'physically' non-identical. However, I avoid using this phrase due to my later rejection of the word 'physicality'. But if an opponent here has the intuition that I am merely describing that Xombies live in a 'physically' non-identical world, then this interpretation is not wrong. These concerns are addressed more deeply in the rest of this chapter, but again, we are currently focusing just on the construction of the Xombie thought experiment while *avoiding* the problematic word 'physical', so that we can get ideas off the ground. Further details will be added in the next chapter.

One advantage of discussing Xombies rather than zombies is that we can engage in further discussions about the important role that having a first-person perspective can play in various other metaphysical or ethical discussions. As mentioned earlier, perhaps common sense tells us that there is nothing wrong with harming a videogame character (who we take to be Xombified) whereas there is something ethically wrong with harming a real person due to their ability to feel excruciating pain. We know *what it is like* to be in pain, and perhaps we will want to say that we do not wish this horrible experience upon anyone else. But also, we might have no issue with using car crash dummies for safety tests, since we presume that they cannot feel any pain. Xombies, in essence, are like car crush dummies, but superficially indistinguishable from humans.

It is helpful for philosophical purposes to have a word for non-conscious creatures that resemble human beings, particularly as AI technology continues to develop. We might also wish to say that having a first-person perspective is necessary for any free will discussion to take off the ground (does it make sense to talk about the 'free will' of a non-conscious entity?) or that Xombified entities like rocks, manikins, and videogame characters do not enjoy the same moral standing as human beings. We might even say that the reason we care about a friend who is depressed is because there is something that it is like to *feel* depressed. If all

psychiatric patients were Xombies, then perhaps there would be no reason to help them despite their verbal complaints, and these sorts of discussions can have impact on psychiatric and self-help related discussions.

Postulating the concept of Xombies can be useful not just for ethical purposes, but also in philosophy of science, to highlight some limitations with the current third-person methodology in psychology. For example, in the early days of psychology, introspection was taken seriously as part of the research method (see Boring 1953 for overview on a history of introspective research in psychology). Wilhelm Wundt, for instance, was known for using introspection in his study of the study of 'Geist' (mind) and 'consciousnesses' (Asthana 2015; Boring 1953) which is in stark contrast to the behaviourist traditions where first-person consciousness was explicitly removed from the scope of study to make the psychological research more consistent with other sciences (Watson 1913). Particularly in areas of psychology where the felt experience is central to the purpose of conducting the research, such as in psychiatry or self-help studies, it may be important to have the language available for discussing that humans are *not* Xombies, and that is why studying these disciplines (what it means to *feel* happy, etc.) have meaning. In contrast, if psychiatry treats patients as though they might as well be Xombies, this can be dehumanising for patients, linking to arguments that are made in well-being related literature (Gambrill 2014).

The Xombie Argument, therefore, serves a purpose of linking philosophy of mind to other fields which also handle consciousness. Strictly speaking, discussing Xombies is not particularly useful if our only aim was to undermine physicalism. This is because we have removed the word 'physical' from the thought experiment (and if anything, one might argue that Xombies just are 'physically' non-identical from us), they are not particularly helpful for serving the purpose which original zombies were supposed to serve: discussions on whether consciousness exists over and above the physical. However, in chapter two, we will be taking

a separate route for rejecting physicalism (i.e., rejecting its definability). Xombies themselves do not pose any threat towards physicalism, but rather, their purpose is to allow for a variation of zombies that can be used in cross-field research (as we will see later in this thesis) while circumventing the classic problems of conceivability/possibility, as we shall now see in the next two sections.

1.3 On Conceivability

According to Chalmers's (2003) classification system, Type A physicalists are physicalists who take zombies to be inconceivable, whereas Type B physicalists refer to physicalists who take zombies to be conceivable but impossible, as there is a separate discussion that can be had on whether conceivability must entail possibility (Chalmers 2002). Anti-physicalists, in turn, tend to believe zombies are both conceivable and possible. In this section, we briefly address the issue of conceivability of Xombies, before turning to the issue of possibility of Xombies in the next section.

It was mentioned earlier that the classic zombie argument is often used to try and undermine physicalism. The idea is that if we can imagine two 'physically' identical worlds, yet it turns out that they are still phenomenally distinct, this would imply that there is something more than 'physical' structures which give rise to consciousness, thereby undermining the claim that everything which exists is physical (Chalmers 1996). Importantly, this argument only works if the thought experiment begins by claiming the two worlds are 'physically' identical, then analysing more deeply if 'physical' reality is all there is to fundamental ontology.

The Xombie Argument, essentially, has opted out of this line of thought due to its removal of the word 'physicality'. The thought experiment already begins by explicitly

accepting that the two worlds operate according to different laws of nature, and that their similarities are only superficial. Furthermore, the definition of 'conscious' humans will also require some caution.

Because Xombies have been proposed to be adapted for cross-field purposes, we are using a wider and looser definition of 'consciousness' so that it is more versatile when applied to other fields. Loosely, 'consciousness' here is being used to mean something along the lines of 'whatever that *feeling* of pain is that makes us care about whether the entity in question is *feeling* that pain or not' (what Frankish may wish to refer to as 'quasi-phenomenal properties'; 2016, p.15). At an intuitive level, what we are focused on now is to acknowledge that some entities are capable of feeling agony (e.g., humans, cats, etc.,) and some entities are not (e.g., rocks, tables, chairs). It is not relevant for our purposes whether this 'feeling' reduces to the 'physical' – largely because we later end up rejecting the definability of 'physicality' anyway. Not only this, but when engaging in ethical or wellness-related discussions, it is not always helpful to define consciousness in purely 'physical' terms. For these reasons, we are understanding Xombies as *definitionally* incapable of 'feeling' agony, much like a rock or a chair – but due their world being a very different place from our own world, Xombies are third-person indistinguishable from us. Therefore, *if* science is limited to *only* third-person methodologies, then Xombies are 'scientifically indistinguishable'.

With that in mind, a Type A physicalist who has an issue with the conceivability of classic zombies really has no reason to take issue with the conceivability of Xombies. Because we have allowed that Xombies live in a very different world from ours (with different laws of nature), there should be nothing surprising about their world being phenomenologically different from ours as well. For all we know, the Xombies live in a world of magic! (or, as later chapters will suggest, it is our world that contains Magic and the Xombie world which lacks it – although this is 'magic' in the sense of a card trick, not the supernatural sense). And therefore,

there is nothing inconceivable about imagining an alternate world where entities superficially similar to us reside, and also turn out to be more akin to rocks, robots, or videogame characters than human beings.

At the same time, a Type A physicalist may still take issue with the way in which we have chosen to define 'consciousness'. Perhaps the Type A physicalist may find it uninteresting or disappointing that we have constructed a thought experiment which uses a 'quasi' definition of consciousness and also accepts that the two worlds are 'physically' non-identical to begin with. The opponent may ask: Does this not destroy the purpose of creating a novel zombie argument? In response, the answer is that if anything, Xombies are helpful when discussions are being had between two parties that both believe original zombies are inconceivable. It is not particularly fun or interesting to discuss a thought experiment that we think does not take off the ground to begin with – and so having a novel Xombie thought experiment that does take off the ground and can be used for other academic purposes allows us to re-engage with the topic from a new angle.

There are other potential issues which may need addressing. The Type A physicalist may object that defining consciousness in this sort of loose or common-sense kind of way is either (a) unhelpful for addressing the mind-body problem (i.e., that it is not rigorous enough of a definition) or alternatively, (b) that this kind of pre-theoretical definition is precisely what the Type A physicalist wishes to dissolve and claim to be an illusion.

If it is (a) that they wish to claim, then the reply is that they are entirely correct. Once again, Xombies are not intended to be used for the purpose of defending or debunking physicalism, but the purpose is to use this concept for cross-field research. In a sense, they are not designed to address the 'mind-body' problem as is conventionally discussed, but they are more helpful for discussing a 'mind-spirit' problem. That is, thought experiments involving

Xombies can help us engage in discussion about how there seems to be something ethically valuable about 'feely' creatures. Xombies are basically humanoid creatures who lack 'souls' – but in a metaphorical sense which will link to Buddhist spirituality later in this thesis.

If, however, the Type A physicalist wishes to assert (b), then this would be a problematic objection. The same word (consciousness) can be defined differently depending on the context in which it is used. Therefore, while it may be valid to try and dissolve a 'feely' definition of consciousness in the context of the mind-body problem (we explore this more deeply in chapter three), in the context of an ethical debate, it might sound sociopathic to insist that feelings of agony do not really exist. If someone was being stabbed, we would not tell them that "Hey, it's just stage magic!" (Dennett 2016, p.66); but if someone was hypnotised to believe they were stabbed, this reaction might be appropriate. These are considerations that will be addressed more deeply in chapter five and six (on hypnosis stage shows and the distinction between wonderous Illusions$_{(A)}$ and deceptive Illusions$_{(B)}$). What is important for now is to see that it is entirely acceptable to define consciousness in one way for the context of discussion A (e.g., ethical purposes) even if the same definition is not appropriate in the context of discussion B (e.g., mind-body problem). The part that might be counterintuitive for an opponent is that we are here constructing a *variation* of a zombie experiment for different purposes than what *original* zombies are most commonly used for – but again, there is nothing wrong with creating a variation of an existing philosophical thought experiment in order to later adapt it for a different purpose.

Therefore, we can conclude that conceivability should not be an issue for Xombies unless an opponent is willing to make an extremely controversial claim that no entities ever feel agony – in the sense that human beings have just as much of an emotional life as a rock. This would have ethical implications, suggesting for instance that there is nothing wrong with running over a human in a trolley problem in favour of saving a manikin, since both are equally agony-free

anyway. This would be a very extreme claim to make, we have hopefully seen sufficient reason for our current purposes to conclude that Xombies are indeed conceivable.

1.4 On Possibility

With the above discussion in mind, an opponent may now be wondering why such a complex set up is required to postulate Xombies if the purpose is to compare entities that feel pain or do not feel pain. Presumably, if this ethical purpose were the *only* reason for discussing Xombies, it would be much quicker and simpler to just compare a person with a manikin, or compare living people with videogame characters, and so on. The reply to this objection is that Xombies are useful for more than just ethical discussions. They can be used for other philosophical purposes as well, such as applying them to discussions in philosophy of science, other fields of metaphysics, as well as other disciplines such as psychology and psychiatry.

What the Xombie thought experiment attempts to highlight is that there exists a gap between the idea of 'laws of nature' and 'laws of physics'. The former, we will understand is including all the ontological laws, entities, dispositions, powers, etc., which exist in our world. The latter, we will interpret as being 'the laws *as discussed by* physics' – which would be limited by potential methodological restraints unique to the field of physics. While it is open to discussion what those methodological limits are, and the details of those limitations may change through time as physics develops, it is still important to have two separate concepts to capture (a) all ontological laws, and (b) the laws which are measurable, given our epistemic constraints. Having this distinction allows us to avoid the risk of tautology described earlier in section 1.1. Being optimistic about the development of future physics is one thing, but to definitionally presuppose that 'future completed physics = future completed theory of

everything' is problematic, as this will turn physicalism into a vacuous and semantic argument that is just a tautology (i.e., true by definition but lacks content).

Postulating Xombies is not only helpful for linking metaphysical discussion with ethical discussions, but it is also intended to be thought-provoking for considering what exactly the gap is between 'laws of nature' and 'laws of physics'. For example, one candidate is to say that physics is constrained by things which are mathematically analysable or that it limits itself to third-person observability (see Goff 2017); and also, another candidate is to consider that the kind of *interests* which physicists tend to have. What is noteworthy here is that the concept of 'laws of physics' is at least to some extent, sociological. Something will only be categorised as a 'law of physics' if physicists have postulated the concept in an acceptable setting (e.g., peer-reviewed journal), and in our society, there are certain bureaucratic procedures in place for this to happen. One must go to a university, get a physics degree, apply for funding, publish in relevant journals, and come up with research that is sufficiently accepted by their peer physicists before an idea can be accepted as a 'law of physics'. This means that if these procedures were to change for bureaucratic or political reasons, then the notion of 'laws of physics' is susceptible to change as well. 'Laws of nature', however, we will reserve for discussing the ontological reality more directly. It is important to have terminology that differentiates between these two concepts.

Furthermore, the measurability of newly discovered 'physical' phenomena is also limited by the kind of apparatuses which can be engineered, and of course, that relevant apparatus must be built in the first place before any measurements can be made with it. Once again, this is limited by sociological considerations, such as the interests of the physicists/engineers (and maybe neuroscientists as well) to determine what kinds of machinery and apparatuses get built (or receive funding), which thereby influences what kinds of research can be carried out, and so on so forth.

Now, we might wish to say that in 'future completed physics', we will have had enough time to overcome all of these sociological restraints. Perhaps given enough time, we will have had a chance to overcome all funding constraints, and maybe future physicists and engineers will have an expanded scope of interest as well. Therefore, while contemporary physics tends to focus on problems which are mathematically analysable (and we cannot *presuppose* that consciousness falls into this mathematical domain), it may be the case that future physicists will also consider expanding their field to cover natural phenomenon which is difficult or impossible to analyse mathematically. Thought experiments might be one method that will be employed, or maybe the need for statistical analyses and replicability in future physics research will work differently.

Whatever the situation is, we might wish to hand-wave at 'future completed physics' and assert that in an ideal world, it will at the very least overcome all sociological restrictions, although perhaps other methodological restrictions may remain (such as us being limited by the sensory organs and cognitive powers of a human being). These are all interesting things to discuss, hence why proposing the Xombie experiment is helpful for having this thought-provoking impact. While some of these topics will be touched upon again later in this thesis, a full investigation of the implications Xombies have towards philosophy of science exceeds the scope of the current project. However, we can hopefully see that Xombies are useful for more than just ethical discussions, hence why such a complex set-up has been created to discuss this thought experiment.

Our present focus is not about debating *what* those limitations are of future physics, but to discuss how these ideas relate to the possibility discussion regarding Xombies. In the classic zombie argument, Type B physicalists claim that while zombies are conceivable (there is nothing incoherent with thinking about them), they are still not possible (there could be no alternate world which contains zombies) (Chalmers 2003). To defend Xombies from this

problem, we can give both a short answer and a longer answer to this potential concern. The short answer is that because we have already said the Xombies live in a world that contains different laws of nature from our own, intuitively, it seems like there should be no reason to reject the possibility of Xombies. If we are willing to accept cross-world similarity relations (Lewis 1986) as this thought experiment requires (and as is typically accepted when someone is willing to engage in a zombie argument), then this would entail that we have already accepted that it is possible for two worlds to exist that are similar in some respects but not others. The typical approach that is taken when refuting the possibility of Xombies is to assert that we cannot claim (a) the two worlds are physically identical, AND also claim (b) the two worlds are phenomenologically distinct. But because Xombies never claim (a), we are only being asked to imagine (b) two worlds that are phenomenologically distinct.

The longer answer requires looking at what has become known as the Phenomenal Concept Strategy (PCS), which claims that 'pain' and 'c-fibres firing' are just two ways of talking about the same thing. In the same way that 'water' and 'H2O' are also just two names pointing to the same thing, phenomenal concepts such as 'pain' just is another way of talking about physical concepts such as 'c-fibres firing' (or whatever neural and behavioural responses accompany the phenomenal experience of pain) (Lotfi 2023; Botin 2022; Muideen 2017; Mabaquiao 2015; Balog 2012; Diaz-Leon 2008; Chalmers 2006). This relies on the law of transitivity:

1. C-fibres firing exists in both worlds.
2. 'Pain' *just is* 'c-fibres firing'.

Therefore,

3. Pain exists in both worlds.

Thus, just as a world containing the morning star must also necessarily be a world containing the evening star (because the morning star *just is* the evening star, and they are merely two names for the same thing), PCS claims it follows that any world contains c-fibres firing must also contain phenomenal pain experiences. If the zombie world contains c-fibres firing, then it must also necessarily contain the phenomenal concept of pain, and so zombies are impossible.

The problem with PCS, however, is that the 'physical' concepts can be understood in two separate ways, neither of which pose a threat towards Xombies. Firstly, note that when PCS talks about 'water = H2O', they are not saying that molecular composition is sufficient to capture the entire scientific story of what water is. Some H2O solutions will contain different levels of H+ ions, and when H2O combines with H+ this will form H3O+, which is still considered water. Furthermore, 'heavy water' refers to water that contains deuterium (an isotope of H, meaning it is H but with different atomic mass). Heavy waters is toxic and contains different properties, but is, nonetheless, water (Tobia et al., 2020; Needham 2010; VandeWall 2007).

In essence, it is not strictly speaking true to say that 'water = H2O'. However, the point that is being made by PCS is a more generic one: that the concept of 'water' *just is* 'whatever scientists tell us it is fundamentally at the physical or chemical level'. In the same way that the evening star *just is* the morning star, and there is an identity relation between the two, there is also an identity relation between phenomenal concepts like 'pain' and physical concepts like 'c-fibres firing', where 'c-fibres firing' is understood as 'whatever scientists tell us it is fundamentally at the physical or chemical level, as well as behavioural responses that accompany pain reactions'.

Under this light, we will understand PCS as claiming that 'water = H2O*' where the asterisk after H2O signifies that it is merely a placeholder term for 'whatever scientists tell us it is fundamentally'. In effect, it is claiming that the colloquial language we use for concepts holds an identity relation with the technical language developed by experts which gives us a deeper look into analysing pre-theoretical concepts. This, however, makes PCS vulnerable to different interpretations of 'physical'. As Hempel's Dilemma highlights, it is ambiguous whether they are saying that 'water = H2O*' in the sense of things that current physics talks about, or in the sense of what future physics talks about. If the former, then water likely does not hold an identity relation with H2O* as current physicists discuss, since water might contain subatomic particles that current physics does not know about yet (and therefore H2O* would not include these elements). If the latter, then H2O* could include yet-to-be-discovered concepts like 'phenomon' particles or a Consciousness Field, which are physics-sounding concepts that were made up in this thesis. But if we are allowed to just make up terminology that implies it contains the 'seed' of consciousness – is it fair to say these are still 'physical' concepts of the sort science talks about?

The problem with PCS is that it only carries force if it exploits the ambiguity in the word 'physical' concepts. Both interpretations of PCS end up being problematic:

1. The first interpretation is to understand H2O as referring to the *molecular* composition.

In which case,

a. In World A, 'Water 1 = H2O'
b. In World B, 'Water 2 = H2O'

Therefore,

c. Water 1 = Water 2 = H2O (law of transitivity)

But because this only focuses on the molecular compositions, 'water' no longer holds an identity relation with 'H2O', as the cases of H3O+ and heavy water suggest. Equally, the opponent cannot replace 'H2O = molecular composition' with 'H2O* = whatever chemical things scientists tell us about' without running into Hempel's Dilemma. Papineau, for example, tries to overcome this by claiming that defining 'physical' to merely mean 'non-mental' will overcome Hempel's Dilemma, referring to the No Fundamental Mentality definition of the notion of 'physicality' (Montero & Papineau 2005). However, this again runs into the problem that it is unclear how 'phenomon' particles and a Consciousness Field would be categorised. These considerations are discussed more deeply in chapters two and three, but we can already begin to see that the sociological element of 'the laws of physics' will throw us a curve ball. What if future physicists start employing dualist philosophers to contribute to their knowledge base in physics? Then sociologically, we might end up with a situation where seemingly dualist laws become contained within the statement of 'whatever scientists tell us it is fundamentally', thereby obscuring the distinction between what counts as a 'scientific' explanation and a 'dualist' explanation. These are considerations a PCS-advocate would need to answer to before PCS truly takes off the ground.

2. The second interpretation is to widen the scope of H2O to include *anything* that enters the discussion in future physics. This would be to forgo the No Fundamental Mentality principle and allow that water = H2O could include potential dualist-sounding ideas, biting the bullet that *if* (in a surprising turn of events) future physics were to accept these concepts, then we will concede that these concepts are, indeed, 'physical' concepts. The physicalist can

still maintain the clam that it is extremely *unlikely* future physics will include such concepts (Melnyk 1997).

This approach still runs into the problem of the panpsychist water discussed earlier in this chapter. If our purpose is about comparing cross-world similarities, where the other world turns out to witness the 'surprising turn of events' that was unlikely in our world, this entails:

a. In World A, 'Water 1 = H2O*'
    i. In World A, Water 1 is conscious.
    ii. Therefore, In World A, H2O* is conscious.
b. In World B, 'Water 2 = H2O*'
    i. In World B, Water 2 is not conscious.
    ii. Therefore, In World B, H2O* is not conscious.

Therefore,

c. Water 1 ≠ Water 2

In Conclusion,

d. 'Water 1 = H2O*' and 'Water 2 = H2O*' cannot both be true, and so:
e. Water 1 ≠ H2O*; Water 2 ≠ H2O* *(reductio ad absurdum)*

In World A, scientists have accepted that panpsychist laws do exist, in the surprising turn of events that we decided was unlikely but not impossible. In World B, no such laws have been accepted. Thus, the water of World A and B end up not holding an identity relation.

In particular, when applied to the possibility of Xombies (where the thought experiment is silent about 'physical laws' and focuses only on 'laws of nature'), PCS will lose its force entirely. In the context of classic zombies, we are being asked to consider two worlds which are supposedly 'physically' identical, in which case it becomes important what counts as a 'physical' similarity and what does not. In such a case, it may be problematic if H2O* includes panpsychist laws, as this would go against the No Fundamental Mentality principle of physicalism. However, there is room for rebuttal here where the physicalist can try to further specify what is they mean by H2O*, or what fundamental mentality would entail, or what exactly the limitations of 'physics' are, etc.

The Xombie Argument can avoid all of this, however, as the word 'physical' was never used in the construction of the thought experiment. In the case of Xombies, what we are being asked to imagine is two worlds which contain different laws of nature, but that the differences are indistinguishable using strictly third-person methodology. If we think it is at least *possible* that, say, the mathematical or third-person nature of methodology in science could create a gap between World A (a world of humans) and World B (a world of human-esque manikins called Xombies), then the worry of possibility for Xombies will have been circumvented.

More to the point, unlike in the classic zombie argument, it is also unclear why physicalists would feel the need to argue against the possibility of Xombies. Unlike with classic zombies, Xombies do not pose a threat to physicalism anyway, and this thesis will take a different route to try and undermine physicalism (i.e., attacking its definability). In a sense, Xombies themselves are completely harmless towards the physicalist – since the first-person/third-person distinction is an epistemic difference and not on ontological one, and physicalists typically do not oppose this idea – and can just be thought of as an innocuous version of zombies that help us engage with other topics (ethics, philosophy of science, psychiatry, etc.,) without posing a direct threat to physicalism. Therefore, in the same way

Xombies need not fear the threat of PCS, PCS-advocates also need not worry too much of the possibility of Xombies.

1.5 The Meta-Problem and Xombies

Chalmers (2018) recently discussed the meta-problem of consciousness, which is about explaining why we think the hard problem of consciousness is a problem at all. It focuses, therefore, on asking what sort of problem intuitions we have which motivates us to be stuck on the first-order issue of the mind-body problem. (Botin 2022; Dennett 2019; Frankish 2019; Chalmers 2018). Xombies provide a novel insight to what the problem intuition might be.

If we have presumed until now that 'laws of nature' and 'laws of physics' can be used interchangeably, this underplays the potential problem that sociological, bureaucratic, political, or other methodological limitations can place when we talk about the laws as discussed by physics. Another potential limitation that was not discussed in this chapter, but will be explored in chapter three, is the idea that school-level physics tells a caricatured and inaccurate story about physics, which later gets re-taught properly at higher level physics – for those students who decide to continue in their studies of this discipline. This is not just in terms of correcting the content of principles in physics, but can change the impression or intuitions one has about what physics *is*, or what this tells us about the 'physical' world (Tong 2017; Lancaster & Blundell 2014, Introduction; Squires 1990, Introduction). Therefore, this creates a higher-order problem that the 'laws of nature *as discussed* by physics *as discussed* by philosophers' (e.g., causal closure principle as discussed by physicalists, such as Papineau 2009a) is a separate type of issue from the 'laws of nature *as discussed* by physics' (= what we have called 'the laws of physics') and also 'laws of nature' (= fundamental, ontological phenomena). The first two have a heavy epistemic component to it, and therefore require an analysis of what those epistemic

limitations are, and if they impact our evaluation of ontological reality; the last one is directly an ontological claim.

The next two chapters will further explore the issues which occur when philosophers use the word ‘physicality’ without specifying enough about what it is that they are trying to point to. Xombies are just the beginning of highlighting the limitations of the classic mind-body problem, the limitations of terminology, and our misunderstandings of what ‘physicality’ is. Next, we turn to the deeper analysis of *why* this word is hugely problematic when philosophising about the mind.

Chapter 2. Structural and Scientific Physicalism

Physicalism can be loosely understood as the metaphysical doctrine which claims that everything which exists is fundamentally physical (Crane & Mellor 1990). Although this claim is straightforward to understand at an intuitive level for those who are not directly engaged in the discussion on physicalism (Ney 2008a), a potentially detrimental ambiguity in the definition of physicalism has been addressed in what has become known as 'Hempel's Dilemma' (Hempel 1980; 1966). The dilemma asks us to clarify whether physicalism is committed to current science, or future completed science. Embracing the first or second horn expresses loyalty to the physical sciences in two very different ways, both of which turn out to be problematic for different reasons:

> "...current physics is surely incomplete (even in its ontology) as well as inaccurate (in its laws). This poses a dilemma: either physicalist principles are based on current physics, in which case there is every reason to think they are false; or else they are not, in which case it is, at best, difficult to interpret them, since they are based on a "physics" that does not exist yet we lack any general criterion of "physical object, property, or law" framed independently of existing physical theory." (Hellman 1985, p. 609).

Crane and Mellor have gone as far as arguing that "physicalism is the wrong answer to an essentially trivial question" (1990, p. 206). To understand the nature of the problem here, there are two points to clarify: (1) How to understand the mind-body debate (i.e., what is the question?) and (2) How to understand 'physicalism' (i.e., what is the answer to that question?)

In the debates advocating or responding to Hempel's Dilemma, the focus of discussion has typically been about the latter point (Dove 2018; Ney 2008b; Montero & Papineau 2005;

Gillett & Witmer 2001; Pettit 1993; Crane & Mellor 1990; Hempel 1980). In this chapter, we will take a further step back and re-evaluate what the question was supposed to be in the first place, so we can be clearer about what type of answer needs to be provided. This is a somewhat challenging task, given that there seems to be no single way of understanding the mind-body debate in a very generic sense, and the focus of the debate has also changed throughout time (Ostenfeld 2018). Even terminology which has arisen in the last century or so, such as qualia, is used differently depending on when it was used, the context it was used in, or the individual who used the term (Crane 2012). Philosophy of mind is naturally a complex topic with a very large scope, and we see here that 'physicalism' is not the only ambiguous terminology to appear in the literature pertaining to the philosophy of consciousness.

This chapter attempts to break down our task into more manageable steps. Section one begins with discussing the ambiguity in the word 'physical' which leads to tautological or otherwise problematic versions of physicalism. By covering these clearly problematic examples first, this will set the context for understanding (a) why clarifying ambiguous definitions is a task we should take very seriously; and (b) what sort of problems a physicalist will need to be mindful of when attempting to define their view. Section two then introduces a new classification system, of separating physicalism into two subgroups: Scientific Physicalism and Structural Physicalism. The former is for physicalists who are committed to a pro-science attitude, whereas the latter is for physicalists who are committed to the No Fundamental Mentality (NFM) principle. This section will discuss how a conflict of interest can emerge between the 'pro-science attitude' group and the 'NFM' group, thus placing pressure on physicalists to pick a camp, which thereby divides physicalists into two separate subdivisions based on their core commitments. Hybrid Physicalism is untenable, as it will either be incoherent, or merely will collapse back into one of these camps.

From this point onwards, the two subgroups of physicalism will be treated as endorsing two different theories. Structural Physicalism (the 'NFM' group) will be understood in the more classic sense of physicalism as being an ontological doctrine which makes claims about what types of substances exist or do not exist. In contrast, Scientific Physicalism (the 'pro-science' group) will be understood as endorsing an essentially epistemological thesis about what type of methodology we can trust for gaining knowledge about ontological matters, their answer being that scientific methods are the only reliable epistemic means for revealing truths about the things which exist in our world.

In addition, to analyse the 'correct' way(s) of defining physicalism, we would first need to know what the question is that we are trying to answer. There appear to be two candidates for viable questions that physicalism might be an answer to: (1) The ontological question: "What is consciousness made of?" (i.e., "Is consciousness fundamentally made up of physical or non-physical stuff?") or (2) The epistemological question: "What methods can we use to understand what consciousness is made of?" where these methods can then have ontological implications (but *not* commitments), such as the idea that 'physical' things are extended in space.

Section three addresses the shortcomings of both Structural and Scientific Physicalism, namely, that the former ends up being either vacuous, incoherent, or overly speculative, and the latter is not really an ontological claim to begin with. In particular, Scientific Physicalism will be divided into two further subtypes: Authoritarian Physicalism and Science Norms Physicalism. The former commits itself to the authority of credibility of those who have the right to claim their status as a scientist (e.g., have a publication in a physics journal; have a university degree) where their expertise in physics can be presumed. The latter commits itself to methodological norms of science, and this transposes into a discussion in philosophy of science. Both are essentially epistemological claims, and while they can have implications

towards philosophy of mind, they are likely not what the typical physicalist had in mind when they argue for ontological physicalism.

Finally, the chapter will conclude by claiming that the reason physicalism struggles with defining itself is because it has misunderstood the original question. Asking "What is consciousness?" ought to be interpreted as a question about the 'essence' of consciousness in relation to our human philosophical interests about the topic, rather than being about the far narrower question of whether consciousness is explainable in non-supernatural terms. For the mind-body problem itself to be understood as an interesting (non-trivial, non-vacuous) question, we will end up having to understand the question itself in a way that renders physicalism an uninteresting or uninformative answer.

2.1 Problems with Definitions

Physicalism is understood as the notion that everything which exists is fundamentally physical, but this definition requires that we know what the word 'physical' entails. At an intuitive level, there may seem to be very little ambiguity. We can point at things such as tables and chairs, or bodies and brains, and declare quite clearly that these are 'physical' entities. However, while there seems to be a consensual understanding at this intuitive level, the matter of defining physicalism is more controversial among those involved in the discussion more directly (Ney 2008a). There are disagreements within physicalists about what the doctrine entails (see Dove 2018 for overview), and such issues are pointed out by anti-physicalists as an argument against physicalism (e.g., Crane & Mellor 1990).

In this first section, we will look at several clearly problematic definitions, which will help to illustrate the issues that can occur when physicalism is sneakily *defined* in a tautological or otherwise problematic way. The issue with tautological definitions, as we will see shortly,

is that it will lead to a version of physicalism that is necessarily true, but only in virtue of how words are (arbitrarily) being defined. This entails that such a version of physicalism will be a trivial and vacuous thesis that is only true for semantic reasons and tells us nothing informative about the ontological world. It also entails that the anti-thesis of a tautological doctrine is necessarily false, due to being defined so that it becomes a self-contradiction. The goal of this section will be to illustrate the sorts of problems that physicalists will need to avoid when defining their doctrine, which will lead to later discussions on more complex possible definitions for physicalism.

(A) Physical = Non-Supernatural

The first example is of a version of physicalism that defines physical as meaning 'not supernatural', where supernatural is simply understood in a naïve sense as 'the things of fairy tales that do not actually exist'. This leads to a tautological version of physicalism that is merely true by definition.

i. Physicalism = everything which exists is physical.

ii. Physical = things that are not supernatural.

iii. Supernatural = things that don't actually exist.

Therefore,

iv. Physical = things that do actually exist.

Therefore,

v. Physicalism = everything which exists are things which do actually exist.

By extension, the anti-thesis would be,

vi. Anti-Physicalism = some things which exist are things which do not actually exist.

Again, this form of physicalism is true by definition, and the anti-thesis is equally false by definition. As this is a purely semantic argument, we have learned nothing new about the ontological world.

(B) Physical = Things Covered in Future Completed Science

Some physicalists have made an appeal to the future of science to specify the details of physicalism, which attempts to understand physicalism as a dynamic and flexible view that will develop as science itself continues to advance (Ney 2008b; Dowell 2006; Wilson 2006; Poland 2003). There are several issues with this type of view, such as the worry of vacuity when appealing to a future science that we currently know nothing about (Crane & Mellor 1990; Hempel 1980), or its unfalsifiability (Ney 2008b). More in-depth discussion on this form of physicalism will take place in section three, but we will briefly examine a clearly tautological version of 'futurist' physicalism (this was also mentioned earlier in chapter one):

i. Physicalism = The claim that everything which exists is physical.

ii. Physical = Things which are covered by future completed physics.

iii. Future completed physics = Future completed theory of everything which exists.

Therefore,

iv. Physicalism = The claim that everything which exists are the things which are covered by a future completed theory of everything which exists.

And the anti-thesis would be,

v. Anti-Physicalism = The claim that some things which exists are things which will not be covered by a future completed theory of everything which exists.

The anti-thesis is a self-contradiction, since if 'a future completed theory of everything which exists' in fact does not cover everything that exists, it does not truly qualify as 'a future completed theory of *everything* which exists'. Therefore, anti-physicalism in this sense is necessarily false due to being a self-contradiction, and any physicalist who wishes to defend a 'futurist' version of physicalism will need to ensure this tautology is avoided. We return to this theme in section three for deeper analysis.

(C) Physicalism = Things Supported by Scientific Evidence

A similar variation of the tautological 'futurist' physicalism is a version of physicalism that is understood as 'being about the things that are supported via scientific evidence'. This is a version of physicalism that is unfalsifiable if a pro-science assumption is made as follows:

i. Physicalism = The claim that everything which exists is physical.

ii. Physical = The things that are supported by scientific evidence.

iii. Scientific Evidence = the *only* kind of evidence worth believing in. *(Assumption)*

Therefore,

iv. Physical = The things that we have evidence for believing in.

v. Non-Physical = The things that have no evidence for believing in.

Therefore,

vi. Physicalism = The claim that everything which exist are things we have evidence for believing in.

This is a form of physicalism that is not technically a tautology but is equally unfalsifiable for different reasons. This version cannot be proven false, precisely because the moment we gain evidence about any entity X, by definition, it would now have to be classified as a thing which is a physical entity. The issue in this situation is (iii) making the assumption that only scientific evidence warrants the formulation of a belief. Thus, as soon as new evidence arises in support of the existence of a new entity X, by definition, we would now have to automatically consider this as being 'scientific' evidence, and thus, entity X is treated as though it was 'physical' all along, regardless of how esoteric or fundamentally mental the entity X appears to be (e.g., ghosts, souls, angels, etc. The moment we have evidence for it, we may now claim that science has proven X was 'physical' all along!) In essence, we never will have reason to include non-physical entities into our ontology. There is no conceivable way in which we can have a thing that is both evidenced in a way that warrants belief and also classified as non-physical, thereby entailing that physicalism is never evidenced against. Physicalism always wins, in a way that renders it trivial.

<u>(D) Physicalism = Things Covered by Science</u>

What we will see in a moment is another similar variation of physicalism that is defined in terms of its pro-science attitude. However, this will be an example where physicalism is not directly defined as a tautology, but rather, it is merely ambiguously defined in such a way that can be exploited to create a tautological defence of physicalism. To illustrate this point, we will look at a quote by Keith Frankish (2019, p,4) and see the sort of problem an ambiguous definition of physicalism can lead to. First, starting with a definition for 'physicalism' and 'physical' which might intuitively seem valid:

i. Physicalism = The claim that everything which exists is physical.

ii. Physical = The things that science tells us about.

Then, we can combine the above definitions with the following quote by Frankish (2019, p.4):

iii. "science tells us that objects don't have such qualitative properties [like redness], just complex physical ones of the sort described by physics and chemistry"

When (iii) is taken by itself, it appears to be making an appeal to the idea that because science does not tell us about objects having qualitative properties, this gives us reason to believe that such qualitative properties (such as redness) do not actually exist and are merely illusions$_{(B)}$. Embedded in this is an assumption that scientific evidence carries with it a certain weight of credibility, where if science tells us a property X does not exist, then this is good reason to believe X does not exist at all. The problem, however, is if (ii) and (iii) are combined together, then Frankish's quote would essentially translate to:

iv. "science tells us that objects don't have such qualitative properties [like redness], just [properties of the sort that science tells us about]."

Which then translates to:

v. "science tells us that objects don't have such qualitative properties [like redness], just [about physical properties]."

Especially if 'qualitative properties' is understood as being synonymous to '(seemingly) non-physical properties', then this quote merely amounts to the claim that:

vi. "science tells us [not about (seemingly) non-physical properties] just [about physical properties]."

This adds no new information, and merely seems to be a reiteration of the claim that:

vii. Science tells us [not about (seemingly) non-scientific properties] just [about scientific properties]

Or in other words,

viii. Science only tells us about what science tells us about, not what non-science seems to tell us about.

Or,

ix. Science does not tell us about the kind of properties that science does not tell us about.

Now, an opponent may object that this is a slightly ungenerous interpretation of Frankish's quote, since I have presupposed a definition of physicalism (as defined in (ii)) which does not come from Frankish himself. Presumably, his intention with the original quote was to

emphasise that science debunks the myths about seemingly non-scientific properties, giving us evidence for why we should not take them seriously as part of ontological reality. However, what is important for now is to see that when there is a lack of clarity with definitions, this intention can become somewhat lost. That is, *if* Frankish were to accept definition (ii), *then* combining (ii) and (iii) would lead to the above type of problematic physicalism, for similar reasons as was discussed above in (C) when 'physicalism' is treated as though it just is a doctrine which follows whatever is supported by scientific evidence, and the credibility of 'science' is trivially magnified. Further issues on this topic are explored later in section three when discussing Authoritarian Physicalism.

(E) The Causal Closure of the Physical (CCP) Principle

This last one is, again, not a tautological definition of physicalism, but rather, is a tautological defence of physicalism that can arise if an ambiguity in the definition of physicality were to be exploited. The CCP principle is loosely understood as the claim that physical events are causally closed, and it is a principle that is often invoked in the context of rejecting fundamental mentality, in virtue of there being no room for non-physical causation to exist in a way that is not overdetermined (Papineau 2009a). The argument typically goes as follows:

i. Physicists assert that all physical events are causally closed. (where this is taken to be a well-accepted fact among physicists)
ii. If physical events are causally closed, then either there is no room for mental causation, or mental causation would be overdetermined.
iii. Therefore, we have good reason to reject the notion of fundamentally mental causation.

The tautological version of this would exploit an ambiguity in the way we understand 'physical', where (i) 'Physicists assert that all physical events are causally closed' might be seen as an assertion that:

iv. Physicists are asserting they can explain everything in purely physical terms.

But if we were to define 'physical' in the following sense:

v. 'Purely physical terms' = things which fall under the scope of physics.

Then (iv) and (v) can be combined to create:

vi. Physicists are asserting they can explain everything that falls under the scope of physics.

If CCP is interpreted in this way, then it is transformed into a relatively uninteresting claim that physicists can explain things that fall under the scope of what is defined as the things that physicists can explain. In which case, this version of CCP would not achieve what the physicalist wants it to achieve. There is now a gap between what CCP is supposed to entail – that physicists are making a very strong claim about how *everything* can be explained in physical terms – whereas due to the problematic definition of (iii), now CCP entails the much weaker claim – that physicists are only making claims about the scope of their own expertise (and are silent about things which fall outside their own expertise, which might still be of interest for philosophical purposes). In order to have the effect that the stronger version of CCP was supposed to have, it would need to be combined with another claim that *everything* indeed does fall under the scope of physics, but again, we have seen already in section (C), caution must be exercised if this assumption is accepted. The problem now is that CCP not only led to

a version that was too weak, and therefore, a vacuous interpretation of CCP, but also that adjusting this slightly will suddenly transform it into a version of CCP that is now far too strong. For instance, consider the following:

vii. Physicists assert that all physical events are causally closed. (where this taken to be a well-accepted fact among physicists)

This time, we can try translating this as:

i. Physicists have decisively discovered that non-physical causation/entities cannot exist.

Therefore,

ii. Physicists have decisively proven physicalism to be true.

This is far too strong a version of CCP, since it amounts to the claim that physicists have already proven beyond doubt that physicalism is true, as though this is already 'a well-accepted fact' among physicists. If this were true, then it would thereby make any further metaphysical discussion redundant. Not only is it questionable whether such a strong claim truly is a 'well-accepted fact among physicists', but even if it were, then the job of a philosopher would now be obsolete, and it is confusing why we are still engaged in this discussion at all. Furthermore, it still does not tell us what 'physical' or 'physicalism' or 'non-physical causation' entails, and the risk of returning to problems from (A)-(D) have not yet been addressed properly.

To summarise, we have seen that an ambiguity in the notion of 'physical' can be exploited to create either tautological definitions or tautological defences of physicalism, leading to vacuous or trivial ways of understanding or defending the doctrine. We have seen

why clarifying the definition will be an important endeavour to be taken seriously, and the type of issues we will need to be mindful of while engaging with the task of providing clarity.

2.2 Introducing the Classifications

The issue of ambiguity in the definition of physicalism has been discussed in the literature in the context of Hempel's Dilemma, which asks whether physicalism subscribes to a doctrine that follows current physics, with all its flaws and incompleteness, or to future physics, which we presently know nothing about thus leading to a vacuous doctrine (Hempel 1980; 1968). What is worth noting here is that there is a basic assumption behind Hempel's Dilemma, which appears to presuppose that physicalism does have some type of commitment towards the physical sciences. That is, to ask whether physicalism is committed to current or future science is an already loaded question, as it presupposes there is an agreement that physicalism surely has something to do with science in one way or another.

According to current physics, which makes no reference to fundamentally mental laws, there may be an intuition that a pro-science attitude will always entail a no fundamental mentality principle. That is, the intuition might be that if we ever do discover reasons for believing in the existence of fundamental mentality, then this would also be the day that we discover the insufficiencies of science. However, a physicalist does not need to make such a strong commitment towards the consistency between future science and the NFM principle. As many acknowledged previously (e.g., Dove 2018; 2010; Judisch 2008; Dowell 2006; Wilson 2006; Worley 2006; Earman 1993; van Frassen 1980), it is quite possible that future science may posit fundamental entities which we may take to be 'mental'.

One example of this would be a case where future physics discovers that consciousness plays a role in collapsing quantum indeterminacies (Pylkkänen, 2017). In such a case, on the

one hand, physicalism still seems supported despite there now being evidence for a seemingly fundamentally mental form of causation. In contrast, a type of physicalism which explicitly commits to a rejection of fundamental mentality might potentially be at risk if future science were to discover this type of evidence.

What is important here is that there are two ways to understand physicalism. The first is the pro-science sense, which we will call Scientific Physicalism. This version of physicalism is committed to the view that everything which exists can be explained in scientific terms (e.g., Stoljar 2023; Ney 2008b; possibly Frankish 2016 and Dennett 2016, 2003, 1991 though these seem more like an attempt at Hybrid Physicalism which will be explained shortly). Therefore, an advocate of Scientific Physicalism will directly have to respond to Hempel's Dilemma, of whether their alliance exists in line with current physics, future physics, or some other variation.

Alternatively, the other approach would be to define 'physical' as simply meaning 'non-mental' (Montero 1999; Spurrett & Papineau 1999; Crook & Gillett 2001; Gillett & Witmer 2001; Montero & Papineau 2005; Wilson 2006; Worley 2006), which we will refer to as Structural Physicalism in that its commitments relate to certain structural considerations, and science merely plays a supplementary or secondary role. Therefore, while scientific evidence can still be cited as evidence for a defence of physicalism, what is truly doing the work in the definitional construction of the thesis is that it commits to certain properties we think fundamentally mental substances would have, then claims that 'physical' things must have or lack those properties. For instance, there are accounts which "defines physical objects as objects with spatial locations." (Markosian, 2000, p.375), in which case a fundamentally mental substance might be considered existing outside of space-time. In such a case, while scientific evidence can still help to support our beliefs, the definition of physicalism no longer relies on current or future science, thereby making the reply to Hempel's Dilemma very easy to establish: The Structural Physicalist is committed to neither current nor future science and is

simply committed to the NFM principle. If future physics indeed were to show that fundamental mentality does exist, then a Structural Physicalist would (and should) be willing to give up their beliefs in physicalism.

The tasks of the Scientific and Structural Physicalists are quite different from one another. For the Scientific Physicalist, they are committed to a defence of the power in science, and that science can account for everything that needs to be accounted for. Therefore, technically what we would want from the Scientific Physicalist is a defence on what it is about the scientific method that gives it this trustworthy status. In contrast, what we want from the Structural Physicalist is to provide us with some criteria for what properties a 'physical' entity should or should not have, or alternatively, what exactly a mental substance would be, if it were to exist. (i.e., What type of evidence would we be looking for from the physical sciences? What sort of entities would we need to discover if we wanted to potentially prove physicalism false?) That is, when they claim that no fundamental mentality exists – then what is it exactly that they are saying does not exist?

Intuitively, the idea of Hybrid Physicalism may seem appealing, where the pro-science attitude and NFM principle can be combined into a single doctrine of a unified physicalism. However, the physicalist cannot both have their cake and eat it too. For instance, consider the Buddhist reincarnation thought experiment (which was also mentioned in the introduction):

Imagine that future scientists discover a new type of gaseous substance emitted by recently deceased people, in the very moment of passing. Then, by tracking the trajectory of the gaseous substance, scientists find that the gas travels, sometimes across the globe, to find a pregnant woman, enters her womb, and goes straight into the foetus. Upon further inspection, it is discovered that this gaseous substance actually changes the brain chemistry of the yet-to-be-born child so that the foetus adopts neural traits previously held by that recently deceased

person. Voila! This research has scientifically proven that Buddhists were right all along. Souls exist. Reincarnation is real! But does this evidence undermine physicalism, or would it merely suggest that souls turned out to be 'physical' all along?

