Kinds and Classification in Consciousness Science

Cecily Whiteley

A thesis for the degree of Doctor of Philosophy.

Department of Philosophy, Logic and Scientific Method of the London School of Economics and Political Science.

July 2022.

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work.

I confirm that a version of Chapter 1 "Aphantasia, Imagination and Dreaming" is published in *Philosophical Studies* **178** (2021): 2111–2132. I also confirm that some of the material contained in this chapter was previously submitted as an essay toward an MPhilStud degree at King's College London in 2018.

I confirm that a version of Chapter 4 "Depression as a Disorder of Consciousness" is to be published in the *British Journal for the Philosophy of Science* (Whiteley, forthcoming).

I declare that my thesis consists of 71,046 words.

Cecily Whiteley

Abstract

Understanding the biological basis of consciousness is one of the central challenges for modern science. Is a mature scientific explanation really possible, and if so, how should consciousness science be organized so as to achieve this? This thesis is a collection of four papers which approach these questions via an account of the natural categories or 'kinds', drawn from philosophy of science, to which paradigmatic mental phenomena like consciousness belong. The central claims defended in the thesis are twofold. Firstly, that to facilitate a mature science of consciousness it is crucial to explicitly introduce a new organizing concept of consciousness modelled on the kinds of concepts which underpin previous successful explanations of other natural phenomena – that of a *natural kind concept of consciousness*. This draws on the general observation that some concepts are better suited to scientific theorizing about the world than others, and thus that progress in scientific and philosophical domains can be (and indeed, has often been) made by working to refine our existing conceptual schemes to reflect this. Secondly, that a mature science of consciousness will be one which recognizes the reality and explanatory importance of a particular category of mental phenomena called '*global states of consciousness'* which includes dreaming and wakefulness. Examination of this latter claim leads to a third idea explored and motivated in the thesis which brings consciousness science into closer connection with psychiatry. This is the hypothesis that Major Depressive Disorder involves a change to a subject's global state of consciousness, marking a change from a state of ordinary wakeful consciousness to an altered 'depressive state' of consciousness akin to a dream or psychedelic state.

Acknowledgements

Firstly, I would like to express my gratitude to my supervisors Jonathan Birch and Matthew Soteriou. I have come to believe that a necessary condition on successfully completing a PhD is the acquiring of a kind and supportive supervisor. I feel incredibly fortunate to have had two.

Jonathan has been a wonderful primary supervisor, and his many insights and critical feedback on multiple drafts of my papers have made all four chapters of this thesis innumerably better. While the work I have done during the PhD on animal consciousness and a broader epistemology of natural kinds did not make it into the final version of this thesis, the many conversations I have had with Jonathan on this topic over the years have been an important source of philosophical motivation and inspiration. The past four years have also had their fair share of difficult moments, particularly due to the COVID-19 pandemic and periods of ill-health. Jonathan's continued support and encouragement throughout these times, as well as his advice and navigation of the academic job market, have been invaluable.

Likewise, Matt has been an excellent supervisor not only for the duration of my PhD but also throughout my MPhil at King's College London, and his role in bringing me where I am today - both literally and philosophically - is hard to overstate. Most importantly given the content of this thesis, Matt can be credited with infecting me with an obsession with dreaming and global states of consciousness. His seminar on 'dreaming, sleep and wakefulness' at King's in 2017 prompted the ideas which now are contained in Chapter 1 of this thesis, and his work on these topics has been instrumental in shaping the initial ideas about depression and global states of consciousness explored in this thesis. In '*Waking Up and Being Conscious'* Matt asks us to reflect on the phenomenology of waking up from a dream and, following Brian O'Shaughnessy, he observes that wakeful consciousness here acts as an epistemic 'corrective'. It was reading this that first prompted the connection for me between wakefulness and depression, given the strong phenomenological similarities in my experience between waking up and emerging from a depressive episode. More importantly than any philosophical influence, however, Matt has dedicated more time and energy to kindly helping me rigorously think through and correct my ideas than any supervisee can ask for, and for that I am extremely grateful.

A special thanks should also be given to three other people. First, although not an official supervisor, Liam Kofi Bright (notable Leverhulme prize winner and soon-to-be-twitter blue tick) has read and provided valuable feedback on many of the papers in this thesis. His genuine interest in and support of my research, as well as his willingness to have impromptu zoom discussions about random philosophical topics, have kept me going whilst trying to write a PhD during a pandemic. Second, thanks to the two people who first suggested to me that I

should do a PhD in philosophy when, at the time, I couldn't imagine anything worse: Louise Richardson and Helen Yetter-Chappell. Almost six years later, the topics they taught – consciousness and Nagel's *The View from Nowhere* – continue to be two of my favorite things to think about.

I was lucky enough to spend two semesters this academic year at NYU - a highlight of my PhD experience. Thank you very much to Dave Chalmers for sponsoring me for this visit, and for hosting some of my favorite memories of New York – the time spent weekly drinking wine and talking philosophy with members of the Conceptualizing the World and NYU Philosophy of Mind Discussion Group seminars. Thank you also to the wider NYU philosophy community for making me feel so welcome and making this - usually stressful - time at the end of the PhD so special.

In addition to those already mentioned, many academics at LSE, NYU and elsewhere have generously given written feedback on, or had conversations with me about, earlier versions of the material contained in these papers. In no particular order, thank you to: David Papineau, Harriet Fagerberg, Adriana Alcaraz-Sanchez, Andrew Richmond, Dave Jenkins, Kirstine la Cour, Alex Grzankowski, Christian List, Chloe de Canson, Bryan Roberts, Nick Shea, Michael Strevens, Stephan Pohl, Cristina Ballarini, Soren Schlassa, Adrian Boutel, Claudia Passos-Ferreira and four anonymous reviewers for *Philosophical Studies* and *The British Journal for the Philosophy of Science* whose comments greatly improved Chapters 1 and 4. Special thanks go to those who, in addition to insightful philosophical discussion, also provided much more important friendship: Fabian Beigang, Sophie Kikkert, Nick Makins, Dmitry Ananyev, Bar Luzon, Noga Gratvol, Ali Rezaei, Matthias Michel, Veronica Gómez Sánchez and Lauren Somers.

A huge thank you to my biggest cheerleader and living anti-depressant, Saadhana. Her constant support and infectious happiness have made the past few years so much brighter, and because of this, I have been able to complete this thesis.

Above all, thanks to my parents, Jackie and Geoff Whiteley, for their love and support, as well to my late brother, Matt Whiteley, to whom Chapter 4 is dedicated. The walks around the Yorkshire Dales during lockdown while writing the PhD with my parents and our dog Ernie are amongst my favorite memories. My dad – the original, (and kindest!), Dr. Whiteley – deserves a special mention here for his interest in and knowledge of the history of science which spurned many interesting discussions with respect to the material in these papers. This thesis is for them.

Contents

Introduction 9

1 Aphantasia Imagination and Dreaming 15

1 What are dreams? . . . 15
2 The imagination model of dreaming . . . 17
3 The empirical evidence for an imagination model . . . 20
3.1. Dream development in children . . . 20
3.2. The neuropsychology of dreams . . . 20
4 Aphantasia and dreaming . . . 22
5 The case against Ichikawa's imagination model . . . 23
6 An inactive imagination model of dreaming . . . 27
7 Suggestions for future research . . . 29
7.1. Extreme imagination and competing theories of dreams . . . 29
7.2. Aphantasia and lucid dreaming: a testable empirical hypothesis . . . 31

2 Ameliorative Physicalism and the Natural Kind Approach to Consciousness 34

1 Introduction . . . 34
2 Carnapian explication . . . 37
3 The epistemic defect argument . . . 40
3.1. A 'seemingly wonderful concept' . . . 40
3.2. Empirical disputes . . . 42
3.3. The argument . . . 46
4 Natural kind concepts . . . 47
5 Ameliorative physicalism and the natural kind approach . . . 51
6 Overcoming objections to the natural kind strategy . . . 56
6.1. Foundational objections to the natural kind approach . . . 57
6.2. Taylor's objection: The problem of an undetected kind . . . 60
7 The objection from primitive concepts . . . 67
8 Conclusion . . . 70

3 The Unity of Dreams 73

1 The unity assumption . . . 73
2 The unity assumption at work . . . 76
2.1. Case study 1: The neural correlates of dreaming . . . 76
2.2. Case study 2: Lucid dreaming as a methodology for the study of dreams . . . 80
3 The definitional approach to dream science . . . 84
4 Two objections to the definitional approach to dreaming . . . 87
4.1. The circularity objection . . . 87
4.2. The empirical insensitivity objection . . . 89
5 The natural kind approach to sleep experience . . . 92
5.1. What is a natural kind? . . . 93
5.2. A natural kind methodology for sleep experience . . . 95
6 Conclusion . . . 97

4 Depression as a Disorder of Consciousness 101
1 What is depression? . . . 101
2 Depression in a clinical context . . . 103
3 The phenomenology of depression . . . 104
4 What is a global state of consciousness? . . . 108
5 The depressive state of consciousness . . . 112
6 Empirical evidence and application . . . 116
6.1. Global state and MDD research: A reciprocal relationship . . . 117
6.2. Psychedelic psychiatry . . . 118
7 Conclusion . . . 120

Conclusion 121
Bibliography 126

List of Figures

Figure 1.1 The distribution of VVIQ scores among aphantasic subjects and control participants.

Figure 2.1. The homeostatic property cluster theory of natural kinds.

Figure 2.2. Shea's 'one' versus 'two kinds' scenario.

Figure 2.3. Scenario X: A multiple kinds view of consciousness involving false negative markers.

Figure 2.4. Scenario Y: A multiple kinds view of consciousness involving an equal distribution of false negative and true positive markers.

Figure 3.1. A monistic and pluralistic view of the relationship between dreaming and sleep stages.

Figure 3.2. The definitional approach to the science of dreaming.

Figure 3.3. The multi-dimensional view of global states of consciousness.

Figure 3.4. The natural kind approach to the science of dreaming.

Figure 4.1. The levelled view of global states of consciousness.

Figure 4.2. The multi-dimensional view of global states of consciousness.

Figure 4.3. Depression as a temporal process.

0 | Introduction

Although often framed in terms of 'analysis', many prominent philosophical questions about minds are also interpretable as questions regarding the correct classification of mental categories. Can we make sense of, and illuminate the nature of, a particular mental category – say, beliefs, or imaginative episodes – by classifying them as instances of a broader mental kind – say, representations, or mental actions? Is the correct classification of perceptual phenomena one on which veridical perception, illusion, and hallucination are classified as instances of the same mental kind, or a taxonomy which holds that these are fundamentally distinct kinds of mental categories? This is no less true of philosophical theorizing about consciousness. In the philosophical literature on consciousness, one typically finds central questions posed in terms of how this particular mental kind relates, or fits into, the existing classificatory system we have for categorizing the natural world. Does consciousness belong in the same 'physical' category which includes things like electrons, water, and neurochemical brain states, or does the reality of consciousness force us to recognize a category of 'non-physical' phenomena which current scientific methods are ill-equipped to explain? Similarly, can we classify all instances of phenomenal consciousness – the what it's like aspect of conscious experience - as instances of conscious access, or are these distinct mental kinds which are not necessarily co-extensive? These latter questions are exciting and pressing, for they have important methodological implications for the growing empirical research program which aims to deliver an account of the neurobiological basis, or explanation, of consciousness. Thought to be one of the central challenges for modern science, the ability of philosophical investigation to shed light on foundational questions which bear on this project serve to further demonstrate the importance of philosophy as a research field, and its role in producing a mature scientific understanding of the mind.

This thesis is the beginning of a version of a methodological research program which approaches the philosophical study of the mind via an account of the nature of the *mental categories* of which paradigmatic mental kinds are an instance – an account which places these mental kinds firmly, but not reductively, in the natural world. That is, it is an approach which seeks to analyze the nature, individuation and ontological constitution of paradigmatic mental kinds via an examination of the nature, individuation and ontological constitution of the natural categories to

which these mental kinds belong. Likewise, it is an approach which aims to answer questions of how we can accurately come to *know* about paradigmatic mental kinds – for example, whether and when they are instantiated in ourselves or in members of other species – by examining how one accurately comes to know about the nature of psychological and biological natural kinds more generally. These two issues are closely related and can be used to inform one another, for one can appeal to claims about the common metaphysical structure of these kinds to motivate a detailed epistemology of how one can come to reliably identify instances of these kinds in typical as well as unusual cases. Finally, it is an approach which also looks to answer the methodological question of how to construct a fruitful scientific research program into a given mental category by modelling this on previous successful scientific investigations into other natural kinds. This is done by examining the detailed methodological procedures which give rise to previous successful scientific explanations of other natural phenomena revealed and documented by the history and philosophy of science.

This thesis begins this research program by exploring these issues with respect to one of the most famous and characteristic mental categories: subjective experience. The emphasis here will be on the latter methodological question of how to construct a mature scientific research program which aims at a neurobiological explanation of this phenomenon. With this in mind, the thesis aims to motivate and contribute two central ideas. First, the thesis aims to motivate the idea that in order to bring about a mature science of consciousness it is essential to first explicitly introduce and motivate a new organizing concept (or set of concepts) of consciousness modelled on the concepts which underpin previous successful scientific explanations of other natural phenomena – that of a *natural kind concept of consciousness* (Magnus 2012). This draws on the general observation that some concepts are better suited to scientific theorizing about the world than others, and thus that progress in scientific and philosophical domains can be (and indeed, has often been) made by working to refine our existing conceptual schemes to reflect this. According to the view I articulate in this thesis, the various concepts of consciousness utilized in contemporary consciousness science are no exception. More specifically, I aim to show that many of the on-going methodological disputes and puzzles that currently slow the maturation of consciousness science as a research field arise as a result of the retention of an epistemically problematic conception of consciousness, and thus that once this is self-consciously replaced with an alternative concept of consciousness, these pressing methodological disputes can be dissolved.

The second core aim of the thesis is to motivate the claim that a mature science of consciousness is one which recognizes the reality and explanatory importance of a particular category or family of mental phenomena which, following Bayne et al. (2016), we can call '*global states of consciousness*'. The rich inner world which constitutes our conscious experience undergoes many different types of change. One of the most drastic, yet familiar types of change is that which occurs when we fall asleep each night and enter a dream, or when we awake from a dreamless sleep into wakeful consciousness. In these and related cases, such as when subjects take psychedelics, our conscious experience undergoes a profound shift - the typical experience of dreaming and being awake being very different - which is *global* in character; it affects and alters all aspects of our

conscious experience, from our sense of self and agency to our experience of time and mental imagery. Despite being central to our familiar experience of the world, the concept of a global state of consciousness, understood as a family or category of phenomena *distinct* from individual states like dreaming, has received comparatively little philosophical and empirical attention. I believe that this oversight is a mistake, and that recognition of this category of conscious phenomena will be crucial to a mature scientific understanding and explanation not only of consciousness but a range of other phenomena of interest to philosophers and empirical researchers. In this thesis I suggest we can begin to demonstrate the importance of this category of phenomena by exploring how the recognition and employment of this concept can revolutionize our understanding and scientific study of dreaming and Major Depressive Disorder.

CHAPTER SUMMARIES

CHAPTER 1 ''APHANTASIA, IMAGINATION AND DREAMING'' The clearest example of a global state of consciousness is the state to which one surfaces to when one ceases to be awake and enters a dream. If this is correct, then this suggests that we can get a handle on the nature and constitution of this broader category of conscious phenomena by examining the nature and ontological constitution of dreaming. Philosophical debate on this question has predominantly focused thus far on the question of whether dreams are best classified as instances of imaginative experiences or rather as misleading wake-like hallucinations – a question which is thought to be important insofar as it has implications for the epistemological debate on dream skepticism (Windt 2016, Sosa 2005, Ichikawa 2009). In this chapter I articulate an empirical challenge to the view that dreams can be classified as imaginings, constitutively involving sensory mental imagery. This is grounded in the recent rediscovery of a visual mental imagery generation disorder known as 'aphantasia'. Recent studies of aphantasia and its neurobehavioral correlates reveal that the majority of aphantasics, whilst unable to generate visual imagery while awake, nevertheless retain the capacity to experience rich visual dreams. I argue that this raises a challenge to standard imagination models of dreaming. After presenting this challenge in the context of Jonathan Ichikawa's arguments for the imagination view of dreams (2009, 2016), I argue that this empirical argument can be overcome if the imagination theorist abandons Ichikawa's account of dreaming in favor of a modified version. This involves the claim that dreams are essentially inactive and involve involuntary forms of imagination (Soteriou 2017, Crowther 2018). I conclude with a suggestion for further empirical research which can test the viability of this alternative hypothesis.

While not explicitly discussed in the paper, this chapter serves to highlight and demonstrate two important methodological assumptions implicitly at work in contemporary dream research which are later taken up and challenged in Chapter 3. Firstly, the assumption that there is a *single kind of conscious state* to which we surface when we dream – in other words, that the diverse conscious experiences during sleep we typically refer to as 'dreams' constitute a natural kind. Second, the assumption that we can provide an adequate answer to the question 'what is the nature of dreaming?' by appealing *solely to* specific 'content-level' features of one's conscious experience

while waking – features which include occurrent imaginative episodes, beliefs, and hallucinatory perceptions. That is, the assumption that an adequate account and explanation of dreaming can be provided without referring to the background structural features of conscious experience which characterize a subject's global state of consciousness.

CHAPTER 2 ''AMELIORATIVE PHYSICALISM AND THE NATURAL KIND APPROACH TO CONSCIOUSNESS'' Leaving dreaming aside for the time being, Chapter 2 takes up the first core aim of the thesis of motivating and articulating a new concept of consciousness modelled on previous scientific explanations which can serve as an organizing concept for a mature science of consciousness. It achieves this by articulating a new physicalist position in the metaphysics of consciousness called 'ameliorative physicalism'. According to this view, philosophers and consciousness scientists ought to move away from descriptive projects regarding our phenomenal concepts – projects on which questions of what features our concepts of consciousness *in fact* possess are central - toward *normative* analyses of these concepts which examine the ways in which our current concepts of consciousness can be refined and improved for the purposes of facilitating empirical and scientific success. Just as Sally Haslanger (2000) has proposed that we should aim to develop accounts of race and gender that will be effective in the fight against injustice, I propose that physicalists should aim to develop accounts of consciousness which will be effective in facilitating a mature science of consciousness.

After laying the foundations for this broader program, the chapter offers and defends a candidate for a replacement concept of consciousness in the form of a 'natural kind concept of consciousness' drawn from contemporary philosophy of science. Inspired by the work of Timothy Bayne and Nicholas Shea on the 'natural kind approach to consciousness' (2010, 2012, 2020), the chapter motivates a reciprocal relationship between the ameliorative physicalist position and the natural kind approach to consciousness. That is, it is argued firstly that one can use recent work on the natural kind approach to consciousness to motivate the idea that a natural kind concept is epistemically promising as an organizing concept for consciousness science and secondly, that ameliorative physicalism provides this approach with the conceptual toolkit needed to overcome several pressing objections levied against it. Together, these present a promising way forward for consciousness science.

CHAPTER 3 ''THE UNITY OF DREAMS AND DREAMING'' Taking its cue from the first two chapters, the third chapter of the thesis takes up the idea that consciousness is best studied as a natural kind and applies this to the scientific study of dreaming (or of conscious experiences in sleep). The chapter starts by identifying an unspoken methodological assumption common to contemporary dream science according to which dreaming forms a natural kind. It is argued that this 'unity assumption' plays an active organizing role in recent empirical research, shaping experimental design and motivating orthodox interpretations of empirical data. It then proceeds to show that despite the crucial methodological role afforded to this assumption in consciousness science, this methodological claim is not justified by the existing conceptual and empirical evidence collated and discussed in dream research. I argue that recognition of this fact calls for a

fundamental revision to the way in which sleep experience is studied in consciousness science. That is, I argue that this critical discussion of the unity assumption reveals that the dominant methodology in dream science – one which proceeds by motivating various *phenomenological definitions* of dreams which on go on to form the explanatory targets of empirical dream research (e.g., Windt 2016, Revonsuo et al. 2016, Wamsley and Stickgold 2021) -- is epistemically risky and an ultimately misguided research strategy. In its place, I outline an alternative 'natural kind approach' to the science of dreaming which aims to identify the global states of consciousness which obtain during sleep which meet robust epistemic criteria for natural kinds. This opens up exciting new possibilities for consciousness science which have yet to be fully explored: namely, the idea that 'dreaming' and even 'sleep experience' may not constitute natural psychological kinds.

CHAPTER 4 ''DEPRESSION AS A DISORDER OF CONSCIOUSNESS'' Part of what is involved in coming to recognize the reality of the category 'global states of consciousness' is recognizing that there are some psychological phenomena which resist explanation and elucidation at a particular level of explanation -- one which makes primary reference to changes in the intentional content of an individual's conscious experience. That is, one way of motivating the reality of this mental category is to call attention to the idea that in order to do justice to the nature and phenomenology of some psychological phenomena one needs to posit and make reference to *the background structural and global features* which typically frame a subject's conscious experience; a level of explanation which goes missing if one focuses solely on content-related aspects of a subject's experience. Acceptance of this general idea invites a natural further question: namely, what *other* salient psychological phenomena, beyond paradigmatic states like wakefulness and dreaming, can be best explained via changes to a subject's global state of consciousness? In other words, are there other phenomena which are currently elusive, but could nonetheless benefit from classification and elucidation in these terms? In this final chapter of the thesis, I motivate the novel hypothesis that *Major Depressive Disorder* is one such phenomenon whose nature and phenomenology can be analyzed and explained in these terms. The hypothesis proposed is that when an individual becomes depressed, the individual departs from a state of ordinary wakeful consciousness and enters a distinctive global state of consciousness akin to dreaming and the psychedelic state. After unpacking and motivating this hypothesis in the context of research in phenomenological psychiatry and consciousness science (e.g., Ratcliffe 2015), I outline two of its important implications for the neurobiology of depression and psychedelic psychiatry. The upshot is a promising and conceptually well-motivated hypothesis about depression which is apt for empirical uptake and development.

As I have suggested, there is a clear sense in which the four papers that constitute this thesis lay the foundations for a much longer and comprehensive research program, not just of natural kinds in the mind, but of the idea that the concept of a 'global state of consciousness' can serve as an organizing explanatory concept for modern psychiatry. As such, the work contained in the thesis inevitably leaves a number of crucial questions about the nature of global states and their significance unanswered. I therefore conclude the thesis with a brief discussion of what I see as the

main research questions which need to be answered if the broader research program is to succeed, to be taken up in future work.

A note to the reader before proceeding. The chapters of this thesis have been written such that they can each stand alone as self-contained research papers. As such and given that these papers are united and guided by the common themes and research project outlined in this introduction, some overlap in the content of sections of these papers has been inevitable. These include: [Chapter 2, Section 4 and Chapter 4, Section 5.1.] on the notion of natural kinds I am working with, and [Chapter 3, Section 5.2. and Chapter 4, Section 4] on the background conception of global states of consciousness I have in mind. Given this, the reader should feel free to pass over the repeated material where they see fit.

1 | Aphantasia, Imagination and Dreaming

Aphantasia is a recently discovered disorder characterized by the total incapacity to generate visual forms of mental imagery. This paper proposes that aphantasia raises important theoretical concerns for the ongoing debate in the philosophy and science of consciousness over the nature of dreams. Recent studies of aphantasia and its neurobehavioral correlates reveal that the majority of aphantasics, whilst unable to produce visual imagery while awake, nevertheless retain the capacity to experience *rich visual dreams.* This finding constitutes a novel explanandum for theories of dreaming. Specifically, I argue that the recent dream reports of aphantasics constitute an empirical challenge to the emerging family of views which claim that dreams are essentially imaginative experiences, constitutively involving the kinds of mental imagery which aphantasics, *ex-hypothesi,* lack. After presenting this challenge in the context of Jonathan Ichikawa's recent arguments for this view, I argue that this empirical challenge may be overcome if the imagination theorist abandons Ichikawa's account of dreaming in favour of a modified version. This involves the claim that dreams are essentially inactive and constitutively involve *non voluntary* forms of imagination. I conclude with a suggestion for further research which can test the viability of this alternative hypothesis and move the debate forward.

1. What are Dreams?

An emerging debate in the philosophy and science of consciousness concerns the nature and ontological status of dreams. Much discussion has focused on solving what Jennifer Windt (2015) calls the "conceptualisation problem": how, if at all, ought we to account for dreaming using the standard psychological terms (such as perception, hallucination, thought and emotion) used to characterize wakeful consciousness? According to one growing family of views, dreams are best understood in ontic terms as instances of *imaginative experiences.* The so-called 'imagination model of dreaming' has been influentially developed by Jonathan Ichikawa (2009, 2016), who argues that dreams constitutively involve both sensory and propositional forms of imagination. In addition to conceptual lines of argument, Ichikawa maintains that the imagination model of dreaming receives considerable support

from various forms of *neuropsychological evidence* which suggest dream imagery and waking imagery share a common neural basis (Solms 1997, Foulkes 1999).

Here, I present an empirical case against the imagination model of dreaming as presented by Ichikawa (2009, 2016). I analyse a series of recent neurological studies which identify a particularly pure case of imagery generation disorder known as *aphantasia* (Zeman et al. 2015, 2016). Subjects with aphantasia - 'aphantasics' - are characterised not just as having mental imagery of a reduced vividness, or the incapacity to produce imagery of certain visual kinds, but as lacking the capacity or power to produce visual forms of mental imagery altogether. Whilst the recent findings on aphantasia have previously been considered in relation to the ongoing 'imagery debate' in cognitive science which concerns the nature of mental imagery and its cognitive function (Zeman et al. 2015, Kosslyn et al. 2006), the full theoretical significance of aphantasia and its implications for debates about the nature of dreaming have yet to be considered. Bolstered by the fact that these preliminary studies of aphantasia reveal that the condition can be both acquired (Zeman et al. 2010) as well as a lifelong stable condition (Zeman et al. 2015, 2016), this paper starts from the claim that subjects with aphantasia and their dream reports provide an interesting test case for philosophers and psychologists who posit a tight ontological connection between dream mentation and sensory forms of imagination.

The plan for the paper is as follows. Sections **II** and **III** outline Ichikawa's imagination model of dreaming (2009, 2016) and the empirical case offered in support of it. Section **IV** introduces the recent studies on aphantasia and the notable neurobehavioural features of the condition that these highlight. In section **V** I bring these together and argue that the dream reports of aphantasic subjects recorded to date present an empirical challenge to the imagination model of dreaming presented by Ichikawa. My argument, which results in a dilemma for Ichikawa, is based on the following claims: (i) Ichikawa is committed to the claim that dreams often involve visual forms of mental imagery - which he characterises as essentially *agential* phenomena - as central and constitutive components, (ii) that aphantasic subjects *ex hypothesi* lack the capacity for these kinds of agential experiences and (iii), that aphantasic subjects nevertheless report having rich *visual* dreams. Insofar as this goes against the predictions made by the imagination model, I argue that these studies thus appear to disconfirm Ichikawa's theory, and in doing so, problematically lend support to alternative accounts which the imagination model was meant to replace. In section **VI** I consider a way in which the proponent of the imagination model can respond to this empirical challenge. The aphantasia studies notably suggest a strong neurophysiological dissociation between voluntary and involuntary forms of mental imagery, with only the capacity for the former being lost by the majority of aphantasic subjects. I argue that the strongest response available to the imagination theorist is to take up the recent calls for adoption of a variation of the imagination model according to which non-lucid dreams essentially involve involuntary, passive or inactive forms of imagination (O'Shaughnessy 2002, Soteriou 2017, Crowther 2018). I conclude in section **VII** with a discussion of the broader implications of aphantasia for theories of dreaming. This includes a proposal for future dream research which can shed light on the empirical viability of this alternative model of dreaming and move contemporary debate on the conceptualisation question forward.

2. The Imagination Model of Dreaming

The claim that dreams in some sense involve imaginative experiences leaves open a number of more sophisticated ontological analyses of dreaming[1]. Motivated by the rejection of the orthodox alternative to the imagination view, which claims that dreams are essentially *hallucinations* which involve misleading, wake-like sensory experiences (or 'percepts') leading to false beliefs, Ichikawa's formulation of the imagination model of dreaming involves a commitment to two central claims[2]:

(i) Imagery: dreams essentially involve *mental imagery* - experiences of the kind which occur when we imagine what something looks, feels, smells or tastes like. 'Imagery' implies, for example, that when I dream of the white cliffs of Dover, my experience is of the same kind as that which I have when I close my eyes and imagine them in front of me when awake (and not, contra the hallucination account, the same kind of experience I would have if I were to visually perceive them). (2009;105).

And

(ii) Imaginings: the belief like states that we take towards the content of our dreams (whilst dreaming) are instances of propositional imagination. 'Imaginings' implies that when I dream that I am swimming the English Channel, I *imagine that* I am doing so; my dream does not involve my having the belief that I am in fact swimming the English Channel (2009;111)[3].

The viability of this now-standard contrast between the hallucination and imagination models of dreaming depends on the acceptance of a number of substantive philosophical assumptions about the nature of imagination, perception and hallucination and the relations between them. Primarily, the presentation of the hallucination and imagination models as mutually exclusive theories depends on the presupposition that there is a difference in *kind* between imaginative phenomena on the one hand and perceptual-hallucinatory experiences on the other. That is, it requires acceptance of the view that there is a phenomenally determinable or otherwise *essential* difference between the categories of imaginative and perceptual-hallucinatory experiences, such that the two are considered distinct kinds of mental phenomena. The rationale for this is as follows: if imaginative and perceptual-hallucinatory experiences differ not in kind but in *degree*, as argued by Hume, then plausibly there ceases to be a substantive disagreement between proponents of imagination and hallucination theories of

[1] Variations on this claim can be found, for example, in Sartre (1940), Walton (1990), O'Shaughnessy (2002) McGinn (2004), Sosa (2005, 2007, 2009), Ichikawa (2009, 2016), Thompson & Batchelor (2014), Soteriou (2013, 2017), Crowther (2018). The assumptions that dreams (a) are phenomenally conscious experiences that occur during sleep and (b) that these are reported accurately upon waking have not gone unquestioned - Malcolm (1956), Dennett (1976), Rosen (2013) - however they are now widely accepted, and I do not press them here. A recent defense of these claims can be found in Windt & Metzinger (2007) and Windt (2015)).

[2] Versions of the hallucination model can be found in Descartes (1641), Hobson (2003) and Revonsuo (2005). Alternatives to hallucination and imagination accounts include the view that dreams involve neither perceptual or imaginative experiences but are sui-generis *immersive spatio-temporal hallucinations* (Windt 2010, 2015), and the pluralist thesis that dreams essentially involve multiple different kinds of experience (Rosen 2012, 2019).

[3] See Sinhababu (2016) for a discussion of how so-called propositional or attitudinal imagination, as a distinct cognitive attitude, can be distinguished from belief.

dreaming[4]. Acceptance of this claim gives rise to a second, crucial, observation: any theory which seeks to provide an analysis of dreaming in terms of imagination needs to supplement this claim with a philosophical account of what imagination *is*. That is, if it is to be properly informative, an imagination model of dreaming ought to include within it an account of what *it is to be* an imaginative experience, as opposed to a perceptual or hallucinatory one. In addition to securing its status as a distinct mental kind from perception and other cognitive phenomena, this must also have the resources to accommodate the various commonalities and differences thought to hold between the two[5]. Ichikawa thus combines his endorsement of imagery and imaginings with a further claim which serves as the final tenet of his theory:

(iii) Subject to the will: The crucial distinction between imagery and percepts is to be made on the basis that the former is necessarily "subject to the will". That is to say, imagination is fundamentally an *agentive* phenomenon (2009;106).

As a type of agentive theory, Ichikawa relies here upon the intuitive idea that the generation or production of mental imagery, as opposed to perception, essentially constitutes a *mental act* over which we have direct control. That is to say, the inception and content of our visual imagery is typically determined *by us*; experimental subjects can choose to visualise an apple as opposed to other pieces of fruit, for example, when asked to "imagine a piece of fruit" or an "inanimate object" in the lab. This is thought to contrast with the traditionally passive nature of veridical or hallucinatory perception, over which we exercise little control. While the agential nature of imaginings is well recognised, agentive theories claim not only that the agential origin of imaginings is common to the majority of imaginative episodes but is rather *constitutive* of them. That is, that every imaginative experience is such that it is necessarily subject to the agent's will (2009, McGuin 2004)[6].

Does this commit Ichikawa to the implausible view that our dreams are under our control, and relatedly, that there can be no instances of so-called 'passive imaginings' - unbidden mental images, or earworms? Whilst other agentive theories of imagination, which differ in subtle ways from Ichikawa's account, are clear in their commitment to this latter claim - and thus are directly opposed to an analysis of dreaming in terms of imagination - Ichikawa doesn't think so[7]. This is due to the fact that Ichikawa claims there is a distinction to be made between an experience having an agential origin

[4] Rosen (2019) includes a recent endorsement of this view. It is worth noting here that the vindication of this set up not only requires that imaginative and perception differ in kind (Kriegel 2015), but also that perceptual and hallucinatory experiences do not. That is, the dialectic here is further complicated by certain disjunctivist theories of perception, which deny that veridical and hallucinatory experiences are of the same fundamental kind (Martin 2004, 2006). This is particularly pertinent in light of the recent attempt to provide a positive disjunctivist analysis of hallucination in terms of sensory forms of imagination (Allen 2015). Again, on this view, hallucination and imagination models of dreaming would amount to the same thesis.