According to a pro-science attitude, these souls are straightforwardly considered 'physical' in virtue of it still being a scientific concept. According to the NFM principle, it might still be argued that these soul-like entities are 'physical', but for different reasons. It might be claimed that what these studies discovered isn't really a 'soul' in any spiritual sense, but it is merely something which *resembles* a soul. Fundamentally, this entity is still classified as a gas, and therefore, does not violate the NFM principle. But then – if the discovery of what looks like Buddhist souls and reincarnation is *still* not enough to undermine physicalism, then what possibly would be enough? Or is physicalism a necessary truth that is ultimately unfalsifiable (like God)? Assuming we do not wish to accept a version of physicalism that is just a tautology such as those seen in section one, or a version of physicalism that is just as unfalsifiable as other religious concepts (we discuss this more in chapter five on falsifiability) the way in which these questions will be answered must depend on whether one has a heavier commitment towards Scientific or Structural Physicalism. That is, IF future science were to diverge from our current intuitions or the currently accepted methodologies (such as if physicists started employing dualist philosophers to contribute to thought experiments to be published in physics journals), then will the physicalist commit to their structural commitments, or their pro-science commitments?

If the NFM principle is the main driving force for the physicalist, physicalism entails:

(1) **Structural Physicalism** = Everything which exists are things with certain kinds of properties, dispositions, or structures (e.g., no fundamental mentality, etc.).

And we would want to ask this type of physicalist a follow-up question about what exactly fundamental mentality would look like, if it were to exist. In contrast, if the pro-science attitude is what is driving the physicalist, then physicalism is to be understood as the claim that:

(2) **Scientific Physicalism** = Everything which exists are things which can be studied using the methods of the physical sciences.

In which case, we would wish to ask a follow-up question about what methods counts as 'the methods of physical sciences'. For instance, is replicability of essence to this? What about mathematical analysability? What about cross-field research where a philosopher or psychologist were to be employed in neuroscience research of the future, and this is taken to have implications in physics? If some methodological constraints can be articulated, then this can serve as a buffer against Hempel's Dilemma, where the physicalist can commit to certain trusted methodologies, and also enjoy the fruits of any ontological implications that come with it, even if we do not yet know the exact nature of the ontological implications as future science might reveal. Finally:

(3) **Hybrid Physicalism** = Everything which exists are things which can be studied using the methods of the physical sciences AND ALSO have certain structural commitments.

The only way in which Hybrid Physicalism would be tenable is if it can be argued that there is a direct overlap in the two definitions and that conflicting interests cannot conceivably occur. Unless this can be done, we cannot begin to specify the definition of 'physicality' in a way that

does not contradict itself. Arguably, the reason that Type A physicalists believe that classic zombies are inconceivable might be because 'physical identity' as defined in the Hybrid way leads to an incoherent or incomprehensible understanding of 'physicality'. If this situation were to occur, then classic zombies will appear inconceivable in virtue of the definition itself being problematic. We return to this discussion in section four when introducing "The Tuned Deck" problem (Dennett 2003, p.17) and see that Hybrid Physicalism is akin to this magical card trick.

2.3 Scientific Physicalism

We have seen so far that there are two different interpretations of physicalism (not including the hybrid form which will be addressed separately later), depending on whether we take the pro-science attitude or NFM principle more seriously. Another way to frame this is to see that the two types of physicalism are actually a response to two slightly different questions.

The mind-body problem, loosely, is understood as asking the question "What is consciousness?" There are two different ways we can approach answering this question. (1) To answer the question directly, and (2) To answer the question indirectly.

One example of the direct approach would be to postulate what type of properties consciousness might have, such as claiming that it is reducible to things which have location in space (Markosian, 2000). This would be the Structural Physicalist approach, as it makes commitments to what properties or dispositions a physical (or non-physical) entity is supposed to have (or not have). In contrast, an example of the indirect approach would be to answer an intermediary question first: "What methods can be used to understand consciousness?" The Scientific Physicalist would answer this intermediary question with the reply "We can use scientific methods (and *only* scientific methods) for answering ontological questions about

fundamental reality". Then, with this question answered, the Scientific Physicalist can use the information provided to us via this methodology to make further ontological assertions.

The two approaches, while they may appear similar, are quite different in the sense that they approach the mind-body problem from two opposite directions. The 'direct' approach of the Structural Physicalist requires that the physicalist comes up with a definition for 'physical' properties or substances in an *a priori* way, then perhaps uses evidence (including, but not necessarily limited to) scientific evidence to justify their claims. While the *a priori* definitions they give can be inspired by current science, notably, the definitions are not *dependent* on current or future science. In other words, Structural Physicalism should be understood as a sort of hypothesis in it of itself: The hypothesis that everything which exists has (or lacks) a certain kind of feature. Then, we go on to test this hypothesis by doing experiments that have the potential to falsify the hypothesis, and if the hypothesis survives these tests, it strengthens the claims made by that theory (Popper 1963; see chapter five of this thesis for deeper discussion on falsifiability).

To contrast this, the 'indirect' approach does not come up with any *a priori* definitions about properties, dispositions, or any other features that 'physical' substances must or must not have. Scientific Physicalism will now be divided into two further subcategories:

(a) **Authoritarian Physicalism** = a commitment to follow wherever the scientists take us.

(b) **Science Norms Physicalism** = a commitment to follow certain research methodologies.

The Scientific Physicalist, in general, is asserting that scientific methods are the only type of method that can be used to understand fundamental ontological reality, and consciousness is no exception. From this point, they can diverge into the two camps: of either just leaving it up to the scientists (either current scientists or future scientists) to decide for us what is or is not scientific; or alternatively, the philosopher of science could take charge of analysing the methodologies that one must use in order for something to be considered a science.

When Scientific Physicalism is understood in the purely authoritarian light of simply signposting to us that we should listen to scientists (and *only* scientists), this turns into a relatively problematic version of physicalism. Alyssa Ney, for example, describes physicalism as essentially an ontological 'oath' that: "I hereby swear to go in my ontology everywhere and only where physics leads me." (Ney 2008b, p.5) This would be a more faith-based interpretation of physicalism, where we are simply taking it on trust that physics provides us with the right means for understanding any ontological question, including matters of consciousness, and therefore we swear to follow wherever physics happens to lead us. While Ney later tries to defend that her proposal does not amount to 'scientism' (Ney 2019), we can see there is some risk if taking the Authoritarian route. 'Scientism' (in its unflattering usage; the usage which Ney explicitly denies applies to her) is a word that can be used to describe an almost religious following of science, which arguably, goes against the 'pro-science' attitude of valuing empirical evidence over authoritarian dogma (see Sorell 2013 in "Philosophy and the infatuation with science"). However, unless the Scientific Physicalist can be a little more specific about *why* we are trusting scientific endeavours, what types of empirical evidence can be valued and for what purpose, or any other comments about the credibility of scientific methodology, there certainly remains a risk that Authoritarian Physicalism turns into 'scientism' in the unflattering sense of the word from above (see Ney 2019 for discussion).

In contrast, Science Norms Physicalism is somewhat less problematic, as it would expresses a loyalty to certain methods rather than a faith in certain kinds of credible people or institutions. An example of this might be Dove's Research Program Physicalism (2018) that combines the idea of a Lakatosian (1978) research program with a Cartesian mechanical philosophy (Garber 2002). The idea here appears to be an advocation of a research program with the 'hard core' commitment of understanding things in terms of mechanical concepts, and trying to defend a view where consciousness is fully explainable according to this research program. (The idea of research programmes is discussed more deeply in chapter five of this thesis). While Science Norms Physicalism will avoid accusations of 'scientism', it should also be noted that a proper investigation of Science Norms Physicalism will largely involve a discussion in philosophy of science and epistemology rather than ontological discussions. Therefore, while Science Norms Physicalism can still have ontological implications, it is essentially an epistemological doctrine at its core, and perhaps not quite what metaphysicians specialising in philosophy of mind might be intending (see also Papineau 2009b). Again, this is explored more thoroughly in chapters four through six when discussing details about 'pro-science' methodologies become our main focus.

For now, we return to the issue of what the question was to begin with, which physicalism was supposed to be the answer to. In the broadest terms, we will phrase the question as follows:

i. Q1: What is consciousness made of?

Where this is synonymous with,

ii. Q2: Is consciousness fundamentally made up of physical or non-physical stuff?

Where 'physical' for the Scientific Physicalist is synonymous with,

iii. Physical = things which can be studied using scientific methods.

Therefore, the question as written in (ii) can be combined with (iii) to create the question:

iv. Q3: Is consciousness fundamentally made up of the things which can be studied using scientific methods?

What needs to be done next is to try and answer Q3 in a way that does not lead to question-begging, tautologies, vacuity, or other issues such as putting the cart before the horse. Ultimately, someone who claims to be a physicalist will want to answer Q3 by claiming:

v. A: Yes, consciousness is purely physical (i.e., stuff scientific methods tell us about)

However, recall that Scientific Physicalists do not have any *a priori* commitment towards certain properties or entities, and are more open-minded about where scientific methods will end up taking us. Therefore, one issue we immediately run into is that answering Q3 does not necessarily tell us the kind of answer that Q1 seemed to be looking for. While Scientific Physicalism informs us it is science we should turn to, it ends up being silent about answering what consciousness *is*.

Scientific Physicalism might at least tell us about what sorts of things we might expect, such as by referring us to current science, and making the further assumption that future science will resemble current science in all the important ways for the purpose of this current debate (Melnyk 1997). However, if we were to ask the Scientific Physicalist why they believe current science provides any good indications about what future science will look like (maybe modern science is *very* incomplete, and this is precisely why philosophers are stuck on the consciousness debate!) the Authoritarian and Science Norms Physicalists will give different

responses. The former will merely trust the authority or credibility of scientists, perhaps because science has been successful in answering a variety of other questions (e.g., rejection of vitalism, see Masi 2022). The Authoritarian Physicalist (in the way we have defined it) has nothing more to say about why we should trust science other than to appeal to its perceived authority. Under this light, Authoritarian Physicalism is an extremely unattractive position, and it may be rather ungenerous to label any existing physicalist as falling into this category. For instance, if we are feeling more generous, we might say that Ney is not truly an Authoritarian Physicalist at heart, but that she merely has not had the chance yet to spell out the methodological norms that have inclined her to make the 'oath' that: "I hereby swear to go in my ontology everywhere and only where physics leads me." (Ney 2008b, p.5). And so, we will now turn to Science Norms Physicalism to see how it fairs instead.

Unfortunately, Science Norms Physicalism also struggles to provide a justification for why they believe current science provides any good indications about what future science will look like. Unless the Science Norms Physicalist is planning to claim that current science is already following a perfect methodology for investigating everything there is to investigate about ontological reality (and this would ideally be justified through engaging with philosophy of science), any suggestion on the incompleteness of current methods risks undermining the current ontological power of modern science as well. Perhaps the Science Norms Physicalist might instead engage in working out in what ways (if at all) our current scientific norms do not meet the ideal norms and see whether this might impact how we understand consciousness. In fact, consciousness might be one candidate answer here where the shortcomings of scientific methodologies become a noticeable potential issue. For example, one might claim that science *should* include "qualitative properties [like redness]" that science does not currently tell us about, and that it *should not* be limited to "just complex physical ones of the sort described by [current] physics and chemistry" (Frankish 2016, p.4). It is much harder, therefore, for a

Science Norms Physicalist to try and support (or reject) the existence of certain ontological concepts on the basis that current science does not talk about them (or that current scientific methodology would not support them), as it is the main task of the Science Norms Physicalist to judge whether current science is suitable for this very purpose. Furthermore, it is unclear why an advocate of Science Norms Physicalism (being an epistemic thesis) would necessarily need to provide any ontological answers. Instead, they could simply leave it open that we will follow wherever future science takes us, but still avoid vacuity by at least telling us what methods this endeavour must follow.

To summarise, because Science Norms Physicalism is an epistemic thesis, it would be question-begging if it presupposed that current scientific methods just are suitable for investigating consciousness, since this is the very task that they are supposed to be investigating in the first place. The positive side is that Scientific Physicalism is able to embrace the 'futurist' horn in Hempel's Dilemma while still avoiding the issue of vacuity. The negative side is that the Scientific Physicalist must forgo its status of being an ontological thesis, and while they can still claim that their doctrine will likely have ontological implications, importantly for our purposes, it has not answered the heart of Q1 (what *is* consciousness?) which we wanted to answer in the first place. While it does answer "Q3: Is consciousness fundamentally made up of the things which can be studied using scientific methods?" but it achieves this by defining the said 'scientific methods' as things which will be able to address consciousness fundamentally, as opposed to deriving it as a conclusion. Thus, Scientific Physicalism is best understood as a theory *resembling* ontological physicalism, but it is not directly engaged in trying to tell us what consciousness *is*.

2.4 Structural Physicalism

The issue of how to define physicalism is most interesting in the context of Structural Physicalism, which is probably the version that most physicalists are likely already committed to: A version of physicalism that is defined in terms of an NFM endorsement but is *supported* via scientific evidence (as opposed to the scientific element being central to the definition, and NFM being merely supplementary). In other words, 'science' need not be a part of the definition for Structural Physicalism, although Structural Physicalists are still free to support their thesis by appealing to scientific evidence. Thus, it is 'pro-scientific' in the much weaker sense that it does take science very seriously (i.e., it does not trivialise scientific findings) but it is also not *bound* by where current or future science will take us. Because the Structural Physicalist commits to physicalism in such a way that removes its allegiance towards the sciences, it can easily avoid Hempel's Dilemma. Instead, regardless of what current or future physics might say, Structural Physicalism will be either true or false, independently of scientific methodology. (E.g., a rigorous philosophical justification or rejection of Xombies might support or reject the NFM principle regardless of whether scientific methods like experimentation or statistical tests help to justify it.)

One common feature among Structural Physicalists is the endorsement of the NFM principle, which has led some physicalists to simply understand 'physical' as meaning 'non-mental' (Montero 1999; Spurrett & Papineau 1999; Crook & Gillett 2001; Gillett & Witmer 2001; Montero & Papineau 2005; Wilson 2006; Worley 2006). However, there are many reasons why this is problematic. One reason that was covered earlier in this current chapter is that there may be ambiguous situations – such as in the case of the quantum indeterminacy issue – where it is unclear whether NFM has been undermined or not. However, a more pressing issue is that defining a concept P as simply meaning ¬Q begs the question of how Q is itself supposed to be defined, which merely brings us back to where we started. Furthermore, some have argued that other concepts can run into its own variation of Hempel's Dilemma as well

(e.g., Judisch 2008). For instance, do we define 'fundamental mentality' in the way that current metaphysicians understand the concept, or do we define it according to future completed metaphysics (Dove & Elpidorou 2021)? This claim can be applied to the definition of any terminology and highlights the inherent problematic nature of defining key terminology based on how some experts with presumed credibility happen to use that term at any given time (Firt, Hemmo, & Shenker 2022).

There are two approaches that can be taken by the Structural Physicalist. One is to commit themselves to certain *a priori* criteria of what a 'physical' entity must look like, such as by appealing to solidity or that it must be composed by atoms, and so on. However, this approach sets the bar quite high for the physicalist, as it requires a great deal of knowledge in contemporary physics to make any kind of realistic hypothesis. If the physicalist merely wishes to play middleman and convey to us what hypotheses are currently made by scientists, this will collapse into Authoritarian Physicalism, as the physicalist here lacks the expertise to make their own judgements about what the physicists are discussing. Therefore, to truly make a feasible hypothesis (of one's own) about fundamental physical reality is not an easy task to engage in. Arguably, we may wish to leave the specific hypothesis formulations to the physicists themselves.

A more approachable form of Structural Physicalism would be to make a more generic *a priori* claim that does not commit us to specific entities or structures, but it is more of an attitude of scepticism towards certain kinds of concepts – such as psychic powers, ESP, and other 'spooky' concepts. What will be helpful is if we could introduce a third concept that allows us to capture the essence of a 'no fundamental mentality' principle without using either of the words 'physical' or 'mental', which will help us to avoid issues of tautologies that were covered in section one. If this can be achieved, then we would be able to establish a definition for both

physicalism and NFM, without invoking problematic circularity, ambiguity, or tautologies in the definition itself.

The third concept that will be introduced now is the notion of 'magic', where the word 'physical' will be understood as being an 'anti-magic' mentality. In other words, it is not so much a pro-science attitude as it is an anti-esoteric kind of attitude. It is a rejection of all things fundamentally 'spooky'. To borrow terminology from the Harry Potter series, the Structural Physicalist will be understood as a physicalist who claims that everything ontological can be explained in terms of 'Muggle logic' where 'Muggle' means 'non-Harry Potter kind of magic'. (Rowling 1997). This allows the Structural Physicalist to reject the kinds of entities that are taken seriously only in fairytales or among certain esoteric or religious circles, while remaining agnostic about what specific properties or entities we accept into our fundamental ontology.

To reiterate, in section two, we defined Structural and Scientific Physicalism as follows:

(1) **Structural Physicalism** = Everything which exists are things with certain kinds of properties, dispositions, or structures (e.g., no fundamental mentality, etc.).

(2) **Scientific Physicalism** = Everything which exists are things which can be studied using the methods of the physical sciences.

To be more specific now, we can understand Structural Physicalists as being sceptical of certain kinds of properties, dispositions, or structures which we currently tend to associate with unscientific belief systems, such as psychics, supernaturalism, esoteric, or religious beliefs. The Structural Physicalist, therefore, can make a weaker version of a 'pro-scientific' claim, in that they are expressing a scepticism towards beliefs which currently tend not to be taken

seriously in scientific communities as they tend to lack empirical evidence according to current scientific methodology. This also allows the Structural Physicalist to avoid getting too caught up in the epistemic considerations about scientific methodology, or authoritarian limitations about *only* trusting wherever physics takes us. That is, the Structural Physicalist can express their scepticism for 'spooky' entities, cite empirical evidence for why we have no reason to believe in such 'spooky' concepts, and still maintain that science tends to provide us with robust methodology for investigating these ideas, all without committing to overly specific ideas from current science.

The problem now, however, is that this form of physicalism becomes an extremely narrow focus, where it is essentially a negative thesis understood as 'anti-supernaturalism' rather making any positive claims about what consciousness actually *is* fundamentally. The issue becomes clearer if we imagine for a moment that one day, Hogwarts is discovered as being real. There seems to be no *a priori* reason to rule out the possibility of starting a scientific investigation about the mechanics of how spellcasting works, such as investigating brain scans of Hogwarts students and comparing them against Muggle brains. We might even come up with a completely 'physical' (i.e., Muggle-accessible) explanation for why wizards appear to have special causal abilities, like being able to move objects at a distance by waving around a wooden stick and chanting Latin-based words. We may meet ghosts such as Nearly Headless Nick and discover that his decomposed brain particles from afar still emit neural activity that act upon him from a distance in a way that is still analysable even through Muggle investigation.

The above creates a somewhat awkward situation for the Structural Physicalist. On the one hand, the defying of non-Muggle logic and the introduction of 'spooky' concepts was *defined* as being non-physical, and so it seems that witnessing spellcasting would prove physicalism false. On the other hand, the entire situation stills seems analysable with typical Muggle concepts. If anything, the only thing that was missing prior to the discovery of

Hogwarts was that we did not have access to measuring the existence of these entities (because wizards kept it a secret from us, and so we failed to find evidence for Real Magic). But importantly, our fundamental ontological picture has not really changed. All we have learned is that some surprising causal powers turn out to exist, but just because this is foreign to our own physiology does not make it 'supernatural'. It is merely unfamiliar or surprising, not truly 'spooky' – at least once we get to understand it.

There is a quote by the science fiction writer Arthur C. Clarke, that "Any sufficiently advanced technology is indistinguishable from magic." (1962, p.250) It seems that the case of discovering Hogwarts is among this category. It only sounds 'spooky' (i.e., non-physical) at first because we are unfamiliar with how it works – but ultimately, once we become less surprised by it, it ceases to be 'spooky' in this kind of way.

The physicalist at this stage, has two options about how to proceed. One option is to continue to define 'physical = non-spooky' in which case even proving the existence of wizardry and mental causation in the form of spellcasting would not undermine physicalism. This, however, runs the risk of essentially creating an emotionally driven definition of physicalism, where 'physical' merely means 'emotionally non-shocking once we get used to it'. Presumably, the physicalist wanted to say something a little bit more than this, if they have come this far to claim 'physical = non-spooky' (or non-magical, in the non-esoteric sense).

At this stage, a physicalist might try to insist that 'physical' might still mean something to do with science (e.g., 'non-physical' = exceeds the scope of science? 'non-physical' = defies the causal closure principle? 'non-physical' = lacks anything that seems 'feely' or 'qualia-esque'?) To define non-physical as meaning that it exceeds the scope of science would require a discussion about what science is, returning us to Scientific Physicalism. Defining non-physical in its relation to the causal closure principle or qualia requires a non-tautological

construction of these concepts. We can keep delaying the issue, but if we want a clear idea of what we mean by 'physical', we will have to put our foot down at some stage and point to a concept where we can clearly state what this concept X means, then explain the relation of 'physical' entities against this concept X.

One candidate way of doing this is to dig deeper with esoteric or psychical belief systems, and articulate what might be wrong with it, why it would entail a (problematic) kind of fundamental mentality, and why we should believe everything which exists is *not* that. This will be our task in chapters four and five, using pseudo-psychics and astrology as examples.

2.5 The Tuned Deck

To summarise our discussion thus far, we have begun to see the elusiveness when attempting to wrap our heads around what physicalism is, and what question it is supposed to be the answer to. Finally, before wrapping up this chapter, we will look at why Hybrid Physicalism is not a tenable option.

To understand this, it will help to look at 'The Tuned Deck' problem discussed by Dennett (2003, p.17), though we reconstruct a variant of the same idea for our current purposes. To summarise briefly, Dennett describes a famous card trick which used to be performed by the magician Ralph Hull. The trick was said to deceive even professional magicians, despite its simplicity. A card would be lost in a deck and after a few flourishes, etc., the magician would then find the selected card. Despite the trick being performed repeatedly in front of the same audience of professional magicians, no one could figure out how the trick was performed.

As it turns out (spoiler alert), the misdirection was in the name of the trick, as Dennett explains. The name of the trick includes a presupposition that what is being performed is 'THE' Tuned Deck card trick, implying that it is the same trick being performed repeatedly. In reality,

the magician uses a different method each time, and so each time the spectators become suspicious of a certain methodology, the magician will proceed to showing another trick (using different methodology) that rules the first hypothesis implausible. Having ruled out a certain method behind 'THE' card trick, the magician proceeds to ruling out other hypotheses his audience may have been formulating – by jumping between various methods and producing the same kind of effect, thereby fooling his audience to thinking that it might be 'real magic'.

> "By *real magic* people mean miracles, thaumaturgical acts, and supernatural powers. No, I answer: Conjuring tricks, not real magic. *Real magic*, in other words, refers to the magic that is not real, while the magic that is real, that can actually be done, is *not real magic*. I suggest that many, e.g., DAVID CHALMERS has (unintentionally) perpetrated the same feat of conceptual sleight-of-hand in declaring to the world that he has discovered The Hard Problem of consciousness. It is, however, possible that what appears to be the Hard Problem is simply the large bag of tricks that constitute what CHALMERS calls the Easy Problems of Consciousness." (Dennett 2003, p. 7)

What Dennett tries to claim in this paper is that 'THE' hard problem of consciousness is much like 'THE' Tuned Deck. It is actually composed of a variety of different tricks that are packaged and sold as though it is a single problem, thereby making it essentially unsolvable. From this, he claims that consciousness is like an illusion. It is not *real magic*, but more like stage magic.

What has been argued in this chapter is, essentially that the sleight of hand happens with the word 'physical'. If classic zombies seem inconceivable, then perhaps the real issue is not so much its inconceivability, but its incomprehensibility. If different philosophical papers keep using the word 'physical' in a myriad of different ways that are mutually incompatible, then we might end up forming of confused image of what THE word means in a way that is common

or consistent across all contexts. Furthermore, the notion of 'real magic' as used in the above story seems to imply a lack of analysable causation. It refers to a situation when a card spontaneously and causelessly (aside from the magician's willpower!) appears in an impossible way. Using this impression of the notion of 'magic', Dennett tries to claim that consciousness is not THAT.

This, however, would be misrepresenting the anti-physicalist position as something that implies 'pro-impossibility'. The spectators of 'The Tuned Deck' probably did not conclude from their inability to find a solution that the trick must therefore be causally impossible. (They did observe the trick in front of them so it must have been causally possible.) Even if it turned out that psychical telekinetic powers were used (e.g., casting a spell so the magician can re-paint the face of the card to become whatever card they want it to be), again, we run into the problem that this is 'spooky' only because we did not think this kind of psychical power can exist in human beings. Whether we still find it 'spooky' just becomes an emotional sort of question about how willing or unwilling a person happens to be for accepting novel concepts.

The Hybrid Physicalist tries to have their cake and eat it too, by claiming both that (a) the methods of science are the most rigorous methods for learning about ontological reality, and (b) the methods of science point towards Muggle logic (of the 21st century). Superficially, it might sound as though (a) and (b) will always go hand-in-hand, but this will just depend on how liberal we are willing to be with what we take to be 'scientific' and how easily we happen to *feel* spooked.

Ideally, the physicalist will want a definition that is not dependent on emotional shock or naïve preconceptions on what we think 'science' is. Thus, if we wish to analyse "Q1: What is consciousness fundamentally?" it is not sufficient to rephrase this as "Q2: Is consciousness fundamentally made up of physical or non-physical stuff?" or "Q3: Is consciousness

fundamentally made up of the things which can be studied using scientific methods?" What is required instead is to make it much clearer what kind of answer it is that we are looking for. This is what we turn to next, to dig deeper into the concept of both physics and stage magic.

Chapter 3. Misconceptions in Physics

The previous two chapters have highlighted the ambiguity in the word 'physical'. One potential solution to this problem would simply be to use 'physical' in whatever way physicists tell us to use it. If it turns out that the word 'physical' can be used in a way that satisfies (a) structural considerations (e.g., No Fundamental Mentality) that encapsulates a physicalist attitude, but also (b) is consistent with the way that scientists use the word, then we might think that we might still be able to salvage a way of making Hybrid Physicalism work, despite the lack of clear unity in the question it is trying to answer. This chapter will argue that this still does not work out in the physicalist's favour, as there is a gap between the way physicists use the word 'physical' and the way that Descartes conceptualised the mental-material dichotomy which makes it relevant towards the mind-body debate (Ostenfeld 2018; Wheeler 2005).

The recurring central problem we encounter once again is that, as Dennett (2003) has argued, there is an inherent problem with understanding 'THE' hard problem of consciousness: It is unclear if there really is just one single problem that can be addressed. As described in the previous chapter, Dennett has likened the mind-body debate to a card trick called 'THE' Tuned Deck where a magician creates an unsolvable mystery just by creating a different problem every time and passing it off as 'THE' same trick each time. Similar to this, when philosophers debate consciousness, it is unclear whether we are all on the same page about what it is that we are trying to explain. The remnants of the Cartesian language about substances imply that we are concerned with what types of (ethereal-ish/non-ethereal-ish) structures exist in the world. But at the same time, in Descartes's time when clockwork automata were considered cutting-edge scientific technology, with a growing sense of scientific authority during the time of the enlightenment, perhaps Descartes's original fascination was in part motivated by wh*ich authority* we should be listening to – religion versus science (Wheeler 2005, Ch2). Was Descartes's original question really about investigating the structural ontology of the world, or

was he asking a more sociologically/politically motivated question about which type of institution we should affiliate with (churches versus scientists) when seeking out knowledge?

Whatever the original Cartesian question was grounded in, to say the least, the original curiosities would have been grounded in physics from four hundred years ago which is now clearly outdated. There is a separate question of whether there is still anything to be salvaged given the advancements in physics that we have already seen since then, and to debate this, we need to engage in clarifying some common misconceptions about physics today. This chapter will attempt to set the context for how physicists tend to use the word 'physicality', and why this falls short for the context of the *meta*physical debate that we are trying to apply it to.

In section one, we begin by examining the debate between physicist Sean Carroll and Buddhist scholar Alan Wallace, where Carroll makes several very confident assertions citing himself in his own capacity as a physicist (Carroll & Wallace 2022; see also Carroll 2016), and Wallace's response where he cites several other physicists and mathematicians who directly contradict Carroll's assertions (Linde 2005, 1990; Zeilinger 2004; Dyson 2004; Wheeler 1996; Heisenberg 1989; Penrose 1989; Planck 1944). In section two, this is further supplemented by referencing a few more physicists and mathematicians (Tong 2017; Lancaster & Blundell 2014; Squires 1990) to examine where the source of confusion seems to be arising from. What we will see here is that the kind of description of 'physical' reality that is often given at school-level, or in physics books directed at non-physicists, is actually quite misleading due to its oversimplified and rather inaccurate nature. This thereby causes misunderstandings, as this level of superficial explanation of physics is insufficiently detailed for our current metaphysical purposes. In section three, we will reiterate what the more accurate description of 'physical' reality as described by modern physicists essentially amounts to, and we will see that it is unhelpful for the purposes of supporting Hybrid Physicalism. Finally, it will be concluded that physicalism is an untenable position because it tries to balance a misunderstanding about

physics with a pro-physics attitude, which is logically incompatible. In fact, Carroll himself endorses a kind of 'poetic naturalism' sort of attitude, which he loosely explains as a combination of (a) a pro-scientific claim, but also (b) that there is something poetic and wonderous about the natural world, and that the pro-science attitude does not have to lead to dry disenchantment about romance or *meaning* that human beings can find in the world (Carroll & Wallace 2022, p.33-34; also Carroll 2016; Crease 2016). This will lead into the next chapter, where we introduce Magicalism as an alternative to physicalism: that is, the novel claim that consciousness is fundamentally Magical – like stage magic – and that this fundamental sense of awe and wonder can be captured *without* over-reducing it to physics nor supernaturalism.

3.1 Carroll and Wallace's Debate

Recently, there has been a debate between physicist Sean Carroll and Buddhist scholar Alan Wallace on discussing the nature of reality (Carroll & Wallace 2022). In this discourse, broadly summarised, Carroll advocates a position where he argues for a layered interpretation of the world, with the physical layer at the bottom, a computational level in the middle, and a more everyday kind of understanding at the top. Using iPhones as an exemplar, he discusses that most fundamentally, the phone is made of fundamental particles; that the middle layer refers to the coding used in the software; and the highest layer refers to the way in which we human beings understand how to use the apps and other functions offered by the phone (*ibid,* p. 27-35). In response, Wallace argues there is no consensus among scientists that the layer of physical particles is the bottom layer philosophically speaking, and this is open to interpretation. Mathematicians, for instance, may claim that *mathematics* is the bottom layer (he cites Penrose 1989 as an example). Alternatively, *meaningful information* could be more fundamental than 'physical' things like matter or energy, where physical things reduce to meaning, not vice versa

(he cites physicist Zeilinger 2004). Finally, if we are willing to say, "that information science is the most fundamental of all sciences" (Carroll & Wallace 2022, p.40), then we might as well go a step further and say *consciousness itself* is the most fundamental layer of reality (physicist Linde 2005, 1990 is cited) (Carroll & Wallace 2022, p.35-45).

Carroll objects in return that he simply does not believe consciousness is the fundamental layer of reality. He claims that due to what physicists call the core theory, we have too good of an understanding already of physical reality at the brain level, and we already know that this cannot be what is at the most bottom layer:

> "Nobel laureate Frank Wilczek has called [it] the core theory, the theory that takes into account these elementary particles, and the forces of nature, including gravity, and electromagnetism and so forth, and my claim is that we understand that one layer. That doesn't mean we understand the bottom layer of reality." (*ibid,* p.32).

He goes on to assert that while there is no consensual understanding amongst scientists about what that bottom layer of reality looks like, he claims we do already know the gist of activities at the level of neurons (i.e., the activity of everyday sizes). He tells us that current physics does have an immensely good understanding of what is involved at this level, to the point where we cannot feasibly expect to learn of the existence of any new (relevant) forces or particles.

> "We understand certain elementary particles, in a certain regime of what they can possibly do. It's just that that regime includes everything that you and I ever do, and everything that we see around us in our everyday lives. The way that I say it is the laws of physics underlying our everyday lives are completely known." (*ibid*, p.33).

While admitting that it is a "a very grandiose, bombastic-sounding claim" (*ibid*), he goes on to essentially justify that when he says we 'understand' neural activity in this 'physical' sense, this does not mean that we understand much at all about higher-order matters. He gives an example of going on a blind date and being asked "Well, tell me about yourself." (*ibid*) to which responding to this by listing all the known fundamental forces and particles would be an inappropriate reply (though he adds as a joke that if you are dating a scientist or someone with a reductive attitude, then giving this reply anyway "might be kind of sexy", *ibid*). Essentially, he concludes that while we do have an extremely high understanding of all the 'physical' factors at the everyday level which can feasibly affect neurons, there is still much left to understand. Of course – now we return to the issue of how this affects metaphysical discussion.

The issue that is relevant for us while doing metaphysics is to ensure that his usage of the word 'physical' is compatible with other instances when this word is used by metaphysicians, such as in the causal closure principle of the physical. Basically, the thing we are trying to avoid is committing the equivocation fallacy, when the same word is used differently in two separate premises and is used to derive a logically invalid conclusion. We would not want a situation where philosophers are quoting physicists who claim that "the laws of physics underlying our everyday lives are completely known." (*ibid*, p.33), taking this out of context, and using it as a premise for further arguments in a way that misrepresents the notion of 'laws of physics' to mean something different from the original quote, thus leading us to logically invalid philosophical conclusions.

In response to Carroll's "very grandiose, bombastic-sounding claim" (his own quote, *ibid*), Wallace returns this with a story: An example taught by Buddha, of the blind men and the elephant. He begins by building upon the claim made by physicist Marcelo Gleiser who claims "we are essentially blind to what exists at the very core of physical reality. All we have

is our measurements, and they give an incomplete picture of what's really going on." (Gleiser 2011, quoted in Carroll & Wallace 2022, p.41). In Wallace's Buddhist story, we are asked to imagine multiple blind men who are presented with an elephant and are asked if they can work out what it is that has been presented in front of them. Each man feels a different part of the elephant – its body, its foot, its back, its tail, its trunk, etc. – and the men begin to form their own theories about what they think is in front of them. They come up with different incompatible theories from one another based on the features that they happened to observe. Each blind man arrives at different conclusions, believing the others to be misguided. Wallace argues that this is itself an *empirical* exercise which is similar to how the sciences work. We are faced with a phenomenon, and we ask ourselves, "What is that phenomenon?" or "How does it work?" and run various tests while we check the validity of our hypotheses. Importantly, we are never directly reading facts off of reality. It is not as though brain regions come with convenient labels when we run brain scans about what those brain regions do, and clever methods must be devised by the scientists to feel the elephant, as it were, to deduce what the concept is that they are trying to observe. Empirical conclusions are always based in inference. Once again, this is why it is both conceivable and possible for Xombies to exist who have the same brain structures as far as observation can tell, but *if* the brain region had come with convenient labels that we could have just read off of, then we could simply read that Xombie brain regions have slightly different functions. Namely, that it lacks the ability for consciousness. There is a sense in which observation is a superficial means of gaining insights about how the world really works, but in many cases, it is still the best means of inference that we can manage. The question is whether this is true in the case of consciousness as well.

There is a similarity between the activity of the blind men trying to figure out the elephant in front of him, and the activity of a scientist trying to figure out the workings of neural mechanisms. The similarity is that we are confronted with the unknown where we cannot

simply just 'look' to obtain an answer, and we need to get clever about how to 'dig deeper'. If we also want to add the claim that 'digging deeper' = providing a more 'fundamental' explanation, then we are *defining* the word 'fundamental' to just mean 'whatever theories are explanatorily deepest'. However, what explanation we take to be 'deepest' or 'most satisfying' will depend on the kind of interests of the scholar reading into that set of explanations. This contains somewhat of a risk that a physicist and philosopher may disagree on what is the most 'fundamental' description, assuming we can accept the definition of 'fundamental' that:

Fundamental $_{(A)}$= digging deeper into the causal story.

Where it may turn out that what is most ontologically fundamental is not necessarily what is most fundamental according to a physicist's interests (i.e., 'physically' fundamental…?) Furthermore, a neuroscientist might feel that a neural explanation of consciousness just *is* most fundamental, but in virtue of that being the most interesting story *to them*. Similarly, physics enthusiast might say that a physics-y explanation is 'most fundamental'. But whether a particular kind of scholarly exercise's interests gain authority may very well be a sociological issue, and once again, this is something we will need to be cautious about. That is, if neuroscientists respect physics as 'THE' most fundamental science – and this is a matter of contextless assumption even when physics does not provide the deepest causal explanation suited for that purpose – then of course 'physical' explanations will be most fundamental. But again, this is only because we will have defined physics as the most fundamental science, in which case now we are using the word 'fundamental' in a different way:

Fundamental $_{(B)}$ = the realms of physics, as opposed to chemistry or biology.

The *meta*physical question is whether the definitions (A) and (B) align for the purposes of the mind-body debate. Now, what is important here is that if we define 'fundamental' in a way that does not specify what it means to 'dig deeper' into a concept, then we risk this concept

becoming overly dependent upon what a particular individual happens to be prone to attending to. A physics enthusiast might insist that 'physical' reality is 'most fundamental' while also allowing that other higher-level explanations can be useful in their own right; whereas a dualist might insist that if other higher-level explanations indeed are more useful for explaining consciousness, that *just is* what it means for something to be fundamental$_{(A)}$.

Wallace explains that problematically, the response which is most fundamental$_{(A)}$ depends on the kind of story that we are trying to tell. He then discusses that an essential feature of Buddhism is to acknowledge that our own assumptions, beliefs, and axioms are inherently shaped by our interests, and the places that we tend to focus our attention. While we might think that our own assumptions and beliefs are objectively fundamental because we believe we have obtained the information from a credible source, there is always a sense in which we are trapped inside the stories we tell ourselves (*ibid*, p.36). The stories that we tell ourselves are determined by the places where our attention is naturally drawn to, and the stories that we do not tell ourselves become blind spots to us, since by definition we are not interested enough in the subject to engage in any storytelling about those topics to begin with. If we mistake our own 'stories' for being 'facts', then this is what leads us to endorsing an illusory and biased image of 'fundamental reality' that ends up being in favour of our own personal interests. Therefore, Wallace claims that it is unsurprising a physicist such as Carroll will be drawn to explanations that imply a *mechanistic* analysis of consciousness. "This is the kind of assertion one would expect from a physicist. This is one of the many stories we tell ourselves about the nature of reality. But it's not the only story, not even the only scientific story. Our notions of what underlies what are based on our starting assumptions, the questions we pose, the focus of our attention, and the conceptual framework in which we make sense of our observations." (*ibid,* p.37) *For Carroll*, the story of mechanistic causation is most fundamental, but only because he has decided physics is the place where interesting 'deeper' explanations happen.

But this is not universalizable, and from Wallace's Buddhist perspective, the 'physical' explanation is not the most interesting nor deepest analysis to understand consciousness.

Interestingly, Carroll does not appear to disagree with this. Aside from their differing usages of the word 'fundamental' (roughly A versus B as described a couple pages earlier), they do seem to be agreed upon the general point that the 'story' of physical reality is just a story, albeit a *very good* story. To use Carroll's language, physics proposes a set of 'models' or representations that we have come up with, which can help us to understand the mechanisms governing our world. In a separate paper, Carroll claims:

> "The standard model [of particle physics] is a partial description, a representation that captures some aspects of how reality behaves, and only an approximate representation at that – extremely accurate within a certain domain, but completely inapplicable in others. What I am a realist about is reality, by which I mean the totality of the physical universe. The standard model of particle physics, like general relativity or Newtonian mechanics, provide useful ways of talking about reality in certain circumstances, but I would not describe them as fundamentally "real."" (Carroll https://philarchive.org/rec/CARRRK, year unspecified, p.1).

The paper is titled "Reality Realism" and he argues that he is not a 'realist' about the existence of entities as described within specific models of physics (particles, waves, etc.,), but rather, that he is a realist about reality as a whole. He believes that something exists rather than nothing (this is not very controversial), and beyond that, we can just let the universe *be* (*ibid*), essentially just focusing our epistemic efforts on questions that are empirically or theoretically evaluable. There is, of course, a further question about whether non-mechanistic questions ever are empirically or theoretically evaluable. Perhaps there is a sense in which we can only ever feel the elephant and describe it in terms of its functions or dispositions, since that is the only

parts which we directly come in contact with. The word 'elephant' is just a name or label that we have given to group a variety of measurable dispositions together – and maybe it can be argued that consciousness is similarly just a set of mechanistic functions. But to say the least, if Xombies are conceivable and possible, we can safely safe that the ability or disposition to *feel* pain, that is the sort of thing that the consciousness debate is trying to talk about. If a psychopathic physicist were to claim that we are merely being stupid for caring about the wellbeing of our family-members because they are made of nothing more than indifferent particles and consciousness is just an illusion, something would be getting missed.

Now, it is important to clarify that Carroll does not reject the existence of particles or waves and other physics-y concepts, and he is, as ever before, confident that the kind of entities as discussed by physics are real and that they do exist. But rather, we can understand his claims in the same way that the blind men might understand the elephant. Physicists know they are examining something which is real, and they can give it all sorts of labels or descriptions to capture the traits or properties of that entity. But at the end of the day, what they are describing is a *model* of that entity. They are making blind inferences about a set of observable traits, and the 'elephant' just is a collection of such inferences.

Oddly, if we put Carroll's various assertions together, it becomes somewhat unclear whether he really is a physicalist in the philosophical sense, although he does claim to be one. He describes himself as a poetic naturalist: that science is important, but also, the poetic wonders that provide romance in meaning to human existence are important too (Carroll & Wallace 2022, p.33-34; Carroll 2016; Crease 2016). He is also not a realist about the kind of entities which physicalists would claim consciousness reduces to. It is all just 'models' or representations that we use to understand mechanistic questions. He does not claim that everything reduces to mechanistic analyses, nor does he claim that everything reduces to the kind of entities that he turns out not to be truly a realist about (i.e., things described in models).

Perhaps what he is, is more of a *meta*physics-sceptic: That he does not believe there is more to ontology than what is directly measurable or inferable through scientific examination. He seems to be against being realist about purely abstract *meta*-physical entities (including 'physical' entities in the *meta*-physical sense), and that abstract concepts such as these do not warrant being taken so seriously. But at this point, we are digging deeper into his intentions than what he has explicitly stated, and so these discussions should be left for another occasion to be addressed once after Carroll has had the chance to respond to the discussions in this chapter. In any case, what is important for our current purposes is to see that even when a physicist makes very strong claims about how we already understand everything that happens at the neural level, it still does not help physicalism in the *meta*physical sense. We need to be cautious about distinguishing 'physicalists' in the sense of being sceptical of supernatural or dualist entities, and 'physicalists' in the sense of being sceptical of non-physics including philosophy and metaphysics. In the next section, we will see that this philosophical conundrum runs even deeper than what the Carroll and Wallace debate already started to highlight.

3.2 David Tong's Lecture

The word 'particle' is especially prone to causing misunderstandings, and the usage of this word in contemporary physics requires further clarification, so we can get to the bottom of where the confusion about 'the physical' seems to be coming from. If we are philosophising about the mind under a naïve impression that the 'particle' can be used interchangeably with the idea of 'the small building blocks of the world', then this would be a usage of the word 'particle' that is not consistent with the way that physicists use the term. For instance, Carroll clarifies that "when we say "particles" in physics, what we really mean are vibrations in these quantum fields." (Carroll & Wallace 2022, p.32); physicist Tom Lancaster states 'particles' are

defined as “excitations in a quantum field” (Lancaster 2019, p.275); and in a lecture on “What is the universe made of?” by physicist David Tong, it is explained that a quantum field can be understood as an abstract idea of a “fluid-like substances” that are “threaded throughout space” (Tong 2017, 10:00-12:30). While a more detailed and technical description of fields can be found in physics books (e.g., Lancaster & Blundell 2014), we focus on discussing elements from Tong’s lecture (2017), as the ideas here are explained in such a way that makes it more accessible to non-physicists, but at the same time is explicitly targeted at correcting the common misconceptions that take place in lectures made accessible to non-physicists.

In the lecture, Cambridge physicist David Tong begins by describing the ‘atomist’ model as first proposed by Democritus in ancient Greece. This is the idea that the world is constructed of Lego-like building blocks, and that once we understand all the different types of building blocks which exist, we can then understand how these different types combine to create further lager structures – but fundamentally, they can all be reduced to the smallest ‘Lego’ unit. In the first ten minutes of the lecture, he describes the periodic table of elements which contain all atoms that exist in the world, and as school-level chemistry teaches us, every object that we see are composed out of a combination of those atoms. Then, at the end of the 1800’s, physicist JJ Thompson discovered that there are particles which are smaller than the atom, which were later referred to as electrons and a nucleus that are further building blocks of all atoms. Subsequently, it was discovered that the nucleus can be broken down further into protons and neutrons, and then in the 1970’s, physicists discovered that inside these are smaller particles called quarks. So in essence, this would seem to suggest that “there are three particles [electrons, up quarks, and down quarks] in which everything we know is made of. […] Everything we see in the world, all the diversity in the natural world, you, me, everything around us, is just the same three particles with slightly different rearrangements repeated over and over and over again.” (Tong 2017, 7:30-8:30). This suggests that there is no room for a fundamental ‘soul’

existing over and above the three fundamental particles, creating the illusion that 'physicalism' is quite easy to grasp (i.e., physicalism = the view that everything which exists are the things made up of these three particles) and that fundamental mentality is out of the question.