[5] See Nanay (2016) for the difficulties associated with this task.

[6] This 'agentialst' characterisation is forcefully captured in Ichikawa's assertion that "to imagine is to act - our imagery is an important sense under our voluntary control; this is not so with percepts" (108). Ichikawa is thus rejecting *epistemological* and *dependency* theories of imagination (O'Shaughnessy 2000, Peacocke 1985, Martin 2001) which I return to briefly in Section VII.

[7] See, for example, Dorsch (2012, 2015).

or being 'subject to the will', and that same experience being under an agent's *voluntary control*. That is, he claims, following Wittgenstein (1967):

"it is possible for something to be subject to the will, and yet not 'under voluntary control'—the annoying song that runs through your head is an example of something like this. It is subject to the will because it *makes sense* to try to banish it; it is not under your voluntary control because you are unable to succeed. We do not always voluntarily control the things we do; this does not stop them from being things we do" (2009;116, italics added).

and,

"Unwelcome imagery is more like an unwelcome habit or addiction than an unwelcome set of chains…. even when our imagery is unwelcome and we cannot banish it, we can *try* to banish it; we know what it is to banish it. We are failing to perform an act." (2009;107).

This distinction between mental agency and voluntary control which Ichikawa appeals to here, and the criterion which he claims is sufficient to demarcate involuntary dream imagery from perceptual experiences, are somewhat vague and deserve further scrutiny and elaboration. For now, it is necessary only to state the notable implication of Ichikawa's model of dreaming, *if* this can be provided: that dreams and dream imagery, like involuntary imaginings, are essentially *agential experiences* over which we lack voluntary control; that is, whose content we can attempt - and fail -to banish. To use Ichikawa's terminology, dreams on this view are like pervasive unwelcome habits or addictive episodes which we, qua agents, do whilst asleep. They are, in virtue of being imaginative experiences, *mental acts.*

In virtue of endorsing these three claims, an imagination model of dreaming is thought to have direct implications for contemporary debates in the philosophy of mind, psychology and epistemology. Thus far, the imagination model has been discussed primarily within the context of (ii) imaginings, and its epistemological consequences in relation to Cartesian dream skepticism[8]. However, recent discussion has concerned the potential of the imagination model of dreaming as an available contrast case to illuminate and motivate an analysis of the metaphysical constitution and epistemic function of *wakeful consciousness*, as a crucial, relatively neglected topic in the philosophy of mind (O'Shaughnessy 2002, Soteriou 2017, Crowther 2018). This, along with the potential of the imagination model to serve as a guide to empirical research in the science of dreaming (providing the much-needed conceptual clarification of the target phenomenon of this research) offers a strong mandate for discussion into the viability of the imagination model that Ichikawa presents, and the success of the arguments offered in support of it.

[8] Indeed, this is the primary motivation for Ichikawa's (2009;104, 2016;157-8) and Sosa's (2005) presentations of the view. Here, the initial proposal is that taking an imaginative, as opposed to doxastic, attitude towards the content of our dreams provides the proponent of the imagination model with a compelling response to dream skepticism. For a more extensive discussion of the prospects for dream skepticism in this context however see (Soteriou 2017) and Ichikawa (2008).

3. The Empirical Evidence for an Imagination Model

3.1. Dream Development in Children

Whilst there is much to be said about the case for the second of Ichikawa's central claims - that dreaming involves taking an imaginative attitude towards the content of our dreams - and whether the imagination model of dreaming ought to commit to this claim, my focus here is on the arguments Ichikawa puts forward for (i) imagery, and in particular, the empirical evidence from cognitive and behavioural psychology he presents in favour of this thesis. Adopting an empirically informed approach, Ichikawa presents two psychological considerations in favour of 'imagery' (2009;9-11, 2016;154-155,). The first concerns a series of long-term studies led by David Foulkes on dream development in children in the 1960s to the 1980s. In the first extensive longitudinal study of dreams (1982) - later documented and analysed in his influential book 'Children's Dreaming and the Development of Consciousness' (1999) - Foulkes and his team analysed dream report data collected over five years from children aged 3-15 in conjunction with performing a variety of cognitive and personality tests. Their findings, based on reports systematically collected in the lab and at home over a five-year period, were radical. In addition to consistently supporting the claim that young children and infants lack the capacity to dream - the leading scientific and folk hypothesis at the time - the results indicated that dreaming is a high-level cognitive process which develops *discretely* over time in accordance with a number of identifiable stages. In addition to tracking changes in the frequency of dreams reported, these stages also mark discrete differences in dream content throughout childhood development which correspond to the type and quality of dream imagery, narration, and active self-representation[9]. Here, Ichikawa cites the following observation from Foulkes' in support of the imagination model:

"From all my data, the suggestion is that dreaming best reflects the development of a specific cognitive competence, indexed by certain kinds of tests of visual-spatial imagination, leading to the conclusion that *imagination must be a critical skill in dream-making*" (1999, 90 italics added)[10].

That is, that children's visuospatial skills - which he and his team tested via tasks typically thought to require *visual imagery* and spatial imagination - was the only waking cognitive variable which correlated with, and developed alongside, high dream frequency and richness across the study.

3.2. The Neuropsychology of Dreams

[9] These were later confirmed in Foulkes' cross-sectional study (1999, 116).

[10] Dream frequency and rich dream content did not correlate for example with increased verbal or memory skills, which seems to repudiate the alternative hypothesis available viz. that dreams in younger children and infants *were* occurring frequently yet were forgotten or not reported due to poor linguistic or recall abilities. See Foulkes (1999, 1996), and, for a nice discussion of the methodological issues raised here, Sutton (2009;532-536) and Rosen (2019;2.2.5).

The second set of studies Ichikawa presents in favour of imagery moves to cement this neuropsychological connection between dreaming and imaginative abilities in children to adults. Here, Ichikawa draws heavily on the work of neuroscientist Mark Solms, whose systematic review of the empirical literature on dream abnormalities - the first of its kind - and subsequent clinico-anatomical study of dreams, is presented in 'The Neuropsychology of Dreams' (1997). Investigating dreaming under neuropathological conditions, Solms aimed to identify changes in dreaming systematically associated with focal cerebral pathology and to describe their clinical and anatomical characteristics. In conjunction with his study of 361 neurological patients, Solms identifies four broad categories of dream abnormalities: those in which patients report (a) a global cessation of dreaming (b) the cessation or reduction of *visual imagery* in dreams (c) increased frequency or vividness of dreams (dream-reality confusion) and (d) recurring nightmares.

According to Ichikawa, Solms' research provides empirical support for the imagination model of dreaming. This is based on the following finding on the behavioural correlates of (b): in cases in which a cessation in visual dream imagery was reported, the same patients reported a further cessation of *waking visual imagery* or, in cases in which dream imagery was reduced as opposed to absent, that their waking imagery was *analogously deficient* (74, 93-102, 228)[11]. Ichikawa maintains that such cases provide a useful means of testing the empirical credibility of the account of dreaming he defends: ''Considering patients with brain damage resulting in imaginative deficits is particularly illuminating: such subjects tend to exhibit precisely analogous deficits in dreaming'' (2016;254). Insofar as this study and similar work (Solms & Turnbull 2002) appear to support a structural connection between dreaming and imagination, Ichikawa concludes that these, in conjunction with the conceptual arguments he presents (2009;108-109), provide a compelling case in support of imagery, and by extension, his imagination model of dreaming.

There is more to be said about these two empirical considerations and their relation to Ichikawa's central thesis. For my purposes, the important feature of Ichikawa's discussion here is the methodological commitment it highlights. In drawing on these psychological considerations to support his imagination model of dreaming in the manner just described, Ichikawa commits himself to the claim that neuropsychological findings about the neural basis of dreaming and mental imagery serve as important *test cases* for philosophical accounts of dreams. That is, a natural redescription of Ichikawa's reasoning here is as follows: support and accommodation of empirical cases like these - specifically, cases in which waking imaginative capacities are reduced or *impaired* - is an important desideratum on an adequate philosophical theory of dreams[12]. The aim in the following sections is not to argue against this claim - it is widely accepted - but rather to argue that this methodological commitment, in the context of a new imagery generation disorder, provides reason to reject one of Ichikawa's central claims, and thus turn Ichikawa's psychological case on its head.

[11]Notably, and unacknowledged by Ichikawa, out of an effective sample of 186 patients only two reported a cessation or restriction of visual dream imagery, leading Solms to conclude that the presence of this disorder in the neurological population was very low (1.1%). I return to this point below.

[12] Where this is understood as an inference to the best explanation (Lipton 2003).

4. Aphantasia and Dreaming

The psychological case serving as the focus of this paper is the contemporary rediscovery of a particularly pure case of imagery generation disorder known as 'aphantasia' (Zeman et al. 2010, 2015, 2016, 2018)[13]. Whilst it is well known that introspective reports regarding the vividness of mental imagery vary significantly amongst groups of individuals (McKelvie 1995, Philips 2014), aphantasia is characterised by subjects - 'aphantasics' - who typically describe themselves as not just as having mental imagery of a reduced vividness, or impaired capacities for mental imagery generation limited to certain visual kinds, but as lacking the capacity to generate (typically visual) mental imagery *altogether*[14]. The work on aphantasia, mental imagery and their neural bases is ongoing (Dawes et al. 2020, Winlove et al. 2018, Fulford et al. 2018), however, the condition has been documented in a series of recent studies (Zeman et al. 2010, 2015) which aim to identify its central neurobehavioral features[15]. Produced as part of '*The Eye's Mind'* AHRC Project at the Medical School at the University of Exeter, the first examines a form of imagery generation disorder exemplified in a single patient - MX - who lost the ability to generate visual imagery abruptly, aged 65, four days after a coronary angioplasty (see Zeman et al. 2010). As part of the study, MX was subject to extensive psychometric testing in which he fell well within the normal range and tested negative for further cognitive deficits. What was surprising however and served as the focus of this first study of aphantasia, was MX's normal performance on a range of visuo-spatial tests, typically thought to require conscious experience of mental imagery[16].

In response to the publication of MX's case in the popular press (Zimmer 2010), a growing number of people have reported recognising MX's condition as a lifelong, as opposed to acquired, impairment[17]. In the follow up study, Zeman and colleagues examine and describe the main neurobehavioural features of this congenital, life-long form of aphantasia (2015, 2016). Studying a group of 21 subjects who reported similar imagery deficits to MX's, the features of aphantasia were elicited via completion of the well documented Vividness of Visual Imagery Questionnaire (VVIQ)[18]. In addition to the expected finding that aphantasic subjects rated significantly lower on VVIQ scores compared 121 controls (see figure 1), the main finding was that aphantasic subjects, whilst reporting a sustained incapacity for imagery generation, nevertheless reported experience of *involuntary* mental imagery. This was evidenced both through the reported experience of wakeful involuntary flashbacks, such as

[13]Previous documentation of aphantasia include Galton (1880), Charcot (1889), James (1890) Faw (2009) and Solms (1997) who refers to a similar condition known as '*irreminiscence'*.

[14] Aphantasia thus understood is an imagery generation disorder (Farah 1984) distinguishable from visual memory disorders and visual agnosias (Zeman et al. 2015).

[15]A larger systematic study of the experimental and neuropsychological features is currently being undertaken which will add to this data (Zeman, correspondence).

[16]Taking an explanation of so-called 'blind imagination' as its focus, the study found that MX adopted different cognitive strategies from control subjects when performing these tasks. This has direct consequences for the long-standing 'imagery' debate (see Zeman et al. 2015;153) which I do not discuss here.

[17] Zeman et al. (2018) report 11,000 individual cases of extreme imagination (including aphantasia and *hyper*aphantasia - characterised by comparably high VVIQ scores, see below).

[18] Marks (1973). See the supplementary materials online http://dx.doi.org/10.1016/j.cortex.2015.05.019. Aphantasia is operationally defined as performance of 16/81 on this scale (Zeman, correspondence).

unwanted visual images or earworms (48%) but also, crucially, in the dream reports of aphantasic subjects, 81% of whom reported *rich visual dreams*[19]. This led the authors to the following conclusion: that there is a *"significant dissociation* between voluntary and involuntary forms of mental imagery (p<.01 McNemar Test)" (379, italics added). This is reflected in the author's previous discussion of patient MX who, while initially reporting a permanent loss of imagery generation, nevertheless regained the capacity to experience involuntary flashbacks as well as visual dream content within the same time period (2010;146-7). Here, Zeman et al. make a similar conjecture:

"The initial loss of and subsequent recovery of dreaming in MX while voluntary imagery remained impaired, suggest that these two forms of imagery involve partially but incompletely overlapping neural networks, which is plausible in both cognitive and neural terms" (154;).

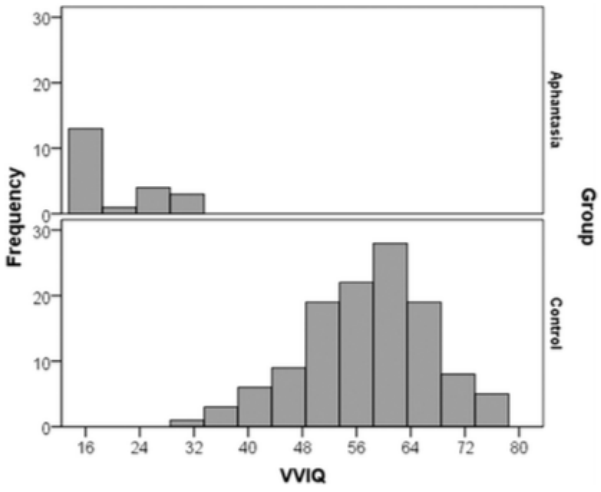

Figure 1: Distribution of VVIQ scores among aphantasic subjects and control participants (Zeman et al. 2015; 379).

5. The Case Against Ichikawa's Imagination Model

As previously stated, Ichikawa's imagination model of dreaming is made up of three key components which jointly amount to the claim that dreams involve instances of agential sensory mental imagery as essential and constitutive components. This commits the proponent of Ichikawa's imagination model to the following: that to the extent that any given dream has visual content at all, this can be accounted for solely in terms of such agential sensory mental imagery[20]. In light of this, the problem

[19] Other notable findings include reported difficulties with autobiographical memory (68%) and low effects on mood.

[20] That is not to say that Ichikawa isn't entitled to claim that there are non-visual (tactile, auditory etc.) forms of mental imagery involved in dreams, nor that there cannot be dreams devoid of visual content. This is only to say that, to the extent that they *do,* these can be accounted for solely in terms of visual mental imagery generation. If the visual content were to be accounted for only partly by imagery and in part by other kinds of experience, this would result in a version of a pluralist thesis. I take it that, as a variety of a

that dream reports of aphantasic subjects raise for this view is straightforward. Faced with a description of aphantasic subjects - characterised in the congenital form as lacking the capacity to generate visual mental imagery in any agential capacity from birth - an imagination model of dreaming as outlined above would predict that these subjects would have dreams with *little or no visual content*. That is, if aphantasic subjects are correctly described as lacking the capacity to experience visual mental imagery when waking, and we grant Ichikawa's claims that (i), per imagery, visual dream content is to be understood as an instance of this kind of wake-type experience and (ii) per the methodological assumption raised by the psychological case, that this ought to be reflected in empirical case studies - then we should expect that subjects with congenital aphantasia would report similar deficits with regard to the visual imagery experienced when dreaming. However, the findings of the current aphantasia studies go against this, providing evidence of multiple cases in which subjects report a sustained, lifelong *loss* of waking imagery whilst the capacity for rich visual dreaming is retained. The preliminary studies thus seem to disconfirm Ichikawa's theory, and in doing so, problematically lend support to alternative accounts which do not make such predictions (such as the hallucination model of dreaming) which the imagination model was meant to replace.

A similar pattern emerges on examination of acquired aphantasia in patient MX. Whilst initially MX's dream reports seem to accord with the predictions made by the imagination model viz. that a sustained loss in the generation and experience of visual mental imagery would be accompanied by a parallel reduction in visual dream content, the long term finding that MX r*egained* the capacity for visual dreams and involuntary imagery *without* also regaining the former (in conjunction with the dream reports of congenital subjects) provides an explanatory burden for the proponent of the imagination model of dreaming: If the imagination model of dreaming is correct, why is it that aphantasic subjects, who lack the capacity to generate and exercise agency over their visual mental imagery, are able to have dreams with visual content?

The empirical challenge that individuals with aphantasia and their dream reports raise for Ichikawa's imagination model relies on the acceptance of four broad claims:

(1) Subjects with aphantasia lack the agential capacity to generate and consciously experience sensory mental imagery.

(2) Visual imagery in dreams is to be understood and accounted for solely in terms of imagery of this kind.

(3) Dream reports accurately reflect dream experience such that aphantasic dreams are correctly described as having visual content.

(4) An adequate ontology of dreams must have the resources to account for empirical considerations similar to those raised by the dream reports of aphantasics.

'reductive' account of dreams (Rosen 2012), the proponent of the imagination model is opposed to this latter claim.

If the challenge - namely, that recent empirical studies provide psychological consideration better accommodated by rival accounts of dreams - is to be deflected then one of these claims must be rejected.

A full assessment of the arguments for and against these claims, and the underlying methodological principles these depend upon, goes well beyond the scope of this paper. However, (1) reflects an uncontroversial claim regarding what the VVIQ studies of subjects with aphantasia confirm, along with a plausible assumption about the general nature of introspection (a claim which, it should be stressed, is relied upon in much of contemporary consciousness science)[21]. (2) as discussed, follows from the endorsement of Ichikawa's imagination model of dreaming. This leaves the proponent of Ichikawa's imagination model with the possibility of rejecting claims (3) or (4). In addition to the fact that both enjoy broad acceptance in the surrounding literature, and thus whose rejection would require significant independent argument, the more pertinent problem for Ichikawa when faced with the prospect of rejecting these claims is that both are relied upon in the existing psychological and - in the case of (3) - *conceptual* arguments Ichikawa utilises in order to motivate the position to begin with[22]. As discussed at the end of 3.2, Ichikawa's endorsement of his psychological case conveys a commitment to the claim that it is an expectation of philosophical accounts of dreams that they are supported by, or at least able to accommodate, empirical cases like these. As such, the rejection of (3) or (4) comes at a great cost to Ichikawa, in removing or significantly weakening the case for the position that we started with. Collectively, the discussion here can be seen as resulting in the following dilemma for the proponent of the imagination model: to either claim that these kinds of empirical considerations are irrelevant to discussions of dream ontology - on the basis of denying either (3) or (4) - and thus abandon the previous psychological case in favour of the model, or grant instead that aphantasia constitutes a compelling empirical case which alternative models of dreaming can better accommodate.

Could a proponent of Ichikawa's theory not object here that (2) does *not* follow from Ichikawa's theory, given his claim that dream imagery results from the exercise of a type of mental agency which does not require voluntary control? That is, that the imagination model is *not* committed to the claim that visual dream mentation is to be accounted for solely in terms of the sort of mental imagery which is lacking in the aphantasaic case? The lack of elaboration and defense of this account of mental agency

[21] The claim also has empirically plausibility - see Pearson & Keogh (2018). Pearson & Keogh attempt to refute an alternative explanation for congenital aphantasia according to which aphantasia results from poor metacognition as opposed to a visual imagery deficit (that is, that aphantasic subjects *do* have visual images in their minds but are in some sense 'blind' to them). They conclude that congenital aphantasia *is* correctly characterised as a disorder of visual/sensory mental imagery, and I grant this here.

[22] See Windt (2015, 2013, 2010) and Windt & Metzinger (2007) for an independent defense of the reliability of dream reports in ideal conditions. Denial of (4) is also independently implausible for several reasons, the most notable of which concerns the recent debate over the methodology of metaphysics and its relationship to science (see Paul 2012). On this view, accommodation of empirical considerations would not be a mere or optional *virtue* of competing theories of dreams but is in fact demanded by acceptance of the claim that the metaphysics of mind proceeds via inference to the best explanation.

as that which is subject to the will - and the criterion of 'making sense to try to banish' which is meant to account for it - makes it difficult to respond to this line of argument[23]. However, an objection to this line of response can be raised which targets the independent plausibility of Ichikawa's voluntary-involuntary agency distinction and its ability to secure an agentialist theory of imagination (and thus, the appeal to this distinction in response to this challenge). This results from the following: any agentialist account which aims to accommodate dreams and involuntary imaginative episodes requires a notion of agency which is (i) *weak* enough to not rule out dreams and involuntary imaginings as *bona fide* instances of imagination and (ii) is simultaneously *robust* such as to be sufficient to distinguish involuntary imaginings from other typical non-agential mental phenomena (perception, hallucinations etc.). The worry then, simply stated, is that it is not obvious that Ichikawa is successful in providing such an account. That is, one might think that it is *neither* clear that the voluntary-involuntary distinction is weak enough to secure dreams within its purview (how plausible is it, for example, that dreams are really things that we *do* involuntarily?) nor that the 'making sense to attempt to banish' criterion is *sufficient* to distinguish between instances of involuntary imagination from other mental phenomena (does it not similarly make sense to attempt to banish *hallucinatory* experiences?). As such, the objection goes, appeal to this account here is unwarranted. In the absence of further elucidation and defense of this notion of agency, Ichikawa's claim collapses into the thesis described in (2)[24].

The foregoing argument can thus be seen as presenting a dilemma for proponents of Ichikawa's theory, resulting from a distinction between a strong and weak view of the relationship between mental agency and visual imagery in dreams. That is, either the proponent of the imagination model of dreaming can claim that notion of agency at play here is *strong* - being comparable to that which is lacking in subjects with aphantasia - which leaves Ichikawa vulnerable to the empirical challenge, or the relevant notion of agency is *weak*. While this deflationary analysis may avoid the empirical challenge, this account - grounded in the notion of agency as that which is 'subject to the will' - leaves the proponent of Ichikawa's view of dreaming with a theory of imagination which lacks independent plausibility; being unable to satisfy the independent requirements on an adequate philosophical theory of imagination. Either way, the argument goes, this model of dreaming ought to be rejected[25].

---

[23]See Ichikawa (2009; 107, 116).

[24]I return to this objection in the final section, where I suggest an empirical means of testing its viability. However, it is worth noting here that one might think that concerns remain *even if* a relevant notion of agency can be conceptually secured. For example, one might argue that the "significant dissociation between voluntary and involuntary forms of mental imagery" Zeman et al. report (379;2015) places significant pressure on Ichikawa's claim that involuntary forms of imagery *shares and inherits* the agential nature of the voluntary imaginings which drives agential accounts of imagination. At the very least, this raises a question for Ichikawa: if recent empirical data pushes us to accept a view on which voluntary and involuntary imagery are, to a large extent, *neurally distinct*, why think that the claim that involuntary imagination shares the agential nature of voluntary imagination remains a plausible one?

[25] Here the following thought may arise. It is commonly held that the capacity for visual dreaming can be preserved in blind subjects if eyesight is lost after the age of 7 years. Does adherence to this argument against Ichikawa's model of dreaming also commit one to an analogous argument with respect to hallucination models of dreaming - that is, commit one to the claim that empirical findings about the dream reports of blind subjects similarly threaten theories which state that dreams are essentially percept-involving

6. An Inactive Imagination Model of Dreaming

Are there other alternatives open to the proponent of the imagination model of dreaming? I have argued that the dream reports of aphantasic subjects pose a problem for *Ichikawa's* model of dreaming insofar as Ichikawa is committed to the claim (2) that dreams essentially involve, or constitute instances of, experiences of agential forms of sensory mental imagery. That is, visual mental imagery which is in an important sense under our agential control. This follows from Ichikawa's characterisation and individuation of imagination more broadly, as that which is necessarily *subject to the will*. Recent literature however suggests that Ichikawa's formulation is not the only way of characterising the imagination model of dreaming. Here, I argue that the best response to the case studies presented in section 2 for a proponent of an imagination model of dreaming is to reject thesis (2) above in favour of a modified claim. This allows the proponent of this model to avoid the dilemma just presented and retain the psychological case Ichikawa outlines in favour of an imagination model of dreaming[26].

Motivated by the need to provide a philosophical analysis of the state of *wakeful consciousness,* as an explanatorily preeminent condition relative to non-wake states*,* the alternative account of dreaming receiving attention in recent literature starts from an analysis of the wakeful condition as necessitating a kind of practical self-knowledge which results from the exercise of autonomous *mental agency* (O'Shaughnessy 2000, 2002)[27]. According to this analysis, the constitutive difference between wakeful and non-wakeful conditions (in this case, dreaming) is given in terms of the (in)capacity to *exercise agency over our mental lives* - that is, a capacity or ability to perform mental actions - which, it is claimed, is responsible for enabling the distinctive and immediate sort of self-knowledge (knowledge

hallucinations? This is an interesting line of thought. It does, however, elicit an obvious line of response on behalf of the hallucination theorist, which suggests a negative answer to this question. This takes as its basis the well-known distinction between 'perception' (as implicated in blindness) and the wider category of 'visual experience'. Here, the thought is that, *unlike* the case of aphantasia and sensory imagination, the fact that one's capacity to visually perceive has been disabled (in blindness) does not entail that the subject in question cannot undergo *visual experiences of any sort,* because these are, crucially, two distinct phenomena. That is, there are percept-involving visual experiences - namely, *hallucinations* - which are not (according to most theories) perceptions. The hallucination theorist can thus accommodate this finding via a line of argument which is unavailable to the imagination theorist, as there is no obvious, analogous distinction in the case of sensory imagination and aphantasia. Thank you to an anonymous reviewer for bringing my attention to this issue.

[26] A note of clarification: the suggestion in this paper is not that this inactive imagination model of dreaming is incontrovertibly *correct*. That is, I do not take myself to present an adequate defense of this view here. Rather, the claim is that there is - pending further work (and examination of the arguments for it presented in the cited articles below) - a viable response to the empirical challenge which can be taken up by the proponent of the imagination model.

[27] See O'Shaughnessy (2000), Crowther (2018) for an explanation of this connection in terms of the 'internal intelligibility' of the stream of consciousness.

of 'what we are up to') characteristic of wakeful consciousness (Soteriou 2017, Crowther 2018)[28]. In O'Shaughnessy's words (2002;426, see also Soteriou 2017) the suggestion here is that "Consciousness [wakeful consciousness] necessitates an overall *mental activeness*, for the reason that the conscious [i.e., the awake] are in control of the overall movement of their own minds, and the dream is an essentially inactive phenomenon" (emphasis added).

How does this help the imagination theorist with respect to dreaming? From this analysis of wakeful consciousness, a general claim regarding the nature of dreaming follows: that dreaming involves the temporary incapacitation of our ability to exercise mental agency. It is thus an *'essentially inactive'* phenomenon. When combined with (i) imagery, this provides us with the following:

(2*) dreaming essentially involves instances of passive, non-agential forms of imagination[29].

This is usually combined with the further claim that the episodes of inactive or non-voluntary imagining constitutive of dreaming are characterised by an *appearance* or illusion of being mentally active. Soteriou (2017) sums the proposal as follows: "In short, [in the dream state] one suffers from a sort of mental paralysis - a form of mental paralysis that doesn't even allow for the possibility of failed attempts to exercise agency over one's thinking. But it is a form of mental paralysis that is accompanied by the *illusion of agency* - the illusion of seeming to affirm, seeming to judge seeming to decide, seeming to be mentally active (13; emphasis added).

Insofar as this model rejects Ichikawa's characterisation of imagination as necessarily subject to the will, the alternative account of the imagination model of dreaming is not committed to the central claim of Ichikawa's account viz. that dreams constitutively involve the same form of *agential* imagination which is lacking in the aphantasia subjects. Given that on this model, the capacity to imagine and the capacity to exercise agency over our mental lives comes apart, it allows for and, to the extent that it defines dreaming in terms of this incapacity for mental agency, *predicts* that agents who lack the capacity to perform mental acts with regard to visual imagination (namely, aphantasic subjects) may nevertheless retain the capacity to visually dream. The implication of this claim should be obvious: on this variation of the imagination model - where (2) is replaced with (2*) - the empirical challenge raised by aphantasia fails to hold.

---

[28] This is meant to be neutral regarding whether this claim is to be accounted for solely in terms of dispositions of agency or broadly in terms of differences in general capacitation not limited to mental agency (Crowther 2018). The latter claim is all that is required.

[29] See the aforementioned articles for discussion of this account, which is proposed as a theory of *non-lucid* dreams. That is, in order to account for lucid dreams (see later) this view is usually accompanied by the claim that dreaming - and sleep mentation generally - is a *non-unitary* phenomenon which allows for hybrid states of consciousness (Voss & Hobson 2014, Howhy and Bayne 2016). That is, they claim that "during lucid dreams, the dreaming subject becomes aware that the events she is imagining are not real, and this is *precisely because* at that point during sleep the subject's ability to exercise agency over her mental life is reinstated, albeit in a limited, degraded form" (Soteriou 2017;12). I leave it open whether this argument is ultimately successful.

As this amounts to a rejection of Ichikawa's (iii) claim - namely, his agentialist theory of imagination - proponents of the inactive imagination model need to supplement this account of dreaming with an alternative theory of imagination which, as stated earlier, can adequately account for the various desiderata on contemporary philosophical accounts of imagination[30]. While the foregoing argument depends on the availability of such an account, it also serves a broader purpose - highlighting the role and dialectical importance of philosophical theories of imagination in the recent debate over dreams which has previously been under-appreciated.

7. Suggestions For Future Research

7.1 Extreme Imagination and Competing Theories of Dreams

Thus far, I have been considering the implications that the dream reports of subjects with aphantasia have for Ichikawa's agentialist imagination theory of dreaming. But what, if any, is the broader theoretical significance of these findings for the philosophy and science of consciousness? This section includes a discussion of two proposals which serves to situate the recent studies of aphantasia in the wider context. Turning first to the implications of the foregoing discussion of the philosophy of mind and consciousness. Assessing these requires us to answer the following questions: first, how does the empirical challenge I present against an active imagination model and in favour of *in*active account of dreaming relate to the previous empirical considerations Ichikawa offers in support of an imagination model of dreaming? Are the results from the aphantasia studies and those from Solms' and Foulkes' studies on dream development and dream pathology congruous - that is, do they, when considered together, provide a strong case for an inactive imagination model of dreaming, or are they instead suggestive of a more disparate evidential picture? Second and relatedly, we are now in a position to ask how the empirical considerations raised by aphantasia relate to alternative theses concerning the ontological constitution of dreams. In the previous section, I argued that the empirical argument based on dream reports of aphantasia could be *offset* if an inactive imagination theory of dreaming was endorsed, but can this empirical consideration do any further positive work in this context? In other words, what does the finding that aphantasic subjects have visual dreams mean for competing analyses of dreams?

At first glance, there seems to be a tension between Ichikawa's psychological case for the imagination model and the dream reports of aphantasics discussed here. This arises as a result of Solms' clinico-anatomical study of dream pathologies described in section 3.2. There, to recall, Solms reported that patients with dream pathologies in which visual dream imagery was lost describe *precisely analogous deficits* in waking visual imagery (Solms 1997;93-102, Ichikawa 2016;254). Insofar as the dream reports of aphantasics constitute multiple cases in which waking visual imagery is lost while dream

[30] Recent literature suggests that this ought to be supplemented by a dependency theory of imagination, which provides an analysis of imagination in terms of their phenomenal resemblance and ontic dependency on perceptual experiences (Martin (2002), Peacocke (1985) Soteriou (2013) cf. Nanay 2016, Nordhoff 2002). However, it seems that an epistemological account of imagination (O'Shaughnessy 2000, Sartre 1940) would serve the imagination theorist just as well here.

visual imagery is *retained*, these results appear to be in tension with this new data; thus, when considered collectively, these appear to be suggestive of a disparate evidential picture with respect to the thesis that dreams constitutively involve instances of sensory imagination.

This initial impression is, however, misleading. Interestingly - and unacknowledged by Ichikawa - out of the effective sample of 186 patients Solms examined, only *two* reported a cessation or restriction of visual dream imagery (and the parallel loss of visual waking imagery), leading Solms to the early conclusion that the presence of this dream disorder in the neurological population was very low (1.1%). The small number of patients examined constitute an obvious, and previously unnoted, limitation to Ichikawa's psychological argument. Nonetheless, this feature of the study provides a means of rendering Solms' results compatible with Zeman et al.'s finding that *81%* of the aphatnasics studied report rich visual dreams. That is, Zeman et al.'s finding here is consistent with the correlational data Solms reports in his study, in which it is acknowledged that this applies only to a small number of patients. Instead of opposing Solms' research findings, the dream report data of aphantasics can thus be seen as a welcome addition to Solms' early results on visual imagery dream disorders and its relation to waking visual imagery deficiencies, demonstrating that the correlational pattern Solms describes is characteristic of *exceptions* to aphantasic cases and not the norm[31].

Turning now to the second question. What significance, if any, does the finding that aphantasic subjects have visual dreams have for competing analyses of dreams? A number of more comprehensive studies on aphantasia and other forms of extreme imagination are currently underway. Thus far, these studies on extreme imagination are suggestive of several notable findings. Most importantly, they reveal that a positive correlation holds between both (i) the absence of waking visual imagery and loss of dream imagery and (ii) the presence of hyperaphantasia and particularly *vivid* visual dreams, compared to normal controls[32]. When considered in conjunction with the main finding of Zeman et al. discussed in this paper, this places an explanatory burden on alternative prominent theories of dreams - such as the hallucination (Hobson 1988) and recent sui-generis theories (Windt 2010, 2015): why, if there is *no tight ontological connection* between dream mentation and sensory imagination, as these theories maintain, are subjects with aphantasia more likely than those with average or vivid imagery to report absence or reduction of visual dreaming? Answering this question and collecting further data which sheds light on its viability, is the next obvious step for philosophers working on dreams who aim to provide satisfactory and empirically adequate solutions to the conceptualisation problem.

---

[31] What explains these cases? Should we not expect that all aphantasics lack visual dream imagery? The imagination proponent owes us an explanation here. One suggestion - open to future empirical confirmation - is that the terms 'aphantasia' and 'hyperaphantasia' do not represent binary categories but are best understood as poles on an imagery vividness continuum (a hypothesis supported by McKelvie 1995 as well as a recent study on individual differences in aphantasic subjects from Dawes et al. 2020). The suggestion then is that those aphantasics reporting no visual dream imagery fall at the extreme end of this spectrum, while those reporting visual dream imagery correspond to more 'moderate' forms of aphantasia.

[32] Zeman (correspondence).

7.2 Aphantasia and Lucid Dreaming: A Testable Empirical Hypothesis

An obvious datum for any conceptual theory of dreaming, and one which is pertinent to the issues discussed in this paper, is the phenomenon of *lucid dreaming* - dreams in which subjects become aware that they are dreaming and, in some cases, are able to manipulate and control their dream content (La Berge 2004, Voss et al. 2009). Insofar as they provide opportunities in which subjects can reflect, introspect and subsequently report the content of their dreams in greater detail, experimental paradigms investigating lucid dreams continue to be at the forefront of empirical dream research[33]. As an *agentive* state of consciousness, in which a subject's capacity for mental agency is clearly retained in some form, lucid dreams also play an important role in motivating Ichikawa's agentialist account of the imagination theory and serve as an obvious counterexample to the inactive imagination model of dreaming[34]. In addition to this initial role however, the rediscovery of aphantasia presents lucid dream research with a further empirical function. When lucid dreaming is considered in the context of the recent studies of aphantasics and their visual dream reports, the following question arises: can subjects with aphantasia *also* lucid dream in a visual sense? Currently, no cases of lucid dreaming in subjects with aphantasia have been reported[35]. However, the availability and empirical examination of such cases could greatly improve, - or, alternatively, weaken - the case in favour of the inactive imagination model of dreaming presented in this paper. That is, if it found that aphantasic subjects generally report an *inability* to lucid dream this would - on the assumption that the capacity to (or otherwise learn how to) lucid dreaming is generally available in typical samples of the population - lend empirical support to the thesis that dreams constitutively involve inactive forms of imagination[36].

---

[33]Voss et al. (2018), Windt and Voss (2018) Saunders et al. (2016) include recent meta-analyses and summaries of this work. Here, lucid dreams are characterised by the three dissociable features: (i) the presence of subject insight (a conscious awareness that one is dreaming) (ii) control (manipulation of dream content and narrative) and dissociation from one's dream experience (such as seeing the dream as if it were playing on a screen or viewing oneself from the outside).

[34] Ichikawa (2009;116-117). See footnote 28 for a discussion of lucid dreaming and the inactive account of dreaming which involves the hypothesis that lucid dreaming is a hybrid phenomenon (Voss & Hobson 2018, Soteriou 2017).

[35] Zeman (correspondence). It would also be useful to conduct further dream studies of aphantasics who don't lucid dream. That is, the following question also appears to be empirically testable: do subjects with aphantasia report an ability to attempt to banish - in Ichikawa's sense - their visual imagery whilst dreaming? This would, if the finding is negative, lend greater support to the argument presented above. Whilst research on the phenomenology and sense of agency in dreams is currently in early stages (see Rosen 2015 for a review), this presents a promising line of work for future research.

[36] Early support for this claim is found in a recent study on aphantasia from Dawes et al. 2020, who report that 'aphantasic individuals also reported experiencing lower awareness and control during their dreams [compared to control subjects]' (6). The most relevant phenomenon here of course are lucid dreams in which the subject is able to *control* and manipulate her visual dream imagery to some degree (whether or not lucid dreams where insight is present but not control could be included in this sort of study will largely depend upon the - currently open question - of know exactly insight should be understood, especially with respect to mental agency - Voss 2014, Kuhle 2014). While control of this kind is less frequent in the 'naturally occurring' lucid dreams found mostly in adolescents (Voss et al. 2012, Voss and Hobson 2014) there is currently good evidence that control-involving lucid dreaming in adults is a learnable skill or ability which can be improved with practice (LaBerge 2000, Schredl & Erlacher 2011, Kuhle 2014). The proposal I am making would thus require researchers to

Conversely, if aphantasia subjects generally *retain* the capacity for lucid dreaming (relative to its uptake and presence in control groups) - and those lucid dreams contain visual content - the inactive imagination model would - on the assumption that dreams *are homogeneous* in nature- be empirically disconfirmed. This potential to facilitate further empirical work, along with the philosophical applications discussed in the previous section imbues aphantasia with a wider theoretical significance in the context of the conceptualisation problem which goes beyond the repudiation of Ichikawa's theory of dreaming and provides philosophers and psychologists with an exciting new avenue of dream research.

compare the ability to *uptake* this form of lucid dreaming in a large sample of aphanatasics, and compare this to the percentage of people able to learn this skill in the general population (that is, my proposal here is consistent with the claim that a proportion of aphantasics and non-aphantasic subjects will not exercise lucid dream control for reasons which have nothing to do with Ichikawa's hypothesis). Thank you to an anonymous referee for prompting me to clarify this proposal.

2| Ameliorative Physicalism and the Natural Kind Approach to Consciousness

The most compelling route to physicalism currently is to adopt a version of the phenomenal concept strategy. Concept strategists have thus far been engaged in a descriptive project vis-a-vis phenomenal concepts. The aim has been to draw attention to features that our phenomenal concepts currently possess, to block the inference from epistemic to metaphysical dualism highlighted in anti-physicalist arguments. This has spurned various attempts to formulate an adequate account of phenomenal concepts which can deliver on this promise. This paper outlines a new route to physicalism which removes the requirement for a defensible formulation of the concept strategy. Inspired by work on conceptual engineering, I argue it is time for physicalists to move away from descriptive accounts of phenomenal concepts to normative analyses which specify how our concepts of consciousness ought to be refined for the purposes of facilitating a mature science of consciousness. After motivating the claim that our existing concept is epistemically defective for this purpose, I propose that we explicate our concept of consciousness as a natural kind concept which refers to the homeostatic property cluster(s) of physical properties associated with our initial samples of consciousness. I argue that this view - *Ameliorative Physicalism* - works to compliment and motivate recent calls to adopt a 'natural kind approach' to consciousness science, providing it with conceptual resources to overcome several objections levied against it.

> "The task of making more exact a vague or not quite exact concept used in everyday life or in an earlier stage of scientific or logical development, or rather of replacing it by a newly constructed, more exact concept, belongs among the most important tasks of logical analysis and logical construction. We call this the task of explicating, or of giving an explication for, the earlier concept . . ."
>
> — Rudolph Carnap (1947: 8)

1. Introduction

The phenomenal concept strategy is an argumentative strategy developed for the purpose of defending physicalism against the knowledge, conceivability, and explanatory gap arguments (Jackson 1986;

Chalmers 2010; Levine 1983). Take any conscious experience C, such as the phenomenal experience of redness, these arguments proceed as follows:

1. There is an epistemic gap between physical and phenomenal descriptions of C.

2. If there is an epistemic gap between physical and phenomenal descriptions of C, then there is a metaphysical gap between physical and phenomenal properties.

3. If there is a metaphysical gap between physical and phenomenal properties, physicalism is false.

Therefore,

(C) Physicalism is false.

The phenomenal concept strategy targets premise (2) of this argument. The well-known thought is simple: while these anti-physicalist arguments may have succeeded in highlighting a gap between our *conceptions* of consciousness, there is no need to think this epistemic gap is indicative of an ontological gap between physical and phenomenal properties, for this may arise as a result of an inferential gap between phenomenal and physical *concepts.* Given that this inferential gap can be given a satisfactory explanation in physical terms, the argument goes, physicalism remains unchallenged. The dualism here, the concept strategist asserts, is one of concepts and not of metaphysics (Loar 1997, Papineau 2002, Díaz-León 2010, Balog 2012).

As is also well known, in order to be successful this strategy requires further supplementation with a positive physicalist account of phenomenal concepts which can explain the lack of inferential, a priori entailment from physical concepts to phenomenal concepts. This has been explored at length in the literature, with philosophers arguing that the distinctive nature of phenomenal concepts is grounded in their being, for example, recognitional (Loar 1999, 2003), quotational (Papineau 2002), constitutional (Balog 2009), or self-preserving concepts (Sturgeon 2000). This descriptive project has recently been revived in light of new objections to these versions of the concept strategy (Goff 2011, 2017, Holman 2013, Veillet 2015) which has in turn spurred the development of new and revised accounts of phenomenal concepts (Taylor 2018, Levin 2019). Despite their differences, these accounts each share a common methodological commitment with respect to phenomenal concepts: viz. that the primary aim here is to describe the nature of phenomenal concepts and account for the distinctive cognitive features they currently and in fact possess. This leaves the phenomenal concept strategist vulnerable to a variety of objections concerning the inability of such accounts to account for the way phenomenal concepts actually are, and what they appear to do (Goff 2017). It also leaves the physicalist with a difficult question with respect to the *science* of consciousness which often goes under-appreciated. For, if we are stuck with these distinctive ways of conceptualizing consciousness, how then should we go about studying the nature and neural basis of consciousness empirically, given that these are inferentially isolated from physical and functional concepts?

In this paper, I will suggest that physicalists sympathetic to the concept strategy (so-called 'type-B' physicalists (Chalmers 2006)) needn't concern themselves with this descriptive question, and that another, more attractive, route to physicalism exists on which these descriptive accounts of phenomenal concepts are no longer so important. This is inspired by the emerging literature on conceptual engineering as a methodology for achieving epistemic progress. This asserts that such progress can - and indeed often has been – achieved in philosophy not only via forms of conceptual analysis, but also via the provision of rigorous *normative* analyses of how our concepts can and ought to be refined and improved to best fit our practical and epistemic goals. The growing literature on conceptual engineering opens up new types of research questions in the vicinity of phenomenal concepts which have yet to be properly explored — namely, are our phenomenal concepts well suited to the empirical and philosophical projects in which they appear? and, if they are not, how can our concepts of consciousness be improved so as to better suit these aims? This paper provides the first examination of this new physicalist strategy, providing a new argument in favour of the strategy, as well as a positive proposal about what kind of concept ought to replace our existing concept of consciousness[37]. Together, these form a new research programme for empirically informed physicalists. Just as Sally Haslanger (2000) proposed that we should aim to develop accounts of race and gender that will be effective in the fight against injustice, this programme proposes that physicalists should aim to develop accounts of consciousness which will be effective in facilitating a mature science of consciousness.

The plan is as follows. The following section outlines Rudolph Carnap's method of explication for scientifically defective concepts. This serves as a guiding framework for the proposal is this paper (**Section 2**). Drawing on well-known methodological work from consciousness science and animal sentience research, **Section 3** provides an argument for the claim that our concept of consciousness is epistemically defective and actively works to hinder scientific research by underwriting intractable methodological disputes. This generates an interim conclusion that we have warrant to conceptually re-engineer our phenomenal concepts for scientific purposes. In **Sections 4 and 5,** I argue that these epistemic defects can be rectified if we re-engineer our concept of consciousness as a *natural kind concept* which picks out a homeostatic cluster(s) of physical properties. I argue this can be situated within and used to motivate recent calls for the adoption of a 'natural kind approach' to consciousness which has been proposed as a solution to these methodological problems (Shea and Bayne 2010, 2020, Shea 2012). A pressing question for existing proposals of this kind concerns why we should think that adopting this approach is likely to be successful, if consciousness is not a natural kind, or otherwise not conceptualised in natural kind terms (Caruthers 2019, Shea and Bayne 2020). The explicative strategy here offers a new answer: the central claim of the natural kind approach is not just that we can and should study consciousness *as if* it were a natural kind, but rather that refining and explicating our concept of consciousness to a natural kind concept is epistemically *useful*, allowing us to achieve

[37] A recent publication published since writing (Schomaker 2021) also explores consciousness in the context of conceptual change. However, this does not explore or defend the that conceptual engineering can be used as a strategy for defending *physicalism*, nor the idea that we can use explication as a way of motivating the natural kind approach to consciousness and facilitating a mature science of consciousness. As such, I see these as together as providing first examination of the issues thrown up by phenomenal concepts in the context of conceptual change and conceptual engineering.

inductive success within the science of consciousness. Before considering an objection to this claim in **Section 7**, **Section 6** considers how this explicative strategy for motivating the natural kind approach to consciousness can work to overcome and alleviate a recent further methodological objection to existing formulations of the natural kind approach according to which the natural kind approach cannot solve the methodological puzzle of consciousness (Taylor 2021). I argue that this objection relies on a conceptualisation of consciousness as a non-natural phenomenon and thus, that if the version of the natural kind approach outlined here is taken up, this objection can be avoided. The position in this paper thus serves a dual function, offering a new metaphysical position with respect to consciousness which can do concrete methodological work in the foundations of consciousness science.

2. Carnapian Explication

'Conceptual engineering' or projects of 'ameliorative analysis' refer to the process of improving our conceptual and representational devices for empirical, theoretical and practical purposes (Cappelen 2018, Thomasson 2021). The central idea is that, in addition to analysing our concepts - projects which have preoccupied much of philosophical theorising in the last century (for example, Grice 1987, Strawson 1959, Beaney 2014, Margolis and Laurence 2019; S5) - philosophers can also be engaged in the provision of normative accounts of what concepts we *ought* to possess, given our practical and theoretical aims. From a conceptual engineering perspective, for example, the philosopher of truth can not only be interested in attempting to define, via the provision of necessary and sufficient conditions, our existing concept of truth, but also in asking what our concept of truth should *be like* given our aims and objectives (Eklund 2002, Scharp 2013). Thus described, conceptual engineering projects have two key components, which we can call the identification of c*onceptual defects* and the provision of *development strategies* for improving a concept, or creating a new one once a conceptual defect has been recognised.

Concepts can be defective in numerous ways[38]. An influential view holds that a central, and perhaps preeminent, sense in which our concepts are defective is in an epistemic sense; a claim which is grounded in the observation that many of our inherited concepts work to hinder theoretical and scientific inquiry in various ways (Cappelen 2018, Simion 2018). This idea is typically spelt out via the claim that many of our concepts are too vague for theoretical and scientific purposes (Carnap 1947) and/or that they otherwise work to support intractable empirical disputes or paradoxes (Cappelen 2018, Pinder 2020, Eklund 2020). This reflects a more general observation about the nature of concepts and their role in enquiry — namely, that some concepts are more useful for and better suited to some theoretical domains than other concepts. This in turn suggests that defective concepts with respect to a particular domain ought to be improved or replaced by new concepts which are better suited to achieving epistemic success within that domain.

[38] For taxonomies of conceptual defects, which include semantic, epistemic, cognitive, moral and practical deficiencies, see Cappelen (2018) and Chalmers (forthcoming). I return to the idea that phenomenal concepts may be defective in more than one sense briefly in §3.

Once a concept has been identified as conceptually defective, several 'improvement' strategies present themselves. Recent literature suggests that here one faces a choice between two options: to either 're-engineer' or improve an existing concept or create a new concept 'de novo' to replace the defective concept, a choice which typically coincides with a decision as to whether to retain the same lexical item (e.g., 'truth') or introduce a new one (Chalmers 2020). A popular strategy for re-habilitating epistemically defective concepts is what Herman Cappelen (2018) refers to an 'topic improving' engineering: the improvement and refinement of a concept for the purposes of advancing empirical enquiry. The method of 'explication' or 'explicating' concepts developed by Rudolph Carnap is a central example of this kind of engineering. In basic terms, Carnap's method of explication refers to the process of self-consciously translating a vague or inexact (that is, epistemically defective) concept known as the 'explicandum' into a more exact, epistemically useful concept - the 'explicatum' - for the purposes of solving scientific disputes and facilitating scientific progress. Here, the 'explicandum' is a 'folk' concept or a concept utilised in an earlier stage of scientific theorising, while the explicatum is a concept which can be demonstrated to better facilitate empirical progress and success. For Carnap, this latter claim involves demonstrating that the explicatum satisfies four theoretical criteria - (i) similarity, (ii) precision, (iii) frutility and (iv) simplicity (1947). More recent literature however suggests that an explication of this sort can be justified on more general epistemic grounds, by demonstrating that the explicatum significantly improves on — in the sense of unravelling or diffusing empirical disputes or paradoxes — the existing concept (Simion 2018, Cappelen 2018)[39].

Explication is common not only in philosophy. It is an established method for achieving epistemic progress in science. The history of science is littered with plausible examples of explication or topic-improving engineering (Jorem 2022). A good example here is in the refinement of the concept of 'planet' by the International Astronomical Union in 2006 (Magnus 2013; 1). Prior to this refinement, astronomy was in disarray due a series of empirical disputes relating to the proper extension of the term 'planet'. These resulted from the observation that, in addition to the long-recognised nine planets, there were a growing number of identified celestial bodies with similar properties (sizes, orbits and so on) to the existing nine planets. The debate came to a head with the discovery of Eris, a celestial body with near identical properties to the then-planet Pluto, which raised the following puzzle: should Eris be classified as a tenth Planet, a dwarf planet, or something else? (And, if it wasn't to be classified as planet, what grounds could one use to argue for this assertion given the astronomical properties it shared with the planet Pluto?). The result was a series of polarised disputes about classification in astronomy, which the then-concept of planet was too vague and imprecise to settle. The solution, announced in 2006 by the IAU, was the production of a new and refined definition of planet, which led to the exclusion of both Eris and Pluto. Crucially, the IAU's choice was not made arbitrarily, but rather on the basis that this new 're-engineered' concept of planet *maximised epistemic success* within astronomy, underwriting a greater number of predictions and successful inductive generalisations concerning celestial bodies. The explicated concept provided a clear set of criteria for the classification

[39] See Brun (2016) and Pinder (2020) for a discussion of how Carnap's theoretical criteria should be viewed in a contemporary context.

of celestial bodies and employment of the concept 'planet' which have been used successfully in astronomy since.

Given its prevalence in the history of science, Carnap's method of explication thus provides a useful guiding framework for the proposal outlined in this paper viz. the proposal that it is precisely this sort of project that physicalists - particularly, those concerned with furthering scientific progress with respect to consciousness - should be engaged in with respect to phenomenal concepts. This constitutes a significant departure from current descriptive work philosophers have been engaged in. For this to work, however, it needs to be the case not only that such an engineering project is coherent with respect to phenomenal concepts (see §7 below), but also that our concept of consciousness in fact warrants a process of explication; that is, that our concept of phenomenal consciousness is in fact *epistemically defective* in the sense that would place consciousness researchers in an analogous position to the astronomers in 2006. That we are in such a position may not strike one as obvious from within philosophy - there are disputes, but what makes these any *more* intractable than other philosophical debates? However, that our concept of consciousness *is* epistemically defective becomes salient once we step away from the metaphysical debates just mentioned and consider how phenomenal concepts feature in and are employed in scientific research. Or so I shall argue in the following section.

A word of clarification before I proceed. I do not anticipate or intend that the argument presented in the following section will be compelling to those who already accept the truth of premise (2) of the argument in the introduction and, by extension, non-physicalism. For they will agree with most of the steps in the argument but offer an alternative explanation for why our phenomenal concepts are defective in these empirical domains. That is, because phenomenal concepts stand in a special epistemic relation to their referent, the latter of which is not the sort of thing which can possibly facilitate inductive success within existing scientific frameworks (Goff 2017). As will become clear, I agree with the non-physicalist here that this is what our phenomenal concepts currently appear to be like (see §3.1). Although there may be way of re-framing this argument so that it targets non-physicalists, I do not take myself to be presenting this sort of case here[40]. Rather my primary target are those philosophers who are already committed to the denial of premise (2), and by extension, physicalism on a priori or a posteriori grounds (e.g. Papineau 2002). This also rules out those physicalists of a 'type-A' nature, who deny that our phenomenal concepts are inferentially isolated from physical concepts - for example, Hoerl (2015) and Eilan (2014) and Jackson (2007), Dennett

[40] In recent work, Tim Bayne and Nick Shea discuss a way of formulating the natural kind approach which is theory-neutral about the ontological dependency relationship between the natural kind or homeostatic property cluster(s) which underwrites consciousness, and consciousness itself (Shea and Bayne, 2020). One way of re-formulating the proposal here in this paper which remains metaphysically neutral then would thus be to formulate the explicatum concept in an analogously neutral way regarding the dependency relation between consciousness and the relevant homeostatic property cluster(s). While I am open to this possibility, I think this proposal has a disadvantage of being more susceptible to the methodological challenge I discuss in section 6.2, which the physicalist version described here does not. As such, in this paper I limit myself to the physicalist claim.

(1991), Conee (1994), Nemirow (1990) [41]. Here again, I do not take myself to be engaging directly with these sorts of views in this paper. The extent to which this strategy can be adopted by philosophers who hold these views will depend on the extent to which they disagree that phenomenal concepts are already natural kind concepts, in the relevant sense required to ground a mature science of consciousness [42].

3. The Epistemic Defect Argument

Recent literature on conceptual engineering suggests that a concept is epistemically defective if either (i), the concept is vague, inexact, or otherwise unfit for advancing theoretical or empirical inquiry, or (ii), the concept works to support an intractable dispute(s) or a paradox which the concept is ill-equipped to settle. Here, I'll argue that there are good, well documented reasons to think that the concept of consciousness currently satisfies both criteria in consciousness science. As such, I will claim, a case can be made for explication and conceptual engineering of this concept.

3.1. A 'Seemingly Wonderful Concept'

If you are not a type-A physicalist, then you will typically think that the concept of phenomenal consciousness, regardless of what analysis of phenomenal concepts you prefer, *seems* to refer to properties which are essentially subjective and qualitative and hence, non-functional, and non-informational. Phenomenal concepts are, as Eric Schwitzgebel (2016) observes 'seemingly wonderful': they are inferentially isolated from physical concepts and appear to refer to properties which are not

[41] One might also ask how the view in this paper relates to *illusionism* — the view that consciousness is an introspective illusion brought about by the misrepresentation of experience as having phenomenal properties (Pereboom 2011, Frankish 2016). I think this is an interesting question, the answer to which will ultimately depend on exactly what illusionism as a theory commits to (c.f. Frankish (2016), Garfield (2016), Humphrey (2011)). While I am open to the possibility of an illusionist formulation of the view I sketch here, there a number of clear differences between my proposal and illusionism as it is typically described (Frankish (2016)) which are worth pointing out. First, I am framing my proposal here as a *realist* proposal about consciousness which can be taken up by traditional type-B or reductive realist theorists. If the natural kind approach does in fact discover that there is a single property cluster associated with consciousness (see below for explanation), I take it this would work to ground a form of *scientific realism* about phenomenal consciousness which, prima facie, seems at odds with some formulations of illusionism. This is what I see as a benefit of the natural kind approach to consciousness — if it succeeds in identifying a single natural kind, realism about consciousness comes for free. Secondly, the physicalist view outlined here clearly has different implications for the science of consciousness than illusionism does. That is, the view I will motivate here does not entail that the science of consciousness ought to focus on solving the 'illusion problem' (Kammerer 2016, Frankish 2016, Prinz 2016) in replacement of research into the neural basis of consciousness — rather, it helps provides us with a strategy (Shea and Bayne 2010, 2020) for studying this. Thanks to Philip Goff for prompting me to clarify the relationship between my proposal and illusionism.

[42] Even if one thinks that there are systematic errors in the standard dialectic which frames the contemporary debate on phenomenal consciousness – for example, if one holds that the tension between objectivity and realism about consciousness anti-physicalists typically exploit is non actual, perhaps on account of the objectivity afforded by our spatial thought in relation to common sense realism about the external world (Eilan 2014) – one might still wish to take up the general line of argument in this paper. That is, one could endorse this alternative view of phenomenal concepts whilst also granting that the resulting concepts are nonetheless unsuited to grounding a mature *neuroscience of consciousness* – and thus, grant that engineering our concept of consciousness to a natural kind concept is epistemically worthwhile.

physically or scientifically explicable (Tartaglia 2016)[43]. This idea, which is endorsed by non-physicalists, eliminitvists, illusionists and phenomenal concept strategists alike (Papineau 2019, Frankish 2016), becomes salient when one examines recent discussions of phenomenal consciousness from a physicalist and scientific perspective:

"What could be the point of describing some state or property as phenomenally conscious if it didn't give rise to any hard-problem-type thought experiments? Indeed, it seems to us that it is the mark of phenomenal consciousness (as opposed to access-consciousness) that it naturally raises (or seems to raise) a set of hard epistemic problems - problems that other aspects of our mental lives don't give rise to." (Carruthers and Viellet 2017).

"These days most consciousness researchers would agree that the concept of consciousness does not, *a priori* at least, refer to a particular psychological function or behavior, but to the fact that we have subjective experiences such as the taste of good coffee, the sound of a dog barking, or frustrating thoughts about consciousness… all examples share one feature, that they are subjective. This core feature is what makes consciousness so scientifically challenging" (Overgaard 2015;8).

The fact that phenomenal concepts appear to refer to non-physical properties provides us with a well-known *pro tanto* reason for accepting the claim that our phenomenal concepts are epistemically defective in scientific domains, and so meet the criteria for (i)[44]. If a concept appears to refer to a set of properties which, by their very nature are resistant to standard forms of scientific explanation, we have good reason to expect that the concept will be unfit for empirical employment and facilitating inductive success within scientific domains. In other words, that we have good reason to think right from the outset that condition (i) above is satisfied, and that the concept of phenomenal consciousness is unfit for advancing empirical enquiry[45].

This intuitive claim remains true even if one maintains as a physicalist that this appearance is an *illusion*, or otherwise a result of distinctive and inferentially isolated phenomenal concepts. While the non-physicalist would assert that this appearance is actual and claim that the science of consciousness should be adjusted accordingly[46], the physicalist at the target of this paper typically grants that phenomenal consciousness at least *seems* to refer to these properties. This places them in a strange

[43] Frankish (2016) adds to this that the phenomenal consciousness also comes with further commitments to infallibility regarding our judgements and beliefs about it. I return to this idea as it features in possible objections to the natural kind approach below.

[44] This, of course, is the basic line of reasoning behind the hard problem of consciousness (Chalmers 1996). As below when I discuss the on-going methodological disputes in consciousness science, the aim here is not simply to rehearse standard arguments in this literature but rather to emphasise and examine how these well-known features of consciousness warrant a *new conclusion* about the concept of consciousness when viewed through the lens of explication and conceptual engineering.

[45] There is also a case to be made for the claim that the concept of consciousness is also *inexact* or vague, in the relevant sense as described in (i). One might appeal here to the difficulty of attempting to define phenomenal consciousness which often take the form of definition by example, or synonymy (Schwitzgebel 2016, Frankish 2016b). For an extreme example of this claim, see Mandik (2016) who argues that the concept of phenomenal consciousness is contentless and therefore meaningless.

[46] For example, via the claim that the role of current scientific theorising is limited to the identification of explanatorily empty *correlational data* between brain activity and phenomenal properties, Chalmers (2000, 2010,) Koch et al. (2016); c.f. Owen and Gupta (2019), Klein et al. (2019).

predicament vis-a-vis the science of consciousness which often goes under appreciated. For, while focusing on motivating accounts of phenomenal concepts which explain the inferential isolation of these concepts from physical concepts enables the physicalist to dissolve the force of anti-physicalist arguments, this strategy has the result that the concept of phenomenal consciousness is retained and carried over unchallenged to non-philosophical, empirical discussions of the neural basis of consciousness. In the next section I argue this is a problematic, insofar as this philosophical concept works to underpin substantive methodological disputes in consciousness science which ultimately hinders scientific progress.

3.2. Empirical Disputes

As is well known in consciousness science, the distinctive way we think about consciousness gives rise to a number of pressing foundational puzzles and methodological problems which hinder empirical progress[47]. Here, I'll summarise what I take to be the central methodological puzzle which concerns the observation and *measurement* of consciousness, and its manifestation in two specific problems relating to the detection of consciousness in cases where verbal report is lacking, and the possibility of testing for nomological dissociations between consciousness and conscious access. While a full exposition of these disputes and their significance is beyond the scope of this paper - the literature on these disputes are well known and have already been discussed in the context of the natural kind approach (Bayne and Shea 2010, 2020 Shea 2012)[48] - the emphasis here will be on demonstrating and explaining that these disputes, unlike others which arise within maturing research programmes, are especially problematic insofar as they are thought to hold in principle or result *a priori* from the nature of the concepts involved. For it is concepts implicated in disputes of this kind that most easily warrant explication.

There is a distinctive methodological puzzle which arises for those who attempt to study consciousness scientifically and design experimental paradigms which can elicit scientific knowledge of its neural basis. This is motivated by reflection on the nature of consciousness as revealed by our concept on the one hand, and the nature of neuroscientific enquiry on the other. The general idea can be expressed as follows: given that consciousness is (or at least appears to be) inherently subjective, private, and qualitative how is it possible to observe and measure consciousness using the third-personal, objective methods characteristic of neuroscience? In particular, how can we devise accurate and reliable third-person measures of consciousness, which go beyond subjective reports? In his discussion of methodological issues in consciousness science Overgaard (2015;4) notes that the intractability of this problem is one which holds in principle due to the subjective conception of consciousness:

[47] These are pressing not only from epistemic perspective but also a moral one, given the ethical implications of determining whether (as we shall see) coma patients and non-human animals are conscious. For a discussion of the moral significance of consciousness see Siewert (1998), Kahane and Savalescu (2009), Shepherd and Levy (forthcoming), Levy (2014) Lee (2019).

[48] In (2020) Shea and Bayne refer to this as the 'validation challenge' - referring to the challenge validating of third personal markers of consciousness.

''The problem shares the 'property of being hard' with the explanatory hard problem, as there seems to be no kind of knowledge or technique we can develop which would turn 'first-person observation' into 'third-person observation'. The limitations to observation (in the case of consciousness) are a matter *of principle*'.

According to a standard approach to measuring consciousness, and thereby attributing consciousness to an individual or mental state, when collating evidence for the presence of consciousness our tests for consciousness ought to be grounded in the sorts of standard functional criteria we use to attribute consciousness to each other in everyday cases (Shea and Bayne 2010). Here, the idea is that we ought to ascribe consciousness to a given individual only if we are able to find introspective evidence that an individual or mental state exemplifies a pre-theoretical sign or marker - typically forms of subjective report, evidence of volitional agency and so on - which is taken to be sufficient for the attribution of phenomenal consciousness in healthy waking adults and children[49]. Similarly, according to this standard approach, we should describe a particular mental state as conscious if it reportable in these ways.

So far, this presents consciousness science with a methodological puzzle[50]. What prompts the idea that this way of thinking about consciousness poses an *intractable dispute* is reflection on cases where the standard approach fails to provide an answer, cases in which pre-theoretical markers are absent or otherwise unavailable. The standard problem cases here, which are often morally salient, include infants (Kouider 2013, Trevarthen and Reddy 2017), patients in coma and persistent vegetative states (Levy 2009, Bayne and Shea 2010)[51], and non-human animals. While many would grant that mammals are conscious, heated debate has revolved around the possibility that fish, insects, cephalopods and crustaceans are conscious and can feel pain (Allen 2004, Edelman et al. 2005, Edelmen and Seth 2009,

[49] For example, consider the typical experimental set up in which a stimulus is flashed on a screen followed by a mask and the participant is asked to record the nature of a stimulus (e.g., a triangle, square etc.). These subjective reports, along with participants metacognitive judgements about their confidence in their initial report are often used to determine whether the participant was conscious of the stimuli or not.

[50] Of course, you might question whether even these standard introspective markers are reliable measures of consciousness. For example, you might worry, following Irvine (2013, 2019), Schwitzgebel (2012), Spener (2015, forthcoming) and others, that introspective markers of consciousness cannot be *calibrated* or validated in the sense required for a mature science of consciousness. Michel (forthcoming) includes a detailed refutation of these skeptical worries with respect to introspective measures of consciousness. This is grounded in the idea that introspective measures can be calibrated via a processing of modelling which compares actual outcomes with ideal models of introspection and those with influencing confounding factors. While this methodological problem is tightly connected to the issue below concerning the possibility of producing non-introspective markers or signatures of consciousness (it is *because* of these worries that the reliability of introspective measures is so important for consciousness science) the argument concerning intractable disputes here I limit to methodological discussion of cases where introspective markers are absent (Seth and Bayne 2022).

[51] PVS is a sustained post-comatose disorder, typically induced by a traumatic brain injury or stroke characterised by high levels of wakefulness combined with low levels of awareness - these subjects retain normal sleep-wake cycles whilst lacking reproducible responses to the self and environment (Berna (2006), Schliff (2007)). This latter feature diagnostically distinguishes PVS patients from those in the *minimally conscious state* (MCS) - a separate state in which patients regain the ability to follow simple motor commands and produce other minor responses to external stimuli (see Royal College of Physicians 200; S2.5). When applied in the context of voluntary behaviour elicited typically via mental imagery tasks, employment of the standard approach has produced compelling evidence that some PVS patients, despite showing no overt behavioural signs of consciousness, are in fact phenomenally conscious (Owen et al. 2006, 2013), Boly et al. (2007), Monti et al. (2010), Cruse et al. (2012) and Shea and Bayne (2010), Klein (2017) for a defense of this argument 'from volition' to the presence of consciousness.