Tong explains that while these are not the same fundamental building blocks as originally described by the Greeks, "the spirit of the issue" (8:30) has remained unchanged in that it suggests the imagery of "Lego brick from which everything in the world is constructed" (8:50). After spending ten minutes describing this model that we learn in schools and take for granted, which *seems* to describe the nature of particle physics, Tong goes on to correct where the common error happens:

> "It's a very nice picture. It's a very comforting picture. It's a picture that we teach kids at school. It's a picture we even teach at undergraduate university. And there's a problem with it. The problem is, it's a lie. It's a white lie that we tell our children because we don't want to expose them to the difficult and horrible truths too early on. It makes it easier to learn if you believe that these particles are the fundamental building blocks of the universe, but it's simply not true. The best theories that we have of physics do not have underlying them the electron particle and the two quark particles. In fact, the very best theories we have of physics don't rely on particles at all. The best theories we have tell us that the fundamental building blocks of nature are not particles, but something much more nebulous and abstract. The fundamental building blocks of nature are fluid-like substances which are spread throughout the entire universe and ripple in strange and interesting ways – we call them fields." (Tong 2017, 9:00-10:20).

The Authoritarian Physicalist described in chapter two is now in trouble. If we are making an oath to following wherever physics takes us, then contemporary physics already tells us that

‘particles’ are defined as ‘excitations of a quantum field’ where fields are considered macro-entities that permeate the entire universe. This leads to the peculiar conclusion that following a micro-reductionist attitude entails having to embrace that the micro-building blocks fundamentally reduce to the largest kind of macro-entity (spreading throughout the entire world), which of course, violates the very micro-reductionist attitude that is the spirit of such analysis in the first place. There is a conflict of interest between (a) following this micro-reductionist attitude of reducing entities to smaller and smaller components, and (b) following contemporary physicists. This may be a perplexing claim for readers who are not familiar with the kinds of discussions explained in the quote above and this may come as a surprise, so we will explore a deeper explanation below to add clarity.

Talking about ‘particles’ is merely a shorthand, or a model used by scientists, either when this model is sufficiently helpful for the purpose it is being applied to, or when introducing newcomers to physics in early stages of their education. It is a useful ‘model’ in that it can help with analysing most causal mechanisms that we deal with at the everyday level, such as for engineering purposes, or solving physics/chemistry problems that affect entities of easily human-accessible sizes (Tong 2017). Therefore, if non-physicists wish to think about particles in a way that resembles Lego-like building blocks, for most (non-philosophical) purposes, this is sufficiently accurate. However, problems occur when we evaluate a physicist’s claim (e.g., Carroll) in the context of a *meta*physical debate, where the philosophers may have an incorrect physical theory in mind, while the physicist continues to use more accessible and user-friendly (but misleading) terminology, exacerbating the source of confusion.

To understand the miscommunication, it is important that we correct our mental image about how we imagine a ‘field’ which Tong explains as being the fundamental building block (if we wish to insist upon using the term ‘building block’ – though as we will see shortly, this is already misleading). Tong begins explaining this concept by describing:

> "If you're a physicist, you have a very different picture in mind [than laypeople] when you think about fields. I'll tell you the general definition of a field and we'll go through some examples so you can get familiar with this. The physicist's definition of a field is the following: It is something that is spread everywhere throughout the universe, it's something that takes a particular value at every point in space, and what's more, that value can change in time. A good picture to have in mind is a fluid, which ripples and sways throughout the universe." (*ibid,* 10:30-11:00)

He further describes that the notion of a field was originally inspired by the intuitions of Michael Faraday when discovering the idea that we now refer to as the electromagnetic field. The intuition he began with was an abstract idea of a single sheet or continuous fluid-like entity that "threaded everywhere throughout space" allowing an event taking place at Location A to influence another subsequent event at another far away Location B. Tong gives an example of a long-distance phone call as using this mechanism. Another example he gives is the idea of playing around with magnets, and as we all know, if you move two magnets closer together, you can feel them pushing or pulling together as the force acts upon two objects that are not making direct contact. He explains that "It's one of the most revolutionary abstract ideas in the history of science" (12:30) which not only improved our understanding of 'physical' causation, but moreover, that it corrected our intuitions about the spirit of atomism which was prevalent prior to Faraday's discovery since the time of the Greeks. In essence, it is a shift in mental imagery, going from a Lego-like worldview to a more abstract one of models, ideas, and intuitions that give us a more accurate 'story' about physics.

Once this leap is made, however, now the Structural Physicalist is in just as much trouble as the Authoritarian Physicalist. While we were still allowed to understand physicist's claims

as referring to concrete objects as if it is a Lego-like building block – which was still the mental imagery accepted during Descartes's time – then Structural Physicalism would be easy to grasp. 'Physical' entities are those things that are reducible to the Lego-like imagery, and if anything 'floats' over and above it, like a fog hanging over Legos, then this would imply dualism is true. However, problematically for the Structural Physicalist, once we accept that not all physical entities have to be Lego-like, and that fields are themselves sort of like fluid fog-like concepts that spread everywhere, suddenly 'dualism' starts to look oddly specific. Pluralism might still be feasible in that we can accept multiple different kinds of fogs, liquids, Legos, and so on to enter our physical worldview for the contexts in which proposing such structures are helpful, but now creating a distinction between only *two* entities begins to look rather arbitrary. Even physicists are not committed to 'physical' structures having a certain kind of property *a priori*. The late Durham Mathematician Euan Squires, who wrote much on quantum physics and its relation to philosophy of mind, argued that technically "matter" and "spirit" are like different basic elements (Squires 1990, p.221). He preferred to call this "functional dualism" (*ibid,* p.60). From a physics-informed perspective, it is unclear whether there is any *a priori* distinction between properties which 'physical' entities must have or must lack. The main difference is attitudinal or maybe epistemic rigour, where physicalism implies scepticism of spiritual or faith-based concepts, and dualism does not. If 'dualism' does not imply a major exception to physics, as being like a foggy liquid substance 'floating' over and above Lego blocks, then splitting the world into *two* components looks unnecessarily specific. Why not accept either monism (i.e., reality is unified) or pluralism (i.e., there are many different components to reality)? Without a Lego model, what do we mean when we discuss 'non-physical' structures?

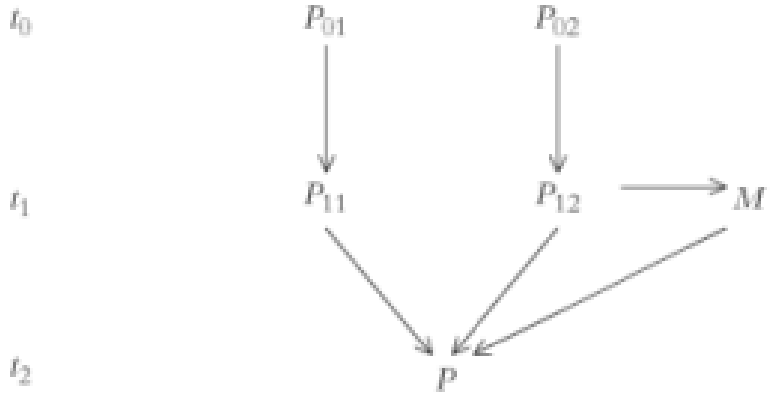
We can imagine, for instance, being in Faraday's shoes when atomists models were taken for granted, and we are trying to explain what is going on when we push two magnets together:

> "[When playing with magnets] It's just a little bit Magical. There's just something special about this weird feeling you get between magnets. And this was Faraday's genius – to appreciate that even though you can't see anything in the space between the magnets and it seems to be empty, [we have reason to believe] there is nonetheless something real there" (Tong 2017, 10:30-11:00)

If Faraday had been a Structural Physicalist who presupposes that 'physical' is defined as an atomist model (as was accepted by physicists at the time), then upon discovering how magnets do not fit into this model, he would either have had to abandon physicalism in favour of magnetic dualism, or alternatively, would have had to conclude that because magnetic interactions cannot be accounted for by 'physical' models, magnetic forces must not exist! It is merely an illusion.

Instead, Faraday takes a more empiricist attitude. There is a phenomenon in front of him which he thinks is worth explaining, and so he constructs a novel scientific model to propose an abstract concept that helps to account for the thing which he wanted to explain. This is not 'dualism' as scientific models are not set in stone. If there is good theoretical reason for adjusting the existing models, then making such adjustments still falls within the scope of physics. This means magnets are still 'physical' entities despite not falling into the accepted model of his time.

The same reasoning could be applied to the consciousness debate. Dualist philosopher E.J. Lowe, for example, described the possibility of invisible mental causation, and illustrates this in the following diagram (Lowe 2008b, p.71, Figure 3.1):



The way to read this diagram is to see that at time t0, physical events P01 and P02 occur simultaneously (such as neural firings). Then, at time t1, each of these in parallel cause physical events P11 and P12. Subsequently, P12 causes mental activity M (the kind of activity that would be lacking in a Xombie universe), and it is the conjunction of P11+P12+M which is required to create the event P at later time t2. What Lowe argues is that while M is necessitated by P02->P12 in *our* world, the laws of nature could have been different in such a way that if M did not occur (such as in a Xombified world), then maybe event P also would not have occurred. The moral of the story is something to be aware about. What Lowe is arguing here is that there are some 'laws of nature' which may be invisible to the empirical methods of physics. As far as physicists are concerned, if M always follows P12 in our world (i.e., where all our empirical observations are happening), then it is causally redundant. The existence of M only becomes important if we could start constructing a cross-world scientific study, since then (and only then) would it be empirically testable that M truly does play a causal role. To summarise more simply, Lowe's claim is that some details about the *meta*physical laws of nature can be omitted from the study of physics to keep it focused on empirically testable questions. If a Xombie world can exist, then the world might be philosophically accessible through thought experiments and logical theorising, but not empirically accessible

through the kinds of tests which physicists use. There is no reason to believe that physics is a completed picture of the laws of nature, even if physics gets the authority in the laws of physics.

What we have seen is that there are two distinct tasks which can be interpreted as being 'THE' hard problem of consciousness. One is the 'physical' question about whether we need to propose more fields, particles, forces, or other entities in order to fully understand consciousness. The other one is the '*meta*-physical' question about whether things discussed in the field of physics is what consciousness should be reduced to at all. These are two separate questions. The first question might best be left to the physicists to handle, since it requires an immense amount of technical knowledge in physics itself to be able to comment at all on how to improve the discipline in any drastic way. The second question is a *meta*-question about whether discussing consciousness really belongs to the domain of physics in the first place. This is the one which the mind-body problem is supposed to be interested in, as the former question would exceed the expertise of most philosophers (and is more of a physics question than it is a philosophical question anyway). On top of this, also consider how the hard sciences typically do not tend to focus ono evaluating emotions, passions, meaning of life, ethics, or other very 'human' concerns. The hard sciences tend to focus on content that are more matters of emotional 'indifferents' for conscious life to flourish. For these reasons, we might even say it's quite likely that physics is the wrong place to look for a fundamental(A) or *deepest* explanation when it comes to understanding the nature of consciousness.

To give an example of this attitude in the philosophical literature, Esfeld (2021) has recently proposed that perhaps we should distinguish between causal functionalism and normative functionalism and accept a dualist ontology in this sense. Causal functionalism loosely refers to the kind of mechanistic explanations that physics and other modern sciences tend to provide pertaining to the causal workings of 'indifferent' variables, whereas normative functionalism would evaluate the functionalities of feeling creatures, i.e., the functions of *non-*

Xombies. This is not so much an issue about fitting consciousness into the world of physics, but rather, it is a more '*meta'* kind of question about how we should handle the idea of consciousness. If we want to grasp its essence, then what sort of explanation will we need to try and dig deeper at its core?

Now, the above reasoning might sound consistent with higher-level theories where it is accepted that even when we say something technically reduces to the physical, it does not always mean that physical explanations are the most appropriate level of explanation. A reductive physicalist might, for example, argue that consciousness is fundamentally physical, and also accept that psychology or neuroscience are better suited than physics for evaluating the functions of consciousness. However, in response to this type of claim, we can recall the confusion about how to define 'fundamentality' as was mentioned earlier. If the word 'fundamental' just means 'digging deeper', then it becomes confusing why we should insist that the physical layer is the most 'bottom' layer. A Structural Physicalist might get away with this if this can at least tell us what properties or dispositions 'physical' entities must have (or lack) and then justify how consciousness fits into this. However, seeing as even contemporary professional physicists struggle with giving us a unified sense of properties/dispositions that all 'physical' entities have or lack (Carroll & Wallace 2022), this task may just be too difficult for philosophers to engage in directly. If the physicalist really wants to insist that consciousness is fundamentally 'physical', it is quite possible that what is motivating them is not so much structural considerations, but rather, it is a scepticism against the lack of epistemic rigorous commonly associated with spiritualism. This will be addressed more in chapter five when we take a closer look at philosophy of science and the problems with pseudo-psychic spirituality.

Moving forward with this thesis, what is important to remember in the meantime is that so long as we are attempting to provide an answer to the *meta*-physical question about "What is consciousness?" we cannot presuppose that the answers can be found in physics, since that

is the very question that we are being asked to evaluate in the *meta*-physical context. Therefore, to advocate physicalism or anti-physicalism, it needs to be made clear what we mean by the word 'physical'; what sort of question we are answering; and what it is we are trying to explain.

3.3 Why Hybrid Physicalism is Untenable

Hybrid Physicalism is a category introduced in this thesis in chapter two, and it is the conjunction of (a) a pro-science attitude, and (b) *a priori* structural considerations such as a commitment to the No Fundamental Mentality principle. The biggest issue here is that with (a), contemporary physics seems to tell us that the smallest entities like particles actually 'reduce' to the largest entities like fields that permeate the entire universe. This makes it rather confusing what we mean when we say that X 'reduces' to Y, or that X is 'more fundamental' than Y, and at the very least, a physicalist would need to clarify what specific position they are arguing for in regard to this potential source of confusion. However, at the same time, supporting (b) the No Fundamental Mentality principle, seems to imply a commitment towards no 'mental' stuff at either the smallest level (particles) or at the largest level (fields). To clarify what is meant by this, the physicalist would need to explain what they mean by (i) fundamental, (ii) 'mental' stuff, and (iii) why we should believe that 'mental' stuff does not exist 'fundamentally', and (iv) why this is *meta*-physically relevant or interesting. In the same way that magnetic stuff can exist 'fundamentally' in a way that integrates this concept with physics, it is not immediately clear why the same cannot happen for 'mental' stuff. Not only this, but it is also not obvious why this is philosophically or *meta*physically interesting. Back in Descartes's time, when physics was much simpler, perhaps there was not much distinction between discussing physics and metaphysics, but this is not true in our contemporary times. Explanations will need to be provided for why this is a matter of interest for philosophers (who lack expertise in physics).

Hybrid Physicalism is untenable, largely because there is not only an abundance of ambiguity with crucial terminology, but also, it is unclear what question we are trying to answer. For those philosophers who are addressing the historically interesting question of how the Cartesian problem has evolved (or devolved) after four hundred years of scientific progress, the mind-body problem turns into a task of correcting outdated Cartesian assumptions. But for those philosophers who have an issue with the semantic usage of ambiguous words – like Cartesian theatre, qualia, phenomenal consciousness, or even 'physical' in Hempel's Dilemma – the mind-body problem turns into a semantic debate about debunking problematic usage of words. Finally, for those philosophers who are addressing the insufficiencies in scientific research methods for investigating consciousness (such as with the purpose of developing the cognitive sciences), the mind-body problem may transpose into more of an epistemic or methodological task that is more in the realms of philosophy of science – which is what we turn to addressing in the second half of this thesis.

Like Dennett's (2003) Tuned Deck example, 'Hybrid Physicalism' seems to be an attempt at trying to provide a single answer to a collection of different questions. It would be great if there is a convenient label we can place on a complex concept that sums up the answer to a series of questions in a single breath, and 'physicalism' appears to attempt just that by stating simply that consciousness just is physical; *everything* just is physical; end of story. It is no wonder that a theory which tries to be generic and all-encompassing will also struggle with issues about vacuity, as we have seen in chapters one and two. It is the equivalent of Tuned Deck situation, where a magician shows us a card trick and challenges us to explain how the trick was done – and we reply by declaring that "You must have used sleight of hand rather than Real Magic!" To this, the magician may reply: "Yes, that part wasn't meant to be a secret. What I'm asking you is can you tell me more specifically how I did the trick?" Just like Hybrid Physicalism, the initial response is too generic and vacuous to answer any question properly.

So now, the remaining question that we address throughout the rest of this thesis is about understanding how to go about getting to the heart of what consciousness is without it becoming too generic, too vacuous, too much about methodology (which is an epistemic discussion), and a variety of other problems that we have so far encountered. The solution that will be offered in this thesis is to leave some sense of Magic in the theory itself, rather than trying to explain *everything* all at once. Magicalism will be the idea that consciousness just is Magical – in the same sense as stage magic. When we try to explain magic tricks in an all-encompassing and methodological way, this 'ruins' the Magic, explaining away the very concept that we were trying to address in the first place. The solution is to simply allow the Magic to remain where it is required. This concept of Magic is what the next chapter turns to.

4. Magicalism: Do You Believe in Magic?

Previous chapters have discussed that there are different ways of interpreting what Chalmers (1996) has referred to as the hard problem of consciousness, making it difficult to pin down what (if any) question is to be considered 'THE' hard problem (Dennett 2003). Roughly, the mind-body debate is still often interpreted in light of a Cartesian nuance of there being a physical/mental dichotomy, where for Descartes, the conception of physical substances involved extension, space, or mechanical motion, and mental substances involved being thinking things (Ostenfeld 2018). While the notion of 'physical' has changed throughout the transition from eighteenth century materialism to modern physicalism (Ney 2008b), the core distinction from Descartes remains that the soul/mind has something to do with subjective or private experience (Ostenfield 2018, p.9) while the 'physical' has to do with the things which can be studied using the methods of the hard sciences such as physics (Papineau 2013).

"Descartes conceptualized mind as a separate substance that was metaphysically distinct from physical stuff." (Wheeler 2005, p.21) There are two questions here that need to be teased apart: (1) The question of whether one, two, or more types of substances exist, and (2) How best to conceptualise the notion of 'mind'. These are two separate questions, and strictly speaking, the second question does not require us to provide an answer to the first question.

In this current chapter, a novel reply to the second question will be given to the mind-body debate by building off an argument that the first question is philosophically uninteresting. The strategy here will be slightly unconventional, in that the chapter does not reject dualism or advocate monism directly, but rather, questions the 'philosophical interestingness' of dividing reality into distinct categories. In the same way we typically take for granted that the existence of magnetic and non-magnetic substances has no metaphysical significance, this section will argue that as a default view, we should accept monism unless we have a good reason to specify

otherwise. Therefore, this will lead us to accepting a 'neutral monist' ontology, not because we have rejected dualism or pluralism, but only because it is advocated as the 'default' analysis of a unified reality, and the burden would be on the opponent to tell us if the default is insufficient.

After section one argues for this monist ontology, section two through four will introduce the three main elements that are important for understanding Magicalism: That it is a 'story' about ontology, told from the perspective of the (a) Eastern (b) Magician's (c) Apprentice, as the title of this thesis will suggest. Section two returns us to the debate between Carroll and Wallace (2022) discussed in the previous chapter, but this time, with a greater emphasis on the more Eastern approach to philosophising about the mind. Section three will then elaborate on the discussion started in the previous chapter on the link between card tricks and philosophy of mind. Finally, section four will discuss a distinction between what we will call the Apprentice Approach and Textbook Approach to knowledge acquisition, where the latter can be more prone to dogmatic reasoning, and the former is more suitable for topics such as consciousness. Finally, the conjunction of the (a) Eastern (b) Magician's (c) Apprentice to philosophising about the mind will be labelled Magicalism: defined as 'a neutral monist view that claims the essence of consciousness is Magic, but in the sense of a card trick, not the supernatural sense'.

4.1 Monism as the Default

This first section will focus on arguing for a neutral monist ontology as a default position, over a dualist ontology. Monism can be understood as the view that claims only one type of substance exists, whereas dualism is the view that there are two distinct substances in existence, neither which are reducible to the other (Crane 2000). In philosophy of mind, it is common for 'substance dualism' to be used interchangeably with the idea of accepting that the two substances which exist are physical and mental substances (Lowe 2010; Swinburne 2009),

though technically, an ontological thesis would still be considered ‘dualist’ so long as two distinct substances are postulated, and the word ‘physical’ does not necessarily need to play a role in advocating dualism (Olson & Seagal 2023, Ch2).

It is important to note that monism does not necessarily always imply physicalism nor a commitment to the physical sciences (Alter & Pereboom 2023). For instance, Russellian monism is a neutral monist view, in that it claims that only one type of substance exists, but that this substance is to be considered neither physical nor mental. Rather, the important distinction drawn by the Russellian monist is in regards to understanding what a substance is, versus what a substance does. The former is known as the ‘intrinsic nature’ of that substance, while the latter is known as the ‘behavioural disposition’ of that substance. Different Russellian monists have different ways of conceptualising these notions. For instance, the panpsychist believes that the ‘intrinsic nature’ of all substances just is consciousness, even in the case of fundamental particles or seemingly inanimate objects; whereas the ‘behavioural disposition’ of a substance are things which scientific observation can often shed light about (Goff 2017, Ch6).

The Russellian monist approach is one method for endorsing a monist ontology while simultaneously capturing the seeming duality that we are trying to explain. One of the largest drawbacks of a dualist ontology is known as the interaction problem, which addresses the difficulty in accounting for how it is that two drastically different substances are able to interact with one another (Lowe 2010). Descartes responded to this originally by postulating that the pineal gland essentially channels ‘animal spirits’ of our mental substances to flow through our nervous systems thus influencing our bodies which are made of physical substances (Wheeler 2008; Cottingham 1985). While the original Cartesian view is now considered to be outdated even according to twentieth century neuroscience (Aldrich 1970), modern Cartesian dualists are still tasked with accounting for the difficulties presented by the interaction problem (Kim 2018; Koksvik 2007; Radner 1985). Non-Cartesian substance dualists attempt to avoid this

issue by rejecting the Cartesian approach of making the claim that mental substances lack any physical component at all. Instead, the non-Cartesian dualist may try to claim that the essence of a person or a self is not reducible to the physical, and yet a person is still an embodied being – and claims interactionism ceases to be a puzzle in the same sort of sense (Lowe 2010).

Monist views, such as Russellian monism (Alter 2012), property dualism, or dual-aspect theories (Macpherson 2006) have the advantage of avoiding the interaction problem altogether. This is because if only one type of substance is proposed in the first place, then there is no mystery about 'how two substances interact' which requires explaining. Monism also has the advantage of being more parsimonious, as only one substance needs to be postulated instead of postulating two distinct substances (Goff 2017, Ch6). For these reasons, monism has a slight upper hand from the beginning, and even just from considerations of Occam's Razor, it can be said that ideally, we would avoid a dualist ontology unless it is absolutely required.

In chapter three, we discussed Faraday's discovery of the electromagnetic field, and how the idea that fields do not reduce to an atomist model does not lead to a metaphysical dualism. In the same sense as this, *even if* we were to discover that ghosts, psychic powers, or other paranormal phenomena turn out to be real, even this kind of strong evidence of 'spooky' entities does not need to compel us to support a metaphysically dualist ontology. If we do not have an issue with accepting magnets or fields under a metaphysically monist interpretation, then why should we have to accept a fragmented reality if ghosts turned out to be real? If we ever do find a way of seeing or interacting with ghosts, then perhaps they might cease to be 'spooky', and while it would be a momentous discovery to learn that ghosts exist, it is not immediately clear why this should fragment our metaphysical reality (aside from the historical factor that the 'apprentices' of Descartes may wish to claim a sort of victory. We return to this idea in section four when discussing the concept of apprentices in philosophy of mind).

Granted, this is merely a commitment we are making to monism being the default, rather than being an argument against dualism. Therefore, an opponent may claim (and rightfully so) that I have not so far provided anything informative in response to the question "What is the essence of consciousness?" I have not yet said what consciousness is or isn't, and all I have said so far is that the fragmented language of dualism seems unwarranted unless we really, really have good reason to do so. The discovery of magnetism did not metaphysically shake us into a dualist ontology – why should the discovery of fundamental mentality (if it ever does happen) shake us in such a way? If it really is just a historical head-nod to Descartes's original way of engaging with the debate, then we will see further in section four why we might wish to move on from this outdated dichotomy altogether rather than continuing to feed into it.

A neutral monist attitude is advocated here as nothing more than a 'default' position, and on its own, I have not intended for this to carry much content. But this should not be a problem for us, since sections two and four will provide the content. With that in mind, we now move on to more interesting discussions to support Magicalism and answer what consciousness *is,* which involves looking at an (a) Eastern, (b) Magician's, (c) Apprentice's take on the matter.

4.2 The Eastern Perspective

Under an Eastern conception, the mind-body dichotomy does not exist to begin with, and trying to explain away a seeming dichotomy would be an awkward move to make, as no 'seeming dichotomy' is there to address in the first place. A key reason for this may come from linguistic differences, shared both in Japanese and Chinese which share much in terms of its written language. In Japanese for instance, the word for 'mind' is the same as the word for 'heart': 'kokoro' (心). For this reason, to say the phrase 'philosophy of mind' in Japanese will inevitably carry the connotation that it is a philosophy about the heart just as much as the mind.

It is not something that is strictly limited to a cognitive exercise, and it can very well involve an appeal to the passions as well. It might even connotate a sense of 'soul' – not in a religious sense, but more in the sense of the phrase 'the artist has put his soul into his painting' and so language that might imply something dualist if uttered in English loses this 'seeming duality' if discussed in the Japanese language. Another notable point is that, the word for 'human being' is written by combining the characters for 'space between' and 'people', and for this reason, Japanese philosophers of mind have argued that the fundamental essence of human consciousness is found, not in the individual brain, but in the *space* that is found *between people*. It is not so much a 'physical' concept as it is a social one. (Kasulis, Yuasa, & Yasuo 1987, Ch1). It is also not a 'mental' concept in the cognitive sense as it is a 'heart-felt' concept relating to the passions, and again, the starting intuition for Eastern philosophers is not about explaining away a seeming duality (which does not exist to begin with in other languages) but it is about wrapping our heads around a 'humanness' that we find when we talk about the heart.

In fact, there is a great overlap between Japanese and continental philosophy, particularly between Heidegger's existentialism and the Kyoto School of Philosophy led by philosopher Nishida Kitaro (Ishihara 2019, 2016; Zavala 2012). "The Kyoto School (Kyōto-gakuha) is a group of 20th century Japanese philosophers who drew on the intellectual and spiritual traditions of East Asia, those of Mahāyāna Buddhism in particular, as well as on the methods and content of Western philosophy." (Maraldo 2024, SEP introduction). Nishida Kitaro provides a discussion on the Buddhist-inspired concept of "absolute nothingness" (called a "mu" state) and the relation of "being" and "non-being" in Western philosophy (similar to Heidegger's discussion on das Sein and das Nichts; see Heidegger 1975, Vol. 9, 103–22; and Vol. 65, 246–47), and the Japanese attitude towards analysing consciousness is captured under this light. Once again, we see a monist ontology: Of things that are 'being' or things that 'exist',

as opposed to things which belong to the category of 'non-being', which tends to play a more abstract aesthetic role, rather than an ontological one.

There is another concept called 'wabi-sabi' – a beauty in impermanence' which is a central concept in Japanese culture. "Wabi-sabi is a beauty of things imperfect, impermanent, and incomplete. It is a beauty of things modest and humble. It is a beauty of things unconventional." (Koren 2008, p.7) This concept of wabi-sabi and non-being play a significant role in much of Japanese art, entertainment, as well as philosophy. Loosely understood, it is the concept that nothing is permanent, but that this is what makes art and philosophy 'Magical'.

A 'sakura' cherry blossom tree is beautiful only because it is not in full bloom forever, and equally, human lives are meaningful or sacred to us only because we do not live eternally. The state of 'being' are points of contact with the real empirical world; whereas philosophy is sometimes concerned with 'non-being', in the sense that we can think beyond what currently 'is' in front of us, and think in terms of abstractions. "All science needs to take some entity or other as its object of study. The point of contact is always in being, not in nothing. The discipline that has to do with nothingness is philosophy" (Tanabe 1964, in Heisig 2001, p. 121).

As such, Eastern philosophical discussions tend to involve a great deal of poetry, again having much in common with continental philosophy – but this kind of poetry is not always the most useful for analytic purposes. Up to a point, we can draw inspiration from such poetic depiction of 'nothingness' and whatnot, but at the end of the day, it leaves wanting a more rigorous explanation that gets to the point of our (analytic) philosophical questions. What is important for us presently is to notice that unlike the Cartesian dichotomy which sets up a contrast between two kinds of substances, the Eastern dichotomy starts by comparing two types of concepts: Empirical 'concrete' entities, and abstract 'poetic' concepts. In essence, if we ever do discovery ghosts or souls in a way that make them empirically addressable, then they would

belong to the former category of empirical ‘concrete’ concepts just as much as the substances discussed by the physical sciences. These concepts by themselves are not particularly interesting for philosophical purposes, however, since it is just about observing what is in front of us. Alternatively, if we are trying to be poetic when we discuss ghosts or souls – such as when we reminisce or grief about our loved ones who have passed away, or try to come to terms with our own mortality and how to handle these deeper concepts – then this would belong to the realm of abstract ‘poetic’ concepts suitable for philosophical or metaphysical discussion.

The former, the empirical ‘concrete’ concepts, are understood as places where “point of contact” (from above quote) happens between (a) external reality and (b) the way we humans conceptualise reality. From the perspective of Eastern philosophers, the error in the Cartesian dichotomy has happened because the “point of contact” empirical approach to understand external reality has become convoluted with a philosophical urge to maintain a sense of poetry and Magic (Kasulis, Yuasa, & Yasuo 1987, Ch1-4; see also Watsuji Tetsuro, in McCarthy 2020; Nagami 1977). This kind of worry is articulated, for instance, in the quote below:

> “In the Christian tradition, nature and the spatial universe are creations inferior in status to human beings, who are endowed with the spirituality of God's image. Consequently, the human spiritual eye can objectify other creations. In the East, however, the essential destiny of human life is to be embraced by life's rhythm in natural space; it is to be together with the animals and plants, with all things that have life, with what the Buddhists call "all sentient beings" or "all living beings" (sattva).” (Kasulis, Yuasa, & Yasuo 1987, p.9)

The idea discussed by the Eastern philosophers is that if one’s initial introduction to philosophy contains a Christian background – where Christianity itself tends to make claims about external reality in a way that reflects a highly ‘Magical’ attitude but taken far too literally – then it is

unsurprising that Western philosophers become preoccupied with the Cartesian dichotomy. That is, it creates an intuitive divide between Christian-esque ideas and science-esque ideas:

> "Hence, there arises in compensation a heretical view that sees the essential human mode of being exclusively in the physiological or physical functions of the body. Out of the Cartesian dualism of mind and body arose the parallelism of mind and matter, the latter eventually developing into a mechanistic view of humanity." (Kasulis, Yuasa, & Yasuo 1987, p.10)

British philosopher Michael Wheeler (2005) also discusses a similar point:

> "According to Descartes, the distinctive feature of explanation in physical science was its wholly mechanistic nature […] namely the commitment that in a mechanistic process, one event occurs after another, in a lawlike way, through the relentless operation of blind physical causation." (p.26)

The underlying idea here is that the mind-body debate is dichotomised in such a way that presupposes 'physical' entities are of a mechanistic nature, whereas 'non-physical' entities are of an ethereal kind of nature. Problematically, mechanistic analyses are not the opposite of ethereal existence, and this leads to a confusion about what is meant by 'physical' and 'mental'. Psychologist Jon Mills (2022) brings up another related point that under physicalism:

> "[T]he individual is reduced to physical substance alone, which gives rise to […] mechanistic views of the human being. In short, the human being is reduced to a thing – a reified biological machine engineered by evolution and stimulated by the environment [thereby stripping] selfhood down to biology." (p.51-52)

Some philosophical questions, such as that of selfhood, finding meaning, or even psychiatry, will require a level of introspection to evaluate normative considerations that play important roles in human life. These issues will be addressed more deeply in later chapters when we take a closer look at psychology, psychiatry, and its relation to stage magic. Right now, what is important is to notice that the idea of 'empiricism' or 'contact points with reality' does not necessitate a *scientific* approach that prioritises replicable and generalisable *data crunching*. The word empirical does not necessitate a mathematical approach, nor does it need to involve being as 'objective' or 'unbiased' as possible. The Eastern approach, if anything, might be understood as an approach of biting the bullet that human matters will contain inherent biases, but that this need not undermine conceptual rigour or accuracy in our philosophical thoughts.

Returning now to the debate between Carroll and Wallace (2022) discussed in the previous chapter, Wallace highlights the biased attitude in Western philosophy to treat all forms of spiritualism as though they are a 'religion' like Christianity. Wallace argues that Buddhism is not exactly a religion in this sense (*ibid*, p.50). "Buddhism doesn't start with a leap of faith in the existence of God, or Nirvana, or reincarnation, or karma, or Buddha." (*ibid*, p.45). "They didn't get there by just believing this or that, or having great faith in the Buddha, but by doing what scientists do: rigorously investigating the phenomena themselves and seeing what comes up." (*ibid*, p.55). He argues here that there is an important distinction between a religion which *starts* with a leap of faith in certain beliefs about external reality (e.g., God exists); and spiritual attitudes such as Buddhism which make no such *a priori* external commitments. Instead, Buddhism is more similar to Socratic questioning where Buddha encourages individuals to discover truths on their own rather than taking someone else's word for it. Following the story of the blind men and the elephant, Wallace continues that each of the men were now in disagreement about what it was that they were observing. Buddha then visits these inhabitants:

> "[The] inhabitants had heard a lot of people propounding the truth, each one with a story incompatible with the others, and each one saying, "My view alone is right." As a result, the villagers had become very skeptical toward anyone who claimed to have answers to the big questions of existence." (*ibid*, p.42). Buddha responds to this: "Rather than arguing that his [own] views were superior to everyone else's, the Buddha surprised them by responding, "It is proper for you . . . to doubt, to be uncertain, for there are good grounds for your uncertainty. . . . Do not adopt views simply because they are the status quo, or because they've long been assumed to be true. Do not accept them based on mere rumor or because they are written in some scripture. Do not accept them based on pure conjecture, some unquestioned assumption, inconclusive reasoning, your own personal bias or others' eloquence, or because it's the view of your teacher." (*ibid*, p.42, referencing 'Kālāma Sutta' III no.65).

The moral of the elephant story is that Buddha himself discourages people from believing in someone else's dogma, just because they said it with confidence. Instead, he advises that the villagers should be sceptical. "[A]nd then on a positive note, he counseled the villagers to test all truth claims with their own experience and determine for themselves whether they were true or false, beneficial or harmful. He was truly an empiricist." (*ibid,* p.42). In essence, the teachings of Buddha are expressed in such a way that the Buddhist is discouraged from believing even Buddha himself, which is very different from Christianity where God or the bible are treated as a trusted source of information that one ought to have faith in. In contrast, Buddha is depicted similar to Socrates in this regard. While both Buddha and Socrates do encourage the wisdom of knowing how much one does not know, their method for convincing other people is to allow them to discover it for themselves (see Husserl & Iyler 2017; Dillon 2000; Courtney 1885 for cross-field discussion between Buddha, Socrates, and Christianity).

The word ‘empirical’ itself means something which is based on observation or experience rather than being based on pure logic or dogma. Wallace argues that Buddha’s teachings are to encourage ‘empiricism’ in this sense – of not just trusting the authority or teacher, and being just as sceptical of these as we are of our own personal biases. The *default* is scepticism, and to foster falsificationism. (This is a thought we will return to more thoroughly in the next chapter when exploring Popper, Kuhn, Lakatos, and philosophy of science on falsificationism.) If empirical considerations (i.e., ‘contact points’ with reality) sometimes lead us towards forming spiritual beliefs, then we would be mistaken to ignore this just out of a brute conviction that spiritual things feel a bit ‘spooky’. Instead, if we want to be cautious, it makes sense to scrutinise and evaluate the exact connection we think we are seeing between empirical matters and a spiritualist conclusion. At the end of the day, it is about deepening the rigour of our reasoning, even if it means the content of that reasoning exceeds the scope of the hard sciences.

The blind men are encouraged to re-evaluate their own claims, for instance, by attempting to check other parts of the elephant and falsifying their own original hypothesis, as they are cautioned against simply taking someone else’s word for something. It is acknowledged that even credible sources of information usually do not reflect the full picture, and even if they did, it is healthy scepticism to double check this by attempting to falsify the claims where possible. “Buddhists, like scientists, need to avoid ideological blockages, concluding that certain evidence can’t be true, because it doesn’t conform to our core beliefs. That’s not scientific, it’s dogmatic.” (Carroll & Wallace 2022, p.55). For these reasons, Wallace claims that physicalists are being just as dogmatic as religious individuals, in that it involves an over-trust in authority.

What the Eastern approach can teach us is that when we do not distinguish sufficiently between ‘facts’ and ‘stories’, (ironically) even a pro-scientific attitude starts running into the same problems as religion. Importantly, philosophy itself is interested in both storytelling and in having accurate ‘contact points’ with external reality. It is not about one or the other, as a

good story without contact points is nothing more than a fictitious myth, but a lack of story with *only* data points risks never fully getting to the heart of the philosophical question. Therefore, scepticism and empiricism must go hand in hand when telling philosophical 'stories'.

4.3 The Magician's Perspective

We have seen in the above section that Eastern philosophy tends to include (i) scepticism, (ii) empiricism, and (iii) a sense of poetic storytelling. As tempting as it might be to see the word 'empiricism' and jump to the conclusion that this is the domain of science, it is important to acknowledge that a pro-scientific attitude is not only one way of attempting to satisfy these three conditions. Another way of satisfying (i)-(iii) is to endorse a pro-magician's attitude. There is a significant overlap between science and stage magic, especially in the sense that for both of these, the *method* is distinctly separate from the *effect*. The magician can tell any poetic story they like about a card vanish, but if the card needs to vanish, then *it is supposed to vanish*. There might be many 'right' ways of performing a trick, and different performers can add their own sense of poetry to it, but there is also a very clearly 'wrong' way of performing that trick. The same goes for science, where an expert surgeon might appear as though they are achieving the impossible by providing miracle cures – but obviously, if the surgeon does not know the method of how to perform the surgery, the 'magical' effect is also just not happening. Both a magician and a surgeon's audience can be in awe that they have witnessed a miracle, but fundamentally, the person doing the tricks/surgery needs to be an empiricist at heart. They need to do the techniques that actually *work*, and the results do not happen as if 'by magic'.

This is not always true in other professions, especially when it comes to artforms that have a strong subjective component. There may be no 'wrong' way for a comedian to tell a joke, and although no one in the audience might find a joke funny, it is harder to tell whether

it is the joke that is the problem, or if the comedian has just been unlucky to find an audience who did not share the same sense of humour. Other artforms, such as writing a novel or painting can have a similar blurry line between 'method' and 'poetic storytelling' where the two might essentially amount to the same thing. Once again, unlike science or magic, the distinction between method and effect are nowhere near as precise. In fact, with some *extremely* subjective professions – such as, say, being a poet – the poetry *just is* the method.

A magician can use poetry and storytelling to *enhance* their work, but if the magician does not know *how* to vanish the card, then we have not gotten as far as discussing the showmanship about how to spin the story of a vanishing card. The method and story are distinct.

A major difference between a scientist and magician, however, is that the former tends to be directed at contributing to a public knowledgebase, whereas the latter tends to be directed at a more personalised performance. Magic, in a sense, is found in the space between our blind spots and focal points of attention. The task of a scientist is to thoroughly enhance the focal points as much as possible and rigorously analyse them, whereas in contrast, the task of a magician is to understand how to manipulate the blind spots better in order to enhance the gap in an entertaining performance. Contrast, for instance, the way that Daniel Dennett describes stage magic, and compare this against the way that a professional magician talks about it:

> "The task of explaining stage magic is in some regards a thankless task; the person who tells people how an effect is achieved is often resented, considered a spoilsport, a party-pooper. I often get the impression that my attempts to explain consciousness provoke similar resistance. Is it not *nicer* if we all are allowed to wallow in the magical mysteriousness of it all? Or even this: If you actually manage to explain consciousness, they say, you will diminish us all, turn us into mere protein robots, mere *things*." (Dennett 2016, p.8)

In contrast, the magician, Derren Brown further expands on this "disappointment of people who find out methods and the confused accusation of 'that magician's a fake' which one sometimes hears might be a pained expression of mystery denied." (2023, p.477) However, for a professional magician, who is not a scholar, but more of a performer or craftsman, there is not much point to 'explaining away' the magic, and instead shows how to avoid over-reduction:

> "Consider: the more a society takes pride in rational, scientific advancement, the more it must organise chaos into controllable, predictable patterns. When it comes to understanding its people, it will seek out a universal psychology, applicable to all. It will view them through a *statistical* filter, imposing a kind of mass-mindedness on its notions of what it is to be human. This indispensable approach, without which the advances of knowledge would be scuppered, suffers from one curious shortcoming, which is so glaring as to be barely noticeable: the *individual* is written out of the picture. You or I become an abstract 'from which all features have been removed', wrote Carl Jung. No *specific* person ever fully participates." (Brown 2023, p.477-478).

The magician then goes on to construct an argument that the inner experience of persons is incompatible with validating it from a scientific, fully generalisable or quantitative perspective:

> "None of us ever precisely encapsulates the statistical mean. […] In our world of people, the individual is the carrier of reality. But individuals bring with them everything that does not conform to the statistical picture. They will constantly elude every norm and break every rule. Everything that makes them concrete is, in the same breath, denied validity in the scientific model. Furthermore, those mysterious, individual aspects have the unfortunate quality of both being deemed 'irrational' (that is, out of the rational, scientific picture) and yet remaining utterly central

> to who that person *is* and therefore what distinguishes each of us from the general. In unfortunate cases, the situation might leave us feeling invalid, unseen, out of sync, ashamed. It certainly leaves us all aware of the tension between our inner experience of ourselves and the version we feel has validity in the [scientific] world." (*ibid*, p.478).

This is the gap between the pro-science attitude and pro-magician attitude when philosophising about the mind. While they both take (i) scepticism, and (ii) empiricism seriously, their *goals* are different. The scientist is directed at making more generalisable claims, and in doing so through the means of statistical analyses, the experience of individuals can get explained away. In essence, a scientist is exploring matters of *universals* rather than *particulars*, whereas for a performer or craftsman, it is the *particulars* – the spectators in front of them – that truly matter.

The remaining question is how to handle (iii) poetic storytelling. Because science tends to be far less about storytelling than, say, the arts, physicalists might form the impression that embracing a pro-science attitude will minimise human biases and 'mere storytelling' when engaging with our philosophical works. In other words, the pro-science attitude may imply that we can remove as much subjectivity as possible from our metaphysical interpretations of the world, thereby making it as objective and unbiased as it can be. However, the problem is that philosophy itself is about the storytelling. A philosophical theory that tells no story and is merely a collection of data is not exactly 'philosophical'. That is just raw data. Even in a science journal, we will see that the journal does not just publish its raw data and statistical analyses in its pure data form, but will tell us some kind of interpretation or 'story' about how to understand this data. Philosophy, essentially, is a more thorough version of this, on providing us with in-depth analyses of how to wrap our heads around various forms of data that we are subjected to. The 'data' can range anywhere from personal experience (e.g., the famous *cogito* argument, "I think therefore I am") to doing a meta-analysis of scientific books and journals (e.g., Dennett

1991). But at the end of the day, as soon as we start giving some kind of explanation for how to analyse raw data, we are, in effect, now telling a 'story'.

If science tends to involve providing us with information about *universals* rather than *particulars*, then telling a philosophical story that is strongly in favour of science does not lead to 'unbiased storytelling'. Instead, it just leads to a different kind of bias: that it is more prone to trivialising the conscious phenomenal experiences of *individuals* in favour of treating people more like statistically analysable *clusters of data* rather than human beings. It explains away the 'personhood', or a sense of our individual minds – the very thing we are trying to analyse. (We will return to these points again in chapters five through seven when looking more into problems from psychology and reducing matters of the mind to data rather than individuals.)

In chapter one, we introduced the concept of Xombies, which are third-person identical but first-person distinct. We can now begin to see how this links together with our current discussion. From a scientific perspective, the distinction between humans and Xombies are trivial. However, for the magician, it makes all the difference in the world whether the audience are actually *enjoying* the performance, or if they are merely 'verbally reporting' that the show was quite lovely. If the audience turns out to be Xombified and have no subjective experiences to begin with, then the magician might as well have performed their show to a wall.

For these reasons, the (i) scepticism, and (ii) empiricism of the more Eastern approach advocated in the previous section does not sit very comfortably with a pro-science attitude. To take an empirical approach and ensure that we have 'contact points' with reality does not necessitate that we must take a maths-y, quantifiable, or statistical approach to understanding the nature of consciousness. Instead, we can learn from the magician's perspective – a 'personal touch' offered by the magician who engages the heart/mind of their non-Xombified audiences.

4.4 The Apprentice's Perspective

This chapter began in section one by arguing for monism as the default metaphysical view in virtue of a fragmented sense of reality seeming unnecessary. To reiterate, magnets might seem Magical to someone who has never seen one before, and the discovery of fields can seem 'spooky' for someone who is committed to an atomist worldview. However, if even Faraday's discovery of electromagnetic fields does not shake our metaphysical worldview into a fragmented dualist reality, we might wish to question why fundamental mentality (if it ever were to be discovered) should shake us into dualism. That is, if ghosts or souls turn out to exist, why should this mean that reality is fragmented into two metaphysically distinct substances?

What we will look at now is the idea of a social allegiance that we might have towards our predecessors, which might shape or influence the way that we philosophise about the mind. That is, for historical reasons, the 'apprentices' of Descartes may wish to claim a sort of victory if it turns out fundamental mentality does exist. In contrast, those who take themselves loyal to the sciences might wish to claim a victory so long as mental substances seem incompatible with contemporary science. It is this concept of 'loyalty' or 'allegiance' that we look at next.

When a philosopher advocates a pro-science attitude (e.g., Ney 2008a), it is tempting to understand this as a *lack* of allegiance towards any authority figure, forming instead the impression that the facts or data will speak for itself. In chapter two, we distinguished between Science Norms Physicalism which does make the discussion much more about scientific methodology (but at the cost of making it an epistemic view involved in philosophy of science rather than on ontological view involved in philosophy of mind); whereas Authoritarian Physicalism explicitly follows the authority of scientists. But even Structural Physicalism which makes a commitment towards to certain entities or properties is still strongly influenced by a certain (i.e., mathematical, generalisable, statistical) approach to knowledge-acquisition.

What is in the background of a scientific approach which can sometimes get overlooked is that scientific empiricism is a specific kind of empiricism, and it is not the only kind. Returning again to Wallace, he discusses that there is a crucial distinction between Buddhist empiricism and scientific empiricism, which he refers to as a differing focus on first-person versus third-person investigation. Buddhism itself is a very personal activity, which people tend to lean into for the purpose of spiritual or emotional tranquillity, rather than for the purpose of gaining objective knowledge. In contrast, science is a very public (i.e., non-personal) activity, which people lean into for the purpose of understanding the world despite personal attachments. What Wallace highlights is that the Buddhist first-person empiricism is still a valid form of 'empiricism' (i.e., it provides 'contact points' with reality) but that its purpose is different from scientific empiricism. Science is typically aimed at third-person investigations that are statistically analysable, replicable, controllable for variables, and overall strives for objectivity, reducing personal biases, generalisable, and so on (Scholtz et al., 2020). How effective the current methods are at achieving this is a separate issue (see related topics on the generalisability crisis, Yarkoni 2022; and replicability crisis, Nosek et al., 2022), but the general idea here is that scientific methods tend to strive *removing* personal biases and try to construct descriptive pictures of reality that are 'objective' as possible, but this means that removing the personal touch can add its own sort of bias – of making reality seem too clean or sanitised, thereby leaving out important nuances where it gets a little too 'human' or messy.

Wallace expresses his concerns that physicalism can lead to a form of 'scientism', a dogmatic acceptance of science as the ultimate (and only) valid source of knowledge, and lead to the exclusion of the more human or nuanced side of reality, though the exclusion of the nuance is driven by conviction or taste rather than genuinely being empirically driven. "Western materialistic theories of mind [...] have one common quality in common: not a single one is empirically testable. They're just beliefs." (Carroll & Wallace 2022, p.57). The spirit of

what Wallace is asking for is a falsificationist one. If physicalists want us to believe in physicalism, then they should propose some kind of falsifiable hypothesis, and show that this hypothesis resists falsification when put to the test against real-world evidence. That is what it would take to truly make physicalism an empirical (and therefore pro-scientific) sort of thesis. Strangely, a pro-science attitude in philosophy amounts to an anti-empiricist (or pro-dogma) attitude *unless* the philosopher is careful to check what their 'contact points' with reality are.

The issue which Wallace has is the idea of 'explaining away' the importance of first-person testimony in favour pushing forward a materialistic agenda. There are several instances where this type of worrying authoritarianism is exemplified. For instance, the Church of Scientific Materialism founded by biologist Thomas Huxley in the nineteenth century accepted science as a religion in its own right (Carroll & Wallace 2022, p.49; Lightman 1983). Similarly, some French Catholic churches were also converted to what were called 'Temples of Reason' during the French Revolution to allow for the quasi-religious worshipping of science (Burnett 2012; Ozouf 1991). These attitudes can lead to a dogmatic form of scientism, and to quote Susan Haack, scientism (in the dogmatic sense) can be understood as "uncritically deferential attitude towards science, an inability to see or an unwillingness to acknowledge its fallibility" (Haack 2012, p.76). In other words, although 'churches' or 'temples' which follow scientific thought will take the *content* of what scientists talk about seriously, it seems to take just as unquestioningly and worshiping an attitude that we tend to associate with religion (Sorrell 2013, Ch1) – and this amounts to essentially replacing the bible with a science textbook.

The notion of an apprentice becomes important here because it forces us to look at how there is nothing wrong with the idea that our allegiance is grounded somewhere, but also to acknowledge we need to be more explicitly aware about how much of our allegiance comes from loyalty, intuition, or faith; and how much of it is justified by the content of discussion. There is a difference between a Buddhist who feels a sense of connection with Buddha's

approach to knowledge-acquisition (e.g., empirical, sceptical, directed at spiritual tranquillity) and someone who believes in every word uttered by Buddha. An Eastern apprentice may be depicted as someone who follows in a master's footsteps of cultivating a spiritual awareness and journeying into the world to gather wisdom, rather than someone who believes everything that a sage tells them. In this sense, the Eastern master is not so much a source of knowledge as they are a source of guidance and inspiration (so arguably, the word 'sage' would be misleading for a Western audience). The Eastern apprentice need not believe everything their *'sensei'* (i.e., teacher/guide/master) tells them, and if anything, is encouraged to rediscover the world themselves. Similarly, a magician's apprentice may respect the master magician's attitude of valuing their audience's concrete experiences, valuing storytelling and showmanship, and taking care to execute the method properly, but the apprentice is free to disagree with the content of the master's beliefs. The *sensei* is not a source of knowledge in the way that textbooks might be when studying for a university exam. The Eastern Magician's Apprentice does not seek dogmatic answers from their *sensei* but instead adopts their attitudes.

This is where physicalists are faced with a fork in the road. If what they are loyal to is scientific methodology (i.e., Science Norms Physicalism), then they are free to contradict physics in any way they like. They could be a Buddhist philosopher advocating souls and whatnot so long as they adopt the pro-empiricist attitude of taking a falsificationist stance seriously (e.g., testing out hypotheses, trying to disprove their own pet theories, etc.) But ironically, this might lead the Science Norms Physicalist to accept entities into their ontology which seem classically dualist (e.g., souls), making it unclear whether 'physicalist' is still a fitting label for this type of person. In contrast, a physicalist who is loyal to the physicists themselves (or the content of what modern physics tells us to believe) is not so much an apprentice who questions their master as they are a student who memorises the right answer that a teacher/textbook is giving them. This itself is not the type of attitude which scientists

have when running experiments and seeking out new knowledge (Sorrell 2013, Ch1), and so ironically, the 'good student' who follows scientific theory can turn out to be a 'bad apprentice' in terms of adopting the core attitude of the very thing which they claim that they are following.

The Apprentice's Perspective reflects this contrast with a more Textbook Approach to knowledge-acquisition. The image of an Eastern apprentice in particular – inspired by the underlying scepticism and empiricism of Buddhism – is someone who is encouraged by their master or *sensei* to question their own affiliations just as much as they question their opposition. The apprentice should not take their master's wisdom/knowledge at face value, but should question the validity of it by going out into the world themselves and testing their beliefs against the real world. And in the event where it turns out that 'spooky' entities need to be taken seriously, this does not need to shatter or fragment our world into two distinct 'substances' that seem to struggle with the interaction problem. As this chapter has tried to advocate, the Magician's Apprentice in particular is someone who takes a more hands-on or personalised approach to investigating the issue of consciousness (such as learning through performance or other forms of direct human contact that still involves some kind of empirical element) rather than reading scientific articles and creating an image of the world derived from data points.

Hopefully, the discussions so far have started to create an image of what Magicalism points towards. Magicalism is not the same kind of ontological view as existing theories like physicalism, dualism, or Russellian monism, in that Magicalism does not strive to give the 'right answer'. Instead, it is a sort of perspective that we are encouraged to take when philosophising about the mind. It is a neutral monist view (i.e., the idea that our sense of reality should not be fragmented) that claims consciousness is much like stage magic: That there are mechanisms and methods behind it, but also the poetic nature of the storytelling does a large part of the work. And to wrap our heads around it, we ourselves need to look at the picture like apprentices rather than students – an attitude of rediscovering the cognitive world through our

own journeys of engaging with the world first-hand; running our own experiments; falsifying our own beliefs; and not overly-trusting books that we read, no matter how credible the source.

If we are working with the assumption that there can be 'THE' perfect source of information (whether it is the bible, or a fully completed scientific textbook) which we can simply read to gain a story-less, bias-less view of ontology, which just is True as an objective fact – then this is simply an overly idealistic goal to have. Even physics itself is limited by various methodological constraints, as well as the interests of the researchers (and funding constraints) governing that path. Perhaps the hard sciences appear to have an edge in their seeming 'robustness' in virtue of the fields dealing with less (if at all) in matters of the passions, heart, or poetry. But it is worth acknowledging that purposefully reducing a nuanced, philosophical topic so that it loses the storytelling – just so that it *feels* more objective to our human minds – would itself be a biased approach. The way in which human beings seek philosophical meaning is itself a nuanced endeavour, and the success of science should not be confused with the idea that nuance in matters of the heart or spirit is our enemy. Instead, what Buddhism offers is a way of wrapping our heads around the idea that we are *only human* – and that we are meaning-based creatures. The reason that we philosophise in the first place is *because* such meaning is important to us, and just because some fields that require less storytelling have experienced great success does not mean that we need to make our own discipline more like theirs. Instead, although it is a difficult task, we can still chip away at constructing better methods of knowledge-seeking that help us reconstruct the cognitive world.

"In the Buddhist view, the defining characteristics of human beings— what makes us special or distinctive—are that we are able to understand meaningful information, and we're able to impart it as well. In short, we are storytellers. We tell many stories. We've heard a very good one here from Sean [Carroll], based on a lot of science, but it is one of many stories. I would suggest that all the stories we tell ourselves, that we tell each other about the physical

universe, life, the mind and so forth, begin with certain assumptions that we believe to be true, but that we cannot prove with empirical evidence or logic. We start with assumptions, beliefs, or axioms. On this basis, we develop theories. And then, throughout the course of our lives, we choose what we attend to, where we focus our attention. If you're a professional physicist, you attend to physical things. And our world as we envision it corresponds directly to what we attend to, and it excludes what we ignore." (Carroll & Wallace 2022, p.36). Our task is not about working out who gets to be in the position of an authoritarian dogmatic figure, but rather, it is to stop engaging in dogmatic forms of epistemology and discover our own wisdom instead.

This chapter has argued that to avoid a dogmatic answer to the ontological question "What is the essence of consciousness", we can reply by giving an answer that retains some sense of ambiguity. If we must give it a label, we can call it Magicalism: defined as 'a neutral monist view that claims the essence of consciousness is Magic, but in the sense of a card trick, not the supernatural sense'. This is a conjunction of three attitudes, the (a) Eastern (b) Magician's (c) Apprentice's perspective. This amounts to a conjunction of three important features of taking seriously (i) scepticism as the default, (ii) empiricism to establish 'contact points' with reality, and (iii) allowing poetic storytelling when the philosophical need arises. With this established, the second half of the thesis will focus on working out the details and boundaries of this philosophical story where we can allow some Magic to be retained but still have a sense of practical rigor. In the next chapter, we further investigate the importance of falsificationism, particularly in the context of parapsychology, and we will see how key figures in philosophy of science – Popper, Kuhn, Lakatos – help further the discussion on Magicalism.

5. Philosophy of Science and Illusions (A)/(B)

In the first half of this thesis, we focused on questioning physicalism and excessive 'pro-science' attitudes, advocating instead a Magical approach to philosophising about the mind. However, going too far in the anti-science direction is itself problematic, and boundaries need to be made clear so that Magicalism does not become too ungrounded, or outright based in false information. When we say that something is Magical, this can refer to an inherent wonder or whimsy contained in the content, and/or it could refer to something which is deceptive. In the introduction, these two concepts were distinguished by referring to them as:

> **Illusion$_{(A)}$** = A surprising or seemingly impossible story that brings a feeling of wonder.
>
> **Illusion$_{(B)}$** = A deception or a falsity. Based in misinformation, or something that is not real.

What Magicalism claims is that consciousness is an Illusion$_{(A)}$, whereas strong illusionism (Dennett 2016; Frankish 2016) claims that consciousness is an Illusion$_{(B)}$. While illusionist philosophers may continue to insist that they by no means wish to claim that consciousness does not exist, and that all they want to claim is to deny consciousness involves our direct acquaintance with non-physical, private mental qualities, it should be noted that illusionists are still making an assertion about this deceptive nuance. They are claiming that the way in which other philosophers often discuss consciousness is not quite as it seems, and in contrast to this, Magicalism makes no claims about the 'deceptive' or 'counterintuitive' nature of private mental qualities. It is this contrast between the two which is of importance to us right now.

To understand the difference, it is important to first discuss where the boundary lies between the two, and we will explore this by looking into some problematic views in astrology,

psychic powers, and faith healing, to contrast these with the more benevolent approach in stage magic. Largely, there are three major problems with holding beliefs in 'pseudoscientific' ideas like astrology:

(I) The Empirical Problem

(II) The Theoretical Problem

(III) The Pragmatic Problem

There are a range of topics where a large consensus exists that a theory has already been proven false, yet some advocates will remain who continue to support the theory. Astrology is one such example. Despite there being good evidence that celestial activities do not actually impact human affairs like our personalities or love interests as astrologists tend to believe, astrology remains popular among certain circles. This leads to a puzzling situation, where a person may continue insisting upon a belief system even when presented with very clear counterevidence (Shermer 2013).

Astrology is an interesting example, as theoretically, it looks as though it is empirically testable. In fact, there is a large amount of evidence showing that celestial activity of one's birth date does not correlate with human affairs (Shermer 2013; Komath 2009; Dean & Kelly 2003; French, Leadbetter, & Dean 1997; Ertel & Dean 1996; Fuzeau-Braesch 1992). This has been empirically tested by looking at 'birth-twins', referring to individuals who are born around the same time, and therefore, under the same celestial conditions, but are not related in any other way. Scientific investigation shows that birth-twins do not have anything in common more than what would be expected among any arbitrary pairing, except in cases where both individuals happen to already believe in astrology, in which case, there is a small bias towards reporting that they have the personality traits their horoscopes suggest that they would have (such as Leos having a warmer personality). If anything, this type of evidence seems to support

the idea that it is not celestial activity that is responsible for the personalities of the individual, but it may be their preconception about astrology which shapes the way they see themselves (Thagard 1987).

(I) The Empirical Problem, as we have called it, refers to such situations where there is clear counterevidence against a falsifiable theory, and yet some people will continue to believe in the concept anyway. In contrast, (II) The Theoretical Problem occurs when a theory is fundamentally unfalsifiable, or accepted as part of a mere assumption. For instance, whenever an astrological reading is inaccurate, the astrologist can simply say "the stars incline, but do not compel" (Komath 2009, p.1571), thus protecting the belief system itself. Finally, (III) The Pragmatic Problem emerges when individuals begin to make real-life decisions based on false information. In these cases, the issue is not just that the information is false, but that individuals are placing themselves in potentially dangerous situations based on a scam-like belief system. In contrast, a magic stage show is presented as a form of entertainment, rather than a 'belief system' where people are encouraged to make real-life decisions based on what they observe. A magic show will often present a seemingly, 'theoretically' impossible situation, but the audience is forced to accept that they are indeed 'empirically' witnessing that situation, thereby creating a gap between expectation and reality in a way that does not create any 'pragmatic' threats. In this chapter, we investigate where the boundary lies between deceptions and wonderous magical whimsy.

In section one, we will begin with a quick overview of some key figures in philosophy of science, which will give us important terminology that will be used throughout the rest of this thesis. Then, in section two through four, we cover the (I) Empirical, (II) Theoretical, and (III) Pragmatic problems, respectively. Finally, in section five, we will discuss stage shows in a little more detail to create a contrast between 'magic' shows and 'deceptive' shows. This will lead to the conclusion that consciousness is like Magic, but that it is not a form of deception.

5.1 Philosophy of Science

There are three key figures who we will explore in this section – Popper, Kuhn, and Lakatos – who have each contributed to the discussion on the demarcation criteria of science, or in other words, determining the boundary between what we can consider a 'science' versus a 'pseudo-science'. Popper advocated the theory of falsificationism, where he claimed it is the falsifiability of a hypothesis which makes something scientific. Kuhn, in contrast, argued that science, in practice, is more about exploring a set of theories as constituting a single 'paradigm' and judge that paradigm more holistically. However, Kuhn has been criticised for having a somewhat relativistic interpretation of paradigms, where different paradigms cannot be compared across one another, thus making them incommensurable. Finally, Lakatos proposed the idea of a research programme in place of Kuhnian paradigms, where the 'progressiveness' of a research programme is evaluated based in how effectively it makes novel predictions (Chalmers 2013). These three key figures are explored in more detail in this current section.

(a) Popperian Falsificationism

Karl Popper (1963) held a view that the demarcation criteria for science (i.e., whether something can be considered a 'real science') is determined by whether the theory is falsifiable or not. According to this idea, a theory which cannot be proven false is not a legitimate field of study for an empirical endeavour such as science. A theory is stronger, not when it is more supported by positive evidence, but when it makes risky predictions and then resists being falsified when the predictions are put to the test. Importantly, hypotheses that make un-risky predictions are not very strong (even if unfalsified) because these hypotheses are relatively vacuous, and so constructing a theory to maximise the chances of it not being falsified does not

lead to a strong theory. A theory is considered stronger not only when it resists falsification, but when can tell us something new about the world by making novel, informative predictions.

One shortcoming of Popper's falsificationism is that it is too stringent of a criterion for realistic implementation. For example, Newtonian physics may make hypotheses about planetary movement, but if we happen to use a faulty telescope, our predictions may be proven false anyway, despite the theory itself having been a good scientific theory. According to a falsificationist approach, this would entail that the theory has been falsified, which of course, would be jumping to hasty conclusions. Popper has been criticised, therefore, that his version of falsificationism is far too stringent, and prone to producing false negatives where it risks throwing the baby out with the bathwater (Chalmers 2013, Ch7).

(b) Kuhnian Paradigms

Thomas Kuhn (1962) discussed the nature of scientific revolutions, which is when the scientific community moves from mainly supporting one type of framework (e.g., Newtonian physics) to a different, mutually incompatible framework (e.g., Einsteinian physics). Unlike Popper, the Kuhnian theory understands a framework as a collection of multiple hypotheses which are integrated and work together, and he refers to this holistic framework as a 'paradigm'. Scientific revolutions occur when there is a 'paradigm shift', according to Kuhnian language, when the scientific community begin to abandon the old paradigm in favour of the new one.

The idea of a paradigm shift is a more accurate representation of how scientific progress works in practice, compared to Popper's view that we can evaluate individual hypotheses independently. The instruments used to measure the hypotheses will often be constructed or engineered using the theories from within the paradigm itself, and also, a paradigm will usually set up its own standards for how the empirical tests are run. For these reasons, while paradigms

do resist the Popperian worry of throwing the baby out with the bathwater just because one hypothesis was proven false, Kuhnian paradigms come with a different kind of disadvantage that fair comparison across different paradigms is difficult, if not impossible. Again, this is because the kind of evidence that is taken seriously by each paradigm might differ, and there seems to be a lack of objective paradigm-less method for evaluating incompatible paradigms at a more meta-level. For instance, a believer of psychic powers can simply disagree that scientific evidence has any value, just as much as the scientist might claim that brute faith does not constitute evidence for forming a belief – and so we end up with an impasse. Thus, different paradigms are referred to as being 'incommensurable'. Due to this, 'paradigms' have been criticised as being too relativistic, or just being a matter of fashion/popularity. It does not help in determining what is a 'real science', as a paradigm can freely define its own criteria for what evidence to take seriously, which makes objective judgement difficult (Chalmers 2013, Ch8).

(c) Lakatosian Research Programmes

A research programme, as discussed by Lakatos (1978) is understood as a set of theories which are characterised by their 'hard core' and augmented by a range of peripheral assumptions and hypotheses referred to as its 'protective belt'. For example, "the hard core of Newtonian physics is comprised of Newton's three laws of motion plus his law of gravitational attraction." (Chalmers 2013, p.131). The 'hard core', also known as the 'negative heuristic' is not meant to be questioned due to it being, essentially, a definitional and integral aspect of the entire research programme. In the event where negative evidence emerges which might potentially threaten Newtonian physics, it is the 'protective belt' or the 'positive heuristics' which must be amended. For instance, if Newtonian physics predicts a particular kind of planetary motion but telescopes fail to measure the predicted results, the 'hard core' of

Newton's three laws must remain untouched, and we ought to investigate other, more peripheral elements first. This may include questioning the efficacy of the telescope or creating entirely new auxiliary hypotheses to save the research programme. (Chalmers 2013, Ch9)

The idea of creating an unquestionable core is important for two reasons. Firstly, Lakatos describes the importance of giving a new research programme a chance to develop, and to work out the details of its various peripheral assumptions and commitments. Because this process takes time, often decades or longer, it would be unreasonable to throw out an entire theory simply because it made a false predictions while the research programme was still in its infancy. In fact, there may be cases where we lack the right technology for verifying the prediction accurately, and poor instruments failing to validate predictions in its early days would not have been a very good reason for throwing out Newtonian physics altogether (Lakatos 1978).

Secondly, there are certain elements of a research programme which can be definitionally important. A Newtonian physics that is stripped of its central qualities, such as being about Newton's three laws of motion, may no longer be deserving of the title of 'Newtonian physics'. At an even more core level, we might even add that we should not question the idea that Newtonian physics is a field of physics, since to question this would merely be violating a definitional aspect of its title rather than being very helpful for evaluating the theory itself.

A Lakatosian research programme enjoys the same advantage as a Kuhnian paradigm, in that it is a collection of theories that cannot be rejected just because one hypothesis happened to be falsified. However, unlike a Kuhnian paradigm, making comparisons across different research programmes is not a matter of fashion or popularity. Instead, Lakatos discusses that some research programmes are 'progressive' in that they lead to novel predictions and gaining new knowledge, whereas other research programmes are 'degenerative' in that they only make

predictions that are *ad hoc* at best, serving only the purpose of protecting its own research programme rather than to contribute something that extends beyond it.

For example, if Newtonian physics makes a few false predictions about planetary motion, up to a point, we can make novel predictions about there being an undiscovered planet near the location of observation which is throwing off our calculations. This is a hypothesis that, if true, leads to the discovery of a novel planet. However, if the predictions keep getting falsified repeatedly, and we have to keep adding more and more 'undiscovered planets' to our equation just to make the calculations add up, and at no point do we ever discover any one of these 'undiscovered planets', then these hypotheses begin to look *ad hoc* (Chalmers 2013, Ch9).

As such, while research programmes do create a lot of room for leeway in making new hypotheses to explain away the failures of old hypotheses, the 'progressiveness' of the research programme relies on at least some of those hypotheses being supported by evidence. A 'degenerative' research programme will either never make any non-*ad hoc* predictions, or alternatively, all their hypotheses will be common sense and contain no element of 'risk'.

A 'progressive' research programme in this sense, is similar to the 'risky predictions' which Popper talks about. A 'progressive' research programme makes novel hypotheses that have implications which extend beyond the *ad hoc* quest of saving its own system, and when these novel predictions resist falsification, it strengthens the framework in a way that makes it objectively more compelling than a 'degenerative' research programme. Thus, unlike Kuhnian paradigms, different research programmes are not taken to be incommensurable (Lakatos 1978). With this terminology now available to us, we will examine what has gone wrong when people form unscientific beliefs when evidence supporting those belief systems is scarce to none.

5.2 The Empirical Problem

We saw earlier in the introduction that astrology is an example of a 'pseudoscience' where empirical evidence clearly falsifies the theory. However, to add more nuance, we saw in the previous section that Popperian falsificationism – of falsifying a single hypothesis – is not sufficient for rejecting an entire framework. That is, if time-twins were the *only* issue with astrology, then the framework itself might still be salvageable by adding further hypotheses to test, thereby further enforcing that 'protective belt' of the astrological research programme.

The issue with astrology is not so much that 'it has been falsified' as though it was proven wrong once and for all. Rather, the issue is that people who believe in astrology do not seem to take a genuine interest in its own definitional 'hard core' – that *astro*logy is supposed to be related to something to do with astro- or celestial entities. At this point, it is useful to know that there is a method behind horoscopes that makes it 'work', and the method itself has nothing to do with celestial activities. The method behind the seeming efficacy of horoscopes can be explained by understanding cold reading techniques, in particular, regarding the usage of Barnum statements (Rowland 2019). A Barnum statement refers to a type of statement that is generic enough to apply to almost anyone, but also feels personal and specific to the person who reads that statement. They tend to incorporate 'double-headed' elements, such as in the statement "At times you are extroverted, affable, sociable, while at other times you are introverted, wary, reserved." (Forer 1949, p.120) This sounds specific in that it seems to make predictions about a person's social attitude, but is 'double-headed' in that it ticks both sides of the two opposite claims: the extroverted and introverted sides. Because people tend to be extroverted at times but introverted at other times (depending on mood or situation), this statement is designed to fit the description for almost anyone (Jain & Mukerjee 2024).

Furthermore, a Barnum statement will typically involve the participant filling in the gaps and doing most of the work themselves to try and fit themselves into the statement, which amplifies the illusion of how personally accurate that statement is. For example, if one is told

that “Security is one of your major goals in life.” (Forer 1948, p.120), different subjects may interpret this question differently. One could think about their choice of stable career; a desire for a stable relationship; or need for financial stability, and these would all make the reading seem ‘accurate’. The technique of cold reading is used by astrologists, fortune tellers, and magicians, where such ambiguous statements are used to make it seem that an individual has psychic or spiritual powers, used to ‘divine’ information that they could not possibly have guessed (Hyman 1981). The effect is especially enhanced if the subject misremembers the vaguer statement that was actually said, and instead remembers the more specific version that they had filled in themselves (Rowland 2019). For example, if a fortune teller makes the statement “you will find security in your life”, the client may recall at a later date that the fortune teller told them that they “will finding a stable partner soon”. If the client reports this to a friend, it then propagates a belief in astrology that makes it seem more accurate than it is.

Because of the nature of Barnum statements encouraging subjects to fill in gaps of vague claims to sound more personal than they actually are, it can lead to a strong illusion of creating a very accurate personality assessment while still remaining generic enough to work for a large population (Farley-Icard 2007). This allows astrological readings to be published in magazines or newspapers that seem highly accurate to a wide and varied audience, particularly when the reader already expects a particular category to be directed directly at themselves, and this can lead to a confirmation bias of filling in the gaps and enhances the effect (Komath 2009). This then leads to a situation where the accuracy of an astrological reading can have nothing at all to do with actual celestial matters, but it still creates a convincing illusion that ‘astrology works’.

With this in mind, let us consider a thought experiment. Imagine that there is an astrology-believing individual who is about to break up with the love of their life because their star signs are incompatible. They are devastated to find out that their partner’s star sign was not what they had hoped, but they are utterly convinced that incompatible star signs entail that

their relationship is doomed to fail anyway. In response to this, a well-meaning friend might encourage this individual to question their belief in astrology. One might point out the scientific studies about how astrological readings do not actually correlate with personality traits, or that there is no physical basis for astrological readings (e.g., no reference to gravitational or radioactive elements; see Bok 1975). Astrology, as its name suggests, is definitionally concerned with something to do with celestial activity. And the friend could try insisting that no such correlations exist between celestial matters and human affairs, such as our personalities or relationship compatibilities, and thus it is unnecessary to break up with the love interest.

The Empirical Problem, as we are investigating now, is when a believer of a system becomes unresponsive to relevant kinds of empirical evidence, and astrology seems to be a topic where many of its believers are relatively unresponsive to being presented with counterevidence to their beliefs. Some have argued that those who claim to believe in astrology usually are not actually concerned with scientific testability or the accuracy of the belief system itself, but rather, the belief in astrology is grounded in the individuals' desperation to seek out solutions to personal problems and thereby reflects their disregard for analysing any potential causal/functional foundations astrology may have or lack (Bok 1975). It has also been argued that the issue with astrology is in its "magical world view" (Bok & Jerome 1975, p.46), or a form of "'magical' symbolism" that enhances an individual's "spiritual understanding" of self-knowledge (Clements 2020, p.259). It is an attempt at satisfying "the existential need to find meaning [by] questioning established thinking and 'rational' cultural practices." (*ibid*). In other words, when people claim to believe in astrology, it is not really about the astro- matters, but rather, what they are interested in is pure poetic storytelling. Empirical evidence plays no role.

When empirical evidence stops playing a role in our belief systems, and we begin to rely *solely* on the poetic value of the beliefs, this can be a slippery slope to other problems, such as the (II) Theoretical problem, and the (III) Pragmatic problem. These are what we turn to next.

5.3 The Theoretical Problem

Here, we will consider a theoretical problem with faith healers. 'Faith healers' refers to the idea of someone who can cure a physical ailment through the usage of some kind of supernatural or psychical power. One example of faith healing is practiced by Christian churches in the US, which professes that one's illness is healed in response to the faith that the client/patient has in God's ability to heal them. This can result in some believers refusing medical treatment for their children's illnesses in fear of this being an expression of bad faith, or to donate most of their life savings to the church as an expression of their good faith in God (Hall 2013). Problematically, the faith healers can use the "heads-I-win-tails-you-lose" technique, so that any positive results will support their own hypothesis while negative results can be explained away (Wiseman 2010). For instance, if the ailment is healed because it was a kind of ailment that was going to heal anyway, the faith healer can claim credit; whereas if the illness does not get healed, the faith healer can claim that it was the patient's own lack of faith which was at fault, thereby leaving the system itself unfalsified.

This is a 'theoretical' problem with the system in that it endorses a self-fulfilling system where all positive evidence is attributed to their own success, whereas all negative evidence is attributed to someone else's failure. This is the "heads-I-win-tails-you-lose" technique (Wiseman 2010), which thereby leads to an unfalsifiable 'perfect' system. Consider, for instance, the following logical argument:

i. If you have enough faith in God, then God will cure your cancer (assumption)

ii. Your cancer was not cured.

iii. Therefore, you did not have enough faith in God (modus tollens).

And if one tries to argue further:

iv. But I believe I *did* have enough faith in God.

Then the reply will be:

v. You must have been mistaken, because (i)-(iii) analytically compel the conclusion that you indeed did not have enough faith in God (modus tollens).

Oddly, this is perfectly logically valid. The only problem is that (i) has been accepted as a premise without having argued for it, and so the conclusion is not logically sound. However, IF (i) is accepted as a premise, then the rest of the argument does logically follow.

Now, the problem with the above is that it is an analytic argument rather than an empirical argument, where it is logically true based on a deductive argument. A characteristic trait of such 'analytic' truths is that because the conditional statement from (i) has already been accepted as an assumption, if there is any empirical element at all, then it is the client's level of faith that is being empirically tested, rather than the system itself being tested. In other words, because the conditional statement from (i) is accepted *a priori*, if there is anything that is being empirically tested here, it is the antecedent of that conditional statement.

This is a theoretical problem with the system, which has framed itself in a way that makes it unfalsifiable, in virtue of it being an analytic argument rather than an empirical one. A similar situation can occur with physicalists who advocate strong illusionism. For instance:

i. If scientists do not talk about qualitative properties, then consciousness must be an illusion (assumption)

ii. Scientists do not talk about qualitative properties.

iii. Therefore, consciousness must be an illusion.

And if one tries to argue further:

iv. But I believe I do have some kind of qualitative experience?

Then the reply will have be:

v. You must be mistaken, because (i)-(iii) analytically compels the conclusion that you *do not* really have qualitative experiences and consciousness is an illusion (modus ponens).

This would lead to a form of physicalism that is an analytical truth, but only because premise (i) has been accepted in a way that makes this conditional statement itself invulnerable. We have seen many other examples such as this in chapters one and two, where physicalism sets itself up as a tautology, or uses circular reasoning in order to create an invulnerable theory.

Now, at this point, a critic might argue that to some extent, there is nothing wrong with assuming the truth of some conditional statements. If anything, it is impossible to construct a deductive argument without assuming anything at all, and so perhaps, an opponent may argue that nothing is wrong with constructing a deductive argument where premise (i) is unfalsifiable.

The issue, once again, is in the question-begging. In circumstances where we are asking *whether* premise (i) is true or not, then it impermissible to simply assume that it is true. In other circumstances where it is not premise (i) that is under scrutiny, then there may be situations where accepting it as an assumption is taken to be permissible. That is to say – context matters.

In the faith healing example, if what we want to know is *"Does* having enough faith in God cure cancer?" then accepting as a premise that "If you have enough faith in God, then God will cure your cancer" would be problematic. However, in this situation, daring to ask the crucial question *"Does* having enough faith in God cure cancer?" might itself be taken as lacking faith in God. After all, if you truly had enough faith in God, then you would not be asking this question to begin with. Therefore, if we want to have any chance at all that God might heal our physical ailments through faith alone, then it makes it impossible to ask this question in the first place, once again, leading to the "heads-I-win-tails-you-lose" situation.

In the physicalism example, if what we want to know is "Is the fundamental essence of consciousness something that is analysable through scientific methods?" then accepting it as a premise that "If science cannot account for consciousness, then consciousness does not exist" would be problematic. Consider, for a moment, an Evil Physicalist who decides that if consciousness is merely an illusion, then there is nothing wrong with harming people, since any 'pain' that is caused to another human being is also nothing more than an illusion anyway. Under these circumstances, it is extremely important to stop and ask, "Is pain Real, or is it an $\text{Illusion}_{(B)}$?" and also ask whether science or physics is really the field we should be turning to in order to have these ethical questions answered. We might also wish to explore the link between ontology and ethics, where once again, it can be important to question whether turning to science is an appropriate move to be making under these circumstances. If, however, the Evil Physicalist were to insist that questioning expert scientific opinion just is "unscientific" and therefore is a sign of ignorance or naivety, it would become impossible to ask the question in the first place without sounding 'ignorant' or 'naïve' in the process of asking that question.

Now, this is not to accuse physicalists as being 'evil', but what we are looking at now is that there is a slippery slope from the (I) Empirical Problem to the (II) Theoretical Problem, to the (III) Pragmatic Problem. That is, (I) If believers of a research programme do not take genuine interest in the 'hard core' of the thing they claim to be interested in, they may overlook important evidence. This goes for supposed astrology-enthusiasts who lack any interest in exploring the causal workings of celestial matters; faith-healers who supposedly are interested in *whether* faith has healing powers or not (but might actually be interested only in the monetary gain that comes with it); and also maybe for a physicalist (physics-enthusiasts?) if they commit to oaths, such as: "I hereby swear to go in my ontology everywhere and only where physics leads me." (Ney 2008b, p.5) without being interested in the *content* of physics, in its own right.

These problems can then lead to (II) Theoretical issues, where a self-fulfilling set of deductive arguments are created to further perpetuate the system in a way that makes it invulnerable to being attacked from the outside. Finally, in some extreme cases, this can risk leading to (III) Pragmatic worries, that people will begin behaving in ethically questionable ways if problematic belief systems are taken far more seriously than they deserve. With this in mind, we now turn to such damaging effects that can be caused by $\text{Illusory}_{(B)}$ belief systems.

5.4 The Pragmatic Problem

We have already seen some instances where a belief in faith healing can lead to damaging outcomes for the believer, such as when a person throws away medication or donates away their entire life savings as proof of their 'faith' in God. We also saw this in the earlier thought experiment about the astrology believer who abandons the love of their life after discovering that they have incompatible star signs. In these types of examples, we can see that forming false beliefs is not just a theoretical vice, but can lead to making big life decisions based on false beliefs, with potentially damaging real-life consequences. That is, it can interfere with a person's ability for informed decision-making, thus adversely impacting one's quality of life.

It is one thing to use astrology as "'magical' symbolism" that enhances an individual's "spiritual understanding" of self-knowledge (Clements 2020, p.259), but it is another thing to use these belief systems to satisfy "the existential need to find meaning [by rejecting] 'rational' cultural practices." (*ibid*). It is important to remain rational when constructing our own beliefs, not just because it is theoretically better to be 'less naïve', but because of its *pragmatic* concerns.

One major difference between a magic show and something which is advertised as a 'psychic' show is that they differ in the level of honesty. Sometimes, 'psychic' shows such as faith healing demonstrations, mediumships (where the psychic claims to be able to put you in

touch with deceased relatives), and other performances that supposedly involve supernatural powers will use exactly the same kinds of techniques used by magicians. However, because it is easier to make money by advertising the show as a psychical performance, this leads to the emergence of fraudulent 'psychics', which are referred to as pseudo-psychics. Again, the biggest difference between a magician and a pseudo-psychic is the level of honesty (Lamont & Wiseman 2005; Ch4). In a magician's performance, we are not told to rethink our entire metaphysical picture of reality and make real life decisions based on what we see on stage. Rather, it is advertised as a form of entertainment. In contrast, psychic performances are presented as demonstrations of genuine 'mental' causal powers which transcends 'the physical'.

If a physicalist is sceptical of this kind of 'fundamental mentality' that is advertised by pseudo-psychics, then this is understandable. When there is no reason to believe in the existence of Harry Potter magic, then we should not be accepting supernatural magic into our ontology all willy-nilly. And if we are in a context where we are not presented with any reason to take souls, ghosts, or spirits seriously, then all of these may be categorised together as 'spooky' supernatural concepts that do not exist, and if we come across any similarly 'spooky' concepts in philosophical discussion – such as a Cartesian mind – we might be tempted to reduce it to something a little more robust and grounded. However, once we take a closer look, we can see that when dealing with these fraudulent 'supernatural' belief systems camouflage as something Magical (but is actually just a false belief propagated for other emotional or monetary reasons), the solution is not to double down on an equally unfalsifiable faith in 'science'. Rather, the solution is to understand what exactly has gone wrong in fraudulent cases.

Take, for example, Rhonda Byrne's popular self-help theory, The Law of Attraction (Byrne 2007). In this self-help theory, an ontological claim is made that there exists a certain frequency of vibration in our thoughts which resonate directly with the universe, which allows our wishes to come true simply by believing in them hard enough. It is a form of fundamentally

'mental' causation, but in a much more literal and stronger sense than what dualist philosophers such as Lowe (2008a) would argue for. Instead, Byrne's theory is asserted in quite a literal sense that there actually are vibrations in the universe that makes our wishes come true, just by thinking about things positively enough. This is an extremely popular book, particularly in the US, as it promotes a positive attitude in advocating the special power found in 'self-belief'. It leads to a (supposedly) inspirational message that if you believe in yourself enough, anything is achievable, and this message can at times be an encouraging thing to hear. However, there is a very big difference between making a psychological claim such as "self-belief can increase resilience so you will keep at your dreams for longer even if you fail initially" and making metaphysical claims like "there are vibrations in the universe that make your beliefs come true."

Once again, we can return to magician Derren Brown, who takes immense issue with this exploitation of 'magical thinking', claiming that Byrnes's argument essentially amounts to the claim: "Believe in yourself, blame yourself" (Brown 2016, p.25). It is all very well when the method happens to work – perhaps there are times in life when having extra confidence is just the boost we need for pursuing our goals – but when it does not work, it leaves the person not only confused about 'why' it did not work, but also inclines them to blame themselves, as though it was their lack of self-belief that *caused* the failure of the action. He gives an example of a woman who wanted to start her own coffee business. So long as the business is working for contingent reasons, it might not matter too much what it is that makes it work. Maybe the coffee shop just happens to be in a good location where customers will come anyway. However, running a business tends to involve a variety of complex decisions needing to be made, and a brute 'self-belief' might not be enough for making all of the necessary logistic judgements. Problematically, if the owner believes The Law of Attraction is *causally* responsible for her success or failures, then she will derive an oversimplified explanation about why something is or is not working, making her unable to improve or adapt to external circumstances by making

helpful predictions. The magician continues to argue, "*any questioning of the system* would be in itself evidence of a lack of faith; not only must one self-deceive but one must also remain fearful of any rational enquiry. A sort of neurotic bravado must be maintained, and to admit to oneself that it could be defective would be to assure the instant failure of the magic." (2016, p.40). This places a believer of The Law of Attraction, or of faith healing, in an awkward position. They are not permitted to question the theory without self-sabotaging in the event the theory is true, as 'lacking faith' actively *prevents* desired outcomes from causally manifesting.

Coffee shops do not manifest just because we wish for them, and the universe does not just decide to hand us what we want because we asked nicely. Most individuals might believe that this is a fairly obvious thing, but the magician warns us that human beings are surprisingly susceptible to being deceived by storytelling, unaware of the limits of our own perceptions:

> "When I perform my day job as a kind of magician, I work with people's capacity to fool themselves with stories. A good magic trick forces the spectator to tell a story that arrives at an impossible conclusion, and the clearer the story is, the better. Normally, everything you need to solve the puzzle happens right in front of you, but you are made to care only about the parts that the magician wants you to. When you join up those dots, so misleadingly and provocatively arranged, you are left with a baffling mystery." (Brown 2016, p.4).

> "We are, each of us, a product of the stories we tell ourselves. […] Some of these stories are consciously constructed, but others operate without our knowledge, dictated by scripts handed to us by others when we were young. We can carry around the psychological legacy of our parents for our whole lives, whether bad or good. Where they have unfulfilled wishes and regrets, these are commonly passed to us as a template for storytelling. Many of these templates make it hard for us to feel happy. […] Similar insidious directives can also come from the Church, our peers,

> classmates and teachers, the cumulative effect of the news media we encounter daily or any number of ideologies in which we find ourselves enmeshed." (*ibid,* p.6)

We have seen in the previous chapter that, by profession, magicians tend to be highly aware of the distinction between 'method' and 'storytelling'. It is all very well to tell a poetic and inspiring story, but when this starts to interfere with being able to make informed decisions about 'how' to achieve our goals, 'how' to run a successful coffee shop, 'how' to cure physical ailments, or 'how' to understand consciousness in ways that avoid ethical problems, then poetry is not enough. The concept of Magic requires both method and nuance for it to work as intended.

> "When we reduce complex ideas to nouns and categories in order to navigate swiftly through them, we start to become mindless." (*ibid,* p.523)

This can become a worry, not just for supernatural belief systems, but for 'scientific' belief systems as well. In the next chapter, we will be seeing how the same issues can affect the scientific field of psychiatry, when complex mental health problems are reduced to nouns like ADHD or autism, thereby creating a confusion about how to interpret such labels that balance a scientific rigor while meeting the social-political needs of the patients who are being labelled.

In the previous chapter, we began to explore the idea of philosophising about the mind from the perspective of the (a) Eastern (b) Magician's (c) Apprentice. This involved the conjunction of (i) scepticism as the default, (ii) empiricism or taking 'contact points with reality' seriously, and (iii) not trivialising the poetic storytelling where it is philosophically required. This was inspired by Eastern Buddhism, as well as the Magician's unique profession of distinguishing between method and storytelling (and enhancing this gap to create $\text{Illusions}_{(A)}$),

but also the idea that there is a sense in which we are all Apprentices to life. We are all living on this planet, navigating our daily lives based on the beliefs and information that we have come to accept. In order to make 'good' decisions in life, this is not just about becoming 'less naïve' in a theoretical sense, but it also involves the practical wisdom of learning what kinds of information to take seriously when making important life decisions. $\text{Illusions}_{(A)}$ can inspire us and enhance a quality of life; $\text{Illusions}_{(B)}$ are by definition deceptive, and can cause trouble.

It is in this sense that consciousness is not an $\text{Illusion}_{(B)}$ in the sense of 'deception' or something which is 'not real', but rather, it is an $\text{Illusion}_{(A)}$ in that it creates room for non-trivial and meaningful storytelling. Without the meaningful storytelling behind the nouns and labels, it can become difficult to see why a theory is philosophically or metaphysically interesting in the first place. However, if it is *only* storytelling, then this can start to impinge on our ability to make informed life decisions based on accurate, credible information about cause-and-effect.

The error with physicalism, as explored throughout the first half of this thesis, seems to be that it focuses exclusively on (I) The Empirical Problem, and (II) The Theoretical Problem of its opponent's theories, but to my knowledge, is almost always silent about (III) The Pragmatic Problem of forming false beliefs. Without the 'contact points' with reality, where we pay attention to pragmatic implications, physicalism itself runs the risk of turning into an overly abstract, vacuous theory. It focuses exclusively on pure 'theoretical' knowledge.

Consider for instance, Aristotle's criticism of Plato who "made a serious mistake in failing to mark the boundaries separating the different sciences and branches of philosophical inquiry. All of them failed to grasp the fundamental distinction between practical and theoretical knowledge" (Devereux 1986, p.483). Aristotle argues that "we cannot regard practical and theoretical knowledge as two species of a single genus, for there is no common definition of knowledge which applies to both. […] Platonic dialectic is a confusing blend of

the practical and theoretical." (*ibid*) The word 'wisdom' as used by Plato refers to a "highest form of knowledge which is at once practical and theoretical." (*ibid*) as though they are one in the same, thus failing to make an important distinction between the 'practical knowledge' that impact our lives in some way, and 'purely theoretical knowledge' like mathematics. (*ibid*)

Aristotle claims that practical knowledge is concerned with *particulars* and not *universals*. While what constitutes 'good' knowledge in theoretical endeavours may invoke an appeal to *universals*, this is not the same sense of the word 'knowledge' that would apply in the context of practical wisdom (or *phronesis*), which involves discussions on how a *particular* (i.e., individual) can act well or live well, according to rational considerations (Devereux 1986).

In this light, what we may wish to do is to create two separate categories of 'knowledge': Those which are required to have 'contact points' with reality in some empirical way (e.g., feedback-responsiveness; falsifiability responsive; hypothesis-testing responsive), and those which are not (e.g., pure deductive logic). These ideas are discussed in greater depth in chapter seven, when we see there are detrimental shortcomings in purely deductive, analytic arguments.

The problem we can take notice of for now is that when we focus only on 'knowledge' acquisition about *universals*, then we can fall into the trap of developing an elaborate theoretical 'knowledge'-base that has little to no bearing on real-world matters. This can range from neutral outcomes of being a rather impotent philosophical theory that has no impact towards any other field (i.e., a 'degenerative' research programme in it of itself) or it can result in an outright damaging philosophical framework that perpetuates logically unsound deductive arguments, which are based in the unquestioned acceptance of problematic premises. This can lead to arguments that might still look convincing, as they are logically 'valid' (but not 'sound').