Elwood 2016, Godfrey-Smith 2016, Tye 2017, Boly et al. 2013, Key 2016, Dinets 2016, Godfrey Smith 2016, Michel 2019, Birch 2020, Birch et al. 2021).

In animal sentience research, much debate has ensued around this question with respect to Teleostei fish which, despite lacking a neo-cortex which underwrites the conscious experience of pain in humans, have nociceptors and exhibit a range of sophisticated pain-like behaviours (Sneddon et al. 2013, Michel 2019). While one side of the debate suggests that the lack of neocortex serves as a defeater against the attribution of consciousness in such cases (Dinets 2016, Key 2016), others appeal to the idea that one can reject this idea given that pain is likely to be *multiply realizable* (e.g., Seth 2016). In more recent work, the central issue here is whether the claim that consciousness is multiply realizable is one which is s*cientifically tractable or in principle verifiable* (Godfrey-Smith 2016, Michel 2019). Michel (2019), for example, provides a compelling argument for the claim that the multiple realization of pain is in principle unverifiable via empirical means. Michel's idea is that in order to have evidence for multiple realizability we need a criterion or necessary condition for demarcating or deciding between cases when consciousness plausibly *is* multiply realized (for example, in fish) from implausible cases (e.g., cases of decapitated frogs which show pain behaviours etc.). However, in order to do this, Michel notes, we need to already know which entities are conscious and which are not (2019;11-12), which we do not. Here, we encounter the idea that the methodological issue is one which holds *in principle*, and - like the debates over the proper classification of Eris - is not susceptible to empirical resolution, due to the nature of the concepts involved:

"one needs to provide examples of multiple realizations of pain in order to provide reasons for believing in its multiple realizability, but one can do so only if one already knows what is conscious and what is not i.e., whether one already knows if pain is multiply realized or not. Therefore, there is *no way to validate or invalidate* scientific claims about multiple realization" (Michel 2019;12).

"The root of the problem in the debate on the existence of fish pain becomes clear. There is a crucial asymmetry in the definitions of nociception and pain. Nociception is defined functionally, as a state caused by a specific set of stimuli which triggers a signal through nociceptors which in turn triggers a set of behaviours. On the other hand, pain is not primarily defined functionally… rather, pain is defined in terms of the specific phenomenology that comes with having a pain" (Michel 2019;14).

The seeming intractability of the debate surrounding the attribution problem within animal sentience research, like the more general problem of measuring and observing consciousness, results directly from the way that we currently conceptualise consciousness as a non-physical phenomenon. This can be demonstrated by contrasting the attribution question with the analogous questions asked with respect to the distribution of other, psychological predicates in comparative psychology. While there are, of course, difficult methodological questions which arise for this research (e.g., Boyle 2019), the difficulties which arise within animal sentience research are of a different *kind* to those within other areas of comparative psychology. That is, Michel's claim above is not that it is impossible to study or empirically validate the existence of multiply realized minds *in general* (for example, that it is impossible to study spatial navigation or working memory in rats), but rather that when it comes to the conscious predicate *pain*, there is a serious *a priori* difficulty to the progression of scientific research; one which arises from the special first-person conception of consciousness (and access

individuals have to their own conscious mental states). The same point applies, mutatis mutandis, with respect to the measurement and detection of consciousness in PVS patients compared to the attribution and detection of other psychological predicates.

The second manifestation of the general methodological puzzle is what is often called 'the methodological puzzle' of consciousness (Taylor 2021, Block 2007; Philips 2018). In consciousness science, much attention has been paid to the distinction between a mental state being phenomenally conscious - there being *something it is like* to be in it - and a state being *cognitively (or consciously) accessible* - where this refers to the content of the mental state being *available* for use in reasoning and rationally guiding speech and action (Block 1995). Do there exist actual cases in which a subject is conscious *without* the subject also having cognitive access to this state? If so, how could we go about establishing this? Recent literature suggests that answering this latter question - producing a series of tests which could provide evidence for or against it - constitutes a serious, and intractable methodological problem or puzzle[52]. This is grounded in the fact that all sample instances of consciousness which form the foundation of contemporary consciousness science are *also* cognitively accessible. That is, in everyday cases in which we come to be aware that ourselves and others are phenomenally conscious (from here on 'conscious') and those cases which we use to start off our scientific enquiry into, for example, the neural correlates of particular conscious contents, the mental states which are phenomenally conscious are *also* cognitively accessible: it is by a subject's report that she is phenomenally conscious of p, and/ or her use of p in motor actions ('verbally report when you see the square', 'press the button if you see a red circle') that we infer that the subject is conscious of p and, in the empirical case, subsequently attempt to isolate and identify its neural signature. Recent empirical literature suggests that this poses an intractable problem, which Shea (2012) summarises as follows:

"The scientific study of P-consciousness has to start with some cases of P to form an object of study, together with some instances of ~P by which to identify what is distinctive about the P cases. The thought is that these initial instances will all be cases where conscious contents are also access conscious, so all the cases of P will also be cases of A, and ~P of ~A…. this presents a methodological challenge: that if there are cases of P without A (or the converse), we could never discover them". (308; italics added).

The methodological puzzle here - highlighted in the last sentence – is a result of the fact that testing for actual (nomological) cases in which cognitive access and phenomenal consciousness come apart seems to require that there be forms of evidence (for consciousness) available which can 'trump' or *override* the evidence used to identify the samples of such - gathered by cognitive access and report - which form the basis of the scientific study of consciousness. The thought is that this is deeply problematic: to produce this sort of evidence seems to be inconsistant with, and would thus serves to undermine, the foundation of this scientific enquiry. This methodological challenge - in this case, demonstrating that the question of whether there are, or *could* be cases of nomological dissociation between phenomenal consciousness and conscious access is empirically tractable - is again one which

[52] Taylor (2021), Philips (2018), Shea (2012), Chalmers (1996;239) and Block (2007).

is thought to hold in principle, given the nature of consciousness[53]. It is *because* consciousness is something which is accessible and validated only subjectively and introspectively, that we seemingly cannot *in principle* establish the existence of consciousness without access using our standard empirical methodology. This suggests that an explication of the concept of consciousness is warranted.

3.3. The Argument

The argument presented thus far can be summarised as follows. The last three premises are re-statements of the main tenets of the conceptual engineering research programme within philosophy. Premises (1) and (2) have been argued for above.

(1) Phenomenal concepts appear to refer to essentially subjective, non-physical properties which are non-functional, non-informational and thus resistant to standard third-person methods of scientific measurement and explanation.
(2) Given (1) our phenomenal concepts give rise to, and work to sustain, difficult empirical disputes concerning the attribution of detection of consciousness when subjective measures are absent; problems which appear to hold *a priori* and cannot, therefore, be settled by further empirical research.
(3) If a concept gives rise to intractable empirical disputes within a domain, that concept significantly hinders scientific progress within that domain.
(4) If a concept significantly hinders scientific progress within a domain, the concept is epistemically defective.
(5) If a concept is epistemically defective, it warrants explication and re-engineering.
(6) By (1-5) our phenomenal concepts are in need of explication and re-engineering.

Thus far I have been arguing that our phenomenal concepts warrant engineering on the basis of their epistemic defects. Before closing however, I would like to mention a further possibility viz. it is likely that our phenomenal concepts are deficient in a further *practical* sense. I have in mind here the claim that the view of consciousness underwritten by our existing concept works to actively harm the credibility of scientific research done within consciousness science; a 'scientific taboo' which limits opportunities for funding consciousness research and the validity of peer-review in the field (Michel and Lau 2019, Dawkins 2001, 2017). This is particularly pressing because answers to the puzzles discussed in the last section - particularly, those relating to the attribution or distribution of

[53] Shea argues that this 'beguiling thought' serves as a placeholder for a deeper and more legitimate methodological problem which concerns the possibility of identifying possible dissociations between access and phenomenal consciousness given the fact that we are at present uncertain whether access is *constitutive* of phenomenal consciousness (321). With this in mind, Shea's argument with respect to the natural kind methodology (below) is best seen as proceeding via two stages, the first of which attempts to dissolve the beguiling thought via appeal to the claim that phenomenal consciousness is a natural property or kind (a probabilistic statement of this is set out in 319-321). The second stage sets out a natural kind framework which can address this second methodological problem regarding the uncertainty of the proper constitution of p-consciousness by cognitive access.

consciousness - are of crucial clinical and ethical importance (Laureys 2005, Birch 2021)[54]. The idea that our phenomenal concepts are practically deficient in this sense, provides further motivation for the claim argued for thus far.

The foregoing presents us with an interim conclusion viz. that we ought to explicate or re-engineer our phenomenal concepts to facilitate empirical progress within the science of consciousness and animal sentience research. This claim is, however, compatible with several claims about the correct explicatum which ought to replace our phenomenal concept of consciousness. Next, I outline a proposal for an explicatum concept drawn from recent literature in philosophy of science and consciousness. This is the claim that our concept of consciousness ought to be re-engineered as a 'natural kind concept'. I unpack this claim below.

4. Natural Kind Concepts

In order for the claim that we should explicate our phenomenal concept of consciousness to a natural kind concept to have purchase we need know more about what is meant by the idea of a natural kind concept. Before outlining the view I favour, it will be helpful to first clarify what I *do not* mean by the proposal that our concept of consciousness ought to be replaced or refined to a natural kind concept. The term 'natural kind' has a long philosophical history (Hacking 2007), and not all the usual connotations of this term apply here.

At first glance, particularly for those outside of philosophy of science, the claim above may appear synonymous with the thesis that phenomenal consciousness should be reconceptualised as *rigidly designating* its referent. This reflects the fact that early research on natural kinds was driven by the causal theory of reference and took as its central objective the question of how - and to what - natural kind terms like 'water, 'gold, or 'tiger' refer. For a long time, this question was answered in terms of the claim that natural kinds function linguistically like proper names - that is, by rigidly designating their referents. Despite its prominence outside of the philosophy of science, this so-called 'Kripke-Putnam' analysis of natural kind terms has since been rejected and superseded in the philosophical literature on natural kinds by views which reject, among other things, the methodological claim that semantic considerations of this sort place strong constraints on what counts as a natural kind[55]. As

[54] Indeed, this sort of reasoning seems to be behind Dawkins' (2012) controversial suggestion that the concept of consciousness ought to be eliminated from animal welfare science. While I am sympathetic with Dawkins' negative claim here, I argue in the following section that scientific progress is best served by re-engineering consciousness as a scientific concept as opposed to arguing for its elimination.

[55] Of the many objections pressed against the Kripke-Putnam analysis of natural kind terms the most prominent include (i), the complaint that the view unfairly assumes a microstructural essentialism which is empirically implausible and unsuited to natural kinds in the special sciences (Khalidi 2012, Magnus 2018), (ii), that it threatens to undermine the status of natural kind claims as a posteriori necessities (Salmon 2005) and (iii), that there is in fact *no semantically privileged class* of natural kind terms (Wikforss 2010, 2020 Haggqvist and Wikforss 2018). For a discussion see Tahko (2015).

such, the Kripke-Putnam analysis of natural kind terms will do no work in this paper and can be put aside going forward[56].

When we reflect on the term 'natural kinds' and its paradigmatic instances — water, gold, electrons — we tend to think of these as categories which have a privileged epistemic and ontological status. That is, they are categories which tend to play a particular epistemic role within a scientific domain and have a particular metaphysical structure. The 'epistemic view' of natural kind terms to be outlined here, developed by P.D. Magnus (2013, 2015) takes as its starting point a distinction between two sorts of questions which reflects this idea. The first we can call the 'naturalness' or 'taxonomic' question:

Taxonomic Question. What distinguishes a category which is natural ('electron', 'planet' 'hepatitis') from an arbitrary category ('round red objects')?

Magnus provides an answer to the taxonomic question motivated by the adoption of three independently plausible theses or assumptions: the principles of 'induction' 'science' and 'domain relativity'. That is, the claims that natural kinds are essentially those categories which (i) support and feature in successful exploratory and inductive inferences, (ii) form components of a successful scientific taxonomy (that is are necessarily *scientific* categories) and (iii) are, crucially, always relative to a *specific domain* of enquiry. The first two principles reflect the standard answer to the naturalness question, on which the naturalness of a kind is grounded in the *projectability* of natural kind categories (Goodman 2000, Boyd 1991). On this view, bona fide natural kinds, in contrast to conventional pragmatic groupings, are those which feature in and explain the epistemic reliability of a large class of enumerative inductive inferences. That is to say, natural kinds explain and justify forms of inference which generalise from a limited number of observed items (*this F is a G)* to a claim about all (unobserved) members of the kind to which the items belong (viz. *'all Fs are likely G*)[57]. Sankey (2018;8), for example, provides the following illustration: "it is because all members of the natural kind platypus have webbed feet that the inductive inference from observed to unobserved platypuses yields a correct conclusion"· This is intended to capture the claim that natural kinds exhibit a *systematic sociability* (Chakravartty 2007;170).

The third of Magnus' principles, preempted in Boyd (1999) and Quine (1959), constitutes a denial of what he calls the 'simpliciter assumption': the claim that there is a fact of the matter whether (for any natural kind candidate X) X is a natural kind (Goodman 2014, Bird and Tobin 2015). Instead, the resulting view asserts that natural kinds are always *domain relative*: natural kindness is not a one place

[56] It is this analysis of natural kind concepts that appears to be is operative in Peter Caruthers' dismissal of the idea that phenomenal consciousness is a natural kind concept: "one thing that everyone has agreed on, at least since Kripke (1980) is that terms referring to conscious mental states aren't used as natural kind terms. In contrast, it is generally agreed that our concepts for substances like water are natural kind ones. Even before we knew anything about chemistry, we used the concept of water to refer to the underlying nature or essence of the recognisable stuff that fills our lakes and rivers…"(Caruthers (2018;53)). I address an updated version of this 'phenomenal concept' objection in Section 6.1.

[57] See Sankey (1997, 2007, 2001) and Kornblith (1993).

predicate '('electron is a natural kind') but is rather a *relational* notion which relates natural kinds to specific domains of scientific enquiry ('electron is a natural kind *for* the domain of particle physics'). Magnus takes reflection on two paradigmatic natural kinds as illustrative of this point:

''even though the kind *predator* does no work in the high-energy or small-scale domains where top quarks matter, *top quark* does no work in the biological or ecological domains in which *predator* is a natural kind. Neither has a claim to absolute importance, but they are important in relation to different contexts and phenomena. The very same kind *top quark* is a natural kind in the domain of particle physics but not in the domain of ecology; and *vice versa* for the kind *predator* ''(Magnus; 2012;3).

These three theses work to form Magnus 'positive epistemic account of natural kinds, on which natural kinds are those categories that scientific enquiry (understood in domain specific terms) is forced to posit to be successful. This more general claim splits into two criteria for natural kindhood which Magnus names the 'success' and 'restriction' clauses:

Pragmatic naturalism: A category k is a natural kind for domain d if the following conditions are met: (1) (*The success clause*): k is a part of a taxonomy that allows scientific enquiry into d to achieve inductive and explanatory success and (2) (*The restriction clause*): any taxonomy that excluded k would not do so.

Magnus goes on to distinguish the taxonomic question from a further question which is often conflated with the former:

Ontology Question. What feature(s) in the word allows natural categories to perform this function?

Magnus' answer to the taxonomic question allows for a pluralistic answer to this question. That is, pragmatic naturalism allows for the claim that different ontological features are picked out by this epistemic role within different domains of enquiry. In particular, this idea is consistent with the claim that natural kinds in given domains will have paradigmatic features and metaphysical structures. In the special sciences, to which the discussion in this paper pertains, the dominant account of the metaphysics of natural kinds is Richard Boyd's homeostatic property cluster view (1991, 1994, 1999)[58]. According to this theory, natural kinds in higher level domains are identical to a *cluster* or repetitions of properties whose reliable co-occurrence in nature is explained and maintained by the operation of a causal mechanism. This general claim, which captures the main features of the view, serves as a placeholder for three more specific theses central to Boyd's homeostatic property theory: (Figure 1)[59].

[58] In addition to its initial development in Boyd's work, the HPC view has also been elaborated and defended in Kornblith (1993); (Griffiths 1999)) and Wilson (1999, 2005).
[59] It is typical to attribute to Boyd a fourth *accommodation* thesis, which I leave out of the main discussion here (1991;139, Craver 2009). This states that to the extent that a given natural category or kind is *indispensable* for these successful inductive generalisations, we can conclude that the kind *accommodates* the causal structure of the world. That is, it is typical to see Boyd as proposing a realist account of natural kinds along the ''no miracles '' defense of scientific realism Putnam (1975) Boyd (1989); Lipton (1994); Psillos (1999).

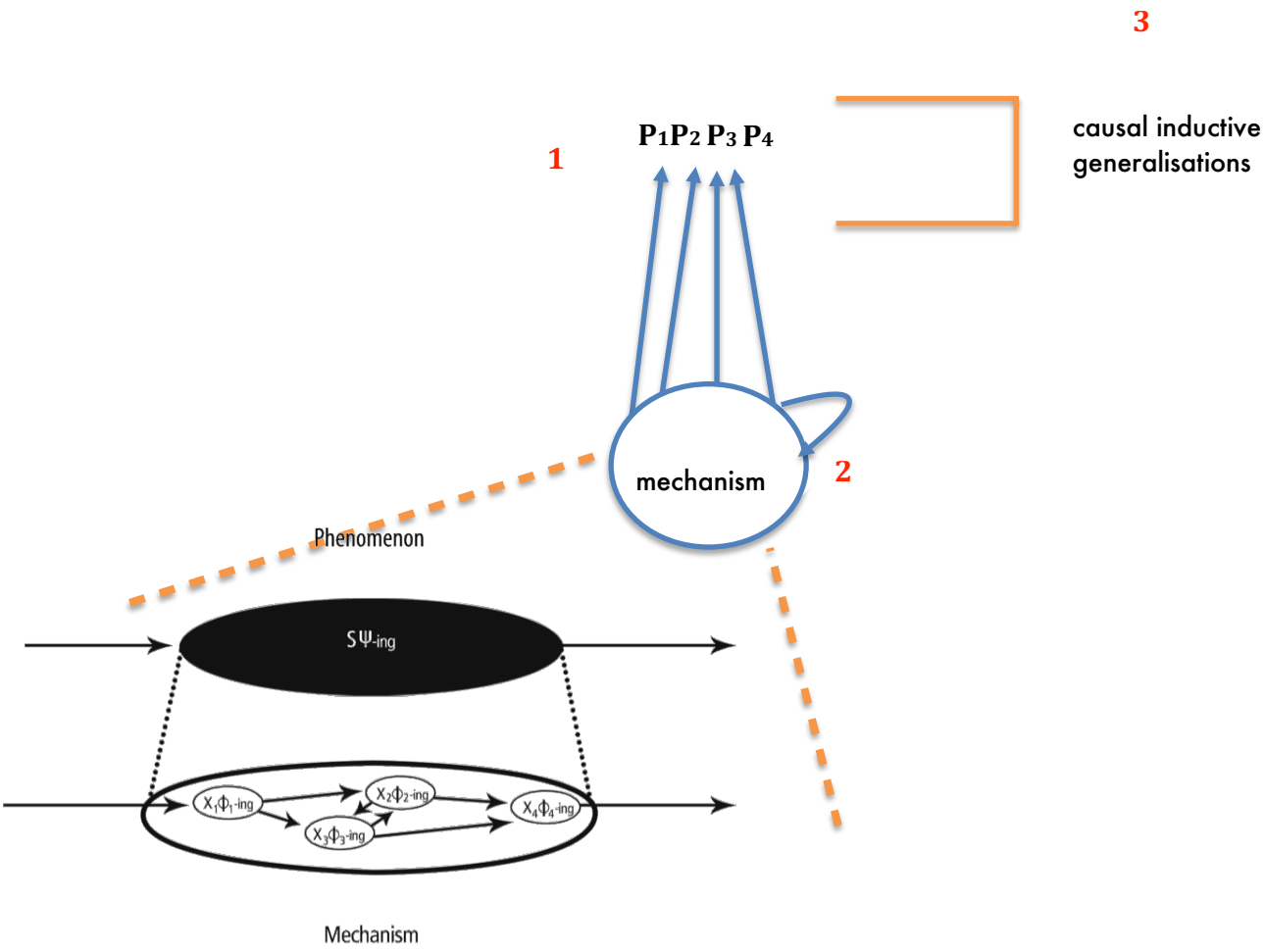

Figure 1: The homeostatic property cluster theory of natural kinds (in the abstract, applied to natural kinds in neuroscience). Here, importantly, the relevant causal mechanisms are *constitutive*. In biological cases such as species (below) the 'causal mechanism' is etiological.

1. *Property Cluster:* natural kinds (in the special sciences) are identified with a cluster of properties that regularly **occur together**.
2. *Homeostatic Mechanism:* a causal mechanism underlies and **explains why** the clustered properties occur together.
3. *Causal Import:* natural kinds, ((1-2) explain) feature in important **causal inductive generalisations**.

The first claim, stated above, is that natural kinds in the special sciences correspond to a cluster of properties which regularly occur together. The paradigmatic cases of such property clusters are thought to be *biological species*, each member of which share important morphological, behavioural and physiological features. Second, while these 'clusterings' regularly occur in nature, none of the clustered properties are taken to be necessary for membership (as essentialist views maintain) but

rather, when enough of the cluster are present, are taken as *jointly sufficient*. Thus, we find several instances of the kind (a given biological species) which, despite lacking some of the common morphological or behavioural properties indicative of the kind, are still correctly classified as such. Third, the common co-occurrence of these properties is not, on Boyd's view, posited as a brute fact but is accounted for by the repeated operation of a causal mechanism which *explains* the repeated or 'homeostatic 'co-occurrence of these clusters in nature. In the case of species, for example, similarity is explained by historical causal structures such as gene recombination, gene transcription, similar selection regimes and common developmental constraints (Boyd 1999, Ereshefsky 2005).

The operation of this causal mechanism here also - via the observation that causal mechanisms often breakdown and fail to function in predictable ways - explains *why* members of the kind exist which lack one or a number of the (seemingly essential) cluster of properties we started with. It thus can account for the forms of variation (Darwinian and otherwise) which essentialist accounts fail to capture. Additionally, this homeostatic property cluster view successfully accounts for the fact that kinds feature in *important causal inductive generalisations* (due to being reliably maintained by a similarity generating causal mechanism). That is, it explains *how* natural kinds meet the two epistemic, 'pragmatic' criteria that Magnus' account posits as necessary and sufficient conditions for natural kindhood.

5. Ameliorative Physicalism and the Natural Kind Approach

The combination of pragmatic naturalism and the homeostatic property cluster account of natural kinds provide a view of natural kind concepts which lend the proposal in this paper clear positive content. This is the claim that natural kind concepts are those concepts which play an indispensable epistemic role within an empirical domain of enquiry, an epistemic role that, in the special sciences, picks out a cluster of co-occurring neural, functional or information properties, underwritten by causal mechanism(s). The positive claim of this paper can thus be stated as follows:

Ameliorative Physicalism: we ought to re-engineer our concept of consciousness to a natural kind concept which, if it refers at all, plays an indispensable role within an empirical domain(s); an epistemic role which is underwritten by a cluster (or series of clusters) of neural, functional, or informational properties, associated with our initial samples of consciousness, whose co-occurrence is explained by a network of causal and constitutive mechanisms.

As described in section 2, if this natural kind concept is to be justified as an explicatum concept for consciousness, in Carnapian or 'topic improving' sense, there must be compelling grounds for thinking that this concept provides a clear epistemic improvement on the original concept of consciousness. That is, that it avoids the conceptual deficiencies outlined and motivated with respect to our original concept of consciousness in section 3. How does this new concept provide the resources for doing so? That (and how) the new natural kind concept provides an epistemic improvement with respect to the issues outlined in section 3.1 should be clear and straightforward. Given the definition

of natural kind concepts offered above it is clear that, in general terms, engineering our concept of consciousness in this way will provide us with a new concept which is well placed to maximise epistemic success within the cognitive neuroscience of consciousness. For ameliorative physicalism asserts that, if it refers at all, the new concept of consciousness is one which maximises epistemic success within the relevant domain[60]. This stands in stark contrast to the 'seemingly wonderful' conception of consciousness outlined in 3.1.

This is a good start. However, to justify the ameliorative concept above as an adequate explicatum for our concept of consciousness it must also be the case that the new concept has the resources to unravel the empirical disputes in consciousness science discussed in section 3.2. That is, the new concept must, like the replacement concept of planet in 2006, have potential to clearly resolve these disputes in a way that our existing concept cannot. To show how this can be done, it is necessary to situate the central proposal made thus far in this paper within a broader approach to the science of consciousness which has been proposed in the last ten years called the 'natural kind approach' to consciousness (Shea and Bayne 2010, 2020, Shea 2012). In this series of papers, Tim Bayne and Nick Shea have persuasively argued that studying consciousness as a natural kind allows for the production of a novel methodology for consciousness science which allows us to resolve the issue of when to attribute consciousness in the absence of verbal report (Shea and Bayne 2010, 2020) and how to test for nomological dispositions between consciousness and access (Shea 2012). That is, it provides a solution to the validation challenge (2020). Reflecting Boyd's homeostatic property cluster theory, this starts by searching for and identifying the functional, neural, or informational properties picked out by first-person reports of consciousness in typical cases, analysing this data for a cluster(s) of properties which regularly and invariantly co-occur using causal modelling, and, eventually, searching for the causal mechanism(s) which underwrite this cluster of properties. The methodological strategy can be summarised stepwise as follows:

The Natural Kind Approach to Consciousness:

1. Search for the functional, neural, and informational properties of consciousness with which it is associated.
2. Identify the way(s) in which these properties cluster together - interrogating the data for property clusters via causal modelling.
3. Having identified the relevant cluster(s), one searches for the underlying causal mechanism that accounts for it, and subsequently determine its proper distribution.

This methodology provides a clear way of guiding research in the science of consciousness which provides a way out of the intractable puzzles associated with measurement and attribution. This is developed at length in Bayne and Shea (2010) in the context of Persistent Vegetative State patients and in Shea (2012) with respect to dissociations between consciousness and conscious access. In the former case, the solution is grounded in the idea that when we have identified the relevant properties in the

[60] As this definition suggests, this view is compatible with the possibility that there may turn out to be no natural kind (HPC) here picked out by 'phenomenal consciousness' at all. See footnote 34.

cluster(s), we can then look for them and employ tests for their instantiation of in the PVS subjects[61]. Here, the central claim is that the detection and presence of such properties can provide strong, albeit inconclusive evidence - as an *inference to the best explanation* (Lipton (1991))- for the presence of consciousness which does not require the employment of our pre-theoretical means of identification. That is, it provides a means of testing for the presence of consciousness in PVS patients which need not involve - and goes beyond - tests for voluntary agency or reports which are usually used to identify consciousness. This key claim reflects a more general and salient feature of scientific investigation into natural kinds, viz. that they allow for and provide us with new ways of identifying instances of the kind which are not limited to, or mediated by, our pre-theoretical criteria[62]. It is easy to see how this general methodology can be carried over and applied to other problem cases, such as infants and non-human animals[63].

Turning now to the methodological puzzle of consciousness and the natural kind approach (Shea 2012). This methodological puzzle arose as a result of the fact that there were no available methods for identifying instances of phenomenal consciousness - and in particular, cases in which it might dissociate from cognitive access - which do not rely on our initial (and crucially, access mediated) means of detection. However, if consciousness is to be identified with a HPC natural kind, Shea notes, there are no in principle barriers against a post-theoretical means of identification of consciousness. The history of science is full of such examples. Importantly, the history of science reveals that post-theoretical means of identification are justified *even* when there is prior uncertainty as to whether the kind is constituted by one of the properties used to initially identify it (consider, *californium*)[64].

Shea suggests that adopting the thesis that consciousness is a natural kind presents us with a straightforward empirical means of testing the hypothesis that consciousness and cognitive access dissociate. When testing for our nomological cluster of properties (in Stage two) two salient

[61] Current candidates for these properties include global availability of information (cognitive access): meta-memory (Hampton 2001), insensitivity to the automatic stem completion effect (Debner & Jacoby 1994, Merikle et al. 2001), Integration of information by prefrontal cortex. Relating to informational processing: susceptibility of grip width to the size contrast effect (Hu & Goodale 2000), trace conditioning vs. delay conditioning (Clark et al. 2002), perceptual discrimination of a novel conjunction of perceptual features (Jack & Shallice 2001). With respect to neural properties: 40 Hz gamma-band neural synchrony, local recurrence in cortex, cortico-thalamic loops and networks involving prefrontal cortex (329-330; Shea 2012).

[62] Compare, for example, the idea that once the chemical composition of water was discovered, it was possible to identify instances of water (e.g., as a gas) which don't meet the pre-theoretical criteria (clear, odorless liquid found in oceans) associated with it (Shea 2012).

[63] See Bayne and Shea (2020; section 6) for a discussion. Whiteley (in draft) also examines how the natural kind methodology can be utilized in animal sentience research and used to provide a way out of the argument Michel (2019) provides concerning the impossibility of validating empirical claims about plausible and implausible cases of multiple realisation of conscious mental states.

[64] 2012;232-233. ''If consciousness were a natural kind, it would be very surprising to discover that our access to it were limited by pre-theoretical measures that we associate with it. In every other case, discovering a natural property allows us to go beyond our pre-theoretic measures. We see *no good reason* why consciousness cannot be investigated in the same way" (2010,470 italics added).

possibilities arise. Either we will find (i) one or (ii) two clusters of properties from our sample data[65]. Shea claims that our two starting hypotheses will make the following predictions (represented in Figure 2). If it is the case that, after having interrogated our data sets using forms of causal modelling (Pearl 2000), we discover *one* cluster of properties, we can, Shea claims, make an inference to the best explanation to the conclusion that there are in fact no nomological dissociations between phenomenal consciousness and conscious access (that is, that the possibility depicted on the left hand side in Figure 1 is actualized). In contrast, if *two* clusters are identified via our causal investigations, Shea maintains, we have good reason to conclude - again via inference to the best explanation - that there *are* instances of nomological dissociation between access and phenomenal consciousness (that is, that the possibility depicted on the right-hand side in Figure 2 is actualised and the two can nomologically dissociate)[66]. The natural kind approach to consciousness, in other words, provides a means of dissolving this persistent methodological problem. As such, the ability and epistemic potential of the natural kind approach - driven by a conceptualisation of consciousness as a natural kind — gives us sufficient reason, as things stand, to think that the natural kind concept of consciousness is justified as a replacement concept for consciousness.

[65] Shea also considers the possibility that there may turn out to be *multiple* property clusters discovered in Stage 2, connected via a complex web of causal interactions (333). I set this possibility aside here, but address this briefly when I consider methodological objections to the natural kind approach in section 6.2.

[66]Here, the inference to the phenomenal and access dissociation is more complicated. Could we not, having discovered two property clusters, nevertheless conclude that phenomenal and access consciousness are always constitutively connected (the left-hand side is still actualized, forming one property cluster), and that the other property cluster denotes a distinct *pre*-phenomenal kind? That is, how are we to tell that one of the two kinds discovered is really *phenomenal*? Shea considers and rejects this objection (334-339).

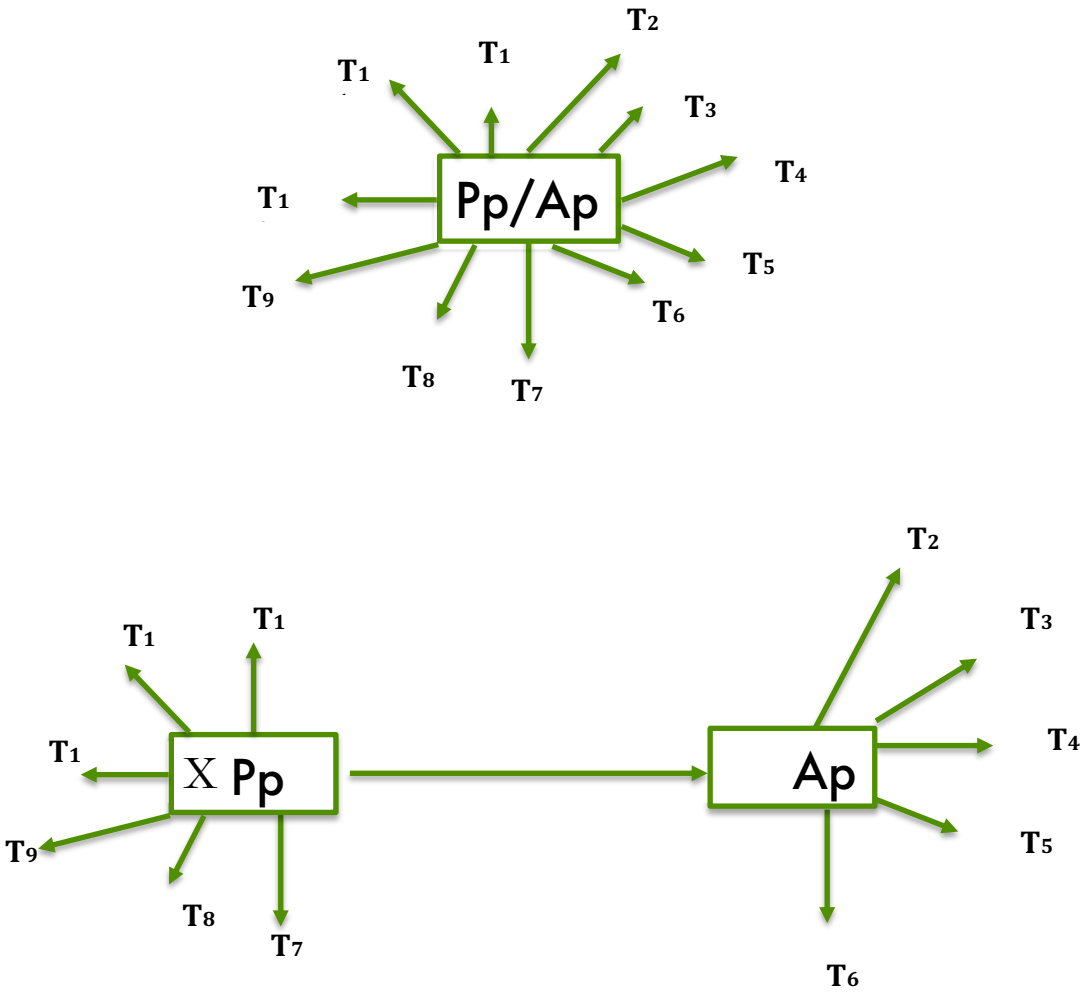

Figure 2: Shea's 'one' vs 'two kind' scenarios. On the left-hand side, one property cluster is discovered, from which the nomological co-instantiation of phenomenal and access consciousness is inferred. The right-hand side represents the possibility that two property clusters are discovered. These can then be identified with phenomenal and access consciousness respectively. T1-10 reflect the tests for property candidates employed in Stage one of the natural kind methodology (adapted from Shea 2012).