5.5 Conclusions/Summary

In the previous chapter, it was argued that consciousness is inherently Magical. In the same way that reducing a magic trick to only its methods can 'explain away' the very thing we are trying to explain, a philosophical interpretation of consciousness also requires method, poetic storytelling, and an acknowledgement of the gap in between to fully wrap our heads around the concept. However, the notion of Magic is often confused with fraudulent belief systems, including pseudo-psychics, or pseudosciences such as astrology, and it is important to discuss in what sense consciousness is $\text{Illusory}_{(A)}$ (i.e., whimsical and wonderous), without it also being $\text{Illusory}_{(B)}$ (i.e., deceptive or something that is 'not real'). In this chapter, three issues were discussed with problematic belief systems: (I) The Empirical Problem, (II) The Theoretical Problem, and (III) The Pragmatic Problem. The first refers to a research programme that is supported despite its hypotheses being very bad at resisting falsification, and it can be caused by its believers not taking the 'hard core' of its own research programme very seriously, thereby resulting in a 'degenerative' and *ad hoc* research programme. This can be a slippery slope towards the developing of the second problem, where deductive arguments accept a question-begging premise as part of its assumption, and derive conclusions in such a way that makes the system itself theoretically invulnerable. This can use the "heads-I-win-tails-you-lose" technique (Wiseman 2010) to create the $\text{Illusion}_{(B)}$ that the belief system is indeed empirically supported – such as by passing off the question-begging premise as something which is empirically justified, even though in reality, it is accepted only as a brute assumption. Finally, this can be a slippery slope towards the third issue: of causing pragmatic problems in people who succumb to a problematic belief system and end up making real life decisions based on it.

The solution to avoiding this slippery slope is to perhaps move in the reverse direction. We can start by being mindful of the pragmatic implications that our own pet theories might cause. Even if one takes oneself to be involved in a purely theoretical pursuit of 'knowledge' (in the Platonic sense, rather than the Aristotelian sense), it does not hurt to be mindful of the

practical impact that our theories can have (or lack) towards other fields of research. Especially if we are to accept a monist ontology, and claim that the universe is fundamentally unified, it is unsurprising that a philosophical interpretation in metaphysics could very well have impact in other academic fields, and that academia, in actuality, is more integrated than what a vertically-divided bureaucratic structure might imply. What we can do, therefore, is to enhance cross-field research and become more attuned to identifying 'contact points' with reality, where we are able to check not just the logical validity, but also the logical soundness of our arguments when our theories come into contact with real-world practical or pragmatic considerations.

This will be the task of the next chapter: To see the impact philosophy of mind can and has had in the field of psychiatry. As we explore the next chapter, we can continue to bear in mind the question: "What is the fundamental essence of consciousness?" Ultimately, these discussions in this chapter and the next few chapters will lead to an argument in the final chapter that consciousness is Magical – that Buddhist 'karma' and 'souls' should exist as a part of our ontological picture in a metaphorical kind of way – but also that philosophy should not lose sight of realism and pragmatism in order to remain grounded and intellectually rigorous.

6. Psychiatry, Hypnosis Shows, and Human Agency

Psychiatry is a field of medicine that is concerned with psychological forms of disruptions (where 'psychiatrist' refers specifically to individuals who have graduated with a medical degree, not a psychology degree). To be struggling with mental wellness, to some extent, is an extension of our 'normal' existence, and some type of conceptualisation is required for behaviours or mental states that fall into the category of a 'special' struggle which deviate from healthy levels of experiencing the ups and downs of life. The concept of 'mental illness' is used to refer to these deviant cases that explores the edges of human psychology, where the difficulties or struggles expressed by the individual is counterintuitive from a pre-theoretical perspective. There are some parallels here with hypnosis shows performed by magicians, where some audience members who witness a participant collapsing on stage might settle for a crude explanation that a hypnosis show is 'all just faked' or that 'people are just pretending to be hypnotised'. A superficial analysis of clinical depression can also yield analogous explanations, such as 'depression is not real' or 'they are merely lazy and just faking it/over-exaggerating'. The commonality between the two is that both a hypnotised participant who suddenly collapses on stage, and a depressed patient who cannot get out of bed in the morning, both appear to be experiencing a lack of agency or control that we normally do not experience in our everyday lives under a typical mental state. When it comes to exploring the edges of human psychology, we can start investigating what is happening when agency does not work as we might expect.

In sections one and two, we will cover some important historical overviews that link mental illness to this concept of agency. This begins with the Victorian asylum days where 'mental illness' was conceptualised as a lack of agency, in the sense of having rabies in that one becomes incapable of rational self-control; and in section two, for contrast, we will explore the modern 'neurodivergence' debate, which stems from similar origins but takes a different turn. In section three, we will spend some time analysing what happens on stage in a hypnosis

show, such as when a participant collapses in response to the command to 'sleep', or a participant experiences catalepsy, and their feet are inexplicably 'glued' to the floor. While the psychiatry examples are heavier and more serious, and the hypnosis examples are light-hearted and fascinating, there is an underlying similarity which highlights just how little we understand when it comes to an analysis of agency within an unusual social environment. In the final section, we tie this back in with our ontological discussion. A view such as physicalism is not particularly helpful for distinguishing when 'I' am in charge of my own agency, and when such self-control or accountability is not warranted because there is a physiological compulsion that *causes* the behaviours. Importantly, this kind of discussion cannot be had unless we can make some headway with what this 'I' entity is. If we reject dualism, we might also reject the somewhat naïve imagery that there is a homunculus controlling the brain or that there is an 'I' watching a Cartesian theatre and using mental causation to interact with reality. But then how do we wrap our heads around 'agency'? There is not a tiny person inside my brain, watching a Cartesian theatre and controlling my behaviours, and yet, in a non-hypnotised lucid state, there is still a sense of ME – as some kind of holistic entity – looking out through my own eyes and controlling my own behaviours. But what exactly is this semi-magical 'ME' entity if it cannot be captured in physical nor dualist terms? This is what we turn to in this current chapter.

6.1 The Origins of "Mental Illness" as Lacking Agency

It has now become a relatively common practice in the US and UK to use psychiatric diagnoses to determine disability support or concessions that a student may receive at schools and universities. There is a growing influence of institutionalised 'mental wellness' campaigns, which have increased in societal impact in recent years (Lipson, Lattie, & Eisenberg 2019). This makes it a societally pressing matter how accurate the information is that is propagated

from official sources of information, including universities, clinicians, and the DSM – the Diagnostic and Statistical Manual of Mental Disorders (APA 2022) – referred to by some as the bible of psychiatry (Horwitz 2021 introduction; Szasz 2010).

While recent years have advocated the normalisation of mental illness to reduce stigmatisation (Stuart 2016), because the topic of psychiatry is typically concerned with 'abnormal' states of mental unwellness, there is a conflict of interest in normalising something which is definitionally abnormal. On the one hand, we may wish to characterise 'mental illness' as being distinct from other more everyday concepts: such as laziness, weakness of will, bad intention, or even stupidity. For instance, a person with depression is not being lazy, and a person with ADHD is not 'stupid' or 'weak willed' as some unconstructive teachers may try to claim. Self-help books are published with these types of messages that help to create a more empowering way of thinking (e.g., Kelly & Ramundo 2006), and the mental health movement focuses on the reduction of stigmatisation, taking a person-centred approach, providing greater levels of help to those who need it, and empowering self-images of the patients (Sunkel 2012; McDaid 2010; Nelson, Lord, & Ochocka 2001). Again, the Brain-Blame Dichotomy highlights the tendency of laypeople to either blame a person for their shortcomings, or to pin the problem on their brain – in the sense that the individual is a patient who is suffering from an issue that falls outside their agential control (Boyle 2013). This has the dualist undertones in its implication that the 'person' is distinct from the 'physical brain'.

> "The idea that insanity was fundamentally different from other illnesses, that it is a disease of the mind rather than the body, only developed towards the end of the 18th century. The scene was set by Cartesian dualism, the dominant philosophical influence at the time" (Kendell 2001, p.490)

What is important about the Cartesian dichotomy for our current purposes is the nuance about agency which it conveys to individuals who are not versed in the philosophical discussions, including members of the public as well as some clinicians. The connotation behind blaming the person is indicative of a moral flaw that the individual is supposed to address in terms of either will-power or religious moral condemnation (Kendell 2001), whereas blaming the brain carries the connotation of releasing the person from such blame and accountability. There is, however, an important distinction between the modern conceptualisation of 'mental illness' and the 'insanity' from the nineteenth century mental asylums.

In the old context, people who were deemed 'insane' were sectioned off as a means for keeping other members of society safe from the cognitively volatile, often at the request of family members who required a method for controlling and confining violent, uncontrollable relatives (Write 1997). The issue here is that precisely because mental illness is not a case of moral shortcomings (where a person can be reasoned with), but it is categorically different in that the individual lacks the agency or rationality to engage in the moral game where individuals are held accountable to their actions (and judged accordingly), drastic measures were required to keep these individuals contained. 'Mentally ill' individuals in this context were perceived as a threat due to a lack of agency, much like in the case with rabies victims:

> "The therapies that were attempted against rabies may be particularly enlightening. By the nineteenth century, physicians had learned through experience to control certain diseases - scurvy, for example - and, of course, people generally recovered from many diseases regardless of what the physicians may have done. But rabies remained confusing and frustrating. The disease was almost always fatal and involved terrible suffering. Confronted by a rabies victim, therefore, physicians were under extreme pressure to do something, but the accumulated experience of the profession simply did not justify any specific therapeutic strategy. In precisely this kind of

> situation - a situation in which one must act but in which experience does not provide clear direction - one might expect to find the influence of shared theoretical assumptions underlying a research programme." (Carter 1982, p.68)

The usage of the word 'research programme' in the above quote refers explicitly to the Lakatosian concept we discussed in chapter five of this thesis (*ibid,* p.68). Carter continues to describe that prior to the work of Pasteur (1885), a rabies cure was searched for in a haphazard manner by administering random treatments that worked for other conditions, hoping something would eventually work. The haphazardness in medical treatment was linked to the medical research programme of the nineteenth century which categorised physical ailments in terms of symptoms rather than by a shared necessary cause. Rabies, for instance, was often referred to as 'hydrophobia', and while the correlation between animal bites and this condition was acknowledged, some physicians believed that the mere idea or the fear of a bite was enough to cause hydrophobia in some patients even without actually being bitten (Carter 1982).

If a disease can be caused in various unrelated ways, then "it is not irrational to expect that unrelated, even contradictory, treatments may be required" (Carter 1982, p.77). Therefore, Pasteur's work on rabies marked the transition from the old research programme that was based in categorising diseases by symptoms, to a new research programme that defined diseases based in shared cause. By assuming that all cases of rabies may have a common cause – and *defining* rabies as a disease that has this particular micro-bacterium as its cause – physicians could disregard the extraordinarily large literature which existed on haphazard theorisations about rabies which were confused and contradicting. As Bacon notes, "Truth emerges more readily from error than from confusion." (Bacon 1869, p.210; from Kuhn 1970, p.18), and the new research programme paved a way of discrediting all the erroneous information.

For these reasons, Pasteur became known as the father of immunology, as his studies popularised the germ theory of disease and "introduced the hope that all infectious diseases could be prevented by prophylactic vaccination" (Smith 2012, p.1). The concept of an immune system did not exist back then during Pasteur's time however, and so the idea that we have a natural system in our bodies that discriminate between different types of micro-entities was an entirely new concept at the time (*ibid*). The symptom-based structure was a degenerative research programme in that it merely applied methods that worked for other diseases, but in a chaotic manner which did not lead to the formation of any novel unfalsified hypotheses. Instead, all the hypotheses regarding rabies cures were rejected, and there was no reason to remain in the old research programme aside from it simply being a matter of convention (Carter 1982).

Not only in the context of rabies, but more generically in the days of mental asylums in Victorian times, the "mentally ill" were understood as people who were a danger to oneself and to others due to their lack of reasons-responsive agential control. Confinement facilities were used, not only to keep the "mentally ill" away from the rest of society, but also to serve as a deterrent for other people to relinquish their agency in inappropriate or destructive ways by attributing a sense of shame to the concept of mental illness (McCandless 1983). In this context, it was preferable to be blamed for exercising agency irresponsibly than to identify as not having agency at all, as under this old framework, 'blaming the brain' for one's behaviours is a slippery slope to claiming a more animalistic and dehumanising state of mind.

The mental asylums of the nineteenth century were confinement facilities designed for the benefit of the rest of society, rather than for the benefit of the mentally ill. "Although premised upon belief in the efficacy of 'moral and humane' treatment, these asylums soon became custodial rather than therapeutic institutions." (Chatterjee 2022, p.1), and "these psychiatric asylums were 'dystopian' spaces, built as part of state-guided 'sanitary' movement which sought to eliminate all forms of deviance from the normative social fabric." (*ibid*).

In our modern understanding of mental illness, the topic of discussion is rarely to do with the violent behaviours and mental states of a rabies victim. However, this can in turn create a novel kind of ambiguity where dangerous states are not easily distinguishable from states that are merely a nuisance (to either the patient themselves, or to others around them). There is a difference between a person with schizophrenia who hears hostile voices in their heads making them inclined to harm others, and someone else who also hears voices, but in a way that is not at all dangerous albeit uncomfortable for the patient (Boyle 2013). Because the current research programme categorises patients based in symptoms rather than cause or levels of danger, this leads to stigmatisation for those who are labelled under the same category as someone who might be a threat towards society (Ferriman 2000). This lack of consistency even within a single disorder is called the heterogeneity of a disorder, and creates a controversy in psychiatry about how to define the disorders more precisely (Koutsouleris & Fusar-Poli 2024).

What is important for our current thesis is to see the relationship between agency and mental illness. According to the historical usage of the term ‘mental illness’, it is clear why this term implies an impairment of agency: That the individual cannot be reasoned with rationally and therefore poses a chaotic or animalistic threat towards other members of society. The notion of agency, therefore, becomes important if we are considering where the burden lies when a problem needs to be addressed. If a person fundamentally cannot be reasoned with, and if we have no other ‘cure’ for the delirium itself, then we cannot expect the ‘insane’ individual to exercise a controlled agency in favour of the moral impact they have on others. It is not so much an issue about how it makes the ‘mentally ill’ individual feel, as ‘mental illness’ was reserved for these more extreme cases, distinct from a more common ‘mental unwellness’.

The modern context, however, is somewhat more ambiguous. The concept of mental health is sometimes handled interchangeably with (a) the medicalised interpretation (i.e., the person is a passive ‘patient’ who cannot but be the way they are due to a chemical imbalance

or biological neural wiring) and (b) a discussion about mental prosperity (i.e., that certain mental states are unhealthy because they bring about stress or other unnecessary disturbances). The former is a descriptive, scientific debate, but the latter is a normative, value-based debate.

The scientific discussion about analysing micro-physical states is a separate debate from the philosophical issue on mental wellness. In fact, there is a further question that must be asked here before proceeding: What is the *goal* when it comes to investigating the matter of mental wellness? The reason this is an important question to ask is because in some situations, one might argue that it is preferable to tell people whatever they want to hear as a means for increasing their level of mental wellness. But if 'prescribing astrology' happened to make people feel less depressed, would we think that this is an appropriate 'treatment'? If we allow for a moment that any form of deception or misdirection is permissible so long as it makes people feel happier, then it would open up the discussion to a philosophically problematic territory: Is the purpose of psychiatry to make someone feel better even if it involves misleading them; or is the purpose about investigating the edges of human psychology so we gain a more accurate understanding of how psychological mechanisms work? This is a thought to bear in mind as we proceed through the rest of this chapter.

6.2 Transition from 'Disorder' to 'Neurodivergence'

The 'neurodiversity paradigm' (in the Kuhnian sense of 'paradigms'; see Hughes 2021, p.50) refers to a way of conceptualising mental atypicality not as a disorder, but as an extension of personality traits which fall outside what we take to be normal or common (*ibid*). In particular, autism spectrum disorder (ASD) has been a highlight in conversations about neurodivergence.

> "Since its initial description, the concept of autism has been firmly rooted within the conventional medical paradigm of child psychiatry. Increasingly, there have been calls from the autistic community and, more recently, nonautistic researchers, to rethink the way in which autism science is framed and conducted. Neurodiversity, where autism is seen as one form of variation within a diversity of minds, has been proposed as a potential alternative paradigm." (Pellicano & den Houting 2022, p.381).

The criticism against the 'medical paradigm' focuses on the overemphasis it places on a person's deficits, which can have damaging impact for one's self-image and exacerbate their mental unwellness. This, however, creates a tension when we consider that a medicalisation of a condition can be necessary to justify why the individual requires specialised care in the first place, as we do not have the resources to provide special care for everyone who feels socially deviant. But to medicalise a condition inevitably ends up focusing on its more negative traits that constitute the individual's deficits, since the purpose of a medical approach is to analyse health problems – and health problems are inherently unflattering. Thus, this 'neurodiversity paradigm' has led to a controversy among parents of autistic children as well as individuals with autism, who "have accused neurodiversity advocates of presenting a sanitized view of what autism can be like and deflecting attention and resources away from the struggles of more severely affected individuals and their families." (Hughes 2021, p.47).

To understand the nature of this controversy, it is important to look at some uncomfortable information about how this 'neurodivergence paradigm' has developed as a response to a dehumanising beginning in the history of autism. What is now known as ASD today is a combination of what used to be called Asperger's syndrome and autism. Prior to this change which took place in the DSM-V (APA 2013), autism and Asperger's syndrome were considered to be different conditions, although part of the same family. Both autism and

Asperger's syndrome referred to "socially unusual children with narrow interests and repetitive behavior" (Ozonoff, South, & Miller 2000, p,29). However, a crucial distinction was that autism referred to children with speech delays or other cognitive difficulties particularly in relation to language development, whereas Asperger's syndrome was defined as an absence of these traits (*ibid*). This led to hypotheses about whether Asperger's syndrome was just autism with a higher-IQ (Szatmari et al., 1995, p.1667), or possibly, it might be the speech delays that *cause* the awkwardness in autistic individuals (since it is frustrating to not be able to use language to communicate one's needs), but the similar awkwardness may have a different cause for those with Asperger's syndrome (Klin 2006; Vokmar & Klin 2000; Kugler 1998).

One of the reasons for combining the categories was related to a discovery of lost archives in 2010 regarding the history of Hans Asperger, who initially coined the term. The archives exposed that he had been a member of the Nazi party, and the 'disorder' was founded in this context. "Hans Asperger (1906–1980) first designated a group of children with distinct psychological characteristics as 'autistic psychopaths' in 1938, several years before Leo Kanner's famous 1943 paper on autism" (Czech 2018, p.1). This was a part of a race hygiene project of 'cleansing' society of social deviants by sending children with problematic traits to a "euthanasia" clinic as a part of the Nazi eugenics programme (*ibid*, p. 4). This unpleasant history of Asperger's syndrome played a role in removing of Asperger's syndrome from the DSM-V (APA 2013) and unifying the condition into autism spectrum disorder (Sheffer 2018).

According to the Nazi view of Asperger's syndrome, this was clearly defined in terms of its negative characteristics in a way that is similar to how "mental illness" was understood in Victorian mental asylums. It was perceived as an absence or impairment of social skills that prevented the citizen from becoming a 'good follower' under the fascist regime (Czech 2018). With this background in mind, the initial definition for Asperger's syndrome was inherently negative, and biased in how it presupposes socially compliant attitudes as being 'normal' and

preferable. This predates Leo Kanner's famous paper on autism that spread the concept to English speaking countries (Kanner 1944), which is thought have developed independently from Asperger's claims about socially deviant children. Therefore, unlike Kanner's definition of autism, Asperger's syndrome was defined based on a negative value judgement against people who were destructive towards a fascist regime and needed to be "cured" (Czech 2018).

It is understandable, under this context, that a gap emerges between two camps. Some wish to reinterpret Asperger's syndrome in a positive light as a set of quirky or unique traits – and this may be referred to as neurodivergence (Giles 2014). The other group wishes to identify as having autism in the sense of requiring a special care from institutionalised systems and has its historical roots more in Leo Kanner's (1944) theory on child development. However, because the two categories are now merged together (and were treated as part of the same 'family' of disorders due to superficial similarities even before their formal integration), it can be difficult to tease apart the issue of 'neurodivergence' from the functional challenges.

The discomfort that individuals on the spectrum report is the idea of being defined in terms of functionalities. To understand Asperger's syndrome as 'high-functioning autism' was seen as an affront to the individual, as it reduces their phenomenal experiences to matters of mere functionality, thereby leading to a dehumanising self-image (Anderson-Chavarria 2022). The link between the reductive definition of autism as a matter of biological nature is also linked to the discomfort of having one's essence defined in terms of functions rather than subjective meaning. "While the neurodiversity paradigm does not [always] challenge the notion that autism is biological in nature, it stresses the need to view autistic people, not as a collection of 'deficits' needing to be 'fixed' but as unique and worthwhile individuals, whose lives have meaning and purpose." (Pellicano & de Houting 2022, p.389).

Neurodivergence was introduced in 1999 by Judith Singer, who described a political view of neurological diversity as being similar to the political categories of class, gender, or race, to nudge autism away from the medical view (Hughes 2021). Walker (2013) uses the Kuhnian term 'paradigm' to refer to this movement, discussing that the neurodivergence framework is like a lens that one may use to view reality. It is socially defined in a way that makes it more about the popularity of the framework, rather than being based in an objective comparison.

> "It would be surprising for a movement that views itself as an extension of the disability rights movement (Walker 2016) to deny that its members are disabled, but analogies drawn between autism and identities such as race, gender and sexuality, and the denial that it is a disorder, may seem to suggest such a view." (Hughes 2021, p.50).
>
> "Advocates of the neurodiversity paradigm are often among the strongest supporters of such [disability support] measures, but the fact that positive interventions are needed suggests that the traits that render them necessary may be intrinsically disadvantageous." (*ibid*, p.54)

The problem with the Kuhnian interpretation, however, is that paradigms "shift for no rational reason, but are sociological phenomena resembling "religious conversions."" (Harman 2019, p.2068). This leads to two problems: (1) Individuals who support the neurodivergent movement will likely be unhappy for their movement to be seen as 'a religious conversion', and (2) Disability support systems tend to rely on the idea that the individual is diagnosed based on some kind of a scientific objective standard, again, due to the problem of limited resources.

If a collection of socially disadvantageous traits turns out to just be a unique personality – and is not even that uncommon (Fombonne 2023; Abdelnour, Jansen, & Gold 2022) – then

it begs the question of in what sense these traits are still considered 'socially unusual' or 'divergent'. Furthermore, as Hughes (2021) argues, changing the social norms in such a way to benefit one community can have adverse impact on other communities, which includes the matter of allocating limited resources, as well as changes in social expectations more widely, as well as the effects this can have on a merit-based social structure which tries to encourage independence and growth rather than encouraging people to feel dependent on institutions because they are convinced they are biologically incapable (or disabled) from standing on their own feet. This is not only unfavourable from some political perspectives, but it can also diminish an individual's sense of freedom to author their own lives, or a sense of self-trust in one's own capabilities and growth-potentials and thus can increase an individual's feelings of helplessness or restrictedness. Therefore, the physically reductive or overly functionalist interpretations of the medicalised framework creates problems for the laypeople who interpret the framework, particularly pertaining to self-image. Equally, the dualist tone of the Brain-Blame Dichotomy also creates a gap between the impressions held by members of the public and the interpretation to psychiatry given by more philosophically informed individuals. These ideas motivate a need for a more philosophically involved approach to psychiatric discussions.

6.3 Hypnosis Stage Magic

We will now give the heavier topics a rest for a moment and consider something that is a little more light-hearted. This will help us to deepen our discussion about agency, particularly when it comes to socially deviant circumstances.

A hypnosis show refers to a type of magical performance where a participant is given commands such as 'sleep' and follows a hypnotist's commands in a way that seems unnaturally compliant. What we are interested in now are the cases in a hypnosis show where the act

involves neither stooges nor conjuring techniques such as sleight of hand, gimmicks, or props that merely create an illusion of 'hypnosis'. Rather, what we are interested in is giving some psychological explanation to genuine social compliance in stage show environments. To explain the functional side of a hypnosis act, two types of classic performances will be explored in this section: (a) The Handshake Induction, and (b) Sticking Feet to Floor. (These sections are written under my own capacity as a certified hypnotherapist trained under the school of Rapid Transformational Therapy; but see also Peer 2022; Jacquin 2016a; b; Brown 2016, 2007).

(a) The Handshake Induction

There is a classic act in hypnosis shows which is known as 'the handshake induction', which is one of the most dramatic ways of putting a participant into a 'sleep state' (see Brown 2011 for demonstration and theoretical analysis). The handshake induction begins with the magician calling a participant onto stage, then offering a handshake as they introduce themselves to the participant. As they shake hands, the magician simultaneously gives the command to 'sleep', and in response, the participant suddenly falls over backwards into a seemingly unconscious state, staying limp and immobile on the floor until directed otherwise.

Breaking this act down more, what we can observe is that the handshake offered by the magician is never fully completed. Instead, as the participant reaches out to shake the hypnotist's hand, the magician grabs the participant's wrist (instead of shaking their hand) and gently but firmly bring their hand up to the participant's face, making them cover their own eyes with their own hand. Then, as the magician gives the command to 'sleep', the participant falls backwards, the magician catches their fall, then gently lays them down onto the ground.

This is referred to as a 'snap induction', in that it is a method for placing a subject immediately and quickly into a 'sleep state', as opposed to taking the time to have them close

their eyes, relax in their chair, imagine walking down a flight of stairs as they 'sleep deeper', and so on. Instead, the handshake induction only takes a few seconds to complete, and when successful, the subject will appear to be fast asleep, receptive to further hypnotic instructions.

While the handshake induction does not work equally well on all participants, when the right type of participant is selected and good technique is used by the magician, it can lead to an impressively dramatic demonstration of seemingly making a person suddenly lose control. The facial expressions and limp muscle movements of the subject also demonstrates to the audience that some form of genuine, spontaneous relaxation has happened in the subject's mind, which defies our expectation of folk psychology that we cannot simply let go of ourselves that easily or quickly just because someone told us to 'sleep', even if we wanted to.

To begin explaining how the act works, it is helpful to first think about the gap which exists between the participant's experience of being called on stage (i.e., an unfamiliar and nervous situation) and what the audience members sitting comfortably in their seats thinks the participant is experiencing. For audience members who have never been on stage, it may not occur to them just how baffling or confusing it may feel to suddenly be called on a stage in front of a large audience, bright spotlights shining in one's face, feeling watched and vulnerable, and not knowing what exactly the social rules are in this unusual situation. The participant will normally want to be cooperative and help to make the show successful, and to avoid awkward embarrassment. Therefore, even before any trick has begun, participants tend to be in a highly suggestible state of looking to whomever is in charge of the situation (in this case, the magician) for clear instructions on how they are expected to behave on stage.

In addition to this state of confusion – of being in the unfamiliar situation, feeling rather self-conscious, and not really knowing what they are supposed to be doing – the handshake induction utilises a further confusion-inducing technique. When we are offered to shake

someone's hand, we tend to initiate an automatic sequence of motions without thinking too much about it, and when this automatic sequence of motions is disrupted unexpectedly, it can create a confusing effect. The subject – who is already confused, nervous, and suggestible just being on stage – is further baffled when the magician does not complete the handshake, but instead, unexpectedly grabs their wrist and moves it towards the participant's own face. Some physical force may be used by the magician here if the participant resists, but largely, the subject will be too surprised and caught off guard to protest and will simply go along with it. Before they have a chance to recover from the shock, the magician gives the command to 'sleep' while applying a small physical force to the participant (if needed), so they lose their balance and are pushed over backwards. The magician then acts as if they are catching them to break their fall, pulling them further to the ground (if needed), and lying them down gently.

Because the participant already begins in a state of being confused about social expectations when they walk onto stage, are nervous about embarrassing themselves in front of a large audience, and will likely have the benign and polite intention of wanting to help make the show successful rather than be a jerk, the participant is already awaiting instructions from the person in charge (i.e., the magician) before anything has even started. By inducing the further confusion of interrupting a handshake and catching the person off guard, the handshake induction is designed to create a sense of relief in the participant from the discomfort of anxious confusion, of not knowing what is going on or how they are meant to act. Particularly if a participant with the right personality type is selected in advance (i.e., a compliant, open-minded, and creative sort of person, rather than an over-analytical or defiant sort of person), the handshake induction creates a situation where it is far easier to just go along with giving up control and let someone else take charge, than it would be to try and resist it (Brown 2007).

Especially because the events happen so quickly in a snap induction, the participant (who may be pre-selected for having a more suggestible, less analytic personality to begin with) will

not be engaging in active explicit analysis as the situation is unfolding. Instead, they will merely be responding to the confusion of the circumstances, the desire for relief from the tension, and a sense of comfort in going along with the hypnotist's soothing but firm instructions. Furthermore, there can be an added sense of catharsis to be in this unusual situation of being allowed to relinquish control in such a drastic way – a situation that does not often happen to adults in everyday life – and the participant might even be experiencing a euphoric sense of liberation as they enjoy the situation of not having to be in control. It is a peculiar situation where the implicit social contract of the circumstances not only gives the participant special permission to relinquish control in a cathartic way, but it is also *preferable* from a social standpoint that they are a good hypnotic subject. They will feel the positive attention in the air coming from both the audience and the magician in charge, the situation is framed as an exciting and magical experience, and they will likely be the centre of attention in their friendship groups after the show as friends ask them later what it was like to be hypnotised on stage. It is in the air, therefore, that relinquishing control is socially desirable (Brown 2023).

Meanwhile, the audience is working with the assumption of a very different social contract. In everyday life, giving up control and suddenly collapsing for seemingly no reason (or just because you were told to) would be a very socially undesirable situation to be in. By default, there is an implicit social contract in our culture that allowing someone else take over one's agency is generally frowned upon, and it can lead to feelings of vulnerability or embarrassment. In everyday life, relinquishing control when you feel confused is usually not the best way to resolve an already-confusing situation, and we feel this implicitly in the air that it might be better to either think for ourselves, or to just stop and ask, rather than to mindlessly follow one-word instructions that are given to us by someone who happened to be in the vicinity. Therefore, the social rules in everyday life about what is or is not desirable differs drastically from the social expectations the participant will feel implicitly when they walk on stage. The

hypnotist exploits this gap in social expectations between the participant and the audience to amplify the drama, thereby creating the illusion that they have 'special mind control powers'.

What is happening here is, in essence, the limitations of folk psychology are being exploited. Even if we accept that the participant was pre-selected for someone who is more suggestible to begin with, we tend to think that agency is not something we can spontaneously give up, even if we so choose. It would be highly unusual in everyday life to succumb immediately to someone else's one-word command just because they happened to make us feel confused. If anything, a person who does something confusing to us, like interrupting a handshake or pushing us over backwards, is more likely to increase our suspicion of them, and it is counterintuitive that in a hypnosis show, this sequence of events might instead increase blind social compliance. If this intuition is our starting point, it can be tempting to explain away the events we witness in a stage show by saying 'maybe they were a stooge' or 'maybe hypnosis only works on more gullible people' and so on. These explanations are more consistent with the folk psychology impression that intelligence correlates with the ability to make our own decisions, particularly in the face of confusion – thereby amplifying the enigma of the show.

(b) Sticking Feet to Floor

'Catalepsy' is a technical term in hypnosis which refers to the state that a participant falls into when they experience a partial immobilisation of the body (usually a limb), typically while remaining in a fully conscious state. A participant, for example, may be told that their foot is stuck to the floor, and the subject, who is fully awake and responsive, will struggle to regain control of their paralysed limb despite showing clear signs of effort to do so (see Brown 2012).

The act works using suggestion and social compliance. Firstly, as the participant is given a sequence of suggestions about their foot being stuck to the floor, there is an overwhelming

peer-pressure implicit in the air which the participant picks up on – that everyone would be disappointed if they were to ignore all instructions and simply start walking. The audience is holding their breath in anticipation of witnessing catalepsy; the magician would be annoyed if the trick fails; and the participant themselves are eager to experience something magical and extraordinary. Because of this, before anything hypnotic has happened, the participant is already experiencing a large incentive to keep their leg in place where it is, rather than to move it and start walking anyway.

Next, the magician eventually suggests that the participant should "try to move their foot but find that they cannot" which essentially is giving the participant two mutually contradicting instructions. On the one hand, they are explicitly being told to "try and move your foot", but at the same time, they are being implicitly informed that they are supposed to "find that you cannot move your foot". The subject experiences the inclination to follow both commands at once, responding to the verbal instruction to "try and move the foot" but also the implicit expectation that they are supposed to find they cannot. The experience of trying to move their foot in opposing directions begins to induce a feeling of tension and rigidity in the leg.

Furthermore, the wording of "try to move your foot" (rather than simply "move your foot") creates the nuance that the participant is supposed to be "trying" to move their leg, rather than actually moving their leg, which involves using different leg-muscles. The participant, therefore, begins to use some of the muscles in their leg that are involved in the leg-moving action, but without using the crucial muscles that are needed in order to actually lift the leg, resulting in further sensations of tension and rigidity in their limb, as different muscles start to work in mutually conflicting sort of ways. Because fighting one's own muscle movements is not a situation we typically encounter in our everyday lives, the participant experiences an unusual muscle sensation that they cannot easily explain, leading to a state of confusion.

The participant, therefore, remains fully awake and conscious, but is confused and bewildered why their leg now feels unnaturally rigid and stuck. In this circumstance, especially if they have already become caught up in the muscle movements, it becomes difficult to break the 'spell' even with conscious effort, as adding further muscle movements simply adds to the tension in the leg, creating further rigidity, and thereby exhausting the leg even further. The hypnotist, at this point, might even suggest to "not resist it" so the participant does not get too flustered from the strain of trying to move an immobilised leg – which only adds to the implicit nuance that there was some sort of "it" to be "resisting" in the first place, thus completing the illusion that a sinister mind control is now possessing the helpless participant.

The common-sense expectation that is exploited in this catalepsy act is our lack of awareness that conflicting social commands can cause a form of decision paralysis. If we are told to both do X and not do X at the same time, especially when there is a slight bias towards not doing the action (i.e., a bias towards inaction), it creates a state of conflict within the agent that makes them act against oneself. It is impossible to comply with two contradicting forms of peer-pressure, one which tells us to not move one's leg because that would ruin the magic and be disappointing for everyone, but also tells us to move one's leg anyway because that is the explicit verbal instruction. Especially for participants who already have agreeable and compliant personalities, they will try to follow both instructions at once, leading to paralysis.

Because most of the audience (and the participant) is unaware that conflicting instructions create a bias towards immobility, the immediate interpretation of the audience (and participant) is of confusion and bewilderment. Exploiting this explanatory gap, the magician can choose to tell whatever story they like, resolving the tension of a confused cognitive state.

It is this explanatory gap between (a) the ontological 'pseudo-theory' which the hypnotist tells the audience – that there is some sort of "it" which is seemingly possessing the participant

who is trying to "resist" the "it" – and (b) the social compliance theories which is the actual method that create the effect. Importantly, a magic show is a context where any ontological ideas are expected to be treated with a pinch of salt. Questioning whether it was 'real' or not is all part of the fun, and a magician is not an authority figure who we take to be the most credible source of information. Of course, psychiatry is a different story, which we now return to.

6.4 Philosophy of Mind in Psychiatry, Psychology, and Magic

For a layperson who believes in naïve dualism, there might be a sense that if they receive a clinical diagnosis for ADHD, autism, depression, OCD, etc., then there is something physiologically compulsive about their behaviour that removes them from being fully accountable in a social sense. This is what Boyle (2013) had referred to as the Brain-Blame Dichotomy. There is a sense in which the social contract changes for a person who has or lacks a diagnosis, and the shift in social contract is often justified on 'scientific' or 'medical' grounds. This can be a cathartic release of control, in the same way that the hypnosis participant is permitted to 'let go' on stage in ways that normal social contracts do not allow, and subjectively, the patient/subject/participant might feel a temporary euphoria as unwanted tension is released.

Philosophically, however, the problem is that if we accept monism (particularly for physicalists), then it is tautologically and trivially true that *all* behaviours are equally rooted in some kind of brain state. As long as we do not believe in a form of dualism that claims there are animal spirits and the like that can cause behaviours, it should be unsurprising to a monist that *every* mental state has an underlying brain state. Therefore, philosophically speaking, we can see that the patient's thought process behind relinquishing a sense of responsibility because a diagnosis was given to them is not fully justified – and it is not difficult to imagine that this nuance can create complications between the patient and their family/friends/colleagues, since

the patient is now being given two incompatible social contracts from institutionalised sources (which tell them that it's okay to not feel responsible) and their friends or family who are still unhappy with their social behaviours. To give an example, a person with ADHD might feel better if they are told that their tardiness is not really their fault, but the friends, family, or colleagues which are inconvenienced by their tardiness might still feel annoyed by it. Meanwhile, the person with ADHD might behave less apologetically after their diagnosis and may even become annoyed with their friends/families/colleagues for not being understanding enough that they are not truly in control of their own actions. This attitude might annoy and alienate their friends and family even further, thus creating an even deeper rift between the patient and their social circle, and in the long term, exacerbating helplessness and alienation. If to combat this, neurodivergent people decide to bond together to make each other feel understood and bond together over these same shared difficulties – ultimately, the issue of tardiness still has not been addressed. People might feel better or worse and social contracts change, but fundamentally, the tardiness itself is understood as a biologically induced disability.

The extent of misunderstandings in this structure will depend on the condition that we choose to focus on. Perhaps ADHD will turn out not to be the best example if plenty of laypeople are already aware that tardiness is not biologically compelled – but 'the best example' will constantly be in flux as social influencers, self-help books, and other sources of information follow and spread different trends when it comes to these topics. The bottom line that is important for our current discussions is to notice the tension in situations where people receive different social contracts from different sources. For instance, if an astrology-believer comes across information that their failed relationship is not really their fault because it was written in the stars, then once again, there will be that temporary sense of relief from blame. But this comes at the cost of not being able to troubleshoot what actually did go wrong in their relationship, so that they can eventually fix the source of the problem. Instead, the astrology-

believer may just become annoyed with their friends who do not believe in astrology, because they are not being understanding enough that the failed relationship really wasn't their fault – and pin the blame on others for not following the social contract they think should be followed.

Which social contract we follow, as a society or a community, is to some extent, a matter of trend, fashion, or fad. However, when a social contract is based in outright mistaken information, this is when Empirical or Theoretical Problems turn into Pragmatic Problems. If the fate of a person's relationship indeed was written in the stars, then it would make sense that there should exist a social contract where we treat that person with compassion for being such a helpless victim. Similarly, if a person's socially unfavourable behaviours are written in their biology, then they too would be a helpless victim that is deserving of compassion. But problematically, this kind of compassion/relinquishing of blame conflicts with one's ability to author their own path – which comes with its own joys, freedom, and bonding with others. At least for philosophical purposes, what becomes important for discussions on agency, accountability, and authorship is to distinguish between behaviours that truly are compulsive, and behaviours which are merely difficult to control – and having an accurate understanding of these factors feeds into the social contracts that have real pragmatic impact on our society.

As another example, we might believe that a seizure falls under the category of 'compulsive' behaviours, and if a person has an unexpected seizure while driving, we might not wish to hold them accountable. Similarly, if an old man has Alzheimer's disease, then we might not blame him for forgetting his wife's name but feel pity and sorrow instead. There is a certain kind of social contract which is invoked when we take someone to be in a position where they lack agency or accountability. While this can involve a greater sympathy for that person, it might also involve a lesser trust in giving them important tasks. A person with epilepsy may not be allowed to drive, and a dementia patient might become unable to live and function on their own. Thus, we see that there is a trade-off with relinquishing agency. The

positive is a cathartic release from responsibility which can be a liberating and empowering experience; the negative is that it comes with a diminished self-image, of someone who is less functional, and is more helpless. What seems to be getting missed here is that people do not feel happy for being 'merely tolerated', and while increasing sympathy for people who are struggling can provide temporary relief, the trade-off needs to be acknowledged on theoretical and academic grounds before political force compels paradigm shifts driven by desperation.

> "while bio-neurochemical-physiology is a necessary condition for mental functioning, it is far from being a sufficient condition for adequately explaining the human being. This becomes especially significant when examining the question of selfhood, freedom, personal autonomy, value inquiry, and the phenomenal quality of lived experience." (Mills 2022, p.29).

The Power Threat Meaning Framework advocated by the British Psychological Society (BPS) encourages such meaning-based explanations when dealing with psychological disturbances (Johnstone & Boyle 2018), in contrast with the DSM 'disease' model advocated by the American Psychological Association (APA). The Division of Clinical Psychology (DCP) of the BPS has released an official position statement against the American 'DSM' system:

> ""The DCP is of the view that it is timely and appropriate to affirm publicly that the current classification system as outlined in DSM and ICD, in respect of the functional psychiatric diagnoses, has significant conceptual and empirical limitations. Consequently, there is a need for a paradigm shift in relation to the experiences that these diagnoses refer to, towards a conceptual system not based on a 'disease' model" (DCP 2013, p.1)". (Johnstone & Boyle 2018, p.12)

Surprisingly, when it comes to psychiatry, the 'medical' model still follows the pre-rabies medical research programme that is largely symptom-based rather than cause-based. In physiological medicine, diseases are typically labelled or categorised based on their presumed cause, such as when diabetes is understood as an abnormality of insulin levels. However, in psychiatric medicine, the categories are symptom-based, such as when autism is understood as a collection of socially deviant traits; ADHD just is an attention-deficit; or depression just is a state of being unusually depressed. Psychiatric diagnosis, therefore, can be a relatively uninformative diagnosis to receive. The patient goes to see a GP knowing they have been feeling depressed, and after answering a questionnaire, the GP confirms, they are indeed feeling depressed. Unlike physiological medicine which adds information about 'causes', psychiatric medicine is merely a labelling system that is silent about causation and can be fairly vacuous (Johnstone & Boyle 2018, Ch1-4).

> "A good example was the discoveries that symptoms including strong thirst and tiredness could be associated with glucose in the urine and that this cluster could be associated with abnormalities in insulin production. Researchers tend to denote these patterns by abstract names, for example, diabetes" (*ibid*, p.22)

The abstract noun 'diabetes' is defined in terms of its biological markers such as insulin production and glucose levels in urine. Subjective complaints about strong thirst and tiredness can be a good sign that the patient should hand in a urine sample to be tested, but importantly, the subjective complaints themselves are not sufficient criteria for receiving a diagnosis, as these may be caused by something completely different (e.g., heatstroke, dehydration, etc.). In psychiatric medicine, where the labels are categorised by symptom rather than cause, there is no guarantee that there is anything meaningfully causal that 'binds' that category together.

Even if research results show that there is an 'underlying brain state' in depression, if the label of 'clinical depression' just is an extreme form of sadness, then this kind of research will not enlighten us to whether the brain state is a 'physical compulsion' or a problem with storytelling.

In fact, the BPS publication even went as far as declaring, "There is no evidence that this process follows any recognised scientific procedure." (Johnstone & Boyle 2018, p.27) and that the diagnostic labels are merely "voted into existence by committees, informed by theories which lack empirical support." (*ibid*, p.26). The concept of an 'abnormal' psychiatric state is something that is decided upon by said committee, rather than being empirically driven.

Considering this, we can come up with some absurd thought experiments, where we imagine that psychiatrists decide to create a new category for a psychiatric condition: Believing Astrology Disorder (BAD). The word is used to describe people who are prone to falling for Barnum statements and making questionable life choices, including donating their life savings away to a group of faith healers because they were charmed by cold reading techniques. We might acknowledge that being BAD is a serious problem. Institutionalised intervention may sometimes be helpful for saving our friends or relatives from themselves, as their brain compels them (?) to make destructive life decisions based on what they believe is written in the stars.

In fact, let us say that BAD becomes a very hot topic with plentiful funding poured into it. Before long, an abundance of publications find evidence for an average brain state which is shared across all BAD members – a reduced activity in the fact-checking region of the brain. Other research might find that most people who believe in astrology will tend to have at least one parent who also believes in astrology, leading to the conclusion that it must be genetic. Of course we might have an issue with this on theoretical grounds. But because the taxonomical nature of psychiatric medicine allows that words can be defined on a symptoms-based basis, as long as there are enough researchers who find it useful to categorise people as having BAD,

the terminology can prevail in psychiatric literature. Further experiments, funded by pharmaceutical companies, might even show that prescribing sedatives reduces the instances of poor decision making (because sedated people tend to not do many things) and this research further 'empirically supports' the medicalizable nature of BAD – despite the community of BAD patients having an issue with this (and also the acronym) because they find it demeaning.

This thought experiment will hopefully highlight the absurdity of haphazardly confusing Structural problems with Scientific problems, in the way that was discussed in chapter two. If the taxonomical lingo we use is defined by non-philosophers, definitions of 'scientific' lingo can carry philosophically problematic connotations, such as the idea that finding an average brain state in a group of individuals is evidence for it being a 'brain' problem that removes the individual from accountability. It can also lead to other social-political problems, where a philosophically problematic definition becomes accepted as though it is a 'scientific' finding.

To conclude, much like an astrology-believing individual who thinks they do not have control over making a relationship work with a partner who has the wrong star sign, a person who believes their brain just is wired differently (and believes they lack the expertise to question this 'scientific' judgement) now loses out on the agency of getting to author their own life story. This makes it difficult for patients to opt out of or reject the 'novel social contract' that comes with being diagnosed with this or that, even if that particular individual does not personally experience its benefits. The real question we might want to ask is *why* the diagnostic system does seem to 'work' for many individuals, and why there seems to be evidence that this is 'empowering' for such patients, improving psychological wellness (Ladd & Churchill 2012).

The answer to this might be similar to the reason that astrology 'works' for improving some people's self-perceived sense of wellness. To be categorised amongst a group of other like-minded people can be a connecting experience that combats one's loneliness. To enhance

this sense of 'belonging' in a social group creates its own kind of social contract – of not being the 'outcast' member of a tribe. For some individuals, the trade-off of accepting that they must relinquish authorship can be a net positive, as it does alleviate a sense of guilt and self-blame which can add to the problems that they already have. Particularly for individuals who embrace a naïve dualism, it can be empowering to be told their problems have a 'physical' origin, since a layperson dualist may not realise *everything* has a 'physical' origin. Instead, they may believe that there being a 'physical' origin is synonymous with the idea that 'it is not a fault with your *personhood'.* And whatever it takes to remove guilt – it probably will feel nice for the patient.

Problematically, however, scientific methods are not particularly well-suited for giving a personalised meaning-based explanation of a social contract that a particular individual happens to be responding to. Analysing human psychology is riddled with an extremely large number of confounding variables that makes it extremely difficult to tease each of these factors apart in a controlled, replicable, and generalisable setting of a scientific test. Therefore, it is not so much that we need to do more science in order to understand psychiatry or neuroscience better, but rather, what we need is a change in research programme at a philosophical level. Instead of having a pro-science attitude and searching for universalisable answers through statistical testing, maybe what we need instead is a pro-magician attitude – a research programme that is better suited for investigating *particulars* responding to social contracts.

Philosophically speaking, human agency cannot be properly accounted for unless there is an 'I' that does the agential work. If this 'I' can be explained away entirely, then every behaviour equally becomes a compulsion, much like a seizure. We do not need to believe in any homunculus inside my brain that does the decision-making, nor do we need to believe in a Cartesian theatre which the homunculus watches from inside the brain. Rather, it is not a homunculus, but it is *me*, deciding on my own actions, looking at the world through my eyes.

No further reduction is necessary, as in the context of certain psychological or psychiatric discussions, sometimes, it can be appropriate to just have 'agency' as its own bedrock.

This does not entail an endorsement of dualism. If to some readers it does sound like dualism, then one thing we might do is to remind ourselves that physics has already had several hundred years to develop, whereas the study of wellbeing, agency, and the human mind is still very much in its infancy. Perhaps because physics is much more developed than other disciplines, this might fuel an elitist impression that physics is 'the' fundamental science. This, however, might just be a sign that psychology needs to catch up in development. It is putting the cart before the horse to presuppose that academia is a place for humans to explore an indifferent 'physical' world. An alternative interpretation is that academia just is a place of human storytelling – with our limited tools (our sensory organs plus cognitive capacities) that we have. We can allow that storytelling and finding meaning is just as 'fundamental' a part of our academic endeavours as exploring the cause-and-effect of 'indifferent' or 'physical' variables. Both are required for investigating matters of 'the passions' or matters about enhancing 'quality of life'. In the final chapter, we will discuss how this can be addressed in a new research programme, which will be referred to as the Magician's Research Programme.

To conclude this chapter, what we have seen is that when we are exploring the unknowns about human agency, reducing our psychology to 'physical' explanations risks leaving out the agency. A separate kind of 'story' needs to be told in order to address the gap between our expectations about how we think agency works, and the blind spot of how social compliance and other factors can affect our psychological wellness, authorship, and the naïve philosophical views held by laypeople which might actually be impacting their psychological lives. In the next chapter, we deepen our understanding of how a magician's approach might help us to overcome some limitations that can be faced according to a more Western 'armchair' approach.

7. Defending Magicalism: The ‘Warm Fuzzies’

In the first three chapters of this thesis, physicalism was criticised from a variety of angles, including the central argument that the defining of the word ‘physical’ is too elusive. As was discussed in the introduction, the professional academic debate when it comes to the mind-body problem can be somewhat different from the types of concerns that laypersons can have relating to similar topics. The academic literature tends to use technical terminology and logical arguments to investigate ontological reality in as unbiased a way as possible. In contrast, a layperson may have other considerations that captivate their attention. What happens to us or our loved ones when we pass away? Could there be a heaven that our souls go to after death, or perhaps we are reincarnated into someone else? What are the ethical considerations in treating other conscious beings? If a person is diagnosed with a psychiatric condition, how much of their problem should be attributed to the *person* and how much of it to the *brain*?

There is a diversity of consciousness-related questions that we could be addressing, and one approach is to simply focus on each narrow topic in its own right. The variety of questions can be divided vertically into different fields or topics, for instance, by having philosophers of mind deal with ontology, psychiatrists with mental illness, and theologists with questions about afterlife or death. For anything that does not fall into any particular category, we might even turn to popular books, self-help material, or have each layperson formulate their own personal theories and opinions based on their own experiences. An alternative to this vertically divided approach, however, is to construct a novel research programme that handles all of these consciousness-related questions as an entangled unit. This is the approach that the remainder of this current thesis will be taking, and we will be referring to the novel research programme as the Magician’s Research Programme, inspired by Eastern philosophy and stage magic.

In chapter four, a novel theory called Magicalism was proposed, where consciousness is understood as something that is fundamentally Magical, but not in a supernatural sense. Magicalism was considered a neutral monist view (simply because we took monism to be the default outlook of a unified reality) that is also anti-physicalist (in virtue of 'physical' being too difficult or elusive to define properly). Then, it was further established that we can think about Magicalism as a conjunction of an (a) Eastern (b) Magician's (c) Apprentice's point of view. To recap, this covered the ideas of a sense of poetry in Eastern philosophy, where it is accepted that all humans are inevitably biased. Our goal is not to maximise the level of unbiasedness that we can achieve, since we are all like blind men feeling around in the dark, but that being said, we must still take empiricism seriously. We should not just follow the lead of an authority figure, including scientists, as though they are telling us some kind of contextless truth, and it is up to us to decide whether the scientific data is even relevant for the philosophical stories that we are trying to tell. Then, the Magician's perspective further highlighted the importance of distinguishing between *method* and *effect*. We can still be empiricists and sceptics when figuring out the mechanisms behind the 'magic', but at the end of the day, we do not want to explain away the magic entirely. Furthermore, the magic is to be found in the idea that magicians are dealing with *particular* individuals – audience members who are indeed non-Xombified and able to experience the joys of the show – rather than *universals* or statistical data, which can lead us to ignoring or trivialising the spectator that is in front of us. Finally, the Apprentice's perspective brought attention to the idea that we all have a sense of loyalty or allegiance when we formulate our beliefs, but we should draw a distinction between the 'good student' who memorises answers from a textbook and is told not to question them, versus the 'good apprentice' who question their own master and discover things about life through their own journeys. To be truly an empiricist, we cannot just follow the lead of another empiricist.

This chapter will further deepen the ideas of the (a) Eastern (b) Magician's (c) Apprentice. Section one begins with a thought experiment about a mediocre magician performing a card trick to a group of people: (A) a psychic believing layperson, (B) a psychic non-believing layperson, (C) a physicalist, and (D) a professional magician. This will allow us to elaborate on the method/effect distinction that is quintessential to the Magician's Perspective, and the attitude the Apprentice ought to have in comparison with the problematic attitude of a physicalist. Then, in section two, we will look at some issues with the heterophenomenological approach as discussed by Dennett (2007) where psychology researchers are interpreted almost like anthropologists, and we will once again see why the issue of reducing *particulars* to *universalisable* statistical data can cause problems, particularly because it elevates the researcher on a pedestal of 'objective observer' which is never quite true in real world research scenarios. Thirdly, we will examine an inherent issue with trying to solve the philosophical problem of consciousness using purely deductive arguments, and a case will be made that what has been missing from philosophy is for us – philosophers – to go out into the real world and gain first-hand experience, much as the Eastern apprentice would. Although this type of personal element can increase the risks of becoming biased, it is a necessary step that armchair philosophers will miss out on, and we will see why the problematic structure of a deductive argument necessitates biting this bullet. Finally, we conclude by introducing the concept of 'the warm fuzzies' which will be elaborated on even further in the next, final chapter.

7.1 The Magician and the Physicalist

Let us start with a thought experiment. Imagine that there is a magician performing a very standard card trick in front of four spectators. The card trick works as follows: One participant is asked to select a random card from a well-shuffled face-down deck. The selected

card is returned to the deck without the magician seeing what the card was. The magician goes nowhere near the deck, and the deck is again, thoroughly shuffled by a spectator. The selected card is lost in the deck, yet somehow, the magician is able to name the chosen card, supposedly by using psychic mind-reading powers!

We will refer to the performer of this trick as Magician X. Furthermore, we will imagine that Spectator A is a genuine believer of psychic powers, and believes that the magician actually did read the spectator's mind using some kind of supernatural ability; Spectator B is a person who is fully aware that (obviously) no psychic powers were used in the trick, but they are nonetheless marvelled at the magical effect of the card trick; Spectator C is a physicalist philosopher who wants to tell everyone else that there is no Magic that floats 'over and above' the cards in some sort of 'spooky' dualist way, and that everything is reducible to purely physical terms; and Spectator D is a professional and experienced magician who thinks Magician X's performance was mediocre at best and could have been performed better.

Spectator A	Believes psychic powers – The trick was Magical!
Spectator B	Does not believe psychic powers – The trick was still Magical!
Spectator C	Does not believe psychic powers – The trick is not *Real Magic*
Spectator D	Does not believe psychic powers – The trick was not very Magical

Let us say that as the trick finishes, Spectators A-D gather together to discuss the trick. Spectator A is excitedly trying to claim that surely this is proof that telepathy is real, and that the no sleight of hand could possibly have been involved. The deck was shuffled by someone else! 'Magic must be real, there is no illusion!' they insist, meaning this in a very literal sense.

Spectator B, in contrast, is fully aware that surely some sort of trickery must have been used. There is no doubt in their mind that a physical explanation of the trick must be possible

– and yet, they also are impressed at the Magical-ness of the performance. Therefore, Spectator B enjoys the childlike wonder, and claims 'It's as though I could believe in Magic again. Magic is real!' but in a figurative sense of referring to a special sense of wonder, not supernaturalism.

Spectator C is instead focused on asserting a metaphysical claim that is responding to to the mind-body literature in academia. They quickly reject the idea of a 'spooky' magical force floating over and above the cards, or over and above the neural activities of the magician and the spectators. Instead, they comfortably conclude that because of problems such as the causal closure of the physical principle or the issue of overdetermination, there is no need to embrace a dualist ontology. Everything he has just observed can be explained and reduced to purely physical or scientific terms, claiming that "Hey, it's just stage magic!" (Dennett 2016, p.66) in the sense of rejecting 'spooky' supernaturalism. Magic is just an illusion. It is not Real Magic.

Finally, Spectator D is rather critical that Magician X's showmanship was not very good and pulls the young magician aside to give them advice. While the trick was not a 'failure' so to say, the Magicalness of the trick could have been enhanced through, say, better storytelling. The professional magician advises, for example, that instead of immediately guessing the chosen card, Magician X could have asked the spectator some personal questions, such as "What is your favourite colour?" or "What is your favourite number?" as though this helps them with the task of guessing the card. Then, Magician X could have gone on on to guess the correct card, giving reasons such as "You hesitated when you said your favourite colour was red, so your card must be a red card!" and so on, making the 'Magic' more believable. From these considerations, Spectator D, the professional magician, teaches that 'the Magic does exist over and above the method', in a figurative sense, indicating sleight of hand is not all there is.

The Magic arises as a response to the spectator's expectations about witnessing something that contradicts how they *think* the world works. It is about achieving something

which we took to be impossible, which creates a sense of curiosity, thrill, and wonder. Without this source of curiosity, we would not be interested enough to philosophise about it in the first place, and so explaining away the source of wonder is counterproductive for philosophical purposes. If a magician spontaneously names the correct card, there is not much sense of 'impossibility' in this situation. We all know that magicians probably are doing some kind of sleight of hand, and this is the first thing that spectators will suspect. But enhancing the showmanship and misdirecting audience members from the obvious answer, this will lead to a better performance (see Brown 2023). Asking the participant for their favourite colour is just the beginning, but Real (Theatrical) Magic arises from the storytelling and showmanship of the magician who appears to achieve the impossible.

Consider, for instance, one interpretation of the concept of 'real magic':

> "By real magic people mean miracles, thaumaturgical acts, and supernatural powers. […] 'Conjuring tricks, not real magic.' Real magic, in other words, refers to the magic that is not real, while the magic that is real, that can actually be done, is not real magic." (Siegel, 1991, p. 425)

But recall also the quote from science fiction writer Arthur C. Clarke (1962, p.250), that "Any sufficiently advanced technology is indistinguishable from magic." An alternative way to phrase this would be to say that any sufficiently familiar technology (or concept) ceases to feel like real magic. In the Harry Potter series, spellcasting is a trivial matter for Arthur Weasley who is so accustomed to Hogwarts-magic that it has stopped being interesting to him. Instead, he is fascinated by Muggle technology, which *to him*, is like Real Magic (Rowling 1998).

Magic seems to be a thing which exists on the borders between things which we are interested enough to attend to properly (e.g., the card trick) and things we are not interested

enough to care about to any extent and so it forms a blind spot (the method or sleight of hand), thereby defying the expectations that are created within our attentional focus. Thus, the moment we bring the blind spot to somewhere that lies within our attention, the Magic disappears.

In the earlier card trick example, the professional magician can give the younger magician advice that by using showmanship or storytelling to *enhance the gap* between the effect and the blind spot, a greater sense of wonder can be achieved. But then, we can imagine that the physicalist (Spectator C) overhears this conversation and has an issue with this idea. If we truly want to explain what happened in front of us, then surely there can be no gap remaining. Otherwise, so long as any gap remains, does this not imply that we have something more to explain? But then, as soon as that gap *is* explained, the magic disappears. So magic is not Real!

It is an interesting thing about magic that the sense of Magic will immediately vanish if the spectator is told the mechanism behind the trick (spoiler alert) — that the entire deck was composed of fifty-two copies of the same card. The trick turns out to be far less impressive than a demonstration of people-reading skills or even sleight of hand. In this particular card trick, there is hardly even a method involved, and realising this, a spectator may feel 'cheated' or 'fooled' – the magician has lied to them by showing them an illusion!

There is a sense in which, the moment we learn how the trick was performed, this undermines the Magic itself, and this is because we have now brought the blind spot to somewhere that lies within our field of attention. Ironically, the same effect occurs when we learn a little bit more about physics and realise that it is much more fluid and versatile than what school-level physics might suggest. As long as it remains a blind spot to us how sciences actually work and what their confident claims mean under a more specific context (as we discussed in Chapter 3), physicalism might retain its charm. An even better solution to addressing our ontological question might be simply to allow some of the Magic to remain, so

we can answer the parts of the question that are most gripping or important to us, without making it trivial or vacuous, or even worse – explaining away the worth and beauty of philosophy! That is, if physicalism were to 'win' at the cost of convincing non-philosophers that philosophy is a bit redundant and stupid, this would be a trivial victory for physicalism. We do not wish to explain away the beauty of philosophy itself. This would be the equivalent of a magician who debunks all his magic, and no one wants to go see another stage show again!

What we have seen is that the essence of Magic cannot be captured in terms of its methods alone, without the Magic itself vanishing and eluding our grasp. That being said, this does not mean that the essence of Magic is inherently uncapturable. For instance, the American magician duo Penn and Teller is a paradigmatic example of magicians who often purposefully give away the method of a trick to play up the irony, surprise value, and comedic effect created when they give away how a trick was done. For instance, one member of the audience might be asked to join the magicians on stage to witness a trick in the more conventional way, while the rest of the audience is 'in' on the methods and experiences a different kind of Magic – of having our expectations defied about how stage magic works.

As long as our expectations are being defied in some sort of way, this can still allow some sense of Magical wonder to flourish. What is 'essential' or 'fundamental' to Magic is not about capturing it in methodological terms, nor is it about misleading the audience, but rather, it is to tell a kind of 'wonderous' story. Oddly enough, to acknowledge that physics too is 'just a story' is itself a Magical idea. While worries about accepting supernaturalism without evidence is still an important problem as we saw in chapter five, there is a sense in which it helps to acknowledge that we are never in a privileged position of directly accessing 'facts'. We are always observing the world through the limitations of our own senses and cognitive capacities, and so, there always will be some blind spots, no matter how knowledgeable or intelligent an individual happens to be. Thus, in answer to the *meta*-physical question, "Will physical

explanations ever be sufficient for explaining consciousness in the ways that satisfy our central curiosities?" the answer is straightforwardly "No – because physics is a discipline that brings our blind spots to the spotlight, and so it is not suitable for explaining a sense of Magic."

But perhaps *meta*physics is not meant to be reduced to the explicit and rigid language of the 'adult world'. Instead, Magic is inherently a childlike concept of wonder and amazement that us grown-ups often forget about as we get older. However, this does not entail that a 'childlike' theory must be 'naïve' or 'uninformed'. In the same way that a professional magician can have a more 'grown up' ways of understanding the Magical childlike wonders, we too can pay respect to the Magic inherent in the world all around us. Otherwise, philosophy itself ceases to be a meaningful endeavour, and if we can explain away the need for *meta*physics, we might as well stop doing it. So now, we turn to how we might execute this in practice.

7.2 The Issue with Heterophenomenology

Dennett's heterophenomenology (2007; 2003; 1991) can be accused of committing the problem of taking 'objectivity' too seriously, to the point where ironically, it ends up creating its own kind of unique bias that makes it no longer 'objective'. To summarise the idea, what Dennett has essentially claimed is that when we explore the psychology of our test objects in a scientific manner (such as in the field of psychology), what researchers are doing is that they are approaching the issue as if they were anthropologists. The analogy he uses is of a scientist who is documenting the beliefs of tribal members who believe in their own God, and as the scientists, our job is not about verifying whether this God is real or not. Instead, we are merely documenting the verbal reports and creating an objective analysis of what they believe to be true while being agnostic in regards to the content of the tribe's beliefs. In other words, the task of the scientist should not be mistaken for studying real properties and features of the God itself,

but rather, the scientist's task is to indifferently and neutrally document *people's beliefs*. Importantly, the scientist does not determine the truth-value of the beliefs and remains neutral.

"In the Heterophenomenological method, we approach the verbal, behavioral report of a subject on her experiences much like an anthropologist would the reports of members of another social group in investigating their religion." (Fletcher 2002, p.106). The test subjects, or the verbal reports we are documenting, is understood as ""fictional worlds" (notional worlds, alternatively) of experience, and it is the role of the investigator to maintain "a constructive and sympathetic neutrality, in hopes of compiling a definitive description of the world according to the subjects" as they record the 'texts' of the authors ([Dennett 1991] CE 83)." (Fletcher 2002, p.106). To summarise, Dennett essentially takes an interpretation of psychological studies where the method of the social sciences like psychology is that about creating a descriptive picture about what people *claim* their beliefs are, which allows scientists to maintain neutrality and objectivity. This approach supposedly allows phenomenological studies to fit in well with the computational or naturalistic structures of other sciences, allowing cognitive sciences to maintain objectivity despite its engagement with seemingly phenomenological considerations.

There are two problems with this approach, however. Firstly, it ignores the idea that the researchers are human beings too, and that their recording of the data might not be neutral, even if they want it to be. This can come in the form of funding biases, publications biases, personal biases, and so on. To draw an analogy with magic again, how 'good' a particular card trick is cannot be evaluated by observing the spectators' reactions alone. The magician's showmanship and the way they engage their audience will inevitably play a role. Now, if the magician tries to remove their own sense of presence as much as possible so that we can focus on the work the card trick is doing all on its own – then as the earlier thought experiment showed, the card trick itself may cease to contain the magic that it otherwise could have had. Observing a card vanish and reappear repeatedly without any showmanship or storytelling from

the magician would get boring very quickly, but this does not necessarily mean that the trick itself was not a good one. This same sort of issue might happen in psychological studies, such as in the psychiatric literature we saw in chapter six. If a psychiatrist encourages the patient to not blame themselves and has a human warmth about them, this itself might help to enhance the treatment. If the psychiatrist objectifies or treats the patient in an otherwise dehumanising way, like they are nothing more than a collection of functionalities rather than an individual human being (Anderson-Chavarria 2022), this might ruin any empowerment that might otherwise have been there. To be truly 'objective' about whether a psychological method is working, we would need to treat the clinicians too as participants in their own right.

But then, if we imagine that we had indifferent observers watching the participant and the clinician, once again, we are faced with the same problem where this new observer is also a human being. The more steps are removed between observer and first-order participant the less bias or personal involvement there will be, but also, if they become too far removed, the observer is no longer really a part of the situation, and so their observations might cease to be accurate. Once again, we can give a magic analogy. Consider this: For a magician, the most difficult kind of spectator to fool is the kind of spectator who is not watching the trick properly. If a magician is performing to a table of people at a restaurant, and there are a couple people at the table who are engaged in their own conversation but watching the magic trick out of the corner of their eye, then these spectators are just engaged enough that they are watching the trick, but not engaged enough so that they are a part of the atmosphere. As a result, they end up focusing on all the wrong focal points that ruins the story or the magic, and when these spectators recount the events later to a friend, they might tell the story in a way that destroys the whimsy of the trick, essentially butchering the performance in the process (Brown 2000).

Similarly, even if we did allow for an almost infinite regress of higher-order observers who become increasingly detached from the psychology participant, at some point, the

observer is no longer really a part of the human story that is unfolding. This leads to a trade-off between 'objectivity' (i.e., non-involvement from the researcher) and accuracy in evaluating the nuances. One such example of a removed or detached approach to research was observed in the behaviourist traditions of psychology in the early twentieth century:

> "The behaviourist views psychology as a purely directive experimental branch of natural science. Its theoretical goal is the prediction and control of behavior. So far, human psychology has been unsuccessful due to the mistaken notion that introspection is the only method available to psychology, and that it is the study of consciousness. Actually, psychology is the study of behavior and therefore need not take recourse to conscious phenomena. […] This suggested elimination of states of consciousness as the objects of investigation will remove the barrier that exists between psychology and other natural sciences, without neglecting the essential problems of introspective psychology." (Watson, 1913, p.158)

What we see here is that Watson had essentially removed the topic of consciousness from psychology, to make it more consistent with physics and other branches of the natural sciences. The inner experience, or what philosophers would refer to as phenomenal experience or phenomenal consciousness, is considered to be a 'black box' that is *assumed* to be inaccessible to any reliable empirical investigation, and therefore, is left out of the picture as an *assumption*. This does certainly help to remove nuance, and therefore biases as well, but at the cost of not really being able to explain the sort of things we want to be explaining.

It is these kinds of detached epistemic approaches which Wallace said does not sit well with his more Buddhist mentality. "What Buddhism brings to this historical situation [on consciousness] is not dogma, but rather methods of rigorous, refined, first-person inquiry into mind as it is subjectively experienced, to compliment the magnificent accomplishments of

objective science, which is the strength of the West." (Carroll & Wallace 2022, p.45). That is, he does not claim that the first-person methodologies of Buddhism are sufficient for establishing true, objective, generalisable knowledge. However, when it comes to consciousness, introspection plays a key role that we cannot do away with. A similar idea was also acknowledged by Western researchers in the early days of psychology, researchers took seriously the idea of training themselves to become introspectively more proficient, in a similar way that psychology students today train to become objectively more analytical. In particular, European experimental psychology reflected Edward Titchener's mindset about using introspection by a trained observer as its primary method for investigating conscious sensory experiences (Titchener 1912). An example of this method is observed in an early experiment by S.I. Franz in 1895 which attempted to measure the limits of visual perception, the method being to make a candlelight dimmer and dimmer (or further and further from the subject) until the trained subject reported a threshold where they could no longer see the light. As this early study shows, nineteenth century psychology took introspection seriously as valid datum for consciousness studies. However, this European attitude in the early days of psychology differed greatly from the American trend at the time of having a more functionalist attitude, represented by major spokesman Angell (1907). While this psychology did not reject consciousness as later behaviourists did, the emphasis was more on the functional adaptability of individuals to their environments, setting a precursor to the behaviourist mentality by placing heavier value on objective observation of individuals, rather than introspection (Hilgard 1980).

The problem with Dennett's heterophenomenology, as well as the more generally accepted method in psychology of recording the self-reports of participants (i.e. third-person documentation of someone else's verbal reports), is that it not only assumes the participants of these studies are fairly good at introspecting and making accurate reports, but it also assumes the researchers are fairly accurate with documenting it in a unbiased way that still captures the

important nuances. Essentially, what this is doing is to record the testimony of untrained introspectors *and* still suffering from some level of bias in the researchers when they document the results. Technically, this might be somewhat permissible if we could accept that there is no such thing as a more trained or less trained introspection, and that we are all equally good at both introspecting and/or documenting someone else's verbal reports. But this would be a questionable assumption to make, as several psychological studies do show the inaccuracy and biases that we tend to fall prey to when forming our beliefs (see Wiseman & Watt 2006 for overview). Therefore, it is suspicious to suppose that (a) having untrained members of the public report their testimony, and (b) having researchers evaluate this (possibly misinterpreting the original meaning behind the reported testimony) will somehow lead to more accurate or 'objective' ways of conducting research.

There is a tension between the more third-person methods of the sciences, and the first-person involvement observed in both Buddhism and Magic performances. The problem is that when we investigate human beings, it matters whether we treat people as *particular* individuals, or *universalisable* data – and this is in virtue of us ourselves being human beings. We tend to feel a greater sense of warmth or connection when we deal with individuals who are in front of us, and a greater sense of detachment when we look at data. It is dangerous to presuppose that one approach is more 'objective' than the other, since we ourselves are only human. We never are fully 'objective' – but we can still range from being warm to detached – and depending on what it is that we are investigating, the preferable approach might be different. As we shall see later, 'the warm fuzzies' are a concept that simply cannot be accessed from a detached angle.

7.3 Deductive Logic in Armchair Philosophy

Before getting to the topic of 'the warm fuzzies', it is helpful to address a certain characteristic with deductive logic, which can trick us into accepting arguments that are logically valid but unsound, creating the illusion that the belief system itself is logically robust, even when its content does not align with the external world. This can affect both supernatural believers who succumb to faulty logic, as well as to more scientifically minded academic philosophers. To begin, consider how a deductive argument takes the following kind of format:

i. Premise X

ii. Premise Y

iii. Premise X AND Premise Y entail Conclusion Z.

A deductive argument is considered to be 'valid' if the conclusion indeed does follow from the premise. In other words, if the conditional 'X+Y -> Z' is true, then the argument is considered logically valid. However, a deductive argument is logically 'sound' only when it is both valid *and* all the premises are actually true. Therefore, if we know for certain that X and Y are both true, then conclusion Z is both valid and sound (Field 2015). Now, also note another interesting feature of deductive logic that it is theoretically impossible to construct a deductive argument without accepting some premises that are not themselves a conclusion to a separate deductive argument. To avoid an infinite regress, at some point, we will need to accept the truth of at least two premises that are themselves supported via a non-deductive approach. If we are trying to validate the truth of Premise X by deriving it as a conclusion to another deductive argument:

i. Premise A

ii. Premise B

iii. Premise A AND Premise B entail Conclusion X.

The above deductive argument does not tell us whether we are justified in accepting premises A and B to begin with, but to derive these as a conclusion in another deductive argument leads to an infinite regress of the same pattern. At some point, there needs to be 'bedrock' premises which can be accepted without deriving it as a conclusion from another deductive argument.

The question now is to address what methods we are allowed to use for justifying premises. There are a wide variety of ways this can be done, including supporting the premise through empirical evidence; making an appeal to authority; using a definitional acceptance of terminology to justify a premise through a semantic argument; appealing to intuition or common sense; or even just accepting some premises as an unargued, brute fact. Broadly speaking, we can separate premises into two categories: *a priori* arguments and *a posteriori* argument. The former includes definitional commitments (e.g., 'bachelor' just means unmarried man), as well as premises we are willing to accept as brute facts or assumptions for the purpose of the argument (e.g., for an ethical argument, just deciding that 'murder is bad').

Furthermore, there is an interesting distinction which occurs when a premise is accepted via (a) an *a priori* argument, or via (b) *a posteriori* argument. In the latter case, this requires an appeal to some kind of information that can be gained only by observing something about the real world that we could not have known prior to observation. The advantage of this is that it provides us with new information about the real world, but it comes with the disadvantage that it is always logically possible to later undermine empirical evidence. To give an example:

i. All swans are white.

ii. I am told that what is in front of me is a swan.

iii. Therefore, (without looking at it), I know what is in front of me is white.

Let us say that the first premise is validated on empirical grounds, using an inductive argument. What this means is that we go out into the world and find as many swan as we can, noticing that they all happen to be white. After a significant enough sample size, we might decide that we have seen enough to tentatively conclude that all swans are *probably* white – but unless we can claim that we have literally observed every swan in the universe, there is always a slim possibility that a non-white swan exists somewhere out there. At best, we will arrive at the conclusion that we have proven all swans are white beyond reasonable doubt, but it is logically impossible to prove something decisively, once and for all, through inductive argument alone.

In contrast, there is another way to interpret premise (i), which is an *a priori*, definitional argument. Perhaps biologists have included whiteness as part of the definition a bird must satisfy in order to be called a swan, and if it is a different colour, then it is labelled as being a different species of bird. In which case, there is no need to go out into the world and observe any birds at all. As a matter of definition, premise (i) is necessarily true. Unlike an inductive argument, the advantage of an *a priori* argument is that once accepted, a definitional acceptance of a premise can never be 'proven false' no matter what evidence might arise later. The weakness, however, is that we can decide on definitions of words in whatever way we like, so definitions alone do not tell us anything about the real world. We are just playing with words.

Ideally, a good deductive argument relies on both *a priori* and *a posteriori* premises to justify not just the validity, but also the soundness of its conclusion. Importantly, there is a trade-off between how likely it is for the deductive argument to be later falsified, and the

likelihood that a deductive conclusion actually reflects real world matters. In order to maximise the ability of a deductive argument's chances to resist falsification, the best strategy would simply be to accept only *a priori* premises. This would lead to a deductive argument that is essentially unfalsifiable no matter what evidence comes up later. For instance, in faith healing:

i. Any successful attempt is attributed to the success of the faith healing procedure.
ii. Any failed attempt is attributed to a lack of faith in the person who needed healing.
iii. Therefore, if you are healed, then God healed you; if you are not healed, then it means that you did not have enough faith.

This is the "head-I-win-tails-you-lose" technique (Wiseman 2010) discussed in chapter five. Importantly, both premises are accepted on *a priori* grounds, and so the conclusion is resistant to any seemingly falsifying evidence that might come up later. If a healing attempt fails, the following deductive argument can be made:

iv. If you have enough faith, God will heal you (derived from conclusion (iii) above)
v. You were not healed (*a posteriori* premise)
vi. Therefore, you must not have had enough faith (modus tollens)

Taken in isolation, premise (iv) looks like this is itself an empirical hypothesis. However, because (iv) is derived from (i)-(iii), (iv) itself is an *a priori* kind of premise. Therefore, combining it later with the *a posteriori* evidence that the method indeed did not work for some people does not undermine the system itself, but instead leads to a modus tollens argument in

(vi) that can be used to derive a conclusion that is perfectly logically entailed. Furthermore, the 'soundness' of this argument can seemingly be (misleadingly) justified by appealing to positive inductive evidence in support of (iv). Maybe there were plenty of people who did experience some sense of healing after the faith healing ritual, though these are probably people who would have recovered anyway, and the faith healing ritual is not causally responsible for the healing. By appealing to inductive evidence only when it is favourable for the system's agenda, this can create the illusion that (iv) is validated on empirical grounds, even though it is an analytic (non-empirical) argument in its core, and it is constructed in a way that makes it empirically resistant.

What is interesting about deductive arguments is that it is theoretically impossible to use a deductive argument in order to derive novel conclusions that contain more information than its premises. A deductive argument is just the process of 'crunching the data' of the premises which have been accepted, but deductive arguments alone do not provide novel information. What this means is that this can create a gap between our understanding of phenomenal consciousness and our ability to explain it academically. If we refuse to include 'personal' or 'subjective' premises in an academic or formal argument because it does not sound 'professional' enough, then for logical reasons, *necessarily*, any deductive argument we make leaves out the personal or subjectiveness that is a characteristic trait of phenomenal experiences.

Once we can acknowledge this logical limitation about deductive arguments, we can turn our attention to an explanatory gap which can arise when engaging in any kind of armchair philosophy. By 'armchair philosophy', what I am referring to now is the sort of philosophising we can conceivably do without ever engaging with the world directly ourselves, such as if we are limited purely to reading books or listening to podcasts, but not through direct engagement. Another way of putting this is if we imagine a world where covid lockdown had lasted a couple of decades. The kinds of youth who grew up in that lockdown might have had access to books and videos, but would never have been a part of the external world directly.

In this structure of 'armchair philosophy' as such, our premises necessarily need to come from other credible sources, since oneself has never had the opportunity to engage with the outside world directly. At best we can judge the credibility of the information we encounter, but unlike the magician who gets to practice their skills and develop their performance philosophies through directly engaging with audience members, the philosopher who is confined to textbooks and podcasts has only second-hand information to work with. It is in this scenario where an important distinction arises between Western traditions (where there is a Christian historical origin) and Eastern traditions (coming from cultures with Buddhist origins).

As we saw in chapter three, while it may be a common misconception that Buddhists believe in reincarnation in a categorically similar way as Christians believing in God, this is not exactly true. Buddhist 'beliefs' are more akin to Greek or Norse mythology, and they are just 'stories' or lore that help us tentatively wrap our heads around the world (Young 2014). Unlike the Christian tradition, where certain spiritual concepts are taken to be *factual* (e.g., "God does exist, the bible says so") and other supernatural concepts are taken to be mere *stories* (e.g., "No young Timmy, there is no such thing as a monster under your bed, don't be silly), the Buddhist tradition does not make a fact-story distinction. Everything is equally just *stories*. If we read a textbook and believe that we are reading 'facts', then it means that we have forgotten this information too was observed, documented, and written by another human being.

It is a part of the human condition to come up with neat and manageable information to wrap our heads around a complex reality, and while some stories have greater predictive value (e.g., science) that help us navigate the world, and other stories might have great emotional benefits (e.g., Buddhist meditation), there is no such thing as 'fact' which is *directly* accessible. (see Heisig, Kasulis, & Maraldo 2011, Ch1). If we believe something to be 'fact', this is often a sign of an overconfidence that is created by the limitation of our own personal attention, interests, and perceptual awareness. We are, by definition, blind to our own blind spots.

Consider, for instance, the following structure of deductive argument which might be motivated by a pro-science attitude:

i. We know from empirical evidence that X is a scientific 'fact' *beyond doubt*.

ii. If X is true, then Y must have (or lack) property Z.

iii. We have overwhelming evidence of X, and so Y must have (or lack) property Z.

Note that the first premise is not superfluous, as we are currently making the following point: Any 'scientific' premise that is phrased in the format of (i) is a massive red flag due to its overconfident nature. Because science primarily uses the method of induction, it is rare, if ever the case, that science will 'prove something beyond doubt'. At best, we might have good reason to believe that something is an extremely good working model for the purpose that we have designed it for (Pigliucci & Boudry 2014). This, however, does not mean that it is just as good of a model when applying it to other purposes. The theory of evolution is an excellent example. It is a very well supported theory, and we have very good reason to believe in it to the point where scientists may tell us it is foolish to believe in any other alternative (such as Creationism). However, while it might be overwhelmingly unlikely that Creationism is true, this does not rule out the possibility that some other theory which is neither evolution nor Creationism will eventually come up, and future scientists might decide that Darwinian evolution – as reliable as it seems in our contemporary times – is a complete hoax. Furthermore, even if the theory of evolution is true, it would be problematic if we drew haphazard philosophical conclusions from this, such as stating that the fundamental purpose of life is to have as many children as possible and spread one's genes as much as possible. Just because the theory of evolution is a good theory for the purpose the model was designed for (a descriptive picture about how our traits

developed), it does not mean that it will have normative implications. This is called the naturalising error, or naturalistic fallacy, wherein people mistakenly infer that because something is true according to a descriptive picture, it is 'better' to live according to that theory (Allchin & Werth 2017).

If any scientific claim makes the exceptionally strong assertion that 'X is factually true beyond doubt' and it seems that X is itself a 'contextless truth' rather than describing a model intended for a certain kind of descriptive exercise, as a general rule, this should ring alarm bells in our mind. Any deductive argument that accepts this kind of authoritarian claim as a premise, where the claim is supposedly justified on empirical grounds, it is very likely committing the fallacy of mistaking a definitional defined truth as an inductive argument. If anyone does wish to use a 'red flag' premise of this kind in a deductive argument, the burden of proof lies with them to spell out what sort of evidence justifies this claim, and in what contexts this claim can be applied. It is not sufficient to appeal to authority, especially if the reference comes from popular books written for non-scientists as the target audience, since sometimes strong assertions like 'X is factually true' is used by scientists or other media personnel to make the material more accessible to the public. To avoid these types of miscommunications, if a scientist ever seems to be making an extremely decisive claim that makes it suspiciously sound too strong for an inductive argument, it is good practice to ask how it is justified, and question whether the strength of the argument transfers to philosophical topics we want to apply it to.

Whether it is referencing psychiatry or physics, it is a red flag if scientists are ever cited as making a supposedly inductive argument that sounds exceptionally strong. The causal closure of the physical (CCP) – the idea that the physical world is causally closed – is supposedly a very decisive fact that physics has already discovered (Tiehen 2015; Papineau 2009). But this is far too strong of a claim to be based in an inductive argument, and it warrants a sceptical examination. It is always possible there are things which exist somewhere out there

in the universe that we just have not measured yet, so it would be more accurate to interpret it as the causal closure of *systems observed thus far*, with an added disclaimer or prediction that because consciousness arises strictly from the brain (assumption), and the brain exists on earth in a familiar scenario that physics has previously analysed, we can (at best) 'reasonably believe with a high degree of confidence' that the system involving mental processes are causally closed in a way that involves only the variables that physics has already discovered. As Wallace has explained, there is an interesting overlap between science and Buddhism in that both value the activity of empiricism. We do not discover scientific facts from an armchair, nor does an Eastern apprentice discover wisdom from a random 'sage' dwelling in a cave. It is up to us to test our beliefs against the world and respond accordingly to empirical concerns in front of us.

To drive this point in, we will look at a Doctor Who analogy before moving on to the final section. In the episode called Midnight (Davies 2008, S4 E10), the Doctor lands on an alien planet called Midnight which has an extonic sun that is so powerful and toxic that no living entity could possibly exist on this planet. Yet, when the bus that the Doctor was on breaks down, and as the passengers await rescue, there is a mysterious noise outside the bus, which appears to exhibit signs of sentience. As the episode unfolds, the alien creature apparently possesses one of the passengers, and also exhibits a capacity for intelligence as it later moves on to possessing the Doctor himself, hijacking his voice and attempting to escape the bus with the rest of the humans. Throughout the episode, the scientist on board continues to insist that there can't possibly be a creature out there – nothing living can exist on this planet! (Note: The next few arguments are not intended as precise deductive arguments, but are used to illustrate how slippery a seemingly logical argument can be.)

i. No 'living' creature could exist outside, given the laws of physics as we know it.

ii. There appears to be a creature outside.

iii. It must be an illusion because we know *a priori* that it could not possibly exist!

The Doctor argues back that the events are indeed unfolding before their eyes, (and though he does not phrase it in quite such philosophical terms), essentially argues there is not much point in rejecting the evidence in front of them on *a priori* grounds, making the deductive argument:

i. No 'living' creature could exist outside, given the laws of physics as we know it.

ii. There appears to be a creature outside.

iii. This is exciting news for scientists because we might learn something new about biology or physics!

What does not get mentioned in the episode, however, is that we have no proof the creature was truly a 'physical' or 'alien' being, and this is merely presupposed since that is the typical pattern of a Doctor Who episode. All we do know is that the creature appeared to exhibit sentience and intelligence, and although no one in the episode raises this point, had a dualist philosopher or a parapsychologist been on board, they might have deductively argued:

i. No 'living' creature could exist outside, given the laws of physics as we know it.

ii. There appears to be a creature outside.

iii. Therefore, that creature must be non-living, i.e., a spirit, a ghost, or a demon!

What is important for our purposes is to see that deductive argument alone does not tell us which of these three interpretations are correct. They are all logically valid, due to each of them interpreting premise (i) differently. The scientist interprets (i) to be an unquestionable 'fact', and therefore the empirical evidence of (ii) is treated as being faulty evidence. The Doctor interprets (i) as tentative inductive reasoning, and so the empirical evidence of (ii) is treated as undermining the tentative theory from (i). Finally, the dualist interprets (i) in a semantic sense of how the concept of 'living' creature is to be defined, and so the empirical evidence from (ii) merely gets categorised as a 'non-living' creature for purely semantic, definitional reasons. All three are logically valid, and the soundness depends on how we interpret premise (i).

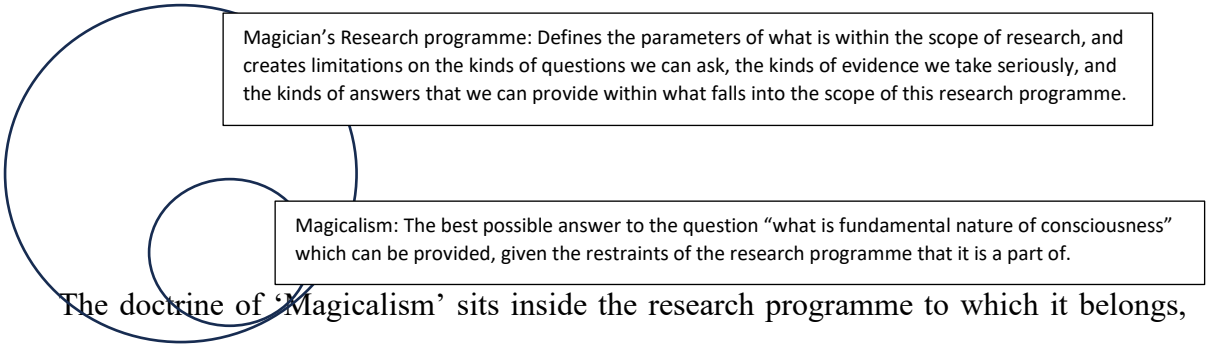
It is a slippery slope where seemingly ontological claims can turn quickly into a semantic argument, and this happens as a result of not specifying what the model is *for*, instead treating the premise as though it is a contextless truth – as though we are all on the same page about what premise (i) entails, and what we are trying to do with it. In the Doctor Who example, the context of the debate is that they are facing a potential threat of an alien attack, and denying the evidence of the entity in front of them simply misses the point of the situation. However, if we believe that premise (i) holds a 'fact' status, then there will be a hesitation or reluctance to question it just because it contradicts our immediate perception in the moment. Things which we believe hold 'fact' status tend to grip us more strongly, and if our own perception does not match the 'fact', we might end up doubting *ourselves* rather than doubting the information that we are working with. It is not unlike the astrology-believer who breaks up with the love of their life over a star sign – believing that it is *written in the stars* in some irrevocable way that their relationship is doomed to fail. It is this sort of conviction that magic shows tend to exploit. Wherever we find ourselves succumbing to non-empirical convictions, a gap of seeming 'impossibility' emerges, and creates the sensation of a magical illusion.

Magicalism is not an advocation of supernaturalism, but rather, it is the idea that 'facts' need to always be treated with a pinch of salt. No matter how credible the information is that we are exposed to, at the end of the day, the 'Magic' in life arises from us personally confronting the limitations of our knowledge, and exploring the blind spots which fascinate us. This type of personal confronting cannot happen unless we treat ourselves like Eastern apprentices who are allowed to question our masters; as opposed to 'good students' who are told we must memorise a textbook and not question the authority. To understand the 'warm fuzzies' which will be discussed in a moment, it is important that we enter the discussion with the right headspace: We are not just looking for a doctrine that tells us what consciousness *is*, but this goes hand-in-hand with the task of looking for a novel research programme which is better suited for dealing with personal and subjective questions. The nuanced and 'biased' aspects of life can still be investigated through a professional and academic capacity. We do not want to leave all the personal, subjective, normative work to be done by self-help gurus and social media influencers. Academia can still play a more credible role in these topics, and this does not have to involve a procedure of statistically analysing quantifiable *universals*.

7.4 The 'Warm Fuzzies'

On a personal level for each of us, there may be a certain kind of 'love' or 'warmth' that we feel towards our friends and family. We feel this way precisely because we feel close; we feel their emotions and experiences unfolding next to us; we feel the 'fuzzies' towards them or coming from them. Their happiness and their lives matter to us *because* they are conscious and alive. It is this sense of the 'fuzzies' that need to be captured to fully encapsulate a theory of consciousness. This is a central factor of what it is to be human, and to have a human mind.

In the next chapter, we will be establishing what the Magician's Research Programme entails, and Magicalism can be understood as a subcategory of this novel research programme:



The doctrine of 'Magicalism' sits inside the research programme to which it belongs, where Magicalism is the word used for the ontological claim, and the Magician's Research Programme refers to the collection of research methods used for engaging in philosophical or empirical debates in consciousness studies, including ontology, psychology, psychiatry, maybe even love, and so on. Magicalism is essentially a *noun* or a *label*, whereas the Magician's Research Programme is a collection of verbs – i.e., an explanation on 'how to' conduct research.

Analysing 'love' exceeds the scope of this current thesis, but some sense of the 'warm fuzzies' needs to be addressed in order to understand *in what sense* Magicalism is still dealing with Real Magic and yet it is not about supernaturalism. There is a scope of storytelling that us humans are capable of addressing – love is perhaps the best example of this – which inherently creates a sense of magical whimsy around it. Unfortunately, these topics tend to be difficult to address head-on in the more 'objective' or 'unbiased' favouring structure of academic discourse, but they are still important topics, nonetheless. In order to understand Magicalism, it is key to acknowledge that Magicalism was never intended to sit comfortably in the Physics Research Programme that other hard sciences are a part of. The Magician's Research Programme is a more humanised endeavour, and although this means it might need to compromise on some fronts when it comes to 'objectivity', it is better to bite the bullet that even a heterophenomenologist is not fully objective anyway – so we might as well take better care of the nuances behind the stories that we are trying to tell.