Return now to the physicalist aspect of this proposal. The suggestion that we ought to re-engineer or explicate our phenomenal concepts as natural kind concepts opens up the possibility of a novel argumentative strategy for resisting the anti-physicalist conclusion; one which, like the phenomenal concept strategy, is attractive because it accepts the plausible claim that there is *currently* an epistemic gap between physical and phenomenal concepts. Recall that such 'type-B' physicalist responses (Chalmers 2010) face a host of objections which are grounded in the way that our phenomenal concepts actually are and what they appear to do (Goff 2017). The proposal here allows a way out of this predicament, and additionally, goes on to explain how and *why* proponents of these views go

wrong in critiquing these accounts for mis-representing the way that phenomenal concepts are[67]. On this view, physicalists who grant that there is an explanatory gap between physical and phenomenal descriptions and attribute this to the nature of phenomenal concepts should not be attempting to describe such concepts in detail, but rather be engaged in describing how these can be best *explicated* for the purpose of furthering scientific research into the neural basis of conscious experience and its distribution. Ameliorative physicalism thus plays a dual function in consciousness research, being attractive to both metaphysicians of mind and philosophers of science working on the neural basis of consciousness.

Thus far, the discussion in this section has concerned how we can use the natural kind approach to support and justify ameliorative physicalism — the view that we ought to re-engineer our concept of consciousness to a natural kind concept. Before moving on to consider an objection to this position, I want to pause to consider first how ameliorative physicalism may in turn support and offer conceptual resources to the natural kind approach which are otherwise unavailable on existing formulations of this strategy. For, if it can do so, this will provide further reason to take the position seriously, for it will further demonstrate the methodological potential of this explicative strategy within consciousness science.

6. Overcoming Objections to the Natural Kind Strategy

Despite being introduced over a decade ago, the natural kind approach to consciousness has yet to be taken up comprehensively within consciousness science. It has also received little philosophical attention and elucidation, with only a small literature examining its methodological and philosophical foundations. Given its - quite significant - potential to provide a way out of the seemingly intractable methodological puzzles detailed above, this is puzzling. Why is this? An obvious answer is that this empirical and philosophical resistance is due to the fact that there are compelling reasons to think that this approach to the study of consciousness is flawed from the outset — that is, that there are compelling objections to the approach which suggest that it is likely to fail. Reflection on the natural kind approach detailed above suggests that this is at least part of the explanation for the current resistance to the natural kind approach. This is because several salient objections present themselves when we consider this approach, while others have been pressed against the natural kind strategy in print (Philips 2018, Shea and Bayne 2020, Taylor 2021). Here, I will argue that the many of these can be shown to result from our current conception of consciousness. It thus follows that if the explicative strategy detailed here is adopted these objections cut no ice against the natural kind strategy. For the ameliorative physicalist will grant that these objections are valid in the sense that they arise from our *current* conception of consciousness but maintain that as objections to the natural kind strategy these are ineffective insofar as the latter is to be understood essentially as a proposal for replacing this problematic conception of consciousness with a better concept. The objections to the natural kind strategy, from this methodological vantage point, are evidence instead for the epistemic defectiveness

[67] This idea also holds for criticisms of some type-A views such as Dennett (2001) which, I think, are better understood in the context of conceptual engineering.

of our existing conception of consciousness, and not evidence against the viability of the natural kind approach. If I am right, this suggests that ameliorative physicalism ought to be embraced as a central component of the natural kind approach to consciousness. That is, in addition to offering a new argument in favour of the natural kind strategy, it also provides the strategy with the resources to overcome several salient objections recently levied against it.

When we reflect on the commitments of the methodology for consciousness science that the natural kind approach recommends, two different sorts of objections present themselves. The first are foundational -- they target the very idea that studying consciousness as a natural kind is likely to be a fruitful strategy. The second are methodological and pragmatic — they aim to show that in practice, the natural kind approach cannot resolve the methodological disputes it purports to. This second set of objections are particularly pressing considering the argument for ameliorative physicalism presented here. For if the natural kind concept cannot work to resolve the intractable disputes considered above, then the justification for explicating our concept of consciousness in this way is under threat. I'll consider these objections in turn.

6.1. Foundational Objections to the Natural Kind Approach

The compelling idea behind the natural kind approach is the claim that consciousness can be studied in the same way as any other natural phenomenon — like water, gold, or hepatitis. Are there compelling reasons for thinking this is, in principle, unworkable? The foundational objections to the natural kind approach are at least threefold. According to an *ontological objection*, the natural kind approach to consciousness ought to be rejected on metaphysical grounds. The idea is that we have good and well-known ontological reasons - most notably, those provided and highlighted by the anti-physicalist arguments sketched above - to maintain that consciousness is *not* a natural kind and cannot be identified with a physical-functional homeostatic property cluster(s) (or the property/properties which underwrites this clustering(s)). This raises the pressing question of why we should expect the natural kind strategy to track consciousness. The second foundational objection to the natural kind strategy is conceptual in nature. According to the '*phenomenal concept objection*' (Shea and Bayne 2020; S3), the natural kind approach ought to be rejected because our concept of consciousness is a phenomenal concept, *not* a natural kind concept[68]. Again, the idea is that for the natural kind approach to succeed, it must be the case that our conception of consciousness is a natural kind concept. But it isn't. Therefore, we ought to reject the natural kind approach as a viable methodology for the study of consciousness. Third, one might object to the natural kind approach on the basis that it demands an openness to several counter intuitive outcomes which we have good reason to reject. Call this the *objection from counter intuitive consequences*.

To get a grip on this latter objection, consider the following historical cases. Doctors diagnose several cases of hepatitis (such as Hepatitis C) in patients who lack any of the typical symptoms which were

[68] The phenomenal concept objection can be found presented in premise conclusion form in Shea and Bayne (2020;12).

used in an earlier stage of theorising to diagnose cases of the disease (Metha et al. 2021). In the chemical revolution in the 17th century, chemists discovered that the reactions under investigation could be better explained by gains and losses in chemical elements like oxygen through processes of oxidation and reduction rather than by phlogiston, leading to the removal of this substance from chemical theory (Weisberg et al. 2019). Medical pathologists examining cases of inflammatory bowel disease conclude that a subset of cases cannot be diagnosed as either ulcerative colitis or Crohn's disease and thus that clinical testing methods produce an *indeterminate result* between these two distinct diagnoses (Tremaine 2011). Prospectors searching for gold deposits discover a mineral found in rocks that looks like gold, but is a distinct, non-valuable chemical entity.

These cases are illustrative of four potential scenarios or outcomes that a natural kind strategy for studying any phenomenon — and thus, a natural kind strategy for the study of consciousness — are committed to. The first illustrates the possibility of identifying an instance of consciousness which lacks its pre-theoretical or 'surface' markers - for example, an instance of consciousness which is not reportable, or one which lacks a determinate phenomenological character. The second represents a potential outcome of the natural kind strategy, if adopted. In this case, there is no natural kind found to be associated with our pre-theoretical markers of consciousness, and thus the category ought to be eliminated from scientific discourse[69]. The third illustrates that there may be cases, some species of non-human animals perhaps, or infants, in which it is indeterminate whether our concept of consciousness applies. That is, that the natural kind approach commits to us the possibility of genuine *borderline cases* of consciousness. Finally, the fourth scenario illustrates the possibility of a case of something which appears to be an instance of consciousness - it shares some of its surface features, and thus we might initially judge it thus - when in fact it is not. According to the objection under consideration, these four outcomes are extremely counterintuitive and at odds with our beliefs about consciousness. That is, according to this line of argument, consciousness is simply *not the sort of thing* that could be eliminated from scientific discourse and fail to exist. Similarly, the objection holds, the very idea of an instance of consciousness lacking a clear or determinate phenomenology, or of phenomenological properties without an instance of consciousness — cases which appear to be at odds with the idea that our introspective judgements about our own conscious mental states are infallible— is incoherent (Antony 2008 Simon 2017)[70]. Any methodology which is predicated on these possibilities

[69] This 'scientific eliminitvism' about consciousness might be motivated by two possibilities: one in which no homeostatic property cluster is found at all, and another on which we find *several* homeostatic property clusters associated with our pre-theoretical markers of consciousness. I discuss this latter possibility with respect to dreaming in the following chapter of this thesis. A further objection to the natural kind approach sketched here asserts that the natural kind approach ought to be rejected on the basis that we have already have good reasons for endorsing scientific eliminitvism with respect to consciousness, and that this has already been established (see, for example, Allport 1988, Churchland 1988, Irvine 2013). Following Shea and Bayne (2020) I hold that this objection is pre-emptive -- it is too early to establish this outcome given that the natural kind strategy has yet to be implemented seriously. As such, I set this issue aside here.

[70] Although for recent discussion and defence of the idea of borderline cases of consciousness see Schwitzgebel (2021), Lee (forthcoming). I leave it open whether these cases succeed on establishing the compatibility of our existing conception of consciousness with borderline cases of consciousness, or only do so by changing our concept of consciousness. If you are compelled by the former idea, you can simply view the objection from counter intuitive consequences as pertaining to the other three scenarios.

then must be rejected as an adequate methodological programme for the scientific study of consciousness.

Given that ameliorative physicalism is a self-consciously physicalist strategy, we can set aside the ontological objection here. Given that the view I have outlined above is presented as an attractive position to those who *already* take the idea that consciousness is an evolved biological phenomenon seriously, this objection does not arise[71]. Consider then the phenomenal concept objection and the objection from counter intuitive consequences. Here it is clear that the explicative formulation of the natural kind strategy offers a new line of response to these objections not available to the existing formulation of the approach. To see this, it is necessary to consider how both objections appear to arise essentially from our existing (and, as I have argued, problematic) conception of consciousness. It is *because* we currently conceive of consciousness in a particular, 'seemingly wonderful' way that the thought that it could not possibly be a natural kind concept is compelling. Similarly, it is this concept that drives the idea that these outcomes are problematic or even incoherent, for these features, as we saw in section 3, are built into this concept. That is, given that this concept appears to refer to the felt phenomenal properties themselves, and not the natural kind that underwrites them, we tend to believe that an instance of consciousness without such properties, or an outcome which eliminates consciousness, is incoherent.

This observation provides the explicative formulation of the natural strategy with grounds for asserting that these objections to the approach simply miss the mark. That is, that these objections misconstrue the natural kind approach as an approach to consciousness which takes this phenomenal concept of consciousness seriously as a concept which is fit to ground a mature empirical research programme. By shifting the central claim of the approach from the idea that consciousness as we *currently conceptualise it* ought to and can be studied *as if* it were a natural kind, to view that reconceptualising consciousness in this way is *epistemically useful* (allowing us to resolve powerful methodological puzzles in consciousness science), we can provide a way out of these foundational

[71] This serves as a key point of departure from the version of the natural kind strategy defended by Shea and Bayne, who respond to this objection by maintaining that the natural kind approach can allow for a dualistic as well as physicalist conception of consciousness (2020;9). The idea here is that the natural kind approach need not commit to the claim that consciousness is a natural kind in a strict sense (that it's *identical* to the associated homeostatic property cluster(s), or the property which underwrites it) but only needs to be tied to such a property cluster with nomological necessity. That is, in other words, the natural kind strategy can remain natural with respect to the ontological dependency relation (Wilson 2021) which holds between consciousness and the property cluster posited by the natural kind strategy, and in doing so, is compatible with more than just physicalist metaphysical views of consciousness. In endorsing physicalism, this explicative strategy I detail here gives up on the metaphysical neutrality (albeit one limited to dualism and physicalism) which is a benefit of Shea and Bayne's formulation. However, as we shall see, I believe this metaphysical neutrality comes at a cost to the existing formulation of the approach, leaving it vulnerable to a series of methodological objections. The removal of a metaphysically neutral conception of the natural kind approach is thus offset by the further benefits that this explicative strategy provides. It is also worth mentioning here that there are also good reasons for thinking that metaphysical neutrality with respect to consciousness is overrated as a desideratum for the neuroscience of consciousness. One might appeal here to recent arguments in the debate over the epistemic status of metaphysics which assert that empirical considerations should come before metaphysical considerations like these, not the other way around (Ladyman 2012, Paul 2012). One might in turn argue that this removes the requirement for metaphysical neutrality in neuroscientific theories of consciousness (e.g., Whiteley 2018).

objections to the natural kind approach. The counterintuitive consequences, on this new view are not bugs of the natural kind approach but *a purposefully built-in feature*; not a counterintuitive bullet to bite (as Shea and Bayne suggest with respect to borderline cases), but deliberately built into an explicatum concept of consciousness on the basis that this can help us facilitate a mature science of consciousness.

This has an advantage over the existing response to the phenomenal concept objection provided in Shea and Bayne (2020) which is worth noting. In their recent paper, Shea and Bayne note that the proponent of the natural kind strategy could respond by pointing out that the success of the natural kind approach does not depend on conceptualising consciousness as a natural kind concept (section 3). While I agree with Shea and Bayne that conceptual or semantic considerations fail to have any purchase against the claim that consciousness *is* a natural kind (nor the claim that it can be studied as such)[72], I think there are reasons for thinking that the natural kind approach will not succeed while our existing concept of consciousness remains operative in consciousness science. That is, that our existing concept of consciousness causes problems for the approach in practice. To see this, let's now turn to the methodological objections pressed against the natural kind approach[73].

6.2. Taylor's Objection: The Problem of an Undetected Kind

The central claim of the natural kind approach, on the view I have sketched, is that conceiving of consciousness as a natural kind is epistemically useful, allowing us to resolve foundational methodological disputes and puzzles within consciousness science in a way which is precluded by our current phenomenal concept of consciousness. As I suggested, I think that this explicative strategy is beneficial because it helps us understand why objections to the natural kind approach which are grounded in our existing concept of consciousness miss the mark in a way that the existing formulation of the strategy cannot. Are there other reasons for resisting the natural kind approach to consciousness? In a recent paper Henry Taylor (2021) raises a new objection to the natural kind approach which challenges even this reformulated interpretation. This objection, unlike the previous ontological and conceptual objections considered, is methodological in nature. It aims to show not that this methodology is ill-suited to studying consciousness in principle (because consciousness is not a natural kind or cannot be conceived of via natural kind concepts). Rather, it aims to show that even if these objections can be overcome and the approach is carried out, that it cannot live up to its central promise of resolving key methodological disputes in consciousness science. In particular, Taylor claims that it cannot resolve the long-standing methodological puzzle of how to determine whether or not there are cases of phenomenal consciousness without conscious access, contra Shea (2012). If Taylor is right, this raises deep difficulties for the natural kind approach I have sketched, as well as ameliorative physicalism.

[72] Given that this goes against the epistemic criteria for natural kindhood now widely accepted in the literature in philosophy of science.
[73] A further foundational objection to the natural kind approach which I don't discuss here is the 'objection from revelation'. For discussion and convincing response, see Shea and Bayne (2020; S4).

Taylor's objection to the natural kind approach centers around a hypothetical but nonetheless 'empirically realistic' scenario about the natural kind structure of consciousness — that is, a scenario which he argues could well turn out to be true of consciousness in our world. In this scenario - let's call it 'scenario X' - Taylor argues that the natural kind strategy would inevitably lead to a false classification of a genuine instance of consciousness (one which fails to coincide with conscious access) as non-conscious, thereby demonstrating that the natural kind approach cannot resolve this methodological puzzle.

Scenario X has two core elements. First, it is a case in which our term 'consciousness' is associated with *multiple kinds*. That is, it is a case in which our term 'consciousness' turns out to be underwritten not by a single natural kind, but two homeostatic property clusters, each of which are underwritten by distinct properties or causal mechanisms. In such a case, 'consciousness' turns out to be like just the categories 'jade' and 'hepatitis' — where we thought there was a single unified kind, we find instead that this category is underwritten by two dissociative natural homeostatic property clusters - let's call them K1 and K2 - which have little in common. As such, each cluster works to underwrite distinct sets of inductive generalisations[74]. Second, X is a case involving many *'false negative'* markers for consciousness. That is to say, markers which fail to be identified as indicative of consciousness in the initial steps of the natural kind methodology, but which nonetheless are causally underpinned by, and form a part of, one the homeostatic property clusters (K2) associated with consciousness. The potential for false negative markers gains plausibility from the thought that when the natural kind approach is implemented, there will inevitably be properties which are missed, or initially misclassified as non-markers of consciousness, when in fact they are[75]. So, proponents of the natural kind approach will agree that many false positives are likely at this initial stage.

Taylor argues that a scenario involving the conjunction of these features presents us with a case where, if the natural kind methodology is implemented, one of these homeostatic property clusters - 'K2' - will not be classified as an instance of consciousness when it is (Figure 3). To see this, Taylor asks us to suppose that in this scenario, *all* of K2's property-markers (T5-8) are false negative markers, i.e., markers which are not identified as initially being causally bound to a consciousness-associated homeostatic property cluster, but which are. That this presents the natural kind approach with a problematic scenario can be demonstrated as follows. Given that in X *none* of K2's property-markers would be identified, adoption of the natural kind strategy would lead to *only* K1's property markers (viz. T1-4) being identified in Step 1 of the approach. Consequently, only K1's markers would be causally modelled and subject to the requisite cluster analysis (as recommended by Step 2 of the natural kind strategy), and thus, only the neural mechanisms underwriting T1-4 being identified (as described in Step 3). But, given the starting stipulation that this is a case in which 'consciousness' is underwritten by two kinds, this methodological procedure would yield the erroneous judgement that consciousness

[74] This is the same as the 'two kinds' scenario discussed in Shea (2012).

[75] Taylor describes two potential examples of false negative markers of consciousness - sensitivity to stem completion (Snogdrass 2002), and delay conditioning (Lovibond and Shanks 2002) - and such cases are common within other historical examples of investigation into natural kinds.

is associated with only one natural kind. Thus, if this potential scenario is actual, Taylor argues, the natural kind approach would erroneously fail to classify K2 as an instance of consciousness.

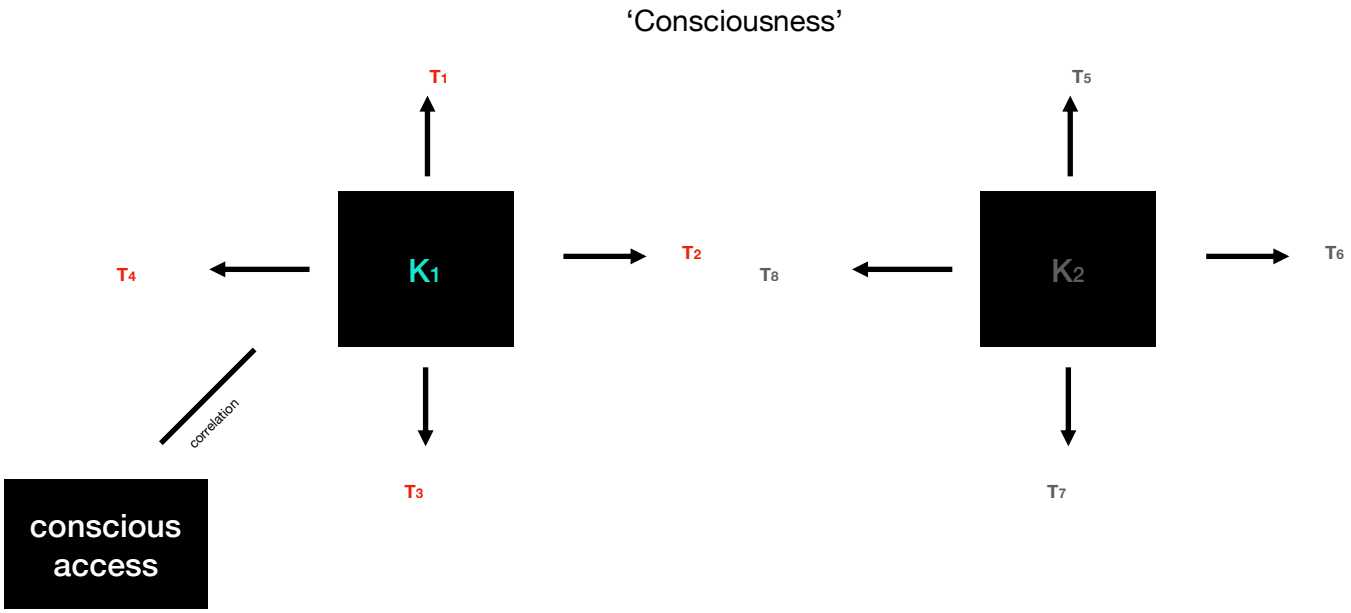

Figure 3: Scenario X: A multiple kind view of consciousness involving false negatives markers (T5-8). Red denotes true positive markers, grey denotes undetected false negative markers. In this scenario K1 is identified by the natural kind approach as an instance of consciousness, whereas K2 is not.

Given this, it is easy to see how this can raise problems in the context of the methodological puzzle of consciousness if one adds to X the following supposition. Suppose that in our scenario X K1 coincides with access whereas K2 does not. The problem, argues Taylor, is that if this was the case and the natural kind strategy was implemented, it would lead to the conclusion that consciousness does coincide with conscious access. Yet this would not, according to Taylor be true, because K2 exists, and is an instance of consciousness without conscious access. The natural kind approach would thus yield a false verdict in this case - it and would seem to do so *necessarily* given that K2's properties T5-T8 are, in this scenario, all false negative markers. Thus, if this objection succeeds, the natural kind approach cannot resolve the methodological puzzle of consciousness in this scenario.

Part of what makes this case so compelling is that a multiple kinds scenario is, for someone looking seriously at the natural kind approach to consciousness, a very real possibility. The possibility that what we think of as 'consciousness' is underwritten by a series of distinct natural kinds is a central component of the natural kind approach to consciousness[76]. Indeed, given the inductive evidence we have from the history of science, we might go so far as to say a multiple kind view of consciousness is reasonably *likely*, if indeed consciousness is a natural phenomenon like any other. This is what Taylor

[76] Indeed, I argue in the next chapter that when this when it comes to the science of dreaming, the openness to a multiple kinds scenario is what ought to motivate an adoption of a natural kind approach to sleep experience over competing methodological approaches to the neuroscience and psychology of dreaming.

means when he says X is an 'empirically realistic' scenario. So, if the natural kind approach cannot distinguish between consciousness and access if this eventuality turns out to be true, then the natural kind approach and, by extension, ameliorative physicalism is in serious trouble.

Fortunately, I think there are good grounds for dismissing scenario X as a genuinely realistic possibility, grounds which are bolstered by a conception of consciousness as a natural kind. The central problem in Taylor's argument, as I see it, is an ambiguity in the assertion that "K2 does not coincide with consciousness access" in X. That is, this claim admits of two distinct readings:

A. K2 sometimes, but not always, coincides with conscious access (thus, K2 is sometimes co-instantiated with K1).

B. K2 never coincides with conscious access (thus, K2 is never co-instantiated with K1).

The essence of my response to Taylor, which takes the form of a dilemma, is as follows. In scenario X, either A is true of X, or B is true of X. If A is true of X, and consciousness is a natural kind concept - as the explicative formulation of the natural kind strategy here has argued - X is a scenario in which the natural kind approach would, over the course of its investigation, eventually classify K2 as an instance of consciousness. That is, Taylor's claim that in this empirically realistic scenario K2 would be inevitably misclassified by the natural kind approach as not conscious is false. We have an empirically realistic scenario here, but not one which is problematic for the natural kind strategy with respect to solving the methodological puzzle of consciousness. Alternatively, if B is true of X and consciousness is a natural kind concept, X is a scenario which is empirically implausible and even *precluded* by a properly naturalistic conception of consciousness. Either way, this response goes, scenario X raises no serious problems for the natural kind strategy as a means for solving the methodological problem of consciousness.

Let's consider the first horn of this dilemma first. Above, I claimed that if A is true of X, and consciousness is a natural kind concept, X is a scenario in which the natural kind approach would, over the course of its investigation, eventually classify K2 as an instance of consciousness. There are two reasons for thinking this. First, if K2 is often co-instantiated with K1 as per A, then we have reason to expect that at least *some* of K2's markers would be identified in the initial stage of an implemented natural kind methodology, and thus that, eventually, K2's remaining false negative markers would be identified[77]. The idea here is that for Taylor's argument to be compelling against the natural kind approach, it needs to be realistic not just to expect false negative markers to exist in this case, but to expect *all* of K2's markers to be false negatives[78]. But given K2's co-instantiation with K1 (at least in a subset of cases) this appears unlikely. When it comes to constructing a realistic scenario, there is a good case for thinking that our scenario X collapses into a more benign 'two kind' scenario of the kind

[77] Recall, these are markers that are we acknowledge as there, but falsely conclude are not indicative of consciousness.

[78] For as Taylor notes, if scenario X is not empirically realistic, the natural kind strategist need not worry about its inability to solve the methodological puzzle in this far-fetched case.

discussed in Shea (2012) described in the previous section — a typical multiple kind scenario analogous to the multiple kind investigation into 'jade' or 'hepatitis' in which our *known* markers in stage 1 of the natural kind methodology are found to track two distinct natural kinds (Y, Figure 4 below).

Second, even if we grant that a scenario in which K2's markers are all false positives - perhaps because K2 happens to only co-instantiated with K1 in a small number of initial samples[79] - there is still good reason to think that the natural kind approach will *eventually* succeed in classifying K2 as an instance of consciousness. That is, while on such a scenario it might be the case than an implementation of the natural kind methodology may *initially* yield an incorrect verdict about K2 (and thus an incorrect verdict about the nomological association between consciousness and access) there is no reason to think that over a longer period of the same methodological investigation K2 would remain misclassified as non-conscious. That is, given that K2 is both causally efficacious (i.e., it's not 'screened off' from K1 p.13) and is found in our initial samples of consciousness, there are grounds for thinking that it will eventually be uncovered and identified in a large number of the initial samples of consciousness. This can then ground an inference to the best explanation to the conclusion that K2 is an instance of consciousness.

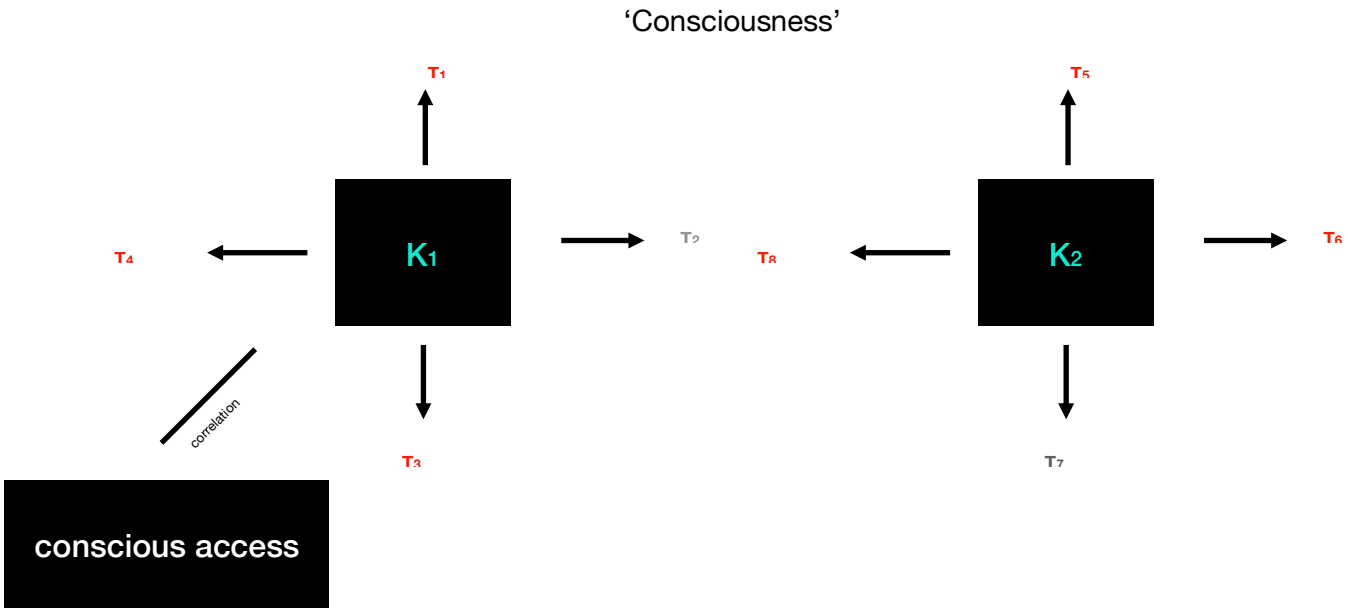

Figure 4: Scenario Y: A 'two kinds' view of consciousness involving an equal distribution of false negative and true positive markers. Red denotes true positive markers; grey denotes undetected false negatives. In this scenario, both K1 and K2 are detected by the natural kind approach.

What an opponent of the natural kind approach would need here is a positive reason to think that K2's properties would *necessarily* fail to be identified at an initial stage and, if they are identified later as

[79] Perhaps K1 is only co-instantiated with K1 in some forms of conscious access, but not all.

just suggested, to fail to be associated with consciousness. What might this be? One suggestion is that this has to do with how we first come to identify our initial markers — that is, how we come to associate our markers with consciousness to begin with[80]. According to this line of thought, the *only* way of verifying our initial markers as *markers of consciousness* is to determine whether they always co-occur with access. If that's correct, then, this would give us positive reason to think that K2's markers would systematically turn out to be false negatives — they would be dismissed as markers of consciousness in X because they (as stipulated) can dissociate at times from conscious access. But the natural kind approach need not commit to this claim about initial markers. It is 'built in' (so to speak) to the natural kind approach that a multiple kinds scenario, and particularly, one in which consciousness may dissociate from access, is a plausible eventuality (Shea 2012). As such, to build in a requirement on initial markers which ruled this out would be antithetical to the spirit of the approach. Instead, the natural kind methodology identifies its initial markers in a way which does not require that these always and in every case co-occur with access. This seems to be Shea's point.

What Taylor may seem to be getting at here is thus a different type of scenario — one in which K2 *never* does and never *can* occur with conscious access (i.e., that B is true of X). This brings us to the second horn of the dilemma. If this were true of X, then I think the stipulation that K2 is an instance of *consciousness* becomes suspect, particularly when one keeps the natural kind concept of consciousness firmly in mind. A natural kind concept of consciousness, recall, is just like any other natural kind concept —like our concept of jade, gold, or hepatitis. That is, if it refers at all, it refers to the homeostatic property cluster(s) associated with our initial samples of consciousness which underwrites a privileged epistemic role within a domain. On this conception of consciousness that I have been arguing for in this paper then, it would be *incorrect* to judge that K2 is an instance of consciousness. The starting stipulation that this is a multiple kinds scenario is a non-starter. The situation here would thus be analogous to one in which it is claimed that an instance of jade or hepatitis exists, but which (i) has nothing micro structurally or biologically in common with known cases of these phenomena, (ii) does not appear in our initial samples of these natural kinds (that is, it is not found to occur in any of the instances we usually call jade or diagnose as hepatitis)[81]. Neither, on this version of Taylor's scenario, would these cases have anything *epistemically* in common with known instances of jade and hepatitis, because the clusters have distinct and non-overlapping sets of property markers. They each, in other words, underwrite non-overlapping sets of inductive generalizations[82].

[80] Taylor, personal correspondence.

[81] For example, one might attempt to envision here an instance of the broader kind 'hepatitis' - hepatitis D which, unlike existing forms of hepatitis, is not found in the liver, causes none of the usual symptoms associated with hepatitis (for example, it doesn't cause the usual yellowing of the skin or eyes and when it present there is no visual signs of inflammation or damage to the cells in the liver). In the analogous scenario with respect to consciousness (ii) would also be true of K2, for it would *never* be present we introspectively attend to, and thus access our conscious experiences, or even when we demonstratively point to the properties we mean to pick out when we talk about consciousness.