Studying consciousness is not the same as studying physics, largely because there are far too many qualitative variables involved in understanding how our conscious experience works, and every study we conduct contains the risk of feeding back into the population we are trying to study. What this means is that a future completed field of consciousness studies might look nothing at all like physics, in which case, we might ask what this entails for the success or failure of 'physicalism'. If, for example, the future of consciousness studies relies on a novel research programme that centres around qualitative research – such as individual case studies or phenomenology – then does this mean that consciousness itself is fundamentally qualitative and non-physical, or does this mean that consciousness is still 'physical' because we believe social phenomena are fundamentally reducible to brain waves and other 'physics-y' concepts? Does it even matter what we call it using Cartesian terminology? Arguably, it might be time to stop treating ourselves as Descartes's Apprentices – and to restart the mind-body debate anew. Consider, for instance, the way that Karl Popper discusses some pitfalls in philosophy:

> "there is much philosophical writing […] which may justly be criticized as meaningless verbiage […] every philosophy, and especially every philosophical 'school', is liable to degenerate in such a way that its problems become practically indistinguishable from pseudo-problems, and its cant, accordingly, practically indistinguishable from meaningless babble. This, I shall try to show, is a consequence of philosophical inbreeding. The degeneration of philosophical schools is in its turn the consequence of the mistaken belief that one can philosophize without being compelled to turn to philosophy by problems which arise outside philosophy." (Popper 1963, p.71)

It has been over four-hundred years since the original Cartesian debate was proposed by Descartes, and since then, this 'myth' of the mind-body problem has taken on a life of its own. In contemporary philosophy of mind, we may have a strong impression that Descartes's main

concern was an almost chemistry-esque kind of quest, of trying to work out what sort of 'substances' exist in the world. However, perhaps the way that Descartes observed clockwork automata might have been similar to a magician observing the mechanism behind a gimmicked magic trick. Is human nature also like a magic trick? And if so, how much of it is a 'gimmick' and how much of it is part of our 'soul' or 'essence'? Or is it *all* merely a wonderless gimmick?

It is characteristic of analytic philosophy that we tend to pursue the ultimate answers to life and existence. We might like to think that these answers are derivable through enough deductive rigour, and by being cautious enough to only accept true premises, this prevents us from including falsities in our analysis. It is as though to treat life and reality like a cookbook. If we know exactly the right formula, and make sure not to add the wrong ingredients, then we will cook something that is promised (by the recipe book) as being a solid dish – one that will satisfy everyone. The task of the philosopher is to cook that dish and answer, "What is the world *like*, fundamentally" as though there is a 'correct' answer that will one day reach universal consensus. In doing so, we try to transcend our own position as 'mere' human observers and reach ultimate objectivity. But problematically, perhaps human beings are far more 'social' creatures than what folk psychology might suggest. It is unclear whether we are even capable of forming 'beliefs' in a vacuum, under an expectation of unbiased objectivity. Perhaps 'beliefs' are a kind of storytelling that we become emotionally attached to, in order to navigate our individual lives and satisfy our own personal interests. This does not mean that we are 'biased' or that knowledge-seeking is a doomed endeavour. It just means that human beings are highly social and relativistic creatures, and that consciousness cannot be accounted for without acknowledging the role of rapport, connectedness, and other 'warm fuzzy' elusive concepts that bind together the collective activity of knowledge-seeking in human communities.

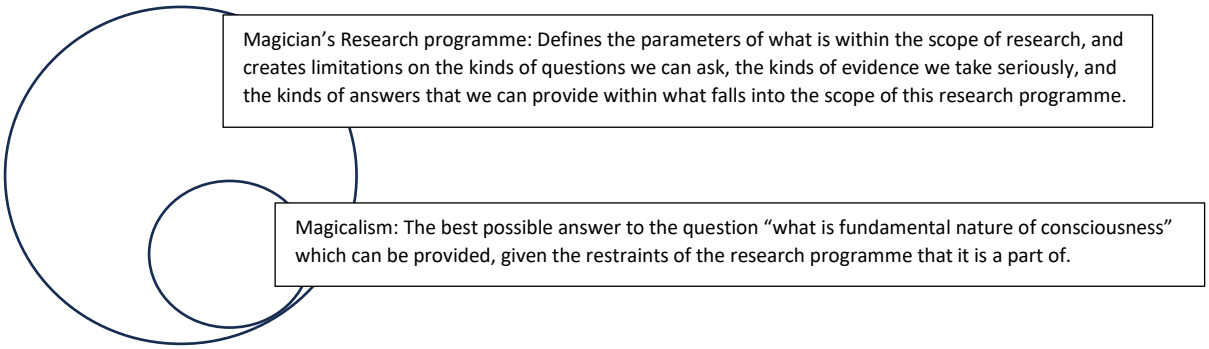
In a deductive argument, if the premises themselves do not include 'fuzzy' concepts, then logically speaking, the conclusion too will not include any 'fuzzy' concepts. A deductive

argument can only ever derive things that were already a part of the premises – and crucially, 'fuzzy' concepts in human affairs come from our own personal experiences. A young philosopher who was raised in a twenty-year version of covid lockdown might never have experienced love, friendship, grief, or other 'human' concepts. And if the books they read do not include these concepts, then what choice would they have but to learn life second-hand, from psychiatry books, definitions decided on by a committee, and a lack of understanding in what it means to take charge of one's own agency? It would be unsurprising if we begin to construct a deductive and logically valid thought process that ends up too sanitised. The Magician's Research Programme strives to discuss a novel academic angle for still including the 'warm fuzzies' into our professional discussions. It is key to this is that each of us, as apprentices in life, gain personal experiences through first-person, empirical encounters. In the final chapter, we will turn to ideas about how to make these 'fuzzy' ideas a little more tangible, once again, by moving away from being Descartes's apprentices who are bound to Cartesian (outdated) lingo, and instead, embracing the mindset of an Eastern Magician's Apprentice.

8. Magician's Research Programme, Karmic Fortune, and Buddhist Souls

In this final chapter, it will be argued that Buddhist 'karmic fortune' is the fundamental causal power when it comes to analysing consciousness, and that consciousness is reducible to 'souls', though this is in the sense of a Buddhist soul, which is *distinct* from supernaturalism. These ideas are *not* to be understood in a dualist way, nor as supernaturalist claims. No psychic abilities are accepted or advocated in this thesis, and the existence of all supernatural entities are denied. Furthermore, while this chapter is written *inspired by* Buddhism, it does not require the reader to accept Buddhism more holistically. Everything claimed in this chapter is, at the end of the day, religion-neutral, and strives to be inclusive towards non-Buddhist readers.

The theory that has been advocated in this thesis has been called Magicalism, the doctrine that the fundamental essence of consciousness is Magic, but in the sense of a card trick and not in the supernatural (Harry Potter) sense. In this chapter, we focus on the bigger picture of the research programme to which Magicalism is a part of.



The general argument will take the following form:

i. Magician's Research Programme is defined as having a 'hard core' of X, Y, Z.

ii. This research programme should lead to the production of novel, non-ad hoc hypotheses that contributes to something outside of protecting its own system,

where these ‘risky hypotheses’ will resist falsification, and it will be able to do this process more effectively than other competing research programmes.

iii. Magicalism is the best answer to the mind-body debate that can be offered using the methods included in the Magician’s Research Programme.

iv. Thus, Magicalism is the best answer to the mind-body debate when compared to answers given by other competing research programmes, in that Magicalism is the best-fit logical answer provided by the most favoured ‘progressive’ enterprise.

And finally:

v. Magicalism points in the direction of accepting (a) Buddhist ‘karmic fortune’ as a fundamental causal power, and also, that (b) phenomenal consciousness is fundamentally reducible to ‘souls’, but in a non-dualist way.

vi. Therefore, we should accept ‘karmic fortune’ and ‘souls’ as a part of our ontology.

To begin, we will need to start with the first process of defining what the ‘hard core’ of the research programme is going to be. Note that because this is a definitional issue, the way we determine the ‘hard core’ is not itself meant to be a falsifiable claim. Instead, it is quite openly an *a priori* commitment that we are going to be making, in the hopes that it will lead us to a ‘progressive’ research programme with great explanatory power in supporting risky hypotheses. Thus, it should be noted that there are three independent processes that will need to happen:

(1) Give a definition for what this research programme is about (*a priori* commitment).

(2) Justify why this *a priori* decision is promising, and why we should give it a chance.

(3) Empirically test how the research programme fairs, once we begin putting it to the test by attempting to falsify its risky hypotheses, which forms its ‘protective belt’.

This current thesis is only concerned with (1) and (2), as (3) requires empirical testing and it can take decades or longer to fully validate and develop a novel research programme. Therefore, if the depiction of the research programme seems somewhat speculative at this current stage, to some extent, this is to be expected. A novel research programme is only just beginning, and we cannot yet expect it to have empirical backing that an old research programme would have.

8.1 The 'Hard Core'

We will begin by defining the 'hard core' of the research programme in a way that is specific enough so that we may contrast it with other competing research programme, yet, it is open-ended enough that it creates room for later developing a 'protective belt' around it. The research programme must be flexible enough that it can grow or adapt as it accumulates empirical evidence, but also just rigid/specific enough that it does not become indistinguishable with competing research programmes. There will be three elements included in the 'hard core':

(a) It is a normative enterprise rather than a descriptive enterprise.

(b) It takes a teleological stance rather than a deontological one.

(c) It is designed specifically for consciousness studies and cannot be applied to an analysis of non-conscious entities.

We will now consider each criterion in turn:

(a) It is a normative enterprise rather than a descriptive enterprise

The Magician's Research Programme makes it explicit that we are looking at a research programme analysing how we *should* do this or that, rather than merely describing reality as bystanders. Note, however, that this does not entail our research programme must be concerned

with morality. 'Normativity' is wider-scoped than just 'morality' or 'ethics'. For instance, to claim "I should get out of bed" is a normative statement, but it is not a moral or ethical claim.

The study of medical research can be given as an example of another research programme that also is inherently normative. Although many tasks in medical research can still make descriptive claims, medicine is a field where the clinician or researcher is not a mere bystander. Rather, they are actively engaged in the giving of advice, recommendations, or treatments (Glas 2012). In contrast to this, an example of a descriptive research programme might be the heterophenomenological research programme described by Dennett (2007; 2003; 1991), touched upon in chapter seven, where the heterophenomenology is a descriptive enterprise where researchers are just bystanders who document what they are observing.

In this context, the Magician's Research Programme advocates that consciousness studies should be closer to the field of medicine than it is to heterophenomenology. However, it is also distinct from the medical research programme of psychiatry in that the Magician's Research Programme is not so much about labelling mental states into distinct categories as approved by credible authority figures. This now brings us to the next 'hard core' element:

(b) It takes a teleological stance rather than a deontological one

Now that we have made it clear we are dealing with a normative enterprise, the next step is to distinguish between two mutually exclusive categories: teleological and deontological. Note here that while we may be accustomed to using the word 'deontology' to mean something synonymous with Kantian ethics, the word itself (with a lowercase 'd') contains a more generic meaning, and by definition, it stands in direct opposition with the notion of 'teleology' (Kymlicka 1988; Vallentyne 1987; Ashby 1950). From here on, we will understand Kantian Deontology to be a subcategory of the wider category called 'deontology' more broadly

construed, and unless specified, the word ‘deontology’ for the rest of this chapter will refer to the broader concept to which Kantian ethics is only one example of.

The teleology/deontology distinction in ethics was introduced originally by C.D. Broad in 1930, where teleology is defined as a normative activity that strives for a certain future purpose or goal, whereas deontology is defined as a normative activity that strives to explain or judge a past event (Vallentyne 1987; Broad 1930). For our current purposes, we are expanding this distinction to apply to normative discussions that are not necessarily ‘moral’ concerns, such as when discussing the distinction between a medical research programme and the novel magician’s research programme that we are currently developing.

The distinction between teleology and deontology can be understood as a ‘forward-looking’ or ‘backward-looking’ way of evaluating normative claims. In ethics, utilitarianism and virtue ethics are both examples of a teleological normative theory. In utilitarianism, the aim is to increase as much utility as possible in the future; and in virtue ethics, the aim is to become virtuous and achieve eudaimonia, i.e., increase our quality of life. According to this teleological interpretation, whether the action or trait is ‘good’ requires looking into the future, rather than looking into the past. That is, if the utility indeed did increase as a consequence of the action, then a utilitarian would say that the agent’s behaviour was ‘good’. Regardless of what the agent’s intention is, or how well they follow various rules of thumb, if they indeed do not obtain the goal of increasing utility, then they have performed poorly relative to the teleological goal they have defined for themselves. (Kymlicka 1988; Vallentyne 1987; Ashby 1950; Broad 1930). In contrast, (unsurprisingly) Kantian ethics is given as a prime example for a deontological ethical theory. In this type of framework, the ‘goodness’ of an action is not dependent on the consequences or outcomes, and the ‘rules’ are worked out independently of the direct feedback which the actions produce. Therefore, if we accept a rule such as ‘one ought not lie’ as a categorical imperative, then whether the agent was successful in their goal to be

'good' does not require taking any consequences into consideration. Rather, the normative value of an action can be evaluated by looking at how well an agent's behaviours align with rules that were created beforehand according to some other consideration that is not about consequences or outcomes of events. To summarise, "A deontological norm is one that evaluates an act by a characteristic that cannot be gathered from its consequences" (McCormick, 1973, p. 62) whereas teleological theories make normative claims in virtue of how well consequences align with the desired outcomes (MacDonald & Beck-Dudley 1994).

As a quick sidenote, an opponent might try to argue that even deontological frameworks are 'teleological' in a sense, in that perhaps Kantian ethics is directed at the *goal* of being as 'good' as possible by satisfying all moral obligations as dictated by categorical imperatives. The bottom line that is important for our current discussion, however, is not whether both frameworks reduce to goal-drivenness, but rather, that teleological frameworks are more about striving for certain outcomes/consequences, whereas deontological frameworks are more to do with categorising actions or behaviours and labelling them with value-judgements. In essence, a teleological view is more about advice-giving; a deontological view is more about judgement.

To link this back to our discussion from the previous chapter, a deontological endeavour is one that is concerned with setting up rules or categories through an *a priori* approach, such as by using deductive arguments or an appeal to authority. In contrast, a teleological endeavour is one that is concerned with striving for a certain outcome, and the efficacy of one's methods can be tested through an *a posteriori* approach, like responding to feedback and adapting to it.

It is in this sense that our research programme is defined as a teleological endeavour, rather than a deontological one. As we shall see in the next section, a magician – who is a craftsman at heart – is engaged in the teleological task of trying to entertain an audience, trying to fool spectators, trying to sell more tickets, and so on. A 'good' magician, therefore, is defined

in terms of how well their consequences align with the desired outcomes, rather than whether they abide to some kind of abstract rule. This claim is further justified later in the section 8.2.

(c) It is designed specifically for consciousness studies

For now, we will return to the final criteria before returning to the teleology/deontology distinction in the next section. This last element of the 'hard core' is that our research programme is specifically designed for the purpose of analysing things relating to consciousness. Note that a research programme is essentially a kind of tool, much like a hammer. How 'useful' or 'progressive' that research programme is will depend on whether it is being used for what it is *supposed* to be useful for, and it is not intended to be the equivalent of an all-purpose, the-only-tool-you-will-ever-need. A hammer, for instance, is an excellent device if we want to wedge a nail into a wall, but it is entirely unhelpful if what we want to do is serve our friend a glass of water. Similarly, the research programme we are developing is designed specifically for engaging in consciousness studies, and it does not claim to have explanatory value in other fields where the task does not involve analysing any conscious entities. Therefore, in the same way that we should not complain that a hammer is 'degenerative' because it is useless at serving a glass of water, we should also expect that the research programme we are developing probably will be useless if we end up misapplying the tool for something it was never designed for. In other words, there is no need to believe that the Magician's Research Programme is a replacement for say, the Physics Research Programme, and we can allow that multiple kinds of research programmes can coexist for different purposes.

8.2 A Magician as a 'Craftsman/Entertainer'

To reiterate, we have defined the 'hard core' as being composed of three elements:

(a) It is a normative enterprise rather than a descriptive enterprise.

(b) It takes a teleological stance rather than a deontological one.

(c) It is designed specifically for consciousness studies and cannot be applied to an analysis of non-conscious entities.

Next, we move on to discussing why this research programme deserves the title of the Magician's Research Programme, which is not explained in (a)-(c). In particular, the concept of a magician becomes important when we further explore the teleology/deontology distinction. While Dennett (2016) has written that, supposedly, a magician's burden is to work out the method behind the trickery, in practice, the career of being a magician is more in the category of a craftsman or an entertainer, rather than a 'professional debunker' or an academic scholar.

Setting philosophy aside for a moment, if someone wishes to make a living as a magician, it is not enough to have an astute eye for working out how other people do their tricks. Instead, it is far more important to be able to perform that trick yourself, and to do so in a way that actually entertains an audience so that they can continue offering their performance as a source of entertainment (leading to selling tickets, etc.) – otherwise, a career as a magician becomes rather unsustainable. No matter how good a magician happens to be at debunking the method behind other people's tricks, this does not mean that they are a 'good' magician themselves; and equally, just because they are very good at deceiving people, it also does not mean that they are successfully entertaining the audience. It is in this sense that being a magician is fundamentally a teleological kind of task, and the magician learns to navigate this through their real-world direct contact with the world. A 'backseat driver' magician who has no idea how to do their own craft might get good at debunking myths on YouTube, but fundamentally, what is needed is the skill of doing the performance, *doing it well*, and providing entertainment.

The magician can pass judgement about other performers all they like, debunking illusions and criticising technique, but this is not the primary task of a magician. In a teleological normative endeavour, the main task is about striving for certain outcomes. This does not have to be monetary or fame-related, but can include things like bringing joy to an audience or creating a successful evening of wonders.

But contrast this now with a slightly comical example of a deontological magician, just to see how silly this would be. We can consider a hypothetical magician takes a deontological approach to their career. Let us say they have decided there are certain abstract criteria which make you a 'good' magician, and they judge anyone who fails to fit into these criteria, such as:

1. If you believe in yourself hard enough, the card *just will* vanish, without any method.
2. If you are open and vulnerable with your audience, they *will* always enjoy the show.
3. A good magician must dress pretentiously: dark capes, top hats, and velvet suits only.

Notice that these are all *a priori* 'rules' rather than *a posteriori,* testable predictions. For instance, if the card does not vanish even when the magician tries their best to believe that it will, this does not falsify the rule, but rather, it just means that the magician did not actually believe hard enough. They are a 'bad' magician, according to the 'rules' that we have accepted *a priori.* The rule itself is assumed to be true and is unfalsifiable. Similarly, if the magician is open and vulnerable with their audience about their insecurities or their misfortunes, and the audience happens to find this very uncomfortable, then it is the audience that is wrong. They are mean and nasty people – but the rule itself is unfalsified. Finally, if the magician dresses pretentiously, that *just is* what it means to be a 'good' magician. They are a fantastic performer.

Recall once again that in this thesis, the word deontology is a broader category than Kantian Deontology. In fact, Kantian ethics does attempt to avoid the formulation of 'absurd' rules such as these by following certain rational constraints in developing those 'rules', such

as the categorical imperative (Reath 1989). In this thesis, we are not attacking Kant, nor do we need to claim that deontological approaches are always ineffective or trivial. Rather, what we need to take notice of is a generic structural problem with deontology that is problematic if we were to use this approach for consciousness studies: That deontology lacks a feedback mechanism (Roache 2024). The main goal of a deontological endeavour is to pass judgements, not to make testable predictions (Kymlicka 1988; Vallentyne 1987; Ashby 1950; Broad 1930).

It is in this sense that modern psychiatry is more of a deontological research programme. It independently sets up certain criteria for what is or is not considered a 'medical' disorder, which is either politically-driven, funding-driven, semantically-driven (when operationalising constructs for the purpose of making them empirically measurable), or just outright decided on by a committee of representatives (Johnstone & Boyle 2018, p.26). According to this independently derived set of categories, we can then plug people into the equation to see how well they fit into this or that. The process of categorising people will then serve a judgemental purpose – to see who we can 'blame' as lacking character, and who can avoid 'blame' and receive special support because they are deemed to have a medical condition. It is a *judgement* passed in regards to someone's agential capacities, and while those who advocate the system might try to insist that psychiatry of course is a teleological endeavour designed to help improve people's mental health, the real question is whether the system does its own heavy-lifting, and ventures into non-ad hoc hypotheses that pose a risk to its own system.

A truly teleological endeavour is more focused on trying to falsify its own hypotheses system (and then hoping their system will withstand the tests) in the way that is discussed by Popper or Lakatos, as we saw in chapter five. Therefore, a truly teleological version of psychiatry would be focused on trying to *falsify* the idea that a diagnostic system is actually helpful for their patients, and if it successfully resists these kinds of 'risky' hypotheses, then we will have an increased confidence in the ability of the research programme to strive for the

kinds of goals and outcomes it claims to be striving for. Consider again, the rabies example from chapter six. As long as physicians were preoccupied with labelling people according to their symptoms, it became impossible to find a cure to the problem because it grouped too many people together that did not have the same cause in common (including people who exhibited 'hydrophobia' due to a fear of dog bites and the mere suggestion that this could make them hydrophobic (Carter 1982). By changing the research programme to a cause-based structure rather than a symptom-based structure, physical medicine was able to transition into a more teleological framework for developing vaccines and fixing the problems it dealt with.

In contrast, psychiatric medicine is still in the same stage as pre-rabies medicine. In fact, with many psychiatric conditions such as autism or ADHD, members of the community sometimes consider it to be inappropriate to say these conditions need 'curing' since patients take their neurodivergence to be an inherent part of their personalities – not something that needs 'curing' (Hughes 2021). Instead, the psychiatric system is primarily useful for helping people to receive benefits or special concessions when required (Hughes 2021), for giving people a sense of relief from the brain-blame dichotomy (Boyel 2013), and giving people language for expressing the kinds of problems they are going through so that this can be communicated better to other people. As psychiatry stands today, it is a quasi-medical programme with an unquestioned 'hard core' that mental illnesses are categorically the same as physical problems (Johnstone & Boyle 2018). This is definitional, not empirical.

Compared to the deontological framework which is not feedback responsive but primarily plays the role of telling us when we are permitted to 'judge' people according to some kind of value-based, normative standard, the teleological framework of the Magician's Research Programme differs from psychiatry. Before moving on, we can give one final note on the psychiatric research programme as it exists today: If it truly wants to claim that it is a *teleological* endeavour that is primarily concerned with achieving the goal of increasing the

psychological wellbeing of members of the public, then the first and foremost task it needs to engage in is to run empirical tests with questioning its own system. This is the hallmark of a 'good' science as discussed by both Popper and Lakatos – an ability to scrutinise its own framework as rigorously as possible to see if its risky hypotheses will resist falsification. This process may be considered politically incorrect or offensive towards some members of certain communities, but if this task cannot be done satisfactorily without regressing into *ad hoc* explanations, the research programme is deemed at best 'degenerative' according to Lakatos, or worse a 'pseudoscience' according to Popper (Chalmers 2013, Ch7 & 9; Lakatos 1978; Popper 1963). Without this rigour of first questioning itself and facing up to testing uncomfortable hypotheses, it is unclear whether it is deserving of the title of 'science' at all.

8.3 Why Magicians are Significant

We have so far explored a few important components to the Magician's Research Programme. To reiterate, its 'hard core' is composed of the following three elements:

(a) It is a normative enterprise rather than a descriptive enterprise.
(b) It takes a teleological stance rather than a deontological one.
(c) It is designed specifically for consciousness studies and cannot be applied to an analysis of non-conscious entities.

Furthermore, it is referred to as the *Magician's* Research Programme because of a special relationship which magicians have when constructing beliefs. We have already discussed the normative and teleological nature of being a professional magician, where being feedback

responsive is an important aspect of making a living as a performer. However, this itself is not unique to magicians, and plenty (if not most) careers will take feedback responsiveness to be an important quality when it comes to making a living out of a successful career path.

What is particularly interesting about magic tricks is that the gap between the method and the effect is incredibly large, in a way that tends not to happen in other kinds of professions. What this means is that – ironically, performing magic tricks might be one of the hardest things to do 'as if by magic' when a person lacks knowledge of method. What is written on the tin – e.g., change the colour of a playing card – hardly gives any hint to the novice performer about how to achieve this effect. If the correct method is not implemented, the effect simply will not happen. That is, if a magician simply stares at a card box expecting the card inside it to vanish 'by magic' without executing the method, then they will be disappointed to find that no amount of waiting will make the card go anywhere. Understanding that effects require a physical cause is a quintessential part of performing a magic trick and *mechanisms* become crucially important.

The importance of method in magic tricks is similar to the structure of science. It is unadvisable to perform surgery without knowing the method as though expecting it to work 'by magic' (though see Wiseman 1997, Ch1 for examples of fraudulent faith healing surgeons who do this anyway). The *method* in science is extremely essential to the success of that endeavour, as science is often engaged in the teleological task of understanding cause-and-effect in order to produce desired consequences – like building a bridge or developing a vaccine.

Secondly, in a magic trick, the storytelling is almost just as important as the mechanism. This makes magic shows not only similar to a science in the sense of having important methods, but magic is also a kind of artform at the same time. If a magician has a horrible stage persona (the *"I will fool you, and badly"* type of attitude that can be off-putting when watching an arrogant magician; Brown 2023, p.288), no matter how well the mechanisms are implemented,

the stage show might not seem very 'Magical'. Similarly, if there is no element of surprise, wonder, or seeming impossibility, it might also lack a sense of 'Magic'. A magic show is a specific type of stage performance, often involving the idea of *achieving the impossible*. If no sense of seeming impossibility is conveyed in the first place, the show is not very 'Magical'.

In fact, some psychologists have explored the idea that learning magic tricks might be beneficial for enhancing the wellbeing of individuals in clinical settings (Wiseman & Watt 2018). Research has shown before that performance arts in general can have a positive impact on both physical and mental wellness (Stickley et al., 2017; Fraser, Bungay & Munn-Giddings 2014; Noice, Noice & Kramer 2014), but magic in particular is a type of performance that involves a conjunction of skills that can be beneficial for self-improvement or self-enrichment:

> "'It doesn't matter that a magic trick is not normally viewed as serious education, it helps to develop a quick eye and a sharp inquisitive mind. A trick, a puzzle, a logical analysis, or a mathematical theorem—everything is good if leads to this aim.... To perform just one or two magic tricks you'll have to do some work: to understand the secret of the trick, to complete all necessary preparations, and to practice until your hand becomes more dexterous and flexible." (Simashko, in Fedorov 2018, p.47-48).

Stage Magic involves honing not just the teleological ability to troubleshoot in order to achieve a desired goal/effect, but it can be a holistic toolset for improving one's ability to think rationally, both in a scientific sense, and in an artistic sense. While the empirical data is still ambiguous regarding the level of efficacy in magic related therapies, this is largely due to small sample size that mainly focus on clinical cases, and other problems with experimental design. (Wiseman & Watt 2018). From a theoretical point of view, magic involves the skills (see *ibid*):

i. Lateral Thinking and Problem Solving

ii. Fostering Trust (the magician is trusted with the secrets of knowing the methods)

iii. Fosters a Spirit of Togetherness (sharing a secret with a group of kindred spirits)

iv. Storytelling and Imagination

v. Practical Skills, Cognitive Skills, Motor Skills

vi. Risk Taking

vii. Adaptability and Resilience

viii. Self-Esteem and Confidence

In essence, to perform a magic show successfully, quite a large range of skills are needed, both mechanistically, and in terms of improving a capacity to *enjoy the impossible*, and bond with fellow human beings over something light-hearted and not-too-serious. Science, poetry, and *heart* is all required, and reducing Magic to just sleight of hand or deception is to miss the point.

The significance of the Magician's Research Programme is that it offers a new way of wrapping our heads around human nature. Instead of searching for the right 'laws of nature' rules, or categories of some kind that describe human beings most accurately, we can instead think in a more teleological way. "How do we improve our collective wellbeing?" The *telos* might be directed at happiness, fulfilment, *eudaimonia*, or just a sense of enjoyment in being a part of a lively community – but whatever the goal is, it is often difficult to work out how to actually strive for these desirable ends. This can also make us overly self-conscious that we are not happy enough, not normal enough, and so on, creating a negative feedback loop where the more we *strive* for improvement, the more we become self-conscious, the more we become distracted and miserable, the more with feel further away from the desired goal of happiness.

It is reminiscent of the hypnosis show technique when the participant has their foot stuck to the floor. To respond to the command that we should 'try' and be happy can create a sense of decision paralysis. What if we do not fit into categories that are typically considered socially 'normal' or 'preferable' – but to try and 'be ourselves' also makes others uncomfortable? We become stuck in our own self-conscious loops, the quest for happiness undermining our efforts.

But consider that conscious life itself, is like a magic show. To escape this loop of chasing happiness in the horizon, the key is to combine both science and art – that is, both mechanism and poetry – and this leads to a qualitative analysis of the human mind that does not rely on wishy-washy or outdated terminology such as Cartesian theatre. To say that the human mind is Magical does not mean that it is without mechanism, but rather, it means that mechanism, while important, is not sufficient on its own. Equally, just telling stories is also insufficient.

Take for instance, what needs to happen if a person wishes to improve their quality of life. There are two possible approaches: (1) address the mechanism to improve consequences, or (2) to address the poetic storytelling to create a more positive headspace. According to faith healers, or 'The Law of Attraction' – the idea that positive thinking is enough to make positive things happen to us (Byrnes 2007) – the *mechanisms* of causal factors are entirely ignored, leaving the agent vulnerable to unexpected misfortunes and reducing the amount of control or authorship they truly have over their own life. This is a mechanism related problem, like a magician who does not realise that a sleight of hand is needed in order for the card to change colour as it should, and is confused that sometimes the trick works (when a hidden collaborator happens to do the method for them) and sometimes the trick fails (when no one else is taking care of the logistics). At the same time, ideas such as clinical depression just being a chemical imbalance in the brain is itself a type of 'story' that helps to alleviate blame or provide a feeling of closure. However, the story itself is not enough. The depression might have a cause that is external to the individual, such as issues with global economy, political turmoil, or family

problems, and to leave this social story untold can make a patient feel unobserved, helpless, dehumanised, and more depressed than they were before the clinical intervention (Boyle 2013).

The Magician's Research Programme offers a way forward with the qualitative 'sciences' in that we should think of them not as a search for quantifiable *universals*, but rather, as a sort of artform. It is like a Buddhist way, or *Tao,* in exploring what it means to have a human mind. And so, this concludes the defence of why this research programme sound like it might be promising and deserves to be put to the test empirically in future studies. The idea that treating our own lives like a magic show – that we can learn to understand the cause-and-effect relationship of psychological questions by investigating *method* and *storytelling* behind our belief systems – can overall lead to an increase in quality of life, thereby combining the theoretical approach of philosophy while still taking care of Problematic Problems in consciousness studies, rather than keeping the two approaches vertically divided. This research programme deserves the chance to be tested out in the field so that the 'protective belt' gets a chance to develop, resist falsification, and eventually lead into a fully developed research programme for the philosophy and psychology on 'quality-related queries of the conscious life'. A further defence of the research programme exceeds the scope of this thesis, as to contribute to the protective belt itself requires empirical testing, which can be left for future occasions.

8.4 'Karmic Fortune' and 'Souls'

So far what has been defended is a research programme that is normative, teleological, and focuses on exploring qualitative issues involving conscious entities. This is a research programme that will hopefully have wide implications, not just in philosophy of mind, but also towards psychology, psychiatry, maybe ethics, and anything else that involves an analysis of conscious beings. However, the main purpose of this current thesis is about ontology, and so

although it is a shame to not be able to dig deeper with the new research programme in other respects, we turn now to our final discussion. 'What is the answer to the ontological question?'

Colloquially, there may be a common misconception in the way the notion of 'karma' is understood as meaning something similar to The Law of Attraction – the idea that if we think positively enough about things we want, then positive things will happen to us (Byrnes 2007). For instance, let us say that someone wishes to win the lottery, but they decide to stop by an orphanage to give a small donation first because it is 'good karma'. What they believe is that because they are making a charitable gesture to the universe, now the universe is more likely to return this charitable gesture back to them, thereby increasing their odds of winning that lottery. Of course, this is not what Buddhists mean when the term 'karmic fortune' gets used.

A better way to understand the concept is to split it into its two components (a) karma and (b) fortune. The latter refers to a sense of luck or chance, the randomness of an indifferent causation. This is similar to the way Stoic philosophers discuss fortune (Dhiman 2021) and in a way, it is not too dissimilar from the way modern physicists may refer to 'physical' causation. It is the push and pull of indifferent various events, intermingling with one another without any overarching 'agenda' as dictated by a god or an intelligent designer. There is no *telos* as such in the concept of 'fortune' itself. Fortune is just a word for the 'unfolding of reality' over time.

Karma, in contrast, is a more humanised concept. The most superficial (and somewhat incorrect) understanding would be to interpret it as the claim that 'good things happen to good people' or that 'what goes around comes around'. However, this is not quite accurately. The "maturing of karmic fortune (*kamma vipāka*)" is described as:

> "The effects of actions that have been performed continuously from an infinite past intermingle with each and every thing that exists, with everything that breathes and lives. These effects are

manifest here and now in the substrate of my whole existence. This is the outer limit of this entity called "I"" (Tamaki 1982, in Ed. Heisig, Kasulis, & Maraldo 2011, p.127).

To make it easier for us to wrap our heads around this, we can use a poetic analogy. Imagine that human agency is like a wind that creates a ripple in the large sea. Each time we take action, we create a small ripple of our own. The sea itself is massive, and most likely, our ripples will not affect anything significant. At best we might nudge a nearby seaweed slightly wayward off its original path, but according to the big picture, each tiny wind is nothing, compared to the sea's natural waves created by non-human factors (such as waves created by the gravitational pull between earth and moon) even when no humans are present. According to this analogy, the waves created by gravitational pull is like the 'physical' causation that exists even without human influence, and the waves created by wind is like the 'mental' causation that occurs as a product of human agency. However, this interpretation is too dualist and inaccurate. It is not that there are two kinds of fundamentally different causal forces – gravity and wind (and maybe other factors as well not included in this analogy) – but rather, it creates one holistic picture in which all causation is tightly intermingled. There is a *plurality* of causes, but just one sea. Karma is not the idea that 'what goes around comes around' or that 'good things happen to good people', but it is akin to a reminder that our actions, even when insignificant can sometimes create large ripples that affect the sea, and at times, come to affect us back in return in the most unexpected of ways. It is not a replacement for 'physical' causation, but rather, it is a way of 'humanising' an otherwise cold sea, just to remind ourselves that the *particulars* of human agency are not forgotten, even if we are small.

We can illustrate this in the following story. Imagine an Eastern child whose grandparents – both with Alzheimer's disease – are scammed out of their life savings by faith healers who promised to cure them of their ailments, and as a bonus, they are promised world peace. Once

the grandparents realise it was just a scam, they are, of course, devastated. But it is not all bad. 'Karmic fortune' works in mysterious ways. Due to this financial crisis, now the daughter and grandchild (i.e., the Eastern child and her mother) move in with the grandparents – and they happily spend the rest of their days surrounded by their loved ones. Although being scammed can be a terrifying event, it brings this family together in a happy realisation. Nowadays, it is less common for people to live with their grandparents, so prior to this scam, it had not occurred to the family to seriously consider this as an option. But being forced into a situation can also be a blessing in disguise. The Eastern child grows up with love, and a fondness for her elders.

The message in this story is not that karma 'punished' the grandparents for their 'gullibility', nor is the message that 'bad things happen even to good people'. It is all a part of 'fortune' that sometimes financial crises or natural disasters occur, and we cannot always control what misfortunes befall us. The world is composed of a complex web of events that push and pull in various directions, almost indifferent to its impact on any particular human being. And yet, in the story above, the grandparents receive a happy ending, not just out of 'luck' but *because* they were lovely people. If they had been nasty individuals, then the family would not have relished in this 'blessing in disguise' and the story would have been sad all the way until the end. The concept of 'karma' is but a reminder that if we cherish humanity, the world is not such an indifferent place, and we can actively search for the warmth in our stories.

Again, note that this story is not meant to be taken too literally. It is not an assertion that bad things never happen to good people, or vice versa. What is being emphasised here is the idea of warmth in the story: to remind ourselves that some 'positive' events in life are only positive because the characters of that story made it a more pleasant event than what it otherwise would have been. This also relates to the concept of 'wabi-sabi' or 'beauty in imperfection'. According to the Eastern picture, the world does not work in terms of deontological rules, but it is an ever-shifting environment as the future slowly unfolds in front

of us. We can analyse the world in a future-directed way of teleological tasks, but if we do wish to analyse the world in retrospect, what we must not forget is that as human beings, we always will be confined to 'storytelling'. Granted, some stories are better than others; some stories outright false; and some stories highly helpful. But at the end of the day, we cannot transcend the attentional limits of our own interests, and it is dangerous to mistake our own unmoving convictions as being 'facts'. Physics is just as much of a story as karma. One excludes humanity, the other does not, and they are each useful for their own purposes (Carroll & Wallace 2022, p.36). If physics seems 'most fundamental', this could just be due to physics having enjoyed a longer history than psychology, so we have a better understanding of it. Once the Magician's Research Programme takes off, it may become clearer that physics and psychology are just *tools*, neither reducing to the other.

With that in mind, let us return to Robert Nozick's thought experiment that we saw in the introduction to this thesis – of the 'seeker of wisdom' who travels to the Himalayas to ask a wise old sage what the meaning of life is. The sage responds that "Life is a fountain" to which the analytic seeker responds with dismay. He presses further, but the sage is unable to give more detail, and if anything, becomes confused that maybe life is not a fountain after all. This leaves the seeker feeling foolish that he had travelled such a great distance just for this. Nozick then goes on to ask, "What did he *expect* to hear"? (Nozick 1981, p.572), claiming that there is already an issue with the seeker's expectation to begin with. A subjective meaning in life is not something we discover by asking someone else for the 'right' answer. The seeker in this story has made two crucial mistakes. An Eastern 'sage' is not someone with all the answers; and also, 'the answer' is in the journey. Karmic fortune is the unfolding of such individual stories that we tell – as *particulars*, not *universals* – to retain a personal element in the story of causality.

Consider for a moment how a magician (or any performer) can affect their audience in profound, sometimes life-changing ways. The light waves from the stage travels to its audience

and hits their retinas; the soundwaves travel through the speakers and vibrates their eardrums; and the show ends within a couple hours. The audience goes home to continue thinking about the show, and several years pass as this information loops and loops in their neural network, maybe staying causally dormant for quite some time before these neural changes manifest in behaviourally observable outcomes. An audience member might adopt a new outlook on life after observing a magic show, and years later, might even decide to become a magician themselves. If we stop and think about it, this is a marvelling idea. The 'physical' causation involves only the transmitting of light energy and mechanical vibration of soundwaves, then 'stuff happens' inside the brain in the form of electrical and chemical activities, eventually creating a massive ripple that is far larger than the size of the original impact. The new magician might go on to affect their own audience members, thereby amplifying the social wave that only started with a tiny ripple, caused by a single magician. There is nothing supernatural about it, but there is something Magical. A human being – the magician – bonds with their audience and impacts future events that manifest somewhere far, far away. It is like the magic of long-distance phone calls or high-speed internet, but there is a *humanness* about it. Unbeknownst to the magician, events unfold elsewhere, in the distant future, across space and time.

Let us continue that earlier story with the Eastern child and her grandparents. The Eastern child – now an adult – one day discovers a Western television programme. A magician is presenting a documentary-style show on faith healing, and she finds herself in awe. Again, 'karmic fortune' works in mysterious ways. The television show is not a reminder of a negative life event, but quite the opposite. It is a warm reminder of a love and sympathy for her late grandparents, and unintentionally, the magician has 'conjured' warmth back into the world. Oddly enough, this is almost like real magic. The magician has created *out of thin air* a warm and positive feeling – the 'warm fuzzies' – across space and time. A show that may have been filmed fifteen years ago exercises causal powers in its future, sending ripples through reality.

The transference of warmth (or cold) in human beings is somewhat similar to the transference of warmth in thermodynamics, but remarkably, it works across longer distances, and across enormous lengths of time. While there is a sense in which technically we can describe 'warm fuzzy' emotions as nothing more than a chemical state, this explanation starts to feel rather contrived. The magician speaks some words, vibrations are created in the air and light waves reach the audience to create neurochemical responses in their brain, and then years and years later, an audience member of a magician might realise that their lives were heavily impacted by one innocuous TV programme or a stage show. It has shaped part of their personality, their career choice, and of course, their actions, causing them to send more ripples of their own back into the great sea to affect yet more human beings. It is for these reason that in the Japanese language, the word 'human' is written as a conjunction of two characters: 'space between' and 'people' – signifying that consciousness exists in the social space between people, and *that,* is humanity (Kasulis, Yuasa, & Yasuo 1987). A 'soul' does not reduce to an individual.

> "A first issue in Eastern metaphysics is how the "soul" is the "inner nature" buried in the corporeal body. The point of departure is to investigate this in light of the inseparability of the mind and body. [...] the foremost consideration in Buddhism must always be the practical issue of guiding the soul. (satori) ['satori' is the Japanese word for 'wisdom' or 'enlightenment']" (*ibid*, p.79)

Buddhist karma is not a supernatural concept, but it is a metaphor to value the social influences of human-related causation. An exploration of psychological wellbeing is not about deciding who is judgeable and who is not, but ideally, it would be a teleological exercise to spread warmth to a wider sphere than the isolating world we live in today. The idea of a Buddhist soul is a reminder that the essence of a person's conscious experience is their 'inner nature' – the feeling, experiencing, and acting *thing* – which in the West we refer to as 'an agent'.

Consider also, that when a loved one passes away, we might continue to carry them with us as we grieve. This rapport can transcend life and death, keeping their memory or 'soul' alive in the minds of loved ones. Consciousness does not reduce to a brain as though it is confined to a single person. It is a wind in the vast sea, all interconnected, and irreducible to an individual.

> "If I were to lose a daughter, and every day I found myself reminded of her, and if I were to think about what she would say or do about this or that, then *my* pattern for a while becomes *hers*. And thus her consciousness – her *pattern* – her *self* – is recreated in me for a while. It's sadly a course, grainy, much-reduced version of her, but it is, in some sense, her. It's like the 'rippling' […] It turns out, then, to be the positive connection between people that provide the mechanism for our 'self' to survive death in any meaningful way. It turns out to be love." (Brown 2016, p.517-518).

It is in this light that we might consider – the fundamental essence of consciousness just is 'love'. It is the warm fuzzies. It is what makes us human, and 'physical' terminology misses the point of these human nuances entirely. It falls outside the scope of science, not because it is supernatural, but because it is not the business of physicists to explain the concept of love.

8.5 The Magical 'Fuzzies' of Consciousness

Let us continue that story of the Eastern youth and the magician. Through time and space, a magician's words affect an individual – a *particular* concrete individual – who then begins to dream of becoming a magician herself. She wants to become the magician's apprentice. She decides to try and get in touch, but upon going on his website, she finds that it explicitly states NOT to send any emails about wanting to be his apprentice, as he gets far too many of these already. If someone wants to be his apprentice, they are to ask him in person, though the answer

will likely be 'no' since he gets too many requests already. If she wants an apprenticeship, something truly impressive will need to be done. A Magical gesture, perhaps. The Eastern not-yet Apprentice writes a manuscript for a magic show as sample writing to impress the magician, and with her partner, journeys to distant lands, far, far away: say, Scotland (which can serve as our Western equivalent of the Himalayan caves) where the magician is on tour. Ripples have traversed the sea of karmic fortune and comes back to affect the magician. The Eastern not-yet Apprentice delivers the manuscript, though out of nerves, she is unable to explain why the documents are significant. The magician does not read them, complaining that the documents are far too long. Like Nozick, the magician responds in confusion, "What was she *expecting*?"

Unlike in the Nozick example, however, the Eastern not-yet Apprentice is not expecting to find a 'sage' with all the answers. Rather, her 'expectation' or 'fear' in these circumstances is – that the magician might be terminally ill. He has dropped hints that this current stage show tour might be his last and looks increasingly ill as time goes on. The stage show itself that he is performing is about grief and coming terms to death. His final message to the audience was: "Remember what is important. Don't live the life others said you should live. Don't let life pass you by just because you were afraid to look like a fool." The magician tells his audience that these are the words of people on their deathbeds, asked to give advice to the youth – the unwell magician looking as though he himself might be headed somewhere close.

The theme of 'death' and how to treat it is an understudied field in the West that is often a taboo topic to discuss, due to its morbid nature. How do we come to terms with the 'wabi-sabi' or the '*impermanence'* of human conscious life? The topic of death is difficult to handle even when it comes to family or friends who we are close to already. It is even more confusing to fear for the wellbeing of a complete stranger – someone one wishes to spend the next several years learning from and building rapport with – but is a currently stranger, nonetheless.

Under a Buddhist mentality, the world is sometimes described as a 'global village' (E.D. Dhammasami et al., 2010, p.460-487). Particularly under a Japanese Buddhist mentality, the beauty in the impermanence of consciousness is found in returning the favour of karmic fortune, while we still have the chance. The idea of reincarnation is just a story or a myth, and while it is a concept that Buddhists may lean into once after someone has passed, it is not a replacement for valuing the limited time that we do have while living on this earth – building rapport in the global village while there is still opportunity. While a magician might be a complete stranger to the not-yet Apprentice, the magician too is a human being, and his 'celebrity status' (if he has one) is not particularly relevant. If a blacksmith looks miserable, a to-be apprentice will try to cheer him up. A magician, who is also a craftsman, may at times deserve the exact same response.