[82] As Taylor writes: "different instances of consciousness are associated with separate underlying natural properties, which have different causal profiles… in such a case, the presence of properties in one cluster does not allow us to infer the presence of properties in the other (they are not 'co-projectable') and the two natural properties feature in separate inductive inferences" (8). As such, these two kinds violate the epistemic criterion of being an instance of the same kind defended by Magnus and discussed in section 4.

In these analogous cases, the stipulation that this new thing (the cluster satisfying conditions i-iii) ought to be classified as a case of jade or hepatitis is puzzling and easily rejected. There are no epistemic, chemical, or biological reasons for grouping these two clusters together and classifying them as instances of the same broader kind. So too, I would argue, on a naturalistic conception of consciousness, we can reject the stipulation that K2 would, if (B) is true of X, be an instance of the broader kind 'consciousness'. The intuition that K2 could be an instance of consciousness is, on this view, again one which owes its plausibility to a problematic conception of consciousness; one which refers to the apparent qualitative properties of experience which are inferentially isolated from physical and functional homeostatic property clusters. If ameliorative physicalism is embraced as a central component of the natural kind strategy, this pressing methodological objection to the natural kind approach can be avoided.

Despite not being ultimately (in my view) successful, Taylor's case against the natural approach is important and useful for those looking to implement the natural kind approach. The discussion of scenario X above highlights the importance of producing and devising further methodological measures within the three-step natural methodology to identify false negative markers and detect potential multiple kind scenarios. It also recommends caution with respect to hasty judgements and conclusions elicited by the natural kind methodology in response to methodological puzzles in consciousness science, particularly those which might be undermined by false negatives and multiple kind scenarios. These are important practical difficulties which the literature on the natural kind approach has yet to properly consider[83]. Moreover, I think the discussion of Taylor's case here is useful insofar as it further serves to demonstrates how retaining the original 'seemingly wonderful' conception of consciousness is epistemically problematic for consciousness science. That is, the discussion here of this potential methodological obstacle to the natural kind approach can viewed as offering further evidence for the epistemic superiority of a natural kind conception of consciousness for the purposes of bringing about a mature science of consciousness.

[83] I take this general point to also apply to other, less pressing methodological objections pressed against the natural kind approach (Philips 2018) which time constraints preclude me from discussing here. These are less pressing given that the main objection presented in Philips (2018) — the worry that there is insufficient agreement on the initial markers of consciousness to get the natural kind approach going — has already been discussed and responded to - I think, successfully - by Shea and Bayne in their recent paper (see 2020; section 5). However, Philips also presses a further objection which is worth briefly responding to here. Philips argues that problems arise for the natural kind response to the methodological puzzle of consciousness when we consider that it is possible - and indeed recognised by proponents of global workspace theory as a theory of conscious access - that there is a category of "preconscious" representations which could correspond to its own distinct homeostatic property cluster. Philips argues that this realistic possibility provides an alternative explanation for the presence of two clusters in the standard two kind scenario; an explanation which Shea (2012) asserts is best explained instead by the identification of these two clusters with access and phenomenal consciousness respectively (p7). Again, while I think that this discussion is useful in demonstrating that researchers need to be careful when implementing the natural kind methodology and interpreting its results, this objection does not work to undermine the potential of the natural kind approach to consciousness as a means for solving the methodological puzzle of consciousness. The idea here is that if it is indeed the case that we have good reason to posit the existence of a 'pre-conscious kind' which precedes and is distinct from conscious access then the one or two kind scenario can be replaced with the question of whether *two kinds or three kinds* are identified by a natural kind methodology.

7. The Objection From Primitive Concepts

Thus far I have been defending the claim that the natural kind approach has potential to resolve key methodological disputes in consciousness science, and thus that a natural kind conception of consciousness is justified as epistemically fruitful explicatum concept. But are there other reasons to reject this explicative strategy with respect to consciousness? The argument in this paper prompts an obvious line of objection which suggests that ameliorative physicalism can and ought to be resisted. This is the compelling thought that, in re-engineering our phenomenal concepts to natural kind concepts there seems to be a clear sense in which I am just *changing the subject* from a phenomenon which is distinctive and philosophically problematic to something which is explicable in scientific terms — a move which seems to miss the point of much recent philosophical discussion and fail to address a real scientific problem. Simply put, the argument thus far may prompt one to ask: is the proposal that we ought to re-engineer our phenomenal concepts to natural kind concepts which are empirically tractable a case of conceptual engineering gone too far? This worry, which expresses a form of skepticism about the limits of conceptual revision, is not unique to my proposal and will be familiar to readers up to date in the conceptual engineering literature. The well-known concern can be traced back at least to P.F. Strawson's critique of Carnap's method of explication:

> "typical philosophical problems about the concepts used in non-scientific discourse cannot be solved by laying down the rules of use of exact and fruitful concepts in science. To do this is not to solve the typical philosophical problem, but to change the subject" (Strawson 1963;505).

This is not the place to defend conceptual engineering from this objection. There is an extensive literature already on this[84]. For now, the claim in this paper remains a conditional one: *if* conceptual engineering is possible for a great number of philosophical and scientific concepts, then it is possible to conceptually engineer our concept of consciousness[85]. We can then ask, is there a way of formulating the Strawsonian objection which targets this more minimal, conditional claim?[86]. In closing, I would

[84] Maher (2007), Justus (2012), Brun (2016), Cappelen (2018), Nado (2019), Prinzing (2018), Simion and Kelp (2020) (Pinder 2021). Responses to Strawson's objection generally come in two types. First, one can offer and defend a criterion for subject preservation in conceptual engineering. Potential criteria proposals tend to come in two forms: (i) semantic approaches which claim that engineering is permitted when the sameness of meaning, or intension, is retained and (ii) functional approaches which argue that conceptual engineering is permitted when the sameness of function is retained. A second 'normative' response asserts that even in cases where subject preservation is desirable, a change of subject may be permitted when balanced against other epistemic and ethical considerations (Pinder 2021, Cappelen 2018;122). Either of these are strategies of response is available to the ameliorative physicalist. For example, in the case of (ii), one might appeal here to the idea that the relevant function is specifying the proper explanatory target of consciousness science.

[85] One might want to appeal here to recent philosophical and psychological arguments for the claim that philosophical concepts (if such a class of concepts exists) are already sorts of natural kind concepts — that is, that they comprise a bundle of replaceable and revisable inductive generalisations about the category they refer to (Strevens 2020). The idea here would be that if this psychological theory of philosophical concepts is correct, and consciousness is one such philosophical concept, there is a clear route to securing the possibility of engineering this concept to a more mature, specific kind of natural kind concept.

[86] According to a second prominent objection to conceptual engineering, projects of this sort are undermined also by meta-semantic externalism, of the kind endorsed by Putnam (1975), Burge (1979) and Kripke (1980) (Cappellen 2018 c.f. Haslanger 2000). For a recent critical discussion of this 'externalist challenge' and its

like to consider and address one such line of objection. This goes as follows. Even if one thinks that projects of conceptual engineering are plausible in general - that is, that a viable response to the Strawsonian objection to explication exists - one might think that there is something *unique about our concept of consciousness* which makes it specifically resistant to conceptual engineering. Here, I have in mind the claim that our phenomenal concept of consciousness, unlike most other concepts utilised in philosophy, is a *primitive concept* (Chalmers 2011, Fiala and Nichols 2019), and primitive concepts are unique in being resistant to engineering (Eklund 2015)[87]. If this is correct, then the proposal in this paper is a non-starter.

What needs to be established to successfully rebut this objection is not the claim that the ameliorative physicalist strategy is guaranteed to succeed, but that we have *no a priori reasons*, grounded in our primitive phenomenal concept of consciousness, to think that it will fail. Here, I will outline two strategies for establishing this latter claim. The first puts pressure on the idea that phenomenal consciousness is a primitive concept. The second offers a strategy for revising the ameliorative physicalist's claim in such a way that it avoids this concern, even if phenomenal concepts turn out to be primitive concepts.

The objection from primitive concepts just outlined relies upon the truth of three claims (i) that a class of primitive concepts exist, (ii) that primitive concepts (if they do exist) are resistant to conceptual engineering and (iii) phenomenal consciousness is one such primitive concept. Here, I grant the truth of the first two claims for the sake of argument[88]. The first strategy targets instead claim (iii). Primitive concepts, if they exist, are plausibly those concepts which are ubiquitous within and across populations. This relates to the compelling thought that primitive concepts are those *basic* or foundational concepts which emerge and stabilize early in childhood development (Fiala and Nichols 2019). In order for the objection from primitive concepts to succeed then, it follows that phenomenal concepts as described in §3.2 must be *widespread* among populations of individuals, in the sense of being possessed by the majority (if not all) of lay people and philosophers alike.

While it may seem obvious that this is the case, particularly from the perspective of philosophy of mind where the existence of phenomenal concepts is widely accepted (Chalmers 2018), the claim that phenomenal concepts are widespread within populations is directly challenged by a growing body of empirical work in experimental philosophy (Systema and Machery 2010, Systema and Özdemir 2019, Díaz 2021, Özdemir 2021) which suggest that lay people do not, by and large, possess the concept of

relevance to projects of conceptual engineering see Pinder (2021) who argues that considerations of meta-semantic externalism bear very little on the plausibility of conceptual engineering.

[87] Eklund refers to these as 'conceptual fixed points' but the general idea remains the same.

[88] That isn't to say that I think either of these claims are currently compelling. Chalmers (2011) and Eklund (2015) include recent discussions of (i) and (ii). By their own lights, neither author presents a definitive argument in favour of the thesis that primitive concepts exist. Eklund's argument for the claim that primitive concepts are resistant to engineering is also brief. This relies primarily on the idea that the engineering of some concepts (such as truth, ought, or existence) would generate inconceivable scenarios and result in *ineffable questions*. While this deserves further scrutiny, Eklund's argument is, by his own lights, inconclusive (2015; §6): ''what I have to say, even if correct, does not immediately show that conceptual engineering is impossible in these cases; it shows only something rather more subtle'' (for example, only that engineering is difficult in such cases).

phenomenal consciousness[89]. These studies come in numerous forms, all of which put pressure on this claim[90]. Of most direct relevance here are a recent set of studies which test the prevalence of 'epistemic' or 'explanatory gap intuitions' in individuals, measured by responses to questions such as "the properties of pain are fully explained in terms of neural activity in the DPI [dorsal posterior region of the insula]" (Díaz 2021). In one study, for example, participants were presented with two stories, one about the association between the subjective conscious experience of pain and neural activity in the DPI, another about the association between water and chemical composition H_2O. After reading each story, participants answered whether they agreed or disagreed with the following two claims: (i) an epistemic reduction thesis: "The properties of pain (like water) are fully explained in terms of neural activity in the DPI (analogous to the chemical composition H_2O)" and (ii) a metaphysical reduction thesis: "The feeling of pain (like the substance water) is just neural activity in the DPI (analogous to chemical composition H_2O)" (metaphysical reduction question)[91]. Contrary to what one would expect if phenomenal consciousness as a concept was ubiquitous in populations, Díaz found that participants agreed with the epistemic and metaphysical reduction theses of consciousness and the DPI to the same extent as the epistemic and metaphysical reduction theses of water and H_20. This was supported by a further three follow up studies[92].

This result is striking and, if reliable, suggests that lay people do not possess a phenomenal concept of consciousness (as of something which appears to be non-physical, non-functional, and non-informational) as the objection from primitive concepts appears to assert. Given the nature of these studies, this result is not conclusive and may well change in light of further follow up studies (c.f. Özdemir 2021). The results do, however give us sufficient reason to *currently doubt* the claim that phenomenal concepts are widely distributed within a population. This, in turn, gives us good grounds to doubt the claim that the concept of phenomenal consciousness is a primitive concept, and thus doubt that it is uniquely resistant to conceptual engineering. From the naturalistic and physicalist perspective from which ameliorative physicalism arises, the upshot of this discussion — that the

[89] One might think that this claim is already challenged by the existence of 'type-A 'physicalists (e.g., Eilan 2014) who deny that there is an explanatory gap between physical and phenomenal descriptions of, for example, feeling pain (footnote 3) however, the empirical argument presented here provides more substantial grounds for accepting this claim.

[90] Wierbicka (2019) further targets this claim on linguistic grounds, on the basis that cross-linguistic data suggests that English terms such as 'consciousness' and 'experience' do not find equivalents in all human languages. Again, while not conclusive, this suggests that phenomenal consciousness - unlike concepts like truth or existence - is not universally shared across populations.

[91] Questions were answered on a scale from 1 ('strongly disagree') to 6 ('strongly agree'). Participants were also asked to briefly justify their responses.

[92] As well as further studies from a different research team (Özdemir 2021). While in one follow up study Díaz found that participant's judgements did not change (when information relating to scientific discoveries of the association between pain and the DPI was removed from the vignette) two further studies did suggest that that a discrepancy in judgements between water and consciousness could be elicited if the vignette contained statements about the history or controversy surrounding the consciousness case ('the relation between them has been debated for centuries' and 'it remains controversial whether feeling pain is the same thing as neural activity'). In these cases, however, Díaz found that the justifications provided by the participants were not related to the nature of consciousness (as one would expect if these judgements reflected a possession of *phenomenal concepts*) but were rather were driven by factors not directly related to consciousness i.e. those relating to the immaturity of the current neuroscience at the time (2018; Sections 5 and 6). That is, Diaz found that to the extent that problem intuitions *could* be elicited in individuals, these were driven by consciousness un-related factors.

viability of the central proposal in this paper is ultimately decidable by further empirical work in experimental philosophy — is not, I think, a problem for the ameliorative physicalist, but rather a conclusion to be embraced. On this view, ameliorative physicalism provides us with a metaphysical position vis-a-vis the ontological status of consciousness whose plausibility can be settled on empirical grounds. This is concordant with a type of 'epistemically kosher' metaphysics which is said to survive the recent arguments pressed against the epistemic foundations of contemporary metaphysics (Ladyman et al. 2010, Ladyman 2012).

Drawing attention to the observation that phenomenal concepts are not widely distributed in the population presents one way of rejecting the claim that there are a priori grounds to reject the viability of ameliorative physicalism. I want to conclude here by outlining an alternative way of resisting the objection from primitive concepts which is compatible with accepting all three of the claims above and thus with the possibility that phenomenal concepts will turn out to be primitive concepts *contra* Díaz and others. Recall that the objection states that *if* phenomenal concepts are primitive concepts, this gives us reason to doubt that a project of conceptual *re-engineering* is untenable and likely to fail. Reflection on the nature of primitive concepts - as concepts which cannot be analysed in terms of other concepts, are incorrigible or *unchangeable*, and implicated in cognitive processes which are informationally encapsulated (Fiala and Nichols 2019) - suggests that such a claim is plausible (c.f. footnote 18). However, this sort of consideration says nothing to the possibility of conceptual *replacement* of primitive concepts with other concepts within a *specific* theoretical domain. That is, that x is a primitive concept may give us reason to think that x is not apt for conceptual revision, but the primitive nature of concepts plausibly has no bearing on whether concepts like x can or should be replaced, particularly within a theoretical domain. A good example in this context is Machery's proposal regarding the elimination (and subsequent replacement) of the concept of a 'concept' itself within neuropsychology with several more specific, empirically informed categories (Machery 2005, 2006, 2009). This strategy is consistent with *retaining* the concept of phenomenal consciousness within other philosophical domains, in accordance with Magnus' domain-relativity thesis with respect to natural kinds[93]. It is thus open to an ameliorative physicalist to adopt a clear 'de novo' strategy (Chalmers 2020) in this context. On this view, the ameliorative physicalist ought to re-interpret their central claim as one which proposes we ought to be engaged in a project of conceptual *replacement* with respect to consciousness, one which is limited to a particular *domain.* This allows them to thereby avoid the objection from primitive concepts[94].

8. Conclusion

[93] Indeed, such a strategy has already been argued for in the case of concept of *pain* (Corns 2016) which this revised proposal can be viewed as complementing.

[94] It is worth clarifying that the claim here is not simply that the ameliorative physicalist should concede the changing of the subject objection in the case of phenomenal consciousness. The claim is still that the engineered concept captures the proper explanatory target of consciousness science, and therefore is a concept *worthy of the name 'consciousness'*. Rather, this response to the objection does so whilst also allowing for a disambiguation from the original defective concept which may be retained for use in other contexts.

Physicalists in philosophy of mind currently face a dilemma between an a 'type A' view which denies the presence of an explanatory gap between physical and phenomenal descriptions of experience, but which is nonetheless useful for scientific enquiry, and a 'type B' view which grants the presence of an epistemic gap but ends up endorsing a concept of consciousness which is epistemically problematic when it comes to producing a mature scientific research programme for consciousness science. My aim in this paper has been to offer a way out of this dilemma by motivating an alternative, more attractive route to physicalism which is consistent with acceptance of an explanatory gap but moves away from descriptive accounts of phenomenal concepts toward normative projects which motivate claims about how phenomenal concepts ought to be. In addition to arguing that such an engineering project is warranted in the case of phenomenal concepts, I have argued that there is a promising candidate for an explicatum concept here in a natural kind concept as detailed by recent philosophy of science. The resulting view - ameliorative physicalism - was shown to significantly improve on existing a posteriori physicalism in two primary ways: first, by showing precisely how and why philosophers go wrong in critiquing these views for ignoring the way our phenomenal concepts actually are, and second, by providing physicalists with a concrete empirical methodology which is capable of resolving long-standing empirical disputes and facilitating progress in consciousness science. This renders the combination of ameliorative physicalism and the natural kind approach to consciousness a new and exciting research programme for scientifically minded physicalists which is apt for further uptake and development.

3| The Unity of Dreams: Two Approaches to the Science of Dreaming

A growing philosophical and empirical literature identifies and debates the nature, neural correlates, and biological functions of dreams. Much of this research operates on the assumption that there is a single kind of phenomenon — 'dreaming' — being investigated in dream science; a unified state of consciousness obtaining during sleep to which it is appropriate to attribute a single set of neural correlates and biological functions. This paper defends two claims. Firstly, that this assumption plays an active organising role in dream science which often goes unnoticed, shaping experimental design and orthodox interpretations of empirical data. Second, that the widespread acceptance of this unity assumption in consciousness science is unjustified — that is, that the unity of dreaming should be a hypothesis for consciousness science to test, and not an a priori assumption on which empirical research is predicated. I argue that recognition of this calls for a fundamental revision to the way in which sleep experience is studied scientifically. In replacement of the orthodox methodology which proceeds by motivating phenomenological definitions of dreams an alternative 'natural kind' approach drawn from philosophy of science is outlined. This opens up several new possibilities - notably, that folk psychological terms like 'dreaming' or 'sleep experience' may not constitute natural kinds.

1. The Unity Assumption

When we are asleep, we are not uniformly unconscious — rather, our sleep is interspersed with bursts of conscious mental activity. This conscious activity takes different forms thought the night: some experiences are complex, bizarre, and emotionally vivid simulations which appear to have long durations. Others are brief and fleeting; a vague visual image of a face or memory, or lone thoughts. In rarer cases, the state of unconsciousness which accompanies sleep is interrupted by a state in which one comes to have an awareness or insight into the fact that one is not awake, and in rarer cases still, a state in which one is able to partially manipulate and control the content of one's experience. When we compare these experiences while awake, we tend to refer to these uniformly as 'dreams'. But are

each of these experiences - the lucid dream I had last night, the non-lucid dream you had - experiences of the same neurobiological kind? What about the different sorts of conscious experience I have throughout the night — do all of the conscious experiences that I had last night while sleeping belong to a common kind, or to several different ones? These questions - which concern the ontological unity, and proper scientific classification of sleep experience - have long taken a back seat in dream science and philosophy of mind and epistemology. This paper considers them and argues for their importance.

The past thirty years have seen the emergence of an extensive empirical literature on dreams and dreaming, which now comprises a flourishing area of interdisciplinary research. The conceptual and empirical issues explored in this literature are vast, but central questions include: What are the phenomenological characteristics of dreams, and how do these differ with sleep stage? What are the neural correlates of dreams and prominent dream behaviours, and what are their biological function(s)? Reflecting the practice above, researchers here have largely assumed that underlying the phenomenological heterogeneity of dreams there is single kind of phenomenon — dreaming — being investigated; a claim which appears to be intimately connected to the dominant practice of specifying and motivating various *definitions* of dreaming from which claims about the unity and taxonomy of dreams follow (Pagel et al. 2001)[95]. In contemporary literature, orthodox views tend to fall within two camps: the view that 'dreaming' is synonymous with any form of sleep mentation (Cippoli et al. 2016, Siclari et al. 2017, Malinkowski 2019, Zadra and Stickgold 2021, Wamsley and Stickgold 2021, Siclari 2021) and the view that 'dreaming' refers to the majority of sleep experiences - but *not all* - viz. those which share a distinctive 'simulation-like' phenomenology, or an immersive, vivid 'here and now ' experience of a self in a world (Revonsuo, Tuominen, & Valli 2015, Revonsuo 2006, Thompson 2014, Windt et al. 2016, Windt 2010, 2016, 2018, 2020, 2021)[96].

As consciousness science begins to emerge as a mature scientific research field (Michel et al. 2019), it is time to consider whether the assumption that there is a single kind of conscious state associated with sleep experience — what philosophers and consciousness scientists refer to as 'dreaming' — is a good scientific assumption to make. This is not necessarily to determine whether this assumption is true, but *whether* its truth should be assumed from the outset as a guiding methodological claim, or instead, left open as a matter for empirical investigation to establish or disprove. My aim in this paper is to

[95]Other prominent issues concern the prediction and manipulation of dream content and its relation to sensory input (Horikawa et al. 2013, Revonsuo 2006, Zadra and Stickgold 2021;Chp2, Windt 2016 chp, Windt 2017, Rosen 2018, Nir and Tononi 2010, Schawartz and Maquet 2002) the nature of the transition from wakefulness to dreaming through via sleep onset, or 'microdream' imagery (Nielsen 2017, Windt 2016 ch11), the nature and significance of lucid dreams (see Section 2), and the relationship between dreams and sleep disorders and parasomnias like REM sleep behavioural disorder, somnambulism and night terrors (Nielsen 2010, Howell and Schenck 2015, Outdiette et al. 2011, Horvath et al. 2016. Remulla and Guilleminault 2004). This growing empirical interest is mirrored in a variety of hotly debated philosophical and methodological questions raised about dreaming, such as the reliability or trustworthiness of dream reports (Windt 2013, 2016 chp3-4, Rosen 2016), the relation between dream and forms of wakeful consciousness such as imagination, perception and mind-wandering (Windt 2016, 2020, Ichikawa 2009, 2016, Rosen 2015, 2020, Soteriou 2017, Irving 2018, Andrillon et al. 2019), and their broader epistemic import (Soteriou 2013, 2017, 2020, Sosa 2005, 2007, 2009, Ichikawa 2009, Crowther 2018).

[96] A comprehensive summary of different definitions of dreaming utilised in the empirical literature can be found Pagel et al. (2001).

argue against the former claim in favour of the latter. This, I shall argue, has far reaching implications. For once it is recognised both that this assumption plays a substantive methodological role in consciousness science and that it lacks the requisite conceptual and empirical justification, we will be forced to reconsider the way in which dreaming, and sleep experience are studied in consciousness science.

The plan for the paper is as follows. *Section 2* argues the assumption that dreaming is a single kind of state plays an active organising role in dream science which often goes unnoticed, shaping not only the way empirical results are interpreted, but also the way in which experiments in dream science are designed. I examine two debates in dream science in order to demonstrate this - one on research on the neural correlates of dreaming and one concerning the significance of lucid dreaming. I examine how orthodox claims and methodological moves made in these debates change if the unity assumption is abandoned or left open.

Given that this assumption plays a crucial methodological role in consciousness science, it is natural to think that this must be *justified* — that there must be strong empirical and theoretical reasons for thinking it must be true. Where issues of kinds and classification in dream science are raised in the literature, this is precisely what is claimed. The arguments here take the form of offering empirical and conceptual considerations in favour of particular definition of dreaming (as above), from which claims about the unity and wider classification of sleep experience follow. Section 3 and 4 examines this argument in detail and argues that it falls short of providing requisite justification for the unity assumption. Specifically, it is argued that once reconstructed, these empirical arguments fail on two accounts. Firstly, I argue that when it comes to justifying the unity assumption, these arguments are circular — the unity assumption itself plays a role in motivating these definitions of dreaming to begin with, and thus cannot be claimed to provide independent justification for it. Second, and most importantly, as empirical arguments these claims are question begging - it is the prior acceptance of the definitions *themselves*, and not, as its proponents claim, empirical data which motivates the acceptance of these definitions of dreaming. In prompting this critical discussion of popular definitions of dreaming, the issues raised in this paper will be shown to have a broader significance, revealing that the dominant methodological framework to dream science which precedes on the basis of these definitions is an epistemically risky and misguided research strategy.

A pressing methodological question for dream science subsequently arises: if the unity assumption shouldn't be assumed from the outset, and yet it currently plays a key role in organising research, how should dream research proceed going forward? In Section 5 I outline a new proposal for an organising framework for dream science drawn from recent work in philosophy of science which helps answer this question. On this view, dream science ought to be in the business of identifying a series of *global states of consciousness* which habitually or pathologically obtain during sleep which meet robust epistemic criteria for *natural kinds*. The key question for dream science on this view becomes the following: are any of the current definitions of dreaming good - in an *epistemic* sense - scientific categories, or should new and more fine-grained categories be proposed which better accommodate the neural, behavioural and phenomenal diversity of sleep experience? This latter suggestion brings

new possibilities - that of scientific eliminitvism about 'dreaming' and 'sleep experience' - into view, which I discuss briefly in closing. The central upshot, however, can be stated as follows: the unity of dreaming should be a hypothesis for consciousness science to test, not an assumption on which research is predicated.

2. The Unity Assumption at Work

2.1. Case Study 1: The Neural Correlates of Dreaming

What are the neural correlates of dreaming? Traditionally, this question has been framed in terms of the sleep-wake cycle: is there a particular sleep stage which is responsible for the occurrence of dreaming?[97]. Prompted by the discovery of REM (or 'paradoxical') sleep — periods of high-frequency, low-voltage activity characterised by rapid eye movements a near-complete loss of muscle tone (Dement 1999: 27–50; Jouvet 1999)— in the 1950s, an initial hypothesis sought to equate the physiological characteristics of REM sleep with the objective markers for the presence of dreaming (Aserinsky and Kleitman (1953), Moruzzi and Magoun 1995, Steriade et al. (2001) Hobson 1988: 154)). This gained support from the observations that individuals awoken from REM sleep report experiences which fit the traditional phenomenological profile of dreaming: vivid experiences which are narratively complex, bizarre, and hyperemotional (Hobson 2009)[98].

However, this hypothesis was soon thought to be largely repudiated by a series of findings, most notably, the finding that dream reports can also reliably be elicited outside of REM sleep in stages of N-REM sleep (e.g., Antrobus 1990; Foulkes 1993; Solms 1997, 2000; Cavallero et al.1990, Hobson et al. 2000, Nielsen 2000, Domhoff 2003; Nemeth & Fazekas 2018)[99]. While there were and continue to be a few dissenters, the idea that these findings show that dreaming and REM sleep are dissociable is

[97] Analogous questions raised in this paper about the proper taxonomy of sleep *experience* can also be - and indeed, are more frequently - raised with respect to the sleep stage architecture itself: what is the correct scientific taxonomy of sleep stage architecture? Are the current sleep stage categories too coarse grained for scientific research, or is the question of how to correctly taxonomise sleep a pragmatic matter? (Nielsen 2000, Dang-vu 2012, Windt 2020). While the discussion here, particularly the account of natural kinds sketched in (S.4), is also applicable to these questions and related issues about the local sleep occurring in N-REM, REM and wakefulness (Andrillon et al. 2019, Bernardi and Siclari 2019, Siclari and Tononi 2017), my focus in this paper is on sleep experience. I briefly discuss how the framework for sleep experience I recommend can jointly shed light on these questions in the final section of this paper.

[98] This and other findings in dream science depend on the prior acceptance of a number of methodological claims about dreaming which have historically been questioned, yet which I, following others, take for granted here – namely, that dream reports in laboratory settings are trustworthy or reliable (c.f. Rosen 2013), and that dreams reports reflect the occurrence of conscious experiences in sleep Malcom (1956), Dennett (1976). For a detailed discussion and defence of these and related methodological assumptions in consciousness science, see (Windt and Metzinger 2007) and (Windt 2016; Chp2-4).

[99] Awakenings from REM sleep tend to elicit reports around 80-100% of the time, while N-REM awakenings elicit reports of conscious experience around 23-75% of the time (Siclari 2021;2) Other findings that put pressure on this claim include: (i) that pharmacological suppression of REM sleep does not eliminate dreaming (Oudiette et al., 2012), (ii) that in a small number of cases, some REM sleep awakenings illicit no dream reports at all (Pagel 2003, Cipolli et al. 2013, Siclari et al. 2012), and (iii) that specific forebrain lesions (Solms, 1997, 2000) were shown to suppress dreaming without affecting REM sleep.

close to scientific orthodoxy in dream science. One finds it stated uncontroversially, for example, in recent studies (Siclari et al. 2017, Cipolli et al. 2017) and reviews of dream research (Baird et al. 2019, Siclari 2021, Wamsley and Stickgold 2021), as well as in contemporary philosophical literature on sleep and dreaming (Windt 2015, 2019, 2020,). For example:

"There is a persistent and popular myth, even within the scientific community, that dream experience originates exclusively within rapid eye movement (REM) sleep… subsequent studies, however, soon demonstrated that dreaming can occur in any stage of sleep, including even the deepest stages of slow wave sleep" (Wamsley and Stickgold 2021;2).

"The discovery of rapid eye movement (REM) sleep—the 'third state of being' besides wake and non-REM (NREM) sleep—led initially to a straightforward view of the neural correlates of dreaming: the wake-like, high-frequency, activated electroencephalogram (EEG) of REM sleep was thought to be associated with the presence of dream experiences and the low-frequency activity of NREM sleep with the absence of dreaming. However, later studies showed that up to 70% of NREM sleep awakenings yield reports of dream experiences" (Siclari et al. 2017;872).

"the proposed identification between sleep and dreaming was found to be flawed...dreams are also recalled with varying frequency following NREM sleep" (Windt 2020;5).

While it is recognised that one could resist this hypothesis by adopting narrower definitions of dreaming, such as 'hyper-emotional, and bizarre sleep experiences' which allow for a closer connection between dreams and REM sleep (Hobson 2008, 2009) or by otherwise positing 'covert' REM-sleep mechanisms operating within N-REM sleep (Nielson 2000), the point I wish to make here is that the orthodox interpretation of the findings above is reinforced by a prior commitment to the claim that there is a single kind of state being studied in dream science. That is, the force and significance of the finding that conscious experiences are reported outside of REM sleep depends on acceptance of the claim that the states being reported in both REM and NREM sleep stages are states of the same neurobiological kind. This is particularly salient in light of the finding that, as a general but not exceptionalness trend, REM and N-REM 'dreams' do in fact have distinct phenomenological characteristics and profiles: REM dream reports tend to be more elaborate, vivid, and emotionally intense, whereas NREM reports tend to be more thought-like, less emotional, non-progressive, repetitive, and more directly related to current concerns (Hobson et al. 2000, Nielsen 2000, Wamsely et al. 2007, Windt 2019, Mutz and Java 2017)[100].

Vindication of an alternative pluralistic hypothesis here would, of course, require further explanation of other important findings on the relationship between REM sleep and dreaming which influence the plausibility of this hypothesis. While I think this is certainly possible[101], the aim here is not to

[100] Siclari (2021;2) includes a useful table detailing the distinctive phenomenological and electrophysiological characteristics of conscious experiences across different sleep stages.

[101] For example, the finding that specific forebrain lesions (Solms, 2000) were shown to suppress dreaming without affecting REM sleep, has been criticised, due to worries about whether dreaming itself, as opposed to merely dreaming recall, is suppressed in Solms' studies (Takeuchi (2005).Similarly, given the very small number of cases in which REM dreams illicit no dream reports (Siclari et al. 2013), one might explain this by appealing to individual differences in dream recall. This gains support from the recent finding that there is large inter-

conclusively argue that the unity assumption is false. The aim instead is more modest: to show how a prior adherence to the unity assumption shapes and alters standard interpretations of empirical findings in dream science.

Adherence to the idea that there is a single kind of conscious state targeted and studied by dream science strongly influences not only how key findings are interpreted, but also how new studies to produce further data on the neural basis of dreaming are designed and constructed. The clearest example of this can be found in the contemporary research programme which aims to identify the neural correlates of dreaming. Here the explicit and self-conscious aim is to look for — and design experimental studies which probe — neural correlates of 'dreaming 'which are common to *all sleep stages* in which conscious experiences are reported (Siclari et al. 2017, Cipolli et al. 2017, Perogamvros et al. 2017, and discussion in Wamsely and Stickgold 2021). These 'within-state' paradigms (in contrast to previous 'between-state' studies (Siclari 2021)) are thus motivated and described as follows:

"Brain correlates of dreaming, then, must include mechanisms active in all sleep stages "(Wamsely and Stickgold 20201;2).

"Traditionally, dreaming has been identified with rapid eye-movement (REM) sleep, characterized by wake-like, globally 'activated', high-frequency electroencephalographic activity. However, dreaming also occurs in non-REM (NREM) sleep, characterised by prominent low-frequency activity. This challenges our understanding of the neural correlates of conscious experiences in sleep. Using high-density electroencephalography, we contrasted the presence and absence of dreaming in NREM and REM sleep" (Siclari et al. 2017;872)[102].

individual ('high 'and 'low recallers') and intra-individual variability (between stages and nights) in dream recall frequency and in perceptual and formal characteristics of dream experience (Cipolli et al. 2017, Schredl et al. 2003). The finding that pharmacological suppression of REM sleep does not eliminate dreaming (Oudiette et al., 2012), suffers from the same difficulties as the N-REM dream reports: it assumes a broad and unitary definition of dreaming. While I think these are promising avenues of argument and investigation, I do not take myself to be defending these claims exhaustively here.

[102] For a critical discussion of this influential study and the significance of its central finding that dream recall is associated with an increase in posterior high frequency EEG rhythms for NCC research, see Ruby (2020).

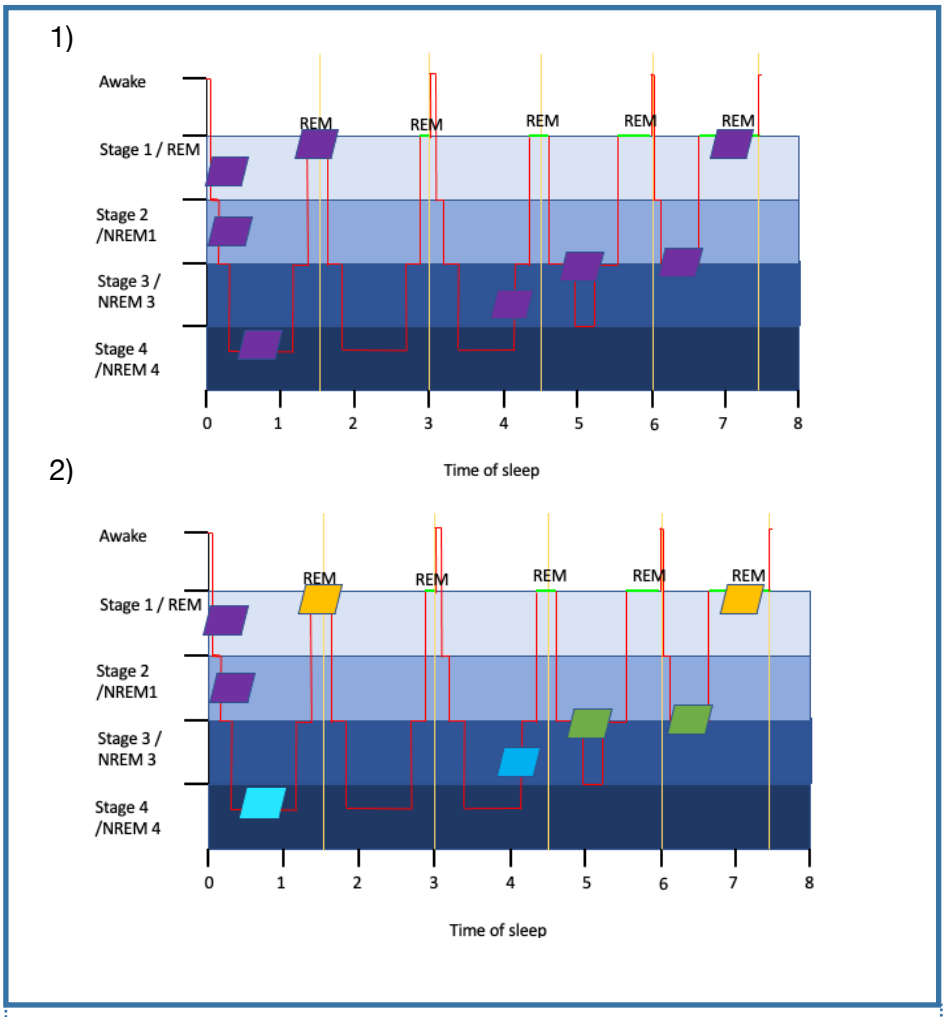

Figure 1: A monistic and pluralistic view of the relationship between dreaming and sleep stages. (1.) represents the standard view that 'dreams' form a unified kind common to all steps stages, in accordance with the 'within state' paradigms in contemporary NCC research. (2.) sketches one of several alternative views made possible by the denial of the unity assumption. Different colours denote different kinds of conscious sleep states.

This is doubly significant. First, focus on such a project makes evidence for the neural heterogeneity of sleep experience difficult to come by. Given that dream scientists are focusing on and actively searching for homogeneity, this may hinder examination of the sorts of neural diversity which would ground an empirical reason to reject the unity assumption (more on this later). Second and more importantly, if dream scientists are doing so without a prior justification for the claim that there is a single phenomenon being studied here (as I shall later argue), there is a significant risk that much of dream science has embarked on a project which is flawed from the outset — that is, a project which searches for a unified set of neural correlates common to all sleep experiences where there may not be any[103]. Once one brings an alternative pluralistic picture into view - or leaves this possibility open - the

[103] This is receives tentative support from the early observation that studies seeking to replicate Siclari et al.'s results have had limited success (D'Atri et al. 2019, Scarpelli et al. 2019, 2020, Wong et al. 2019, and for discussion Ruby 2020). However, it is important to note that the claim here isn't that there would thus not be *any* neural correlates common to these experiences if the unity assumption was false - they are both conscious experiences, after all - but rather that this would not be properly considered a hypothesis about the neural

mandate for an NCC research programme which bases experimental design on the search for a unified set of NCCs is, if not completely undermined, significantly called into question (Figure 1).

2.2. Case Study 2: Lucid Dreaming as a Methodology For the Study of Dreams

Part of what is so puzzling about the widespread acceptance of the unity assumption is the growing evidence that the conscious experiences that occur during sleep are incredibly phenomenologically diverse. Perhaps the most salient example of this is found in the contrast between 'ordinary 'dreams, in which the subject remains unaware that she is dreaming, and lucid dreams in which subjects become aware they are dreaming and, in some cases, are able to partially control the content of their dream experience (Dresler et al. 2012, Voss et al. 2013, LaBerge 1985, 1990, 2015, La Berge et al. 1986, Windt 2015)[104].

Despite not being ubiquitous in the population, lucid dreaming is now a central paradigm and 'experimental methodology' for the study of the nature and neural basis of dreaming [105]. This is because lucid dreaming allows for the removal of several methodological obstacles for studying sleep experience in the lab. Unlike ordinary dreams, whose content and occurrence can be determined only after the fact via dream reports and forced awakenings, lucid dreams can now be objectively verified using eye signaling paradigms— those in which participants are asked to move their eyes in a distinct pre-agreed sequence upon becoming lucid (Dement and Wolpert 1958, Hearne, 1978; LaBerge et al. 1981)[106]. This technique - lucid dream signaling through pre-agreed eye movements or 'PAEM') - also enables participants to 'time stamp' the start and end of instructed experimental tasks performed in lucid dreams, which allows for optimal examinations of the neural correlates, not only of dreaming, but of specific dreamt behaviours and contents (LaBerge 1990, Dresler et al. 2011, Erlacher et al. 2010, Oudiette et al. 2018 Baird et al. 2019, Konkoly et al. 2021)[107]. These exciting experimental innovations have led to the call for more research on how to pharmacologically (or otherwise) induce lucidity in

correlates of *dreaming as such*, as is typically assumed. On this view, the finding then that slow wave activity (spectral power in 1-4 Hz range) in the posterior hot zone is negatively associated with dreaming in NREM and REM sleep (Siclari et al. 2017) may work to support a hypothesis about the NCC of conscious experience generally (Bayne and Howhy 2016), but not of a neural correlate of a specific *state* of consciousness like dreaming.

[104]For an overview of historical references of lucid dreaming see LaBerge (1988), and for a recent in-depth overview of the cognitive neuroscience of lucid dreaming see Baird et al. (2019).

[105] The prevalence of lucidity in dreams is low in the population, with only around 20% of the adult population reporting one lucid dream per month, and only 51-55% reporting a spontaneous episode of lucid dreaming once in their life (Schrefl and Erlacher 2011). Even less (around 1%) have the ability to carry out signaling procedures as required for two-way communication paradigms (Saunders et al. 2016, Zadra and Stickgold 2021b).

[106]This is possible because eye muscles are exempt from the muscular atonia that accompanies REM sleep (Aserinsky and Kleitman, 1953; Jouvet, 1962).

[107] For earlier work in this vein see Tart (1965), LaBerge et al. (1981, 2001), and Fenwick et al. (1984). Recent bi-directional paradigms have been used to study a variety of sleep phenomena such as perception of elapsed time in dreams (LaBerge 1990), motor cortex activation during hand clenching (Dresler et al. 2011), voluntary control of respiration (Oudiette et al. 2018) and pursuit eye movements during tracking of visual imagery (LaBerge et al. 2018) (Baird et al. 2021).

non-lucid dreamers, in order to increase the number of available participants for dream studies (Mota-Romlin et al. 2019, LaBerge et al. 2018, Baird et al. 2019, Konkoly et al. 2021, Zadra and Stickgold 2021b)[108]. The long-term potential of these experimental innovations is often stated explicitly:

"These repeated observations of interactive dreaming, documented by four independent laboratory groups, demonstrate that phenomenological and cognitive characteristics of dreaming can be interrogated in real time. This relatively unexplored communication channel can enable a variety of practical applications and *a new strategy for the empirical exploration of dreams'*. (Konkoly et al. 2021;1417, emphasis added).

"Two-way, real-time communication between researchers and lucid dreamers immersed in REM sleep offers a new and exciting window into the study of *dreams and dreaming*" (Stickgold and Zadra 2021b).

While these experimental breakthroughs will no doubt substantially add to our empirical understanding of the nature and neural correlates of lucid dreaming, their advertised potential to revolutionise the study of dreaming *simpliciter* depends again upon accepting of a version of the unity assumption. The very idea that we can study the nature of dreaming by studying the nature and features of lucid dreaming, relies on an implicit commitment to the claim that lucid and non-lucid dreams belong to a common neurobiological kind[109]. The idea is that, in order for this to work, we need epistemic grounds for thinking that results in the latter case will *generalise* to instances of the former; that is, that we can legitimately infer that the relevant features observed and recorded in studies on lucid dreams will obtain in all relevant instances of non-lucid dreams.

Such a claim seems unproblematic when one is in the grip of the monistic idea that there is a single kind of mental activity involved in sleep experience. For this would give us grounds for thinking that features of lucid dreaming would generalise to non-lucid cases. However, once the unity assumption is called into question, or left open, the validity of this claim becomes dubious[110]. If the unity

[108]Methods to induce lucid dreaming tend fall into three categories: (i.) cognitive techniques, such as forced awakenings and developing intentions to remember that one is dreaming (known as the Mnemonic Induction of Lucid Dreaming – 'MILD' -Levitan and LaBerge, 1986; Neuhäusler et al., 2018) (ii.), external stimulation during sleep such as tMS, and (iii.) pharmacological intervention. For critical discussion of these methods and their effects on sleep quality and health see Vallat and Ruby (2021), and Soffer-Dudek (2020).

[109]It also relies on the further claim that the term 'lucid dreaming 'itself refers to a unified kind of sleep experience, a claim which can also be questioned on phenomenological and neurological grounds. In the latter case, for example, lucid dreams have been recorded not only in REM sleep but N1 and N2 stages of N-REM (LaBerge 1980, Stumbrys and Erlacher 2012, Mota-Rolim et al. 2015). See Baird et al. 2019 2.3 for discussion. Phenomenologically, lucidity has several core elements or features — most notably, insight or awareness that one is dreaming, dissociation (watching one's dream from an 'external 'observer perspective, such as on a screen Norieka et al. 2010) and control (the partial manipulation of one's dream plot) — which frequently dissociate from one another (Voss et al. 2009). Other more fine-grained taxonomies of the phenomenological elements of lucid dreaming are discussed in Holzinger and Mayer (2021).

[110]This point becomes particularly pertinent when one considers the fact that in addition to the substantive phenomenological differences between lucid and non-lucid dreams, lucid dreams also have distinctive neurophysiological characteristics, such as increased gamma band activity in fronto-temporal areas of the brain (Voss et al. 2009, 2014, Dresler et al. 2012, Baird et al. 2019). This has led some to suggest that lucid dreaming constitutes a distinctive 'hybrid state 'between dreaming and wakefulness (Voss et al. 2009, 2014, 2018, Soteriou 2017). This hybrid state hypothesis is one of the only clear examples of an explicit denial of the unitary assumption in the empirical literature, and thus lends support to, the central line of argument defended in this paper.

assumption with regard to lucid and non-lucid dreams is false, the experimental findings in the former case— for example, concerning dream enactment behaviours in lucid dreams — would not necessarily generalise to all non-lucid dreams, for these would be experiences of different kinds, and there is no guarantee that distinct neurobiological kinds will share the particular features in question. Here again, the seemingly innocuous acceptance of the unity assumption with respect to sleep experience is playing a substantive and active role is shaping how dream research is designed and conducted.

As above, the operation of the unity assumption here not only influences how experiments are designed in dream science, but also how experimental findings and their significance are interpreted. A good example of this concerns the use of lucid dreaming in arguments for and against different conceptual or philosophical views about the nature of dreaming. Here, a central question concerns whether dreams are constituted by imaginative or hallucinatory experiences (Windt 2016). Considerations drawn from lucid dreaming have been used to motivate arguments in favour of both views about the ontological constitution of dreams.

In favour of a perceptual model, Baird et al. (2019) present a new empirical argument against the view that dreams are imaginative experiences. This centers around the finding that during lucid REM sleep, subject's eye movements, measured via the intentional slow tracking of visual motion while lucid, mirror the patterns of eye movements which occur during tracking of visual motion in waking perception, while strongly diverging from the eye movements accompanying tracking of visual motion during waking imagination (LaBerge et al. 2018)[111]. Baird et al. take this as good empirical evidence in favour of the claim that dreams involve perception-like hallucinatory experience. The idea is that this finding not only suggests that the visual quality of dream imagery is on par with, or very similar to, the visual quality of waking perception but also allows for a disconfirmation of a key prediction of the imagination model of dreaming [112]:

"Together these findings help address several broad questions within cognitive neuroscience and sleep research. First, the data provide *empirical evidence* for a difficult to test question that has been asked at least since Aristotle: "are dreams more like perception or imagination?" (Nir and Tononi 2010). Based on the smooth tracking behaviour the findings suggest that, at least in this respect, the visual quality of REM sleep dream imagery is more similar to perception than imagination" (Baird et al. 2019, emphasis added).

One salient feature of wakeful imaginings which serves to distinguish them from perceptual and hallucinatory experiences is that imaginative experiences are typically under an agent's voluntary control. That is to say, while an agent typically has little control over what they perceive at any given time, the content and nature of one's imaginings are typically determined by the subject —- I can choose to imagine that I am in London when I am in fact in New York, for example, or imagine that I am in any place I like, when I like, at will. In this sense, the content of our imaginations are up to us,

[111] A Bayesian classification model discriminated both REM sleep dreaming and perception from imagination with greater than 98% accuracy (Baird et al.2019; 317).
[112]This argument both relies on and works to confirm the claim that the eye movements of REM sleep track the gaze of the dreamer. For a review of this "scanning hypothesis" (see Arnulf, 2011).

and this serves to distinguish them — at least initially — from perceptual or hallucinatory forms of experience over which we exercise little control.

This agentive characterisation of imagination, combined with the observation that lucid dreams reliably obtain, serve as one of the central arguments in favour of the imagination model of dreaming as defended by Jonathan Ichikawa (2009; 2016). According to this argument, lucid dreams provide support for an imagination theory of dreaming on the basis that, qua agentive states, lucid dreams demonstrate that dreams are in an important sense tied to a subject's capacity for mental agency. The idea is that, insofar as lucid dreams involve a significant element of control and exercise of one's mental agency, this precludes an analysis of dreams in terms of percept-involving hallucinations, for these latter experiences are not typically under a subject's voluntary control:

"The second reason in favor of the suggestion that dream experiences are subject to the will is that sometimes, some people have lucid dreams — dreams in which the dreamer, aware that she is dreaming, takes active and conscious control over the content of her dream. Percepts and beliefs are never under our active control, so the orthodox model can certainly not describe lucid dreaming; the imagination model fits the experience well. Insofar as lucid dreaming is similar in character to non-lucid dreaming, this provides reason to adopt the imagination model for dreaming in general." (Ichikawa 2009; 116).

These two arguments obviously point in different directions, and I will not attempt to adjudicate between them here. Rather, as before, my aim is to highlight the methodological role that the unity assumption plays in underpinning and supporting these arguments; a methodological role which is exemplified equally on both sides of the debate. In both cases, the authors move from a claim about some observed features of lucid dreams - that they are accompanied by certain patterns of eye movements, and a capacity for mental agency - to a general claim about the nature of dreaming *simpliciter* viz. that the visual quality of all dreams (or at least all REM dreams), and not just lucid dreams, are similar to waking perception, and that all dream states inherit, and are closely tied to, a subject's capacity to exercise mental agency over their dreams. However, when one calls the assumption that every sleep experience belongs to a common neurobiological kind into question, or starts with the view that this is an open empirical question, the validity of this standard argumentative strategy is thrown into doubt.

The preceding discussion ought to be sufficient to demonstrate that the truth of the unity assumption, often overlooked in methodological discussions in dream science, is not a merely a pragmatic or terminological issue which can be dismissed as a verbal dispute (Chalmers 2011). Rather, the metaphysical claim that conscious experiences in dreams are unified does substantive methodological and practical *work* in dream science, work that suggests that if unwarranted, the dominant research strategy for studying sleep experience adopted in consciousness science may be largely misguided.

Once recognised, it is natural to hope that the widespread acceptance of the unity assumption in dream science reflects the fact that there is strong empirical and conceptual justification for adopting this view at the outset of empirical enquiry. At the very least, one might hope that even if there isn't currently strong justification for it, that the standard methodological approach in dream science is in

a good position to be able to recognise this, and thus be receptive to the sorts of empirical evidence which could cast doubt on the truth of this claim. In the next section, I argue that neither of these claims are true. The problem, I will suggest, lies in the standard practice of working to specify and identify a clear *definition* of dreaming which can organise and unite different strands of experimental research. On reflection, this methodological orthodoxy is unable to substantiate or justify claims about the correct taxonomy of sleep experience. Nor, shall I argue, does it allow for the requisite flexibility and sensitivity in light of empirical discoveries which would allow for a gradual convergence on such an improved scientific taxonomy. This calls for an alternative approach.

3. The Definitional Approach to Dream Science

The unity assumption, as we have seen, is widely accepted. What explains this? An obvious answer is that this follows from standard views of what philosophers and dream scientists take dreams to be - that is, the operationalisation or *definition* of dreaming which guides their research. This approach to sleep experience, which proceeds via stipulating a definition of dreaming which forms the explanatory target of experimental research, is standard methodological practice in consciousness science. This was motivated historically by multiple incompatible definitions of dreaming which hindered the comparability of research findings (Pagel et al. 2001, Windt 2010, Revonsuo et al. 2015):

"Given this lack of consensus [on the correct operationalization of dreaming], it is critical for each researcher to provide their own operational definition of dreaming, in order to delineate the range of experiences that they are interested in studying. Here, we define the phenomenon of interest as encompassing all spontaneous subjective experiences arising during sleep, regardless of whether this mentation seems more or less "dreamlike " in the colloquial sense. Our goal is to describe the mechanisms that give rise to spontaneous subjective experiences during sleep, and to contrast those with the processes that support conscious experience during waking "(Wamsely and Stickgold 2021;2) [113].

As noted, two definitions of dreaming dominate the contemporary empirical literature[114]. The first, highlighted in the paper just quoted, states that 'dreaming' refers to any sort of conscious experience which occurs during sleep. The second, growing in prominence in recent work, states that dreams pick out those conscious experiences during sleep which are *simulatory* in nature. According to 'simulation models' of dreaming, a large subset (but not all) of our conscious sleep experiences are unified, and thus warrant classification into a single neurobiological kind 'dreaming', in virtue of sharing a distinctive 'simulation-like' phenomenology (Revonsuo, Tuominen, & Valli 2015, Revonsuo 2006, Thompson 2015, Windt et al. 2016, Windt 2015, 2018, 2020). On this view, the majority of our sleep

[113]See also "the first task of dream science is to come to a consensus on what dreaming is, and in the case that this can be accomplished, formulate a description of the phenomenon that majority agree on" (Valli 2011;1085)
[114]A third historically prominent view involves the idea that we can characterise dreams in terms of the characteristic formal phenomenological features of REM sleep experiences (Hobson et al. 2009, 2001). This view, associated with Hobson's AIM model of dreaming (and the earlier 'activation-synthesis' view), is not widely accepted, and has been abandoned by Hobson himself (Revonsuo and Valli 2015).

experiences are unified in virtue of sharing a 'core 'of phenomenological properties: they each involve immersive, vivid 'here and now 'experiences of a self in a world[115].

These phenomenological definitions of dreaming have different implications for how one taxonomises the conscious experiences reported in sleep. In doing so, each appear to involve an implicit commitment to different claims about the unity of dreams and sleep experience. For example, while the view that 'dreams' refers to experiences during sleep which are simulational in nature does suggest that there are some sleep experiences - such as non-immersive imagery, sleep thinking, isolated bodily sensations, selfless and contentless experiences - which are genuinely 'dreamless' (Windt et al. 2016, Windt 2016, 2018, 2020, Thompson 2016), both this view and the wider definition of dreams as sleep mentation both suggest that the *majority* of sleep experiences are unified in a sense that would justify the methodological practices discussed above. That is, if correct, they both seem to support and vindicate the methodological practice of searching for an NCC of dreaming which is common to all sleep stages, as well as the practice of generalizing findings on lucid dreams to many other 'non-lucid' dreams. This is because, according to these definitions, the same phenomenon 'dreaming' is occurring in all such cases.

This forms the basis of a standard argumentative strategy for defending the unity assumption, which goes as follows. The unity assumption with regard to dreaming is justified on the basis that it follows from conceptually and empirically motivated definitions of dreaming. The claim here is not merely that there are conceptual and phenomenological arguments for these definitions, but also that these definitions have strong empirical support. In, to my knowledge, one of the only other explicit discussions of the taxonomic issues raised in this paper, Jennifer Windt makes precisely this argument in her careful defence of the claim that 'dreams' are immersive spatio-temporal hallucinations (from now on, 'ISTH')[116]:

''Given its heterogeneity, it is important to realise, at the outset, that there may turn out to be no invariant and distinctive phenomenal core of dreaming. Dreaming may not turn out to be a natural kind. On pains of committing the armchair fallacy, one should not simply assume that the concept of dreaming picks out a class of experiences characterised by any single, highly invariant, and distinctive phenomenal property…. clearly having a unitary account of dreaming would be preferable, in view of its greater simplicity and parsimony. Such metatheoretical considerations do not, however justify pressing the target phenomenon of dreaming into a conceptual straitjacket…. *fortunately, as I will argue, it is possible to introduce a unified and distinctive account of dreaming that is maximally empirically plausible* and minimally legislative, suggesting only slight adjustments to the concept of dreaming" (Windt 2016;517; emphasis added).

[115] Different simulation views are distinguished by what specific properties are placed at the phenomenological core beyond this immersive "subjective world-for-me" (Revonsuo, 2006, p. 75). For example, one family of views include additional social elements to the simulation in accordance with the idea that the dreaming functions to simulate and strengthen social skills and bonds (Revonsuo et al. 2015). Others emphasise aspects of 'minimal selfhood' and weak phenomenological embodiment which make these immersive dream worlds possible (Windt et al. 2016, 2015a, b; Windt, 2018). In earlier work, Windt adds two further conditions to this definition: (2) under the assumption of reportability, (3) occurring in sleep (2010).

[116] See also the similar discussion in Windt (2010;299-301).

Windt's and Wamsley and Stickgold's comments here are indicative of a broader methodological approach to the study of dreams which is widely adopted. Detailed most extensively by Windt and other simulation theorists, this methodological approach starts by examining the phenomenological features of sleep experiences revealed by an examination of dream reports elicited under ideal conditions (Windt 2016). These phenomenological features are then examined and scrutinised in order to identify a "phenomenological core" of sleep experience. The idea here is to identify some core phenomenological properties or characteristics of sleep experiences - properties such experiences have in common - which, once identified, forms the basis of a definition of dreaming in terms of necessary and sufficient conditions. The resulting categories of sleep experience which emerge from this approach - viz. 'dreams' which possess this core, and 'dreamless' sleep experiences which lack it - subsequently go on to form the primary explanatory targets of dream research. That is, they guide, via providing operationalisations of 'dreaming' and 'dreamless experiences', empirical investigation into the neural correlates and biological functions of dreams. It is from within this broader methodological framework that the argument from definitions is provided and justified (Figure 2).

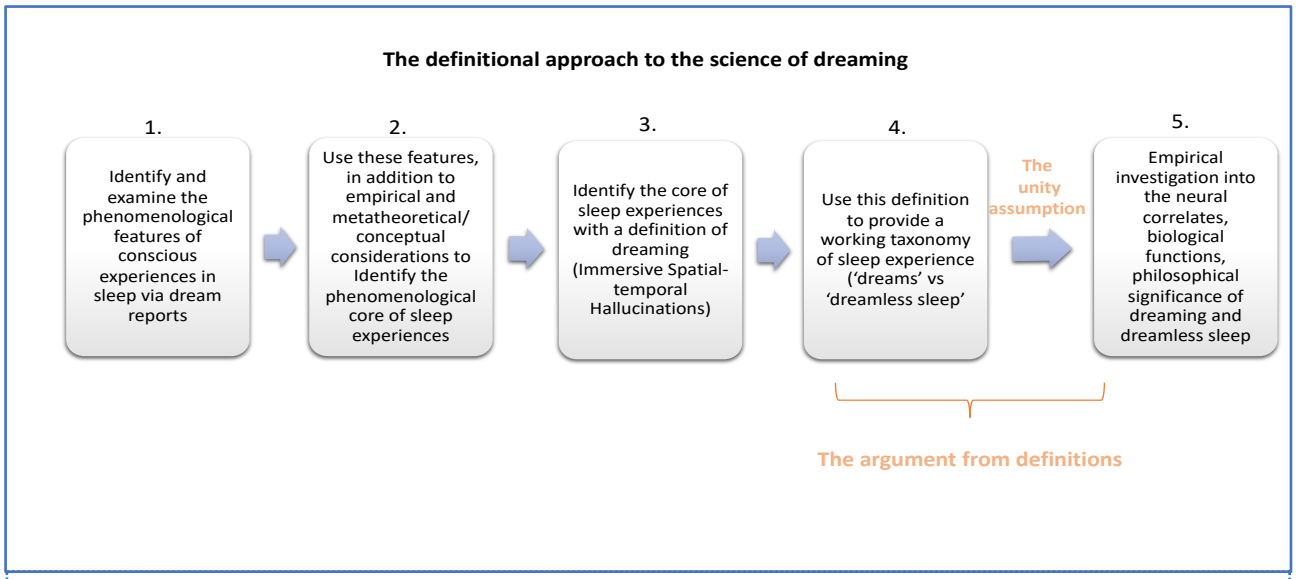

Figure 2: The Definitional approach to the science of dreaming.

If Windt and others are right, then the unity assumption is empirically and theoretically supported. In the next section, I raise two objections to this claim, focusing closely on Windt's arguments[117]. The first takes issue with the metatheoretical claim that definitions of dreaming can be used to justify the unity assumption at all. Rather I will argue that, on closer inspection, the definitions themselves are motivated in large part *by* the unity assumption. The second and most substantive objection takes

[117] Windt's defence and discussion of her ISTH definition of dreaming is most careful and extensive and is thus where the motivations for adopting the definitional approach to sleep experience are most clear and explicit. As such, my discussion here will focus mostly on the arguments she provides, on the assumption that the considerations offered in her work reflect the considerations motivating the phenomenological approach more generally.

issue with the claim that there are strong empirical considerations in favour of these definitions — that is, that empirical data does in fact support or 'tightly constrain' Windt's simulational definition of dreaming as is claimed[118]. This latter claim, in combination with the arguments in the previous section, have broader methodological significance, for they reveal that the definitional approach to dreaming which these definitions form a part is generally inflexible and insensitive to empirical evidence in the manner required if a precise and scientific taxonomy of sleep experience is to be provided. This calls for an alternative approach.

4. Two Objections to the Definitional Approach to Dreaming

4.1. The Circularity Objection

In order for the argument from definitions to work as an *independent* argument for the unity assumption — one which is capable of providing independent support for the substantive methodological role this assumption plays in consciousness science — the arguments provided for accepting a given definition of dreaming must proceed without already assuming the truth of this assumption. That is, if there is to be a compelling argument for adopting the unity assumption in dream science based on a definition of dreaming, it ought to be the case that those considerations of unity do not themselves constitute part of the argument for these definitions. For if this were the case, then the appeal to definitions here would not amount to an argument for *accepting* the unity assumption in consciousness science, so much as a circular argument which examines what follows *from* it. The first objection to the argument from definitions, and the broader definitional approach to dreaming, I wish to raise states that this later scenario is what we find when we examine the conceptual arguments offered in favour of contemporary definitions of dreaming. Thus, the argument from definitions falls foul of circularity.

That the prior acceptance of the unity assumption plays a motivating role in the definitional approach to dream science can be evidenced clearly in two earlier steps of this methodological approach (Figure 2). The first place this can be evidenced is in the very idea of a 'phenomenological core' of sleep experience which is said to track the neurobiological kind 'dreaming'. To see this, consider first Windt's remarks which motivate the idea of a phenomenological core in her book and seminal article on the immersive spatio-temporal hallucination account:

"Rather than attempting to capture the characteristic phenomenological and profile of dreaming [e.g., Hobson et al. 2000], the goal of this paper is to identify its phenomenological core. *The idea is that the target phenomenon ought to be, first, something that is invariant and stable across different types of dreaming*." (Windt 2010, emphasis added).

[118] Windt (2016;518), see also the introduction to her book where she discusses the need for any theory of dreaming to be empirically plausible (2016;1).

"In searching for the phenomenal core of dreaming we are, after all, searching not just for any kind of maximally invariant property of dreaming but for something that will yield the conceptual glue, if you will, to *hold different types of dreams together*". (Windt 2016;519; emphasis added).

The explicit aim of Windt's project is to examine the subjective reports of sleep experience elicited under ideal laboratory conditions in order to identify a set of phenomenological features that all, or most, of these reports *have in common*. The idea is that there is a 'core' of phenomenological features which underlie all cases of dreaming, even if these features are rarely, if ever, experienced in isolation from other aspects of dreaming. But what mandate have we for thinking, if such an invariant phenomenological core exists, that this phenomenological core successfully tracks a distinctive neurobiological *kind?* That is, what motivates the move from Step 3 to Step 4 in the definitional approach to dreaming [Figure 2]? The answer I think, lies in the prior acceptance of, or sympathy to, the idea that there exists – or is likely to exist - a single kind of conscious state associated with sleep. In other words, whether [as Windt claims] we have good reason to investigate which phenomenological features *unite* the diversity of sleep experience depends on whether there is *already* good reason to think that there is, or is otherwise likely to *be*, a deep metaphysical unity to sleep experience -- and moreover, that this is likely to be reflected in the phenomenological features common to the majority of sleep experiences [119]. It is because the unity assumption is already implicitly accepted from the outset that philosophers and dream scientists have thought that Windt's project and those like it are reasonable research strategies to pursue. It is useful to keep in mind here that the idea is not to call into question the idea that a phenomenological core to the conscious experiences one has while asleep *exists* – given that the phenomena in question are conscious experiences this is certainly likely. Rather, the point is to call into question the motivations for thinking that this phenomenological core are of primary relevance to the question of how to correctly taxonomise sleep experience[120].

The second step in the definitional approach where a prior acceptance of the unity assumption can be evidenced is in the conceptual or 'metatheoretical' justification provided by Windt and others for the simulation definition of dreaming (that is, in Step 2). That is, the reasons these authors provide for accepting their definition over other candidates in the first place. Here, simulation theorists argue that the simulational model of dreaming is attractive insofar as *it alone can unite* many sleep experiences within its purview. Similarly, they maintain, it is attractive as an organizing operationalization of

[119] In later work (2016; Chp11) Windt combines the phenomenological core with a functional criterion for sleep (which is at odds with Windt's previous discussion of the possibility of extending the concept of dreaming out of sleep (2010)). However, the point made here vis-à-vis the motivations for identifying the phenomenological core of sleep experience remains the same.

[120]To see this compare how one might go about producing a taxonomy of any three conscious experiences – X, Y, Z – which occur while a subject is awake. Given that these are conscious experiences, one would expect that at some level of description it is possible to identify a core of phenomenological properties these experience share. The important question, for our purposes, is what reason we have for thinking that this phenomenological core is indicative of (in the sense of tracking) the fact that XYZ all belong to a distinctive neurobiological kind, over and above being instances of conscious experiences. The answer here lies, I think in the idea that we already have compelling reason to think that this is the case. That is, that this view only becomes compelling if one *already* has good reason to think that these experiences are each instances of the same neurobiological kind.

dreaming because it provides and makes room for a definition of dreaming which is *independent of sleep stage*[121]:

"a proper theory of dreaming *should be simple yet covering so that the same general principles apply to many types of dreams*, including the pathologies of dreaming, animal dreaming, and other special cases" (Revonsuo et al. 2015 in 'how to make theoretical progress in dream science'; emphasis added).