Whether or not this storytelling is palatable for Western ears, the idea to value the 'souls' or experiences of strangers, and to contribute back to the karmic fortune of ripples in the sea is an integral part of Eastern philosophy. If we think someone is dying, then we may only have one chance to tell them how much good they have done for us. It is unfortunate if this seems 'weird' to Westerners, but if it does seem 'weird' at all, then this is reflective of just how squeamish we are about discussing behaviours that seem 'judgeable' from a deontological perspective. But this is a separate issue from the 'teleological' purpose in a research programme.

It is difficult to *rationally* work out what we should or should not do, when the circumstances themselves are unique or unusual. It might be tempting to try and shift 'personal' matters to be discussed in places outside of academic endeavours. However, this is the very assumption that this thesis has tried to question throughout, and it is fitting that we can finish on the same note. The Magician's Research Programme proposes something in between a 'personal' and 'professional' approach to academic discussion. It is about the 'risky'

predictions that we test out in our everyday lives – as Apprentices to life in the way described in chapter four – and expanding our knowledgebase according to 'contact points' with reality.

Ultimately, agential decision-making might be an individual's own personal choice, but it is helpful to allow for discussion on psychological normativity in a formal capacity, that does not just leave these tasks up to the domain of self-help gurus, social media influencers, capitalism-driven therapies, political psychiatry, or statistical averages in psychology. What has been missing in consciousness studies is the Magic: The *story* of human beings connecting.

To finish the story, the magician turns out not to have been dying. Later he states that he was overworked and accidentally over-dosing on cough syrup, which may have contributed to the 'dying' demeaner. Perhaps the Eastern not-yet Apprentice had only seen what she had most feared, after her experience with her grandparents. The ripples of human connection interlink in strange and unexpected ways. But to say the least, such nuances, subtlety, and the beauty of the human condition can become lost when we force it into a certain kind of written format, and this is a shortcoming of analytic philosophy (Benatar 2010). In particular, if the premises of a deductive argument must be devoid of 'personal' elements in order to create an impression of professionalism, then deductive conclusions necessarily will be devoid of 'personal' elements as well, leading to theories like physicalism that explain away the subjective Magic.

But is it wrong to sometimes step outside of our Westernised stories to explore a different culture's way of navigating the world? Can we not see the entire human condition as a humorous and smile-provoking charm – rather than to look at it with a desperation for eradicating all that is ambiguous, sad, or uncomfortable? In the spirit of decolonisation and the Magician's Research Programme, this thesis attempted a demonstration of this Eastern Magic as a form of empirical exercise – though the attempt was unsuccessful for bureaucratic reasons.

There is a beauty in the idea of 'achieving the impossible' by questioning the stories that we are used to telling ourselves. Imagine: A magician creates ripples years and years ago through making television programmes, debunking the problems of faith healers, and thereby conjures a sense of 'warmth' into faraway Eastern lands. The light waves and sound waves of the performance affects the Eastern not-yet Apprentice, and triggers a sequence of neural activities that make her decide to step outside her own story of being a socially anxious introvert, and trying out new things that she would never have dreamed of achieving before. A thesis is written and completed despite her cripplingly anxiety temperament – a thesis which may never have emerged if the ripples had never begun. If she later goes on to write anything that is published or affects other people, then the ripple will further propagate to wider spheres over time. Ripples that would never have emerged had she not had the push to become a more courageous version of herself. And at the end of the four-year thesis, how poetic would it be if the karmic fortune came rippling back to the magician himself? The magician, who claims to be extremely solitary and introverted who seemingly has no interested in taking a random Eastern apprentice who weirdly approached him with a long document years ago – finds himself travelling all the way to Durham for the Eastern not-yet Apprentice's PhD viva. Accordingly with Eastern tradition, the master magician is there to witness the rite of passage of the apprentice (De Craemer 1983), completing the loop of karmic fortune as a particular kind of 'mental' or intentional causation that ripples back to the original agent over time.

It is not the kind of 'mental' causation that exists distinctly from 'physical' causation. Rather, it is a type of intentional causation that highlights the mysteries of human agency and how it works in unexpected ways. There is a human tendency to try and fit things into neat narratives – though it often becomes a blind spot the kind of narratives and convictions that we ourselves are trapped inside, pertaining to our own tasks, professions, and self-image. We create arbitrary limitations for ourselves all the time, deciding on 'who we are' as though we

are a character in a TV show with distinct, unnuanced, and unchanging personalities. But unlike fictional characters, real human beings have agency. We are able to change (Brown 2016, p.14).

The Magician's Research Programme encourages this activity of self-questioning over and beyond providing positive claims for justifying our own belief systems. "We are all trapped, inside our own heads. And our beliefs and understandings about the world are limited by that perspective. Schopenhauer wrote, 'Every man takes the limits of his own field of vision for the limits of the world'. Of course, then, we mistake that story we've constructed of our lives as the truth." (*ibid*, p.10). If our philosophical goal is about finding accurate 'stories' to tell, then this starts with questioning one's own beliefs about how academic arguments must work. Unfortunately, there is a bureaucratic rule in the UK that 'guests' are not allowed to attend a PhD viva, and so the idea of an 'empirical demonstration of karmic fortune' never got very far, and the request was just met with generic confusion about why an academic examination should involve any kind of 'personal' anecdotes or guest invitations. But Karl Popper writes:

> "Genuine philosophical problems are always rooted in urgent problems outside philosophy, and they die if these roots decay. In their efforts to solve them philosophers are liable to pursue what looks like a philosophical method or technique or an unfailing key to philosophical success. But no such methods or techniques exist; in philosophy methods are unimportant; any method is legitimate if it leads to results capable of being rationally discussed. What matters is not methods or techniques but a sensitivity to problems, and a consuming passion for them; or, as the Greeks said, the gift of wonder." (Popper 1963, p.72)

The line between 'professional' and 'personal' need not be so black-and-white when it comes to the Magician's Research Programme. Magicalism is the view that the fundamental essence of consciousness just is Magical – like a stage show, but not in the supernatural sense. It is full

of ambiguity about method, wonder in the passions, and the best moments are when something which seems impossible is achieved and made a part of reality – as if by Real Magic.

'Karmic Fortune' is not a supernatural concept, but it is the metaphorical intertangling of 'souls' (i.e. inner essences) of different individual agents, each acting as *particulars* rather than *universals*, in the backdrop of the cosmic indifference of the vast cold sea. Imagine if the magician were to find himself travelling to visit the to-be Eastern Apprentice, *to be there in-person* after her 'rite of passage', an event that no one would ever dream would be actualised or made possible as a part of reality. It seems like an impossible thing, especially under a Western 'story' that a PhD is meant to be a 'professional' endeavour. And yet, in the spirit of decolonisation, we might be able to have the best of both worlds: The storytelling and 'soul' of Eastern philosophy, and the analytic rigour offered by Western philosophy. There is no one single 'correct' method for doing philosophy – "any method is legitimate if it leads to results capable of being rationally discussed" (Popper 1963, p.72). There is something grippingly Magical – *impossible* – about the idea that there is so much about being human that we do not understand. It seems a shame not to 'test out' the extent of karmic fortune (i.e., the ripples we can or cannot create) when it is poetic to attempt it, which can lead to further rational analyses of human nature and how to go about investigating it in our capacity as professional academics.

If there is just one take home message that this thesis tries to advocate, then it is the idea that we should allow for 'personal' and 'professional' to start intermingling just a little bit more, and to test the boundaries of a novel Magician's Research Programme in service of a research programme that is designed to investigate the quality of life in *particulars*, not *universals*. Without opening our scope of research to these matters, the issues about consciousness become unsolvable through the methods of strictly impersonal deductive arguments. The meta-problem is not "Why are we stuck on how to understand consciousness?" but something closer to home:

Why are we stuck on creating a black-and-white distinction between 'personal' and 'professional', even though we are studying a thing which by its very nature is extremely private and personal – *consciousness*. "What is consciousness?" is not the kind of question that can be understood by appealing to objective nouns. It is a collection of personal 'doings' or verbs, of qualitative experiences, states of 'beings'. It is something warm and fuzzy, and it is about the 'seemingly impossible'. For the reasons explored throughout this thesis, we conclude here that the essence of consciousness just is *Magic*. Our everyday mental lives are like a stage show, and if we want to capture this in professional academic writing, the writing itself needs to include that sense of Magic. Otherwise, consciousness becomes forever an elusive concept.

> "On this view, momentous questions of ultimate significance cannot be addressed satisfactorily in the coolly dispassionate manner of analytic philosophy. Instead, such questions must be engaged passionately, stirring the heart rather than sharpening the mind." (Benatar 2010, p.2)

In conclusion, consciousness just is something Magical. The Japanese word 'kokoro' refers to both mind and heart, and understanding consciousness is about the gap between the attentional focus of our 'minds' – and the blind spot of poetic intuitions as experienced in our 'hearts'.

Conclusion/Summary

This thesis began by engaging in the typical mind-body debate, exploring what the fundamental nature of consciousness might be. Early chapters discussed the limitations with the word 'physical', that it is not only elusive and too ambiguous, but that it also seems to be motivated by a misleading image of physics. This debate evolved into a discussion about how 'THE' problem of consciousness might not be a single, unified problem after all. With these discussions in place, we moved on to discuss the idea of a novel doctrine called Magicalism: The neutral monist view that the fundamental essence of consciousness is Magic, but in the sense of a card trick, not a supernatural sense.

Beyond the basic idea that Magicalism is a conjunction of something (a) Eastern, something (b) Magician-related, and something from an (c) Apprentice's perspective, the further details discussed in this thesis became more nuanced. The issue of consciousness is not cleanly or purely an ontological question, as it interrelates with other problems such as epistemology (what sources of knowledge are credible), research methodology (how to investigate epistemic concerns), pragmatism (how to treat people who struggle with psychological problems), death and grief (what happens to our 'souls' or 'essence' when our loved ones die), and love (what it means to be human). Magicalism is but one doctrine within the overarching novel research programme, which we have called the Magician's Research Programme. While this thesis primarily focuses on the ontological question, it has been inevitable that we include discussions from some of the other topics which the research programme as a whole deals with, as again, the topics are all interconnected. Most importantly is the claim that if we leave out the more personal/human elements from our philosophy of consciousness, we risk losing out the important nuances in the storytelling, thereby explaining away the very thing that we are trying to get a handle over. The balance of how much 'personal' elements to include in a professional academic discussion is a matter of ambiguity, and it is

quite possible that this thesis has overstepped this boundary, particularly in the final few pages of the last chapter. However, what has been most important for our current purposes is to question the way in which we approach 'human' fields of philosophical enquiry, and pushing the boundaries or taking risks is all a part of the method advocated within the Magician's Research Programme. If we do not push boundaries, take risks, and try new things, concepts such as grief and love may never become truly accessible, and it is worth bearing in mind that the very beginning stages of a novel research programme are expected to be rough and bumpy. Risking mistakes, and not making things 'too sanitary' is all, intentionally, part of the process.

In this thesis, we have tried to undermine the physicalist attitude by highlighting that physicalism not only struggles with the issue of definability, but also that the pro-science attitude observed in physicalism itself clashes with the more empiricist and teleological attitude of the sciences themselves. Human life is messy and nuanced, but this is what makes philosophy magical and an endeavour that is worth doing. If we can explain away all the magic, then we might as well stop engaging with it altogether. Hopefully, this thesis will have contributed to inspiring future studies to retain a sense of Magic, while simultaneously having an empiricist attitude, and pragmatically moving academia forwards as our understanding of consciousness continues to develop. As Popper (1963, p.72) has said, "Genuine philosophical problems are always rooted in urgent problems outside philosophy, and they die if these roots decay." The next stage will be to reground the mind-body problem back into real world issues, such as psychology, psychiatry, or stage magic, and further develop these ideas accordingly.

Bibliography

Abdelnour, E., Jansen, M. O., & Gold, J. A. (2022). ADHD diagnostic trends: increased recognition or overdiagnosis?. *Missouri medicine*, *119*(5), 467.

Aldrich, V. C. (1970). The pineal gland up-dated. *The Journal of Philosophy*, *67*(19), 700-710.

Allchin, D., & Werth, A. J. (2017). The naturalizing error. *Journal for General Philosophy of Science*, *48*, 3-18.

Alter, T., & Pereboom, D. (2023). Russellian monism and structuralism about physics. *Erkenntnis*, *88*(4), 1409-1428.

Alter, T. (2012). What is Russellian monism?. *Journal of Consciousness Studies*, *19*(9-10), 67-95.

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). <https://doi.org/10.1176/appi.books.9780890425787>

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). <https://doi.org/10.1176/appi.books.9780890425596>

Anderson-Chavarria, M. (2022). The autism predicament: models of autism and their impact on autistic identity. *Disability & Society*, *37*(8), 1321-1341.

Angell, J. R. (1907). The province of functional psychology. Psychological review, 14(2), 61.

Ashby, W. (1950). Teleology and deontology in ethics. *The Journal of Philosophy*, *47*(26), 765-773.

Asthana, H. S. (2015). Wilhelm wundt. *Psychological Studies*, *60*(2), 244-248.

Balog, K. (2012). In defense of the phenomenal concept strategy. *Philosophy and Phenomenological Research*, *84*(1), 1-23.

Benatar, D. (Ed.). (2016). *Life, death, and meaning: Key philosophical readings on the big questions*: Rowman & Littlefield.

Bok, B. J. (1975). A critical look at astrology. *Humanist*, *35*(5), 6.

Bok, B. J., & Jerome, L. E. (1975). *Objections to astrology*. Buffalo, NY: Prometheus books.

Boring, E. G. (1953). A history of introspection. *Psychological bulletin*, *50*(3), 169.

Botin, M. (2022). The phenomenal concept strategy cannot explain problem intuitions. *Journal of Consciousness Studies*, *29*(7-8), 7-31.

Boyle, M. (2013.) The persistence of medicalisation: Is the presentation of alternatives part of the problem? In S. Coles, S. Keenan & B. Diamond (Eds.) *Madness contested: Power and practice* (pp.3–22). Ross-on-Wye: PCCS books.

Broad, C. D. (1930). The principles of demonstrative induction (II.). *Mind*, *39*(156), 426-439.

Brown, D. (2023). *Notes from a Fellow Traveller*: The Neat Review.

Brown, D. (2016). *Happy: Why More or Less Everything is Absolutely Fine*: Penguin.

Brown, D. (2012). *Svengali* [Video]. Channel 4. <www.youtube.com/watch?v=huoRpn74d2c>

Brown, D. (2011). *The Assassin with Stephen Fry | The Experiments* [Video]. Channel 4. <https://www.youtube.com/watch?v=owootTAuxic>

Brown, D. (2007). *Tricks Of The Mind*: Channel 4.

Brown, D. (2000). *Pure effect: direct mind reading and magical artistry*. H & R Magic Books.

Burnett, T. (2012). What is scientism?. *AAAS DoSER*.

Byrnes, R. (2007). *The Secret*: Simon & Schuster UK.

Carroll, S., & Wallace, B. A. (2022). 2 THE NATURE OF REALITY: A DIALOGUE BETWEEN A BUDDHIST SCHOLAR AND A THEORETICAL PHYSICIST. In *Great Minds Don't Think Alike: Debates on Consciousness, Reality, Intelligence, Faith, Time, AI, Immortality, and the Human* (pp. 26-61). Columbia University Press.

Carroll, S. (2016). The Big Picture: On the Origins of Life. *Meaning, and the Universe Itself (Dutton, 2016)*.

Carroll, S. M. *Reality realism*. https://philpapers.org/archive/CARRRK.pdf

Carter, K. C. (1982). Nineteenth-century treatments for rabies as reported in the Lancet. *Medical History*, *26*(1), 67-78.

Chalmers, A. F. (2013). *What is this thing called science?*. Hackett Publishing.

Chalmers, D. J. (2018). The meta-problem of consciousness. *Journal of Consciousness Studies, 25,* No. 9–10, 6–61

Chalmers, D. (2017). The hard problem of consciousness. *The Blackwell companion to consciousness,* 32-42.

Chalmers, D. J. (2006). Nine phenomenal concepts and the explanatory gap. *Phenomenal concepts and phenomenal knowledge: New essays on consciousness and physicalism*, 167.

Chalmers, D. J. (2003). Consciousness and its place in nature. *The Blackwell guide to philosophy of mind*, 102-142.

Chalmers, D. (2002). Does conceivability entail possibility?. *Conceivability and possibility*, *145*, 200.

Chalmers, D. J. (1996). Facing up to the problem of consciousness. *Toward a science of consciousness: the first Tucson discussions and debates. MIT Press, Cambridge*, 5-28.

Chatterjee, A. (2022). Spaces of Cure or Confinement? Inside the Walls of the Mental Asylums of the 19th Century. In *Covid-19 in India, Disease, Health and Culture* (pp. 50-61). Routledge.

Clarke, A. C. (1962). *Profiles of the Future : An Inquiry into the Limits of the Possible*: Harper & Row, p.

Clements, P. (2020). Astrology, modernity and the project of self-identity. *Culture and Religion*, *21*(3), 259-279.

Cottingham, J. (1985). Cartesian trialism. *Mind*, *94*(374), 218-230.

Courtney, W. L. (1885). Socrates, Buddha, and Christ. *The North American Review*, *140*(338), 63-77.

Crane, T. (2000). Dualism, monism, physicalism. *Mind & Society*, *1*, 73-85.

Crane, T. (2012). The origins of qualia. In *History of the mind-body problem* (pp. 177-202). Routledge.

Crane, T., & Mellor, D.H. (1990). There is no question of physicalism. *Mind, 99*, 185-206.

Crease, R. P. (2016). Physics: Material to meaning. *NATURE*, Vol 533, p.34.

Crook, S., & Gillett, C. (2001). Why physics alone cannot define the 'physical': Materialism, metaphysics, and the formulation of physicalism. *Canadian Journal of Philosophy*, *31*(3), 333-359.

Czech, H. (2018). Hans Asperger, national socialism, and "race hygiene" in Nazi-era Vienna. *Molecular autism*, *9*, 1-43.

Davies, R. T. (Executive Producer & Writer). (2008). *Doctor Who S4E10*. BBC.

Dean, G., & Kelly, I. W. (2003). Is astrology relevant to consciousness and psi?. *Journal of Consciousness Studies*, *10*(6-7), 175-198.

De Craemer, W. (1983). A cross-cultural perspective on personhood. *The Milbank Memorial Fund Quarterly. Health and Society*, 19-34.

Dennett, D. C. (2019). Welcome to strong illusionism. *Journal of Consciousness Studies*, *26*(9-10), 48-58.

Dennett, D. C. (2016). Illusionism as the obvious default theory of consciousness. *Journal of Consciousness Studies*, *23*(11-12), 65-72.

Dennett, D. C. (2007). Heterophenomenology reconsidered. *Phenomenology and the Cognitive Sciences*, *6*, 247-270.

Dennett, D. (2003). Explaining the" magic" of consciousness. *Journal of Cultural and Evolutionary Psychology*, *1*(1), 7-19.

Dennett, D. (2001). Are we explaining consciousness yet?. *Cognition*, *79*(1-2), 221-237.

Dennett, D. (1991). *Consciousness explained*. Penguin uk.

Devereux, D. T. (1986). Particular and universal in Aristotle's conception of practical knowledge. *The review of Metaphysics*, 483-504.

Dhammasami, K. (Ed.). (2010). *Global Recovery: The Buddhist Perspective* (pp. 460-487): Mahachulalongkornrajavidyalaya University.

Dhiman, S. (2021). More than happiness: A stoic guide to human flourishing. In *The Palgrave Handbook of Workplace Well-Being* (pp. 1433-1472). Cham: Springer International Publishing.

Diaz-Leon, E. (2008). Defending the phenomenal concept strategy. *Australasian Journal of Philosophy*, *86*(4), 597-610.

Dillon, M. (2000). Dialogues with death: The last days of Socrates and the Buddha. *Philosophy East and West*, 525-558.

Dove, G. (2018). Redefining physicalism. *Topoi, 37*, 513-522.

Dove, G., & Elpidorou, A. (2021). A Dilemma about the Mental. *Organon F*, *28*(4), 867-895.

Dowell, J. L. (2006). Formulating the thesis of physicalism: an introduction. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *131*(1), 1-23.

Dyson, F. J. (2004) *Thought-experiments in honor of John Archibald Wheeler,* in Science and Ultimate Reality, 72–89, 88.

Earman J (1993) Underdetermination, realism, and reason. Midwest Stud Philos 18:19–38

Ertel, S., & Dean, G. (1996). Are personality differences between twins predicted by astrology?. *Personality and Individual Differences*, *21*(3), 449-454.

Esfeld, M. (2022). Against levels of reality: the method of metaphysics and the argument for dualism. In *Levels of Reality in Science and Philosophy: Re-examining the Multi-level Structure of Reality* (pp. 47-62). Cham: Springer International Publishing.

Farley-Icard, R. L. (2007). *Factors that influence the Barnum Effect: Social desirability, base rates and personalization*. The University of Texas at El Paso.

Fedorov, A. (2018). The first foreign edition of Hoffmann's modern magic. *The Magic Circular 112:*47–48.

Ferriman, A. (2000). The stigma of schizophrenia. *BMJ*, *320*(7233), 522.

Field, H. (2015). What is logical validity. *Foundations of logical consequence*, 33-70.

Firt, E., Hemmo, M., & Shenker, O. (2021). Hempel's Dilemma: Not only for physicalism. *International Studies in the Philosophy of Science*, *34*(2), 101-129.

Fischer, E., & Sytsma, J. (2021). Zombie intuitions. *Cognition*, *215*, 104807.

Fletcher, G. (2002). Holism, Hermeneutics, and Heterophenomenology: A Critique of Daniel Dennett's Intentional Stance.

Fombonne, E. (2023). Is autism overdiagnosed?. *Journal of Child Psychology and Psychiatry*, *64*(5), 711-714.

Forer, B. R. (1949). The fallacy of personal validation: A classroom demonstration of gullibility. *The Journal of Abnormal and Social Psychology*, *44*(1), 118.

Frankish, K. (2019). The meta-problem is the problem of consciousness. *Journal of Consciousness Studies*, *26*(9-10), 83-94.

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

Fraser, A., Bungay, H., & Munn-Giddings, C. (2014). The value of the use of participatory arts activities in residential care settings to enhance the well-being and quality of life of older people: A rapid review of the literature. *Arts & Health*, *6*(3), 266-278.

French, C. C., Leadbetter, A., & Dean, G. (1997). The astrology of time twins: a re-analysis. *Journal of Scientific Exploration*, *11*(2), 147-55.

Frith, C. D., & Rees, G. (2017). A brief history of the scientific approach to the study of consciousness. *The Blackwell companion to consciousness*, 1-16.

Fujiyama, K. (2013). *Tezuma -- traditional Japanese magic: Kohtaro Fujiyama at TEDxTokyoTeachers* [Video]. TED Conferences. <https://www.youtube.com/watch?v=Ns-O_GX8QTk>

Fujiyama, S. (2010). *Fundamentally, what it a professional magician is*: Tokyodo Publisher.

Fuzeau-Braesch, S. (1992). An empirical study of an astrological hypothesis in a twin population. *Personality and individual differences*, *13*(10), 1135-1144.

Gambrill, E. (2014). The Diagnostic and Statistical Manual of Mental Disorders as a major form of dehumanization in the modern world. *Research on Social Work Practice*, *24*(1), 13-36.

Garber, D. (2002). Descartes, Mechanics, and the Mechanical Philosophy. *Midwest Studies In Philosophy (Wiley-Blackwell)*, *26*(1).

Giles, D. C. (2014). 'DSM-V is taking away our identity': The reaction of the online community to the proposed changes in the diagnosis of Asperger's disorder. *Health:*, *18*(2), 179-195.

Gillett, C., & Witmer, D. G. (2001). A 'physical' need: physicalism and the via negativa. *Analysis*, *61*(4), 302-309.

Glas, G. (2012). Medicine is a science and a normative practice. *Philosophy, Psychiatry, & Psychology*, *19*(4), 285-289.

Goff, P. (2017). *Consciousness and fundamental reality*. Oxford University Press.

Haack, S. (2012). Six signs of scientism. *Logos & Episteme*, *3*(1), 75-95.

Hall, H. (2013, November 19). Faith Healing: Religious Freedom vs. Child Protection. *Science-Based Medicine*. <https://sciencebasedmedicine.org/faith-healing-religious-freedom-vs-child-protection/>

Harman, G. (2019). On progressive and degenerating research programs with respect to philosophy. *Revista Portuguesa de Filosofia*

Heisenberg, W. (1989) *Physics and Philosophy*, 21–22.

Heisig, J. W. (2001). *Philosophers of nothingness: An essay on the Kyoto School*. University of Hawaii Press.

Heisig, J. W., Kasulis, T. P., & Maraldo, J. C. (Eds.). (2011). *Japanese philosophy: a sourcebook*. University of Hawaii Press.

Hellman, G. (1985). Determination and logical truth. *The Journal of Philosophy*, *82*(11), 607-616.

Hempel, C. (1980). Comments on Goodman's ways of worldmaking. *Synthese*, *45*, 193-200.

Hempel, C. (1966). *Philosophy of natural science*. Englewood Cliffs, New Jersey: Prentice Hall

Hendry, R. F. (2019). Emergence in chemistry: Substance and structure. In *The Routledge handbook of emergence* (pp. 339-351). Routledge.

Highfield, R. (2003). *The science of Harry Potter: How magic really works*. Penguin.

Hilgard, E. R. (1980). The trilogy of mind: Cognition, affection, and conation. *Journal of the History of the Behavioral Sciences*, *16*(2), 107-117.

Horwitz, A. V. (2021). *DSM: A history of psychiatry's bible*. JHU Press.

Hughes, J. A. (2021). Does the heterogeneity of autism undermine the neurodiversity paradigm?. *Bioethics*, *35*(1), 47-60.

Husserl, E., & Iyer, A. (2017). Socrates-Buddha. In *The New Yearbook for Phenomenology and Phenomenological Philosophy* (pp. 398-415). Routledge.

Hyman, R. (1981). Cold reading: How to convince strangers that you know all about them. *Paranormal borderlands of science*, 79-96.

Ishihara, Y. (2019). Heidegger and Nishida's transformations of transcendental reflection. *Tetsugaku companion to phenomenology and Japanese philosophy*, 77-94.

Ishihara, Y. (2016). Transcendental Philosophy and its Transformations: Heidegger and Nishida's critical engagements with.

Jain, E., & Mukerjee, A. (2024) Barnum Effect: Influence of Social Desirability, Base Rate and Personalization.

Jacquin, F. (2016a). *Hypnotherapy: Methods, Techniques and Philosophies of Freddy Jacquin*: Freddy H Jacquin.

Jacquin, A. (2016b). *Reality is Plastic: The Art of Impromptu Hypnosis*: Anthony Jacquin.

Johnstone, L., & Boyle, M. (2018). The power threat meaning framework: An alternative nondiagnostic conceptual system. *Journal of Humanistic Psychology*, 0022167818793289.

Judisch, N. (2008). Why 'non-mental'won't work: on Hempel's dilemma and the characterization of the 'physical'. *Philosophical Studies*, *140*, 299-318.

Kanner, L. (1944). Early Infantile Autism. *The Journal of Pediatrics, 25*, 211–217.

Kasulis, T. P., Yasuo, Y., & Yuasa, Y. (1987). *The body: Toward an Eastern mind-body theory*. Suny Press.

Kelly, K., & Ramundo, P. (2006). *You mean I'm not lazy, stupid or crazy?!: the classic self-help book for adults with attention deficit disorder*. Simon and Schuster.

Kendell, R. E. (2001). The distinction between mental and physical illness. *The British Journal of Psychiatry*, *178*(6), 490-493.

Kim, J. (2018). Against cartesian dualism. *The Blackwell companion to substance dualism*, 152-167.

Kind, A. (2018). The mind-body problem in 20th-century philosophy. *Philosophy of Mind in the Twentieth and Twenty-First Centuries: The History of the Philosophy of Mind, Volume 6*, 1.

Kirk, Robert, "Zombies", *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/fall2023/entries/zombies/>.

Kirk, R. (2008). The inconceivability of zombies. *Philosophical Studies*, *139*, 73-89.

Kirk, R. (2005). *Zombies and consciousness*. Clarendon Press.

Kirk, R. (1999). Why there couldn't be zombies. In *Aristotelian Society Supplementary Volume* (Vol. 73, No. 1, pp. 1-16). Oxford, UK: Oxford University Press.

Klin, A. (2006). Autism and Asperger syndrome: an overview. *Brazilian Journal of Psychiatry*, *28*, s3-s11.

Koksvik, O. (2007). Conservation of energy is relevant to physicalism. *dialectica*, *61*(4), 573-582.

Komath, M. (2009). Testing astrology. *Current science*, *96*(12), 1568-1572.

Koren, L. (2008). *Wabi-sabi for artists, designers, poets & philosophers*. Imperfect publishing.

Koutsouleris, N., & Fusar-Poli, P. (2024). From heterogeneity to precision: redefining diagnosis, prognosis, and treatment of mental disorders. *Biological Psychiatry*, *96*(7), 508-510.

Kripke, S. A. (1980). Naming and necessity: Lectures given to the princeton university philosophy colloquium. In *Semantics of natural language* (pp. 253-355). Dordrecht: Springer Netherlands.

Krummel, J. W. (2018). On (the) nothing: Heidegger and Nishida. *Continental Philosophy Review*, *51*(2), 239-268.

Kugler, B. (1998). The differentiation between autism and Asperger syndrome. *Autism*, *2*(1), 11-32.

Kuhn, T. (1970). The nature of scientific revolutions. *Chicago: University of Chicago*, *197*(0).

Kuhn, T. (1962). *The nature and necessity of scientific revolutions*. na.

Kymlicka, W. (1988). Rawls on teleology and deontology. *Philosophy & public affairs*, 173-190.

Ladd, P. D., & Churchill, A. (2012). *Person-centered diagnosis and treatment in mental health: A model for empowering clients*. Jessica Kingsley Publishers.

Lakatos, I. (1978). Science and pseudoscience. *Philosophical papers*, *1*, 1-7.

Lamont, P., & Wiseman, R. (2005). *Magic in theory: An introduction to the theoretical and psychological elements of conjuring*. Univ of Hertfordshire Press.

Lancaster, T. (2019). Skyrmions in magnetic materials. *Contemporary Physics*, *60*(3), 246-261.

Lancaster, T., & Blundell, S. J. (2014). *Quantum field theory for the gifted amateur*. OUP Oxford.

Lewis, D. K. (1986). *On the plurality of worlds* (Vol. 322). Oxford: Blackwell.

Lightman, B. (1983). Pope Huxley and the Church Agnostic: The Religion of Science. *Historical Papers*, *18*(1), 150-163.

Linde, A. (2005). *Choose your own universe*, in ed. Charles L. Harper, Jr. Spiritual Information:100 Perspectives on Science and Religion (West Conshohocken, PA: Templeton Foundation, 2005), 139.

Linde, A. (1990) *Inflation, quantum cosmology and the anthropic principle*, in Science and Ultimate Reality, 426–458, 451

Lipson, S. K., Lattie, E. G., & Eisenberg, D. (2019). Increased rates of mental health service utilization by US college students: 10-year population-level trends (2007–2017). *Psychiatric services*, *70*(1), 60-63.

Lotfi, M. (2023). Phenomenal Concepts Strategy and Chalmers' Argument. *Philosophical Thought*, *3*(1), 1-13.

Lowe, E. J. (2010). Why my body is not me: the unity argument for emergentist self-body dualism. In *Emergence in science and philosophy* (pp. 141-162). Routledge.

Lowe, E. J. (2008a). A defence of non-Cartesian substance dualism. *Psycho-Physical Dualism Today: An Interdisciplinary Approach*, 167-183.

Lowe, E. J. (2008b). *Personal agency: The metaphysics of mind and action*. Oxford University Press, USA.

Mabaquiao Jr, N. (2015). The phenomenal concept strategy and a master argument. *KEMANUSIAAN* Vol. 22, No. 153–74

Macdonald, J. E., & Beck-Dudley, C. L. (1994). Are deontology and teleology mutually exclusive?. *Journal of business ethics*, *13*, 615-623.

Macpherson, F. (2006). Property dualism and the merits of solutions to the mind-body problem: a reply to Strawson. *Journal of Consciousness Studies*, *13*(10-11), 72-89.

Marcus, E. (2004). Why zombies are inconceivable. *Australasian Journal of Philosophy*, *82*(3), 477-490.

Markosian, N. (2000). What are physical objects?. *Philosophical and Phenomenological Research*, 375-395.

Maraldo, J. C., "Nishida Kitarō", *The Stanford Encyclopedia of Philosophy* (Spring 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/spr2024/entries/nishida-kitaro/>

Márton, M. (2019). What does the zombie argument prove?. *Acta Analytica*, *34*(3), 271-280.

Masi, M. (2022). Vitalism and cognition in a conscious universe. *Communicative & Integrative Biology*, *15*(1), 121-136.

McCandless, P. (1983). Dangerous to Themselves and Others: The Victorian Debate over the Prevention of Wrongful Confinement. *Journal of British Studies*, *23*(1), 84-104.

McCarthy, E. (2020). Watsuji Tetsurō. In *The Oxford Handbook of Japanese Philosophy*.

McDaid, S. (2010). Redefining empowerment in mental health: an analysis using Hannah Arendt's power concept. *Journal of Power*, *3*(2), 209-225.

Melnyk, A. (1997). How to keep the 'Physical' in physicalism. *Journal of Philosophy, 94,* 622-637.

Mills, J. (Ed.). (2022). *Psychoanalysis and the Mind-body Problem*. Routledge.

Montero B (2009) What is the physical? In: Beckermann A, McLaughlin BP, Walter S (eds) The Oxford handbook of philosophy of mind. Oxford University Press, Oxford

Montero, B. (2006). What Does the Conservation of Energy Have to Do with Physicalism? Dialectica 60: 383–396.

Montero, B. (1999). The Body Problem. Noûs 33: 183–220.

Montero, B., & Papineau, D. (2005). A defence of the Via Negativa argument for physicalism. *Analysis, 65,* 233-237.

Muideen, A. O. (2017). The problem of consciousness: an assessment of Michael Tye's and David Chalmers' criticisms of the phenomenal concept strategy. *Inkanyiso*, *9*(1), 1-12.

Nagami, I. (1977). *Human existence and the Cartesian ego in the work of Heidegger and Watsuji: an inquiry on ethics and technological rationality* (Doctoral dissertation, The University of Chicago).

Nagel, T. (1980). What is it like to be a bat?. In *The language and thought series* (pp. 159-168). Harvard University Press.

Needham, P. (2010). Water and the development of the concept of chemical substance. *A history of water, Series II*, *1*, 86-123.

Nelson, G. B., Lord, J., & Ochocka, J. (2001). *Shifting the paradigm in community mental health: Towards empowerment and community*. University of Toronto Press.

Ney, A. (2008a). Physicalism as an attitude. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 138* (1)*,* 1-15.

Ney, A. (2008b). Defining physicalism. *Philosophy Compass, 3(5),* 1033-1048.

Noice, T., Noice, H., & Kramer, A. F. (2014). Participatory arts for older adults: A review of benefits and challenges. *The gerontologist*, *54*(5), 741-753.

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., ... & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual review of psychology*, *73*(1), 719-748.

Nozick, R. (2016). Philosophy and the Meaning of Life. *Life, Death, and Meaning: Key Philosophical Readings on the Big Questions*, 65-92.

Olson, E. T., & Segal, A. (2023). *Do We Have a Soul?: A Debate*. Taylor & Francis.

O'Rourke, A. (1998). Caring about virtual perts: an ethical interpretation of Tamagotchi, 1-20.

Ostenfeld, E. (2018). Ancient Greek psychology and the modern mind-body debate.

Overgaard, S. (2006). The problem of other minds: Wittgenstein's phenomenological perspective. *Phenomenology and the Cognitive Sciences*, *5*, 53-73.

Ozonoff, S., South, M., & Miller, J. N. (2000). DSM-IV-defined Asperger syndrome: Cognitive, behavioral and early history differentiation from high-functioning autism. *autism*, *4*(1), 29-46.

Ozouf, M. (1991). *Festivals and the French revolution*. Harvard University Press.

Papineau, D. (2013). The rise of physicalism. In *Proper Ambition of Science* (pp. 182-216). Routledge.

Papineau, D. (2011). What exactly is the explanatory gap?. *Philosophia*, *39*, 5-19.

Papineau, D. (2009a). The Causal Closure of the Physical and Naturalism. *McLaughlin, Beckermann, and Walter (eds.), The Oxford Handbook of Philosophy of Mind*, 53-65.

Papineau, D. (2009b). The poverty of analysis. In *Aristotelian Society Supplementary Volume* (Vol. 83, No. 1, pp. 1-30). Oxford, UK: Oxford University Press.

Papineau, D. (2002). Thinking about Consciousness, Oxford: Oxford University Press.

Papineau, D. (1998). Mind the gap. *Philosophical perspectives*, *12*, 373-388.

Pasteur, L. (1885). Methode pour prevenir la rage apres morsure. *C. R. Acad. Sci.* 101, 765–774.

Peer, M. (2022). *Tell Yourself a Better Lie: Use the power of Rapid Transformational Therapy to edit your story and rewrite your life*: Rapid Transformational Therapy.

Pellicano, E., & den Houting, J. (2022). Annual Research Review: Shifting from 'normal science'to neurodiversity in autism science. *Journal of child psychology and psychiatry*, *63*(4), 381-396.

Pettit, P. (1993). A definition of physicalism. *Analysis*, *53*(4), 213-223.

Pigliucci, M., & Boudry, M. (2014). Prove it! The burden of proof game in science vs. pseudoscience disputes. *Philosophia*, *42*, 487-502.

Planck, M. (1944) "Das Wesen der Materie" (The Nature of Matter), speech at Florence, Ital, Abteilung V, Rep. 11 Planck, Nr. 1797,Archiv zur Geschichte der Max-Planck -Gesellschaft (Archives of the Max Planck Society).

Poland, J. (2003). Chomsky's challenge to physicalism. *Chomsky and his critics*, 29-48.

Popper, K. (1963). Conjectures and refutations. *The Growth of Scientific*.

Pylkkänen, P. (2017). Is there room in quantum ontology for a genuine causal role for consciousness?. *The Palgrave Handbook of Quantum Models in Social Science: Applications and Grand Challenges*, 293-317.

Radner, D. (1985). Is There a Problem of Cartesian Interaction?. *Journal of the History of Philosophy*, *23*(1), 35-49.

Reath, A. (1989). The Categorical Imperative and Kant's Conception of Practical Rationality. *The Monist*, *72*(3), 384-410.

Rigsby, C. A. (2010). Nishida on Heidegger. *Continental philosophy review*, *42*, 511-553.

Robinson, H. "Dualism", *The Stanford Encyclopedia of Philosophy* (Spring 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/spr2023/entries/dualism/>.

Rowland, I. (2019). *The Full Facts Book Of Cold Reading: The definitive guide to how cold reading is used in the psychic industry*: Ian Rowland Limited.

Rowling, J. K. (2007). *Harry potter and the deathly hallows*. Bloomsbury.

Rowling, J. K. (1998). *Harry potter and the chamber of secrets.* Bloomsbury.

Rowling, J. K. (1997). *Harry potter and the philosopher's stone.* Bloomsbury.

Scholtz, S. E., de Klerk, W., & de Beer, L. T. (2020). The use of research methods in psychological research: A systematised review. *Frontiers in Research Metrics and Analytics*, *5*, 1.

Sheffer, E. (2018). *Asperger's children: The origins of autism in Nazi Vienna*. WW Norton & Company.

Shermer, M. (2013). Science and pseudoscience: The difference in practice and the difference it makes. *Philosophy of pseudoscience: Reconsidering the demarcation problem*, 203-223.

Siegel, L. (1991): *Net of Magic: Wonders and Deceptions in India*. Univ. of Chicago Press.

Smith, K. A. (2012). Louis Pasteur, the father of immunology?. *Frontiers in immunology*, *3*, 68.

Sorell, T. (2013). *Scientism: Philosophy and the infatuation with science*. Routledge.

Spiegel, T. J. (2023). Why Naturalism cannot (Merely) be an Attitude. *Topoi*, *42*(3), 745-752.

Spurrett, D., & Papineau, D. (1999). A note on the completeness of 'physics'. *Analysis*, *59*(1), 25-29.

Squires, E. J. (1990). *Conscious mind in the physical world*. CRC Press.

Stickley, T., Parr, H., Atkinson, S., Daykin, N., Clift, S., De Nora, T., ... & Hogan, S. J. (2017). Arts, health & wellbeing: reflections on a national seminar series and building a UK research network. *Arts & health*, *9*(1), 14-25.

Stoljar, D. "Physicalism", *The Stanford Encyclopedia of Philosophy* (Spring 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/spr2024/entries/physicalism/>.

Stoljar, D. (2023). Non-standard Physicalism: The Epistemic Approach to the Problem of Consciousness. In *What is Consciousness?* (pp. 63-131). Routledge.

Stoljar, D. (2009). Hempel's Dilemma. In *From Truth to Reality* (pp. 191-207). Routledge.

Stuart, H. (2017). Mental illness stigma expressed by police to police. *Isr J Psychiatry Relat Sci*, *54*(1), 18-23.

Stubenberg, Leopold and Wishon, Donovan, "Neutral Monism", *The Stanford Encyclopedia of Philosophy* (Spring 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/spr2023/entries/neutral-monism/>.

Sunkel, C. (2012). Empowerment and partnership in mental health. *The Lancet*, *379*(9812), 201-202.

Swinburne, R. (2009). Substance dualism. *Faith and Philosophy*, *26*(5), 501-513.

Szasz, T. (2010). Psychiatry, anti-psychiatry, critical psychiatry: What do these terms mean?. *Philosophy, Psychiatry, & Psychology*, *17*(3), 229-232.

Szatmari, P., Archer, L., Fisman, S., Streiner, D. L., & Wilson, F. (1995). Asperger's syndrome and autism: Differences in behavior, cognition, and adaptive functioning. *Journal of the American Academy of Child & Adolescent Psychiatry*, *34*(12), 1662-1671.

Thagard, P. R. (1978). Why astrology is a pseudoscience. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* (Vol. 1978, No. 1, pp. 223-234). Cambridge University Press.

Tiehen, J. (2015). Explaining causal closure. *Philosophical Studies*, *172*(9), 2405-2425.

Titchener, E. B. (1912). The schema of introspection. The American journal of psychology, 23(4), 485-508.

Tong, D. (2017). *Quantum Fields: The Real Building Blocks of the Universe - with David Tong.* [Video]. The Royal Institution. <https://www.youtube.com/watch?v=zNVQfWC_evg>

Vallentyne, P. (1987). The teleological/deontological distinction. *The Journal of Value Inquiry*, *21*, 21-32.

VandeWall, H. (2007). Why water is not H2O, and other critiques of essentialist ontology from the philosophy of chemistry. *Philosophy of Science*, *74*(5), 906-919.

Van Fraassen, B. C. (2002). *The empirical stance.* Yale University Press.

Vicente, A. (2011). Current physics and 'the physical'. *The British Journal for the Philosophy of Science*.

Volkmar, F. R., & Klin, A. (2000). Diagnostic issues in Asperger syndrome. *Asperger syndrome*, *27*, 25-71.

Walker, S., & Scherer, S. W. (2013). Identification of candidate intergenic risk loci in autism spectrum disorder. *BMC genomics*, *14*, 1-6.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological review*, *20*(2), 158.

Wheeler, M. (2008). Cognition in context: phenomenology, situated robotics and the frame problem. *International journal of philosophical studies*, *16*(3), 323-349.

Wheeler, M. (2005). *Reconstructing the cognitive world: The next step*. MIT press.

Wheeler, J. A. (1990) *"Information, Physics, Quantum: The Search for Links,"* in Complexity, Entropy, and the Physics of Information by Wojciech Hubert Zurek (Redwood City, CA: Addison-Wesley).

Wilson, J. (2006). On characterizing the physical. *Philos Stud, 131*, 61–99.

Wiseman, R. S. (2010). Heads I win, tails you lose: How some parapsychologists nullify null results and what to do about it. *Debating psychic experience: Human potential or human illusion*, 169-78.

Wiseman, R. (1997). *Deception & Self-Deception: Investigating Psychics*: Prometheus.

Wiseman, R., & Watt, C. (2018). Achieving the impossible: a review of magic-based interventions and their effects on wellbeing. *PeerJ*, *6*, e6081.

Wiseman, R., & Watt, C. (2006). Belief in psychic ability and the misattribution hypothesis: A qualitative review. *British Journal of Psychology*, *97*(3), 323-338.

Wiseman, R., & West, D. (1997). An experimental test of psychic detection. *The Police Journal*, *70*(1), 19-25.

Witmer, D. G. (2018). Physicality for physicalists. *Topoi*, *37*(3), 457-472.

Worley, S. (2006). Physicalism and the via negativa. *Philosophical Studies*, *131*, 101-126.

Wright, D. (1997). Getting out of the asylum: understanding the confinement of the insane in the nineteenth century. *Social history of medicine*, *10*(1), 137-155.

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, e1.

Young, K. (2014). General Characteristics of Oriental Mythology. *Bulletin of the Deccan College Post-Graduate and Research Institute*, *74*, 313-326.

Zavala, A. J. (2012). Heidegger in Nishida Philosophy. *Natureza Humana-Revista Internacional de Filosofia e Psicanálise*, *14*(1), 72-100.

Zeilinger, A. (2004) *"Why the Quantum? 'It' from 'bit'? A Participatory Universe? Three Far-reaching Challenges from John Archibald Wheeler and Their Relation to Experiment,"* in Science and Ultimate Reality: Quantum Theory, Cosmology and Complexity eds. John D. Barrow, Paul C. W. Davies, and Charles L. Harper, Jr., (Cambridge, UK: Cambridge University Press, 2004), 201–220, 218–219.