"spatiotemporal situatedness is a promising candidate for identifying the phenomenal core of dreaming for a number of reasons. To begin with, it suggests a unified and state-independent account of conscious experience". (Windt 2016;521).

The reasoning here can, I think, be summarised as follows. What justifies the organising methodological principle that there is a single kind of state being studied in dream science - one which justifies claims such as the orthodox view that there is a single sleep-stage invariant state being studied in dream science? Well, this metaphysical claim follows as a consequence of adopting a prominent definition of dreaming, such as the view that dreams are ISTH reported in sleep. What justification do we have for adopting this definition? Well, these definitions receive justification on the basis that they pick out a phenomenological core of sleep experience which can unite these within a single definition, as well as the fact that this provides a definition of dreaming independent of sleep stage. This line of argument suggests that there is a problematic epistemic circularity in the argument from definitions which gives us reason to doubt whether popular definitions of dreaming should be used to independently justify the methodological practices outlined in Section 2.

It might reasonably be objected here that my argument is somewhat uncharitable in light of the earlier discussion. Is it not stated earlier [and directly in the quotes above] that these definitions are not *only* supported by theoretical considerations like those just discussed, but also tightly constrained by empirical evidence? If this is so, then perhaps the circularity just highlighted is not of a problematic, or vicious kind. For, even if there are some concerns about circularity in the way these definitions have been motivated by Windt and others, if there is nonetheless strong empirical evidence in favour of a definition of dreaming like the simulation view, then the argument from definitions will be capable of providing independent justification for the unity assumption. I turn to arguments of this form in the next section.

4.2. The Empirical Insensitivity Objection

As naturalistic philosophers of mind and scientists working toward an empirical framework for dream science, it is unsurprising that the reasons Windt and others provide for their definitions of dreaming are not only phenomenological or conceptual in nature, but also empirical. What then, is the empirical evidence which is said to support the view that dreams are immersive spatiotemporal hallucinations

[121] This idea also plausibly supports the widespread adoption of the definition of dreaming as any conscious sleep mentation.

— that each dream necessarily involves a basic sense of spatiotemporal location, an experienced self in a world? The empirical arguments Windt offers come in both a negative and positive form, with the negative arguments drawing on empirical considerations to discount or reject other plausible candidates for a phenomenological core of sleep experience, and the positive argument drawing on empirical considerations in order to provide motivation for the view that the phenomenological core of sleep experience is in synonymous with an ISTH (2010; 2016, 2018).

Let's consider the negative arguments first. Before endorsing the ISTH definition of dreaming, Windt considers several further candidates for a phenomenological core of sleep experience drawn from previous literature which could work as a definition of dreams. For example, the view that dreams necessarily involve the integration of multi-modal imagery (Hobson et al. 2000, Solms 2000), that they necessarily involve characteristic cognitive and mnemonic deficiencies, bizarreness etc. (Hobson et al. 2009), and that they necessarily involve emotions. Windt's argument against these views center around the claim that there are empirical studies of dreams which reveal counterexamples to these hypotheses — that is, that there are studies which reveal the existence of dreams which lack these features. Thus, she argues, that (i) a variety of studies show that imagery in non-visual modalities is rarely mentioned in dream reports (Hobson 1988), (ii) that visual imagery can be lost while dreaming remains, as in studies of subjects with congenital blindness (Kerr 1983), suggesting that visual imagery is not necessary for dreaming, (iii) that the existence of *lucid dreams* suggest that cognitive deficiencies are not necessary for dreaming, and similarly (iv), that small numbers of dream reports contain no mention of emotion, and thus that emotion cannot qualify as a central 'core' of dreaming (Merrit et al. 1994)[122]. These arguments all have a similar structure. They work by providing counterexamples to an alternative definition of dreaming (that is, proposal for the phenomenological core of sleep experiences), revealed by empirical studies of dream reports.

Windt's positive empirical argument for the ISTH view proceeds via the claim that empirical studies of dream reports show that a basic sense of self is invariant amongst distinct kinds of dreaming. This centers on the finding that dream reports consistently involve the presence of the *self*, albeit one which is only weakly embodied, which Windt takes to provide positive evidence for the claim that the minimal core of dreaming is self-location centered on a hallucinated world (Strauch and Meirer 1996, Occhionero et al. 2015, Speth et al. 2013). Here, she argues that empirical work has shown that underlying the variability of self in dreams (as it relates to embodiment, beliefs about oneself in the dream etc.) is a core of *minimal phenomenal selfhood* — a minimal sense of self which involves the experience of immersion or being present in a world (519-523). This more minimal form of selfhood, she notes, is compatible with the experience of 'selfless' dreams (i.e., the absence of more robust aspects

[122] In later work (2016), Windt argues that emotions are a more promising candidate given the literature suggesting that emotions are reported in all dreams (Sikka et al. 2014). Here, she offers here two conceptual arguments against the view that emotions form a part of the phenomenological core: (a) that this would mean that isolated emotional experiences occurring during sleep (terrors) would count as dreams, a claim she said would 'obliterate' the distinction between dreams and night terrors. and (b.) the worry that emotions could not form the 'conceptual glue' which explains the core of dreaming (2016; Chp5). I think there are good reasons to be skeptical of these conceptual arguments, but I do not discuss this further here. My focus in this section is on the empirical considerations Windt offers.

of self), and instead picks out a 'spatiotemporal situatedness' — a 'here and now' experience of being present in a world. There is more to be said about the further conceptual arguments Windt provides for this model, however as far as the empirical considerations go, the argument is a simple one viz. that empirical studies reveal that all dreams (via dream reports) possess phenomenological characteristics of a self in a world.

I think both the negative and positive empirical arguments here suffer from the same epistemic problem. In short, these arguments are problematic insofar as they require one to already know or accept a claim about the proper extension of the term 'dreams'. Consider again the negative arguments just sketched and reflect on how one might go about examining or critiquing them. An obvious strategy here is to challenge the idea that the empirical studies mentioned do in fact demonstrate that there are dreams which lack these features. But how does one go about assessing this claim? To be able to assess whether a given empirical study demonstrates that there are dreams which lack feature X, for example, one must *already* have in mind a particular view of what dreams are – that is, *which* conscious experiences in sleep count as dreams - to be able to determine this. Similarly, to be able to assess whether the view that all dreams have at their core an immersive, spatiotemporal situatedness is in fact common to all dream reports or not, one must already know which reports of consciousness in sleep are *relevant* to such an assessment. But, if the aim is to motivate a particular definition of dreaming and subsequent taxonomy of sleep experience which can reliably guide empirical research in the way Windt envisions, then one cannot already assume an answer to this question.

The situation gets worse when one considers how Windt and other simulation theorists respond to potential empirical counterexamples to their view - for example, the presence of conscious experiences in sleep revealed by dream report analysis which *lack* this immersive minimal core. Here I have in mind the isolated imagery, hypnagogic imagery, and the 'white dreams' (dreams which potentially lack determinate contents) reported in sleep. For here it is open to simulation theorists to simply *dismiss* such counterexamples to their definition, on the basis that these are distinct kinds of '*dreamless*' experiences - and this is exactly what they claim (Windt et al. 2016). However, in the context of the negative arguments above, this raises the obvious question: why are we licensed to accept *these* as dreamless experiences, but not the counterexamples raised to other definitions? The answer here lies in the idea that these arguments depend on a prior acceptance of a claim about the proper extension of the term 'dreams'; that is, a prior acceptance of a particular definition of this term. Thus, on closer inspection we find that these empirical arguments are question-begging in a problematic sense. The empirical evidence here only supports an ISTH definition of dreaming if a general view of dreaming and broader taxonomy of sleep experience is already accepted. This applies mutatis mutandis to other definitions of dreaming. This leaves us with the circular justification critiqued in the previous section.

The problems discussed here, are I think, indicative of a wider problem for the definitional approach to dream science which go beyond specific simulation models. The foregoing discussion suggests that popular definitions - and subsequent taxonomies – of dreams and sleep experience which currently guide research are insensitive to empirical evidence and revision in a way that dream scientists and naturalistic philosophers of mind would want to avoid. That is, when aiming to produce a scientific

taxonomy of any phenomenon, one typically expects that one's initial taxonomy to be subject to further (and potentially quite radical) revision in light of new evidence and findings. The wider problem for the definitional approach to dream science that these issues highlight is that this is not obviously possible within the existing definitional framework for dream science. That is, given the arguments just discussed, it is not clear how a definition like ISTH *could* be revised in light of new evidence about the phenomenological or functional features of dreaming. For example, when faced with a potential new counterexample to this definition of dreaming from recent empirical literature, a proponent of a simulational view can simply *define away* these results – they can maintain that this new finding instead reveals features of other, *distinct* kinds of 'dreamless' sleep experience. Similarly, by fixing the extension of dreams in order to determine which evidence is relevant to an assessment of the claim that 'all dreams share a core immersive phenomenology', the proponent of the simulation model of dreaming can continue to maintain that, despite new findings about the phenomenological and neural features of sleep experience, that all positive evidence in favour of their view remains the same. There is a broader insensitivity to empirical evidence here, which suggests that the definitional approach to sleep experience is epistemically risky (given that the taxonomies they produce might well turn out to be incorrect) as well as misguided research strategy (that is, despite being intuitively compelling, this methodological framework is not empirically sensitive such as to reliably guide experimental research). This calls for an alternative approach to the science and philosophy of sleep experience.

5. The Natural Kind Approach to Sleep Experience

The question at the heart of our discussion thus far — is there a single kind of conscious state which we surface to while asleep or several — is a question about scientific categorisation. It asks us to reflect on the following: is the correct scientific taxonomy for consciousness science one which posits a single category 'dreaming' to capture the nature of sleep experience or one which posits several distinct categories? This question itself invites a host of further questions which are likely to have already occurred to the reader throughout the course of this discussion viz. what does it mean to say that one scientific taxonomy is the 'correct' one? And how could we know whether the system of categorisation we have currently in dream science corresponds to this correct classification or not?

These questions are not unique to consciousness science and arise in every scientific domain. In the philosophy of science, questions of this sort are said to revolve around and point to a crucial feature of the world, on which there has been a long and complex philosophical discussion. To ask questions of this sort is to ask about *natural kinds* - that is, those scientific categories which reflect real divisions in nature as opposed to categories (such as 'red and blue striped objects') which are arbitrary or gerrymandered. In this final section, I want to consider how contemporary research on these questions in philosophy of science can be used to help illuminate and make progress on the issue concerning monism vs pluralism in dream science which has occupied this paper. This will involve combing answers to three questions, which I will now consider in turn: what makes a category or kind natural

as opposed to arbitrary? What kind of mental category is dreaming, and can we understand distinct members of this category as forming natural kinds?

5.1. What is a Natural Kind?

When thinking about how best to categorise, and thereby understand, the world, it is natural to start with a distinction between natural and non-natural categories. When it comes to understanding the world, we typically think that there are some categories better suited to the task than others — and this, we might think, reflects the fact that these categories work to approximate a structure and divisions that are *really present* in the world, and not merely contingent ways of carving up reality. The question of what distinguishes natural from non-natural categories - what distinguishes categories like 'electron' and 'planet' from 'libra' or 'pixies'- is at the heart of recent research in philosophy of science. Here, one typically finds the claim that genuinely natural categories can be distinguished from non-natural categories on *epistemic* grounds, insofar as the former but not the latter are *projectable* in the sense that they feature in and explain the epistemic reliability of a large class of enumerative inductive inferences (Goodman 2000, Boyd 1991). The key idea is that natural kinds explain and justify forms of inference which generalise from a limited number of observed items (*this F is a G)* to a claim about all unobserved members of the kind to which the items belong (viz. *'all Fs are likely G'*) (Sankey 1997, 2007, 2001, Kornblith 1993).

This epistemic feature characteristic of natural kinds has recently been developed into a comprehensive theory of natural kindhood by the philosopher P.D. Magnus (2012, 2013), who argues that for any give category k to be natural, it must satisfy two conditions:

(1) *The success clause*: k is a part of a taxonomy that allows scientific enquiry into d to achieve inductive and explanatory success.
and
(2) *The restriction clause*: any taxonomy that excluded k would not do so.

The thought here is a fairly intuitive one. Natural kinds are those categories which are indispensable for a scientific domain, being those categories which science is forced to posit in order to achieve explanatory success. On this view, what distinguishes natural categories like 'planet' or 'Homo Sapiens' from non-natural ones like 'libra' or 'pixies' is that the former are *epistemically fruitful* in the sense that they not only feature in and work a large range of inductive inferences, but are also the categories that *must* be posited, for science to achieve inductive success.

While many contemporary philosophers of science take these epistemic features to be definitive of what it is to be a natural kind, many — including Magnus — argue that there are nonetheless specific ontological structures common to natural kinds which work to *explain* how and why natural kinds

meet these epistemic conditions[123]. In higher level scientific domains for example, which the discussion in this paper concerns, it is natural to combine the epistemic criteria above with the *Homeostatic Property Cluster* theory of natural kinds made popular by Richard Boyd (1991, 1994, 1999)[124]. According to this ontological theory of natural kinds, natural kinds in higher level domains (including neuroscience, psychology and so on) are identical to a cluster or repetitions of properties whose reliable co-occurence in nature — and thereby, their epistemic *projectability* - is explained and maintained by the operation of a network of historical and constitutive causal mechanisms[125].

While this summary of recent work on natural kinds is inevitably cursory, it is illuminating vis-a-vis the issues thrown up in this paper in several respects. First, it suggests that the question is there a single kind of conscious state associated with sleep, or several, and the related debate concerning how to define and operationalise dreaming is one which can be reframed in terms of *epistemic utility*. That is, it allows for a more precise reformulation of the question central to this paper, which goes as follows. Are a wider set of inductive inferences in consciousness science underwritten by a taxonomic system which posits a single category of sleep experience, or by an alternative taxonomic system which posits several categories? In particular, which categories of sleep experience are we, qua scientists, *required* to posit in order to achieve explanatory and inductive success?

Second, the discussion here also suggests that in seeking to answer these questions, dream scientists ought to be looking to identify a series of *clusters* of properties associated with consciousness during sleep. These clusters, as above, are those which reliably co-occur among individuals of our species, and whose joint occurrence can be explained reliably by a network of neural mechanisms. This provides a methodology for adjudicating between monism and pluralism about sleep as experience as follows. If we find that there is a single cluster of properties - one which functions as a whole to underwrite a reliable set of inductive inferences - underwritten by a single set of neural mechanisms, we would have good grounds for concluding, for example, that the standard monistic definition of dreaming (one which equates dreaming with sleep mentation) is the optimal or 'correct' scientific taxonomy. Conversely, if, after rigorous investigation, several *distinct* clusters or groupings of properties were found, clusters which were individually *indispensable* for explanatory success in consciousness science, and underwritten by distinct neural mechanisms, we would be liscenced in rejecting the unity assumption in favour of a pluralistic thesis[126].

---

[123] In doing so, Magnus thereby makes a distinction between a 'taxonomic' question in the vicinity of natural kinds (what distinguishes a category which is natural from an arbitrary category?) and an ontological question (what feature(s) in the word allows natural categories to perform this function). I will not discuss the distinction between these two questions further here. For discussion see (Magnus 2012; chapter 1, 2013).

[124] See also Kornblith (1993); (Griffiths 1997, 1999), and Wilson (1999, 2005).

[125] For example, individual members of a biological species share important morphological, behavioural and physiological features. Here, the common co-occurence of these properties is not a brute fact but is rather accounted for by the repeated operation of a causal mechanism which *explains* the repeated or 'homeostatic ' co-occurence of these clusters in nature. In this case, for example, similarity is explained by historical causal mechanical structures (such as gene recombination, gene transcription, similar selection regimes and common developmental constraints, Boyd 1999, Ereshefsky 2005).

[126] This mirrors and is thus continuous with a recent approach to the study of phenomenal consciousness based on natural kinds, which advocates for a similar cluster-based methodology — Shea and Bayne (2010), Shea (2012).

As it currently stands, this latter suggestion is somewhat vague as it is framed at a very high level. In order to make this proposal more precise as an organising framework for dream science, more needs to be said about the *sorts* of properties dream scientists ought to be looking for; and in order to determine this, more needs to be said about the kind of mental category to which dreaming itself belongs. For knowledge of the *kind* of mental category to which dreaming belongs will help provide us with knowledge about the *sorts* of properties which serve to distinguish members of this more general kind. In the next and final section, I draw on recent theoretical work in consciousness science in order to add more detail to this initial proposal. This takes as central the idea that dreaming belongs to a broader family of mental states known as *global states of consciousness,* whose members are distinguished in virtue of instantiating different groups of consciousness-related capacities.

5.2. A Natural Kind Methodology for Sleep Experience

Simulation views of dreaming seem to be getting at an important point when they seek to identify dreaming with a core immersive phenomenology. The significant phenomenological feature they track is the idea that to dream in the ordinary sense is in part to be immersed in a simulated dream *world* or reality. Another reason to be partly skeptical of this view as a *definition* of dreaming, however, is that these sorts of phenomenological properties, while distinctive of dreaming, are not unique to it. Very similar sorts of simulative immersion, for example, are also associated with other mental phenomena — most obviously, it is associated with being in a state of *wakefulness*, where one's world is [if one is not an idealist] external reality, but it is also associated with various hallucinatory, psychedelic, and perhaps psychotic states (Windt 2016; chapter 12). In these cases, too, it makes sense to talk of finding oneself immersed in a distinctive 'world' of experience, and this immersion seems part of, or at least is characteristic, of what it is to be in such a state. This reflects the fact that these mental phenomena belong to a unique family of mental states known as *global states of consciousness*; so called because the changes implicated in coming to be in one state rather than another (wakefulness to dreaming, dreaming to a psychedelic state and so on) affect the totality or *global* character of a subject's conscious experience as opposed to local, content-related features of experience.

What does being in a particular global state of consciousness consist in? Recent work in philosophy and consciousness science has sought to provide an answer to this question. The resultant view draws on two key ideas. First, that global states of consciousness like wakefulness and psychedelic states are *multi-dimensional constructs*: they differ from one another not along a single dimension, but several (Bayne et al. 2016, Bayne and Carter 2018) (Figure 3). Second, the dimensions along which these global states differ and can be compared are a set of *conscious-related capacities* (McKilliam 2020, Crowther 2018, Soteriou 2017). These are said to correspond to several families of capacities which are framed at both neural, functional, and phenomenological levels.

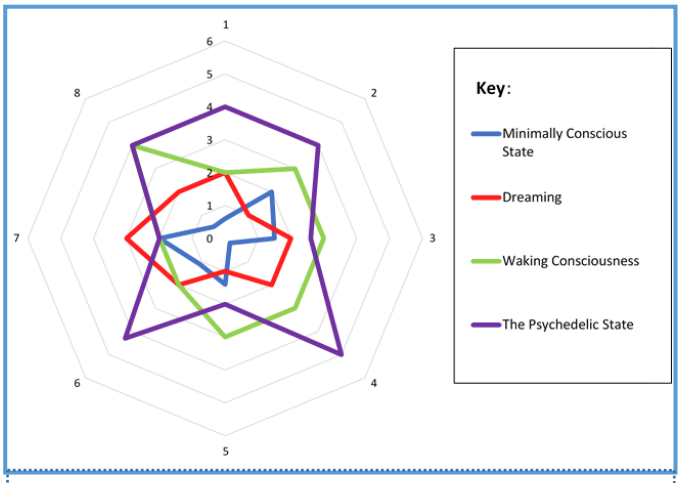

Figure 3: The Multi-Dimensional View of Global States (adapted from Bayne et al. 2016). Dream science, on the natural kind methodology, would be seeking to identify different clusters of capacities which correspond to different global state spaces.

For example, these include (i) *sensory capacities* (corresponding to the intensity and volume of contents experienced (Bayne et al. 2016) and the quality of conscious contents (Fazekas and Overgaard (2016), (ii) *cognitive capacities* (corresponding to mental control and manipulation, attentional capacities and concentration, (enhanced or impaired) capacity to formulate and distinguish novel thoughts (Bayne and Carter 2018), (iii) *subjective capacities* relating to the subject's experience of unity, self and time (corresponding to the experience of time as stopping or slowing, an openness to past and present, altered experiences of self, the boundary between self and environment, the capacity for experiences of 'ego- dissolution 'and experiences of disembodiment (Bayne and Carter 2018) and finally, (iv) *functional capacities* (corresponding to the extent to which information can be globally broadcast, both in terms of the ability to which conscious contents can be broadcast, and the range and number of consumer systems it can be broadcast to (Bayne et al. 2016; McKilliam 2020).

Above, I suggested that recent literature on natural kinds in philosophy of science allows for a reframing of the unity question which has been the primary focus of this paper. The recent work on global states of consciousness just sketched helps to add more detail to this preliminary proposal. In particular, it suggests that the properties in question that dream scientists should be investigating are those *consciousness-related capacities* which correspond to distinct global states of consciousness. It thus suggests a methodological framework for work on dreaming in consciousness science on which the aim is to identity the global states of consciousness which habitually (and, in some cases, pathologically) obtain during sleep. This will involve developing a more fine-grained and comprehensive taxonomy of the capacity-involving dimensions which frame the global state space (by cross examining the neural, functional and subjective dimensions not only of sleep experience, but other global states of consciousness like the psychedelic state) as well as collecting large amounts of data on the subjective and objective properties of conscious sleep experience which can be used as

input in causal models to determine how these properties cluster together in such a way as to underwrite the widest set of inductive inferences (that is, from one sub-set of the cluster, to others). This is a new way forward for consciousness science.

Importantly, the new framework that is brought into view on the basis of the literature discussed above is one which is neutral, in the abstract, on the question of monism and pluralism about sleep experience. That is, unlike the standard methodological paradigms discussed in previous sections (straightforward NCC research, and lucid dreaming paradigms), it does not trade on or require dream scientists to assume ontological assumptions from the outset which have little empirical justification. From a scientific perspective, such a neutral outlook is obviously preferable. Whether there is one cluster to be found or several is not something which can be anticipated prior to the start of a posteriori scientific investigation — it is simply left open[127].

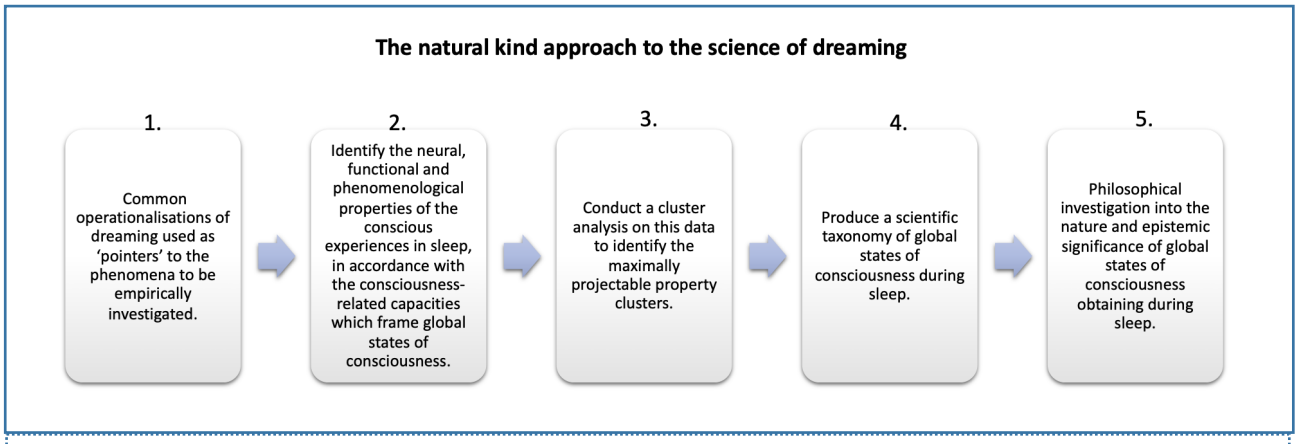

Figure 4: The natural kind approach to the science of dreaming.

6. Conclusion

On this new methodological picture then, the question of whether monism or pluralism is true is one which cannot be determined prior to our empirical investigation. I want to conclude here by briefly raising two questions which arise naturally from this idea: what would it mean for our ordinary concept of dreaming if pluralism about sleep experience turns out to be true? And, what implications does this methodological picture have for philosophical research on dreaming, and how this ought to be conducted?

[127] In this sense, it can be said that this methodology is *continuous* with the use of standard operationalisations of dreaming in the empirical literature, but not beholden to them. On the natural kind methodology, these definitions — sleep mentation, or simulation views — can be used to identify, in the first instance, the conscious phenomena to be investigated and analysed in terms of properties (Figure 4). Unlike the current use of these definitions however, this initial role does not hinder investigation by building in ontological assumptions from the outset.

Starting with the first of these, if pluralism about sleep experience is the outcome of the cluster-based natural kind methodology, this would support a picture on which our ordinary concept of dreaming (or at least one of our ordinary concepts of dreaming) *fails to refer scientifically* in the sense that, where we thought there was a unified kind, there isn't one. If such a view was genuinely empirically supported, it would - I think - work to ground a form of *scientific eliminitvism* about the concept of dreaming - and perhaps further, about the concept of *sleep experience* itself - one which mirrors the scientific eliminitvism about concepts defended by Edouard Machery (2006). These two outcomes are worth discussing here briefly. The possibility of the former idea suggests that while there may be distinct clusters of experience to be found in sleep, and while they may have some properties in common, they are not unified in the sense of forming a natural kind. At most, they may be said to each belong to a broader sub-kind (for example, 'consciousness during sleep'). This is interesting in several respects. Most notably, it suggests that there could be a viable eliminitvist position about dreaming which is grounded in empirical considerations. Unlike current attempts to motivate an eliminitvist position which trade on dubious claims about the lack of conscious experience during sleep (Dennett 1979, Malcom 1959), this scientific eliminitvist view would be one which is supported foremost by empirical data.

The second possibility - that the term 'sleep experience' itself fails to refer to a scientific category - is more radical. The idea here would be that, as part of our natural kind approach to sleep experience (in Step 4) we may find that the global states of consciousness that obtain during sleep have more in common – sharing a larger overlapping state space - to some of the global states of consciousness that obtain whilst waking than they do with each other. On this eventuality, there would be no sub-kind of conscious state specific to sleep to which the global states obtaining during sleep belong. The situation here would be analogous to our epistemic situation with respect to intuitive folk psychological categories which turned out to be scientifically inaccurate— for example, 'aquatic' and 'land' animals. That is, just as whales have more in common with land mammals than with other aquatic animals, so there might be some "dream" states that have more in common with waking experiences than with other kinds of sleep experience[128].

Second, if the unity assumption isn't - as I have argued - in fact justified, how should philosophical research which also by and large assumes it, proceed? I think there is room here for a complementary philosophical research programme which develops alongside the empirical one I have sketched (Figure 4 Step 5). On this view, philosophers ought to refrain from explicit endorsements of the unity assumption until this research programme is carried out, and more evidence in favour of pluralism (or against it) is available. This is the broader lesson. The more specific methodological claim which emerges from this discussion, however, is the claim that philosophers interested in determining the nature and epistemological consequences of dreaming ought to be engaged in a 'ground up' project which starts from an examination of the ontological constitution and epistemic significance of the

[128] One might think that this second possibility hinges on the further question of whether sleep itself is a natural kind, for one might think that there is room for individuating some kinds of global states of consciousness causally on the basis that these are pathological states that obtain due to disorders of the sleep-wake cycle (for example, night terrors). I leave this possibility open here.

consciousness-related capacities which are instantiated during sleep. In relation to discussions of dream skepticism, for example, a central question becomes whether the capacities which underwrite the global state(s) that obtain during sleep — those which are empirically and subjectively identified — and their interactions vindicate or otherwise preclude the sorts of epistemic significance philosophers typically attribute to dreams (see, for example, Soteriou 2017, Crowther 2018, O'Shaughnessy 2002, whose arguments could be interpreted in this vein). This makes way for an exciting new philosophy, as well as a science, of dreaming; one which is in both cases grounded in the idea that conscious sleep experiences form natural kinds.

4 | Depression as a Disorder of Consciousness

First-person reports of Major Depressive Disorder reveal that when an individual becomes depressed a profound change or 'shift' to one's conscious experience occurs. The depressed person reports that something fundamental to their experience has been disturbed or shifted; a change associated with the common but elusive claim that when depressed one finds oneself in a 'different world' detached from reality and other people. Existing attempts to utilise these phenomenological observations in a psychiatric context are challenged by the fact that this experiential 'shift' characteristic of depression appears mysterious and resists analysis in scientific terms. This paper offers a way out of this predicament. The hypothesis proposed is that when an individual becomes depressed, the individual departs from a state of ordinary wakeful consciousness and enters a distinctive *global state of consciousness* akin to dreaming and the psychedelic state. After unpacking and motivating this hypothesis in the context of research in consciousness science, I outline two of its important implications for the neurobiology of depression and psychedelic psychiatry. The upshot is a promising and conceptually well-motivated hypothesis about depression which is apt for empirical uptake and development.

1. What is Depression?

Depression is of interest to philosophy of psychiatry for a variety of reasons. Most prominently, depression (and the clinical category 'Major Depressive Disorder') is useful as a typical example of a diagnostic category or kind. A central question within philosophy of psychiatry in the last twenty years has been whether psychiatry's diagnostic categories, of which depression is a paradigmatic instance, carve nature at its joints in the same way as physical or chemical categories like 'electron' or 'hydrogen' do, such as to grant psychiatry scientific legitimacy (Beebee and Sabbarton-Leary [2010], Tsuo [2013]; [2016], Kincaid and Sullivan [2014], Tekin [2016] Tabb [2019]). This popular question of psychiatry's *natural kind status* can be distinguished from at least one further question concerning the nature of depression as a mental disorder - namely, what *is* depression and what distinguishes it from other

mental disorders? (Kingma [2013]) - which the numerous answers to the question of whether depression is a natural kind underdetermine[129].

Existing answers to this question in the philosophy of psychiatry seek to equate depression either with its current diagnostic symptoms or with their causal aetiology. However, the sufficiency of these descriptivist and causal answers to the question *What is depression?* have recently been challenged by phenomenological analyses of depression (primarily, Ratcliffe [2015]) which suggest that these views are currently incomplete insofar as they fail to capture something essential to what it is to be depressed; something which goes missing if one focuses solely on the existing diagnostic effects of depression or their constitutive causes. This is the observation, as revealed by extensive first-person reports of Major Depressive Disorder ('MDD'), that when an individual becomes depressed, a distinctive phenomenological change occurs. The depressed individual reports that a 'profound shift' has occurred with respect to the way she is 'rooted' in the world or in reality; a distinctive change which is connected to the elusive sense in which depression involves a feeling of detachment from the world and from other people. The distinctive experiential changes which occur when individuals become depressed ought to be taken seriously as an important source of data for psychiatric research. However, the empirical utility of this phenomenological observation for the diagnosis and treatment of depression is severely limited by the fact that it is currently unclear how to elucidate and study this experiential shift empirically. The reports just mentioned raise the question of what this fundamental change *is*, such that it can be empirically studied and operationalised. This thought is exacerbated by Ratcliffe's endorsement of a phenomenological theory of depression as an 'existential shift', motivated by these phenomenological observations, which rejects the possibility that a full understanding of depression can be attained from a scientific perspective.

This paper outlines and motivates a new hypothesis about the nature of depression which, in a departure from Ratcliffe, accounts for and explains the distinctive phenomenology of depression in naturalistic terms. This is the hypothesis that when an individual becomes depressed the individual withdraws from a state of ordinary wakeful consciousness and enters a *distinctive global state of consciousness*. On this view, the changes an individual undergoes when she becomes depressed are profound in the sense that they resemble (being changes of the same kind) the sorts of changes to her experience that occur when she departs from waking consciousness and starts dreaming, or when she enters a psychedelic state. I shall argue that this claim, which has become available only recently due to advances in consciousness science, captures and explains the distinctive experiential 'shift' revealed by Ratcliffe's phenomenological analysis of first-person reports of depression.

The plan for the paper is as follows. **Sections 2** and **3** outline the two views of depression which emerge from current clinical literature on MDD and explain how Ratcliffe's phenomenological analysis of depression - as involving an experiential shift - suggests that these views are currently insufficient or incomplete. The key claim here is that there is a *level of explanation* essential or necessary to understanding what goes on in depression which is left out by views which focus solely on the existing

[129] Compare, for example, how the claim that the biological species *Homo Sapien* is a natural kind, as opposed to an individual (Slater [2013]), underdetermines a number of more specific theses about what it is to be human.

diagnostic symptoms or their underlying constitutive causes (Kendler and Gyngell [2020]). This leaves us with a puzzle concerning how to elucidate this level of explanation in a way which is amenable to scientific investigation. Drawing on recent work in consciousness science **Section 4** outlines a prominent account of global states on which global states of consciousness are multi-dimensional states which regulate the sorts of capacities that are online or available for a creature. In **Section 5**, I use this multi-dimensional view to motivate and flesh out a new hypothesis about depression according to which being depressed involves a change to an individual's global state of consciousness. This can be viewed, I argue, either as supplementing current views of depression, or as offering a new middle ground between existing accounts. Drawing on Ratcliffe's work, here I offer a more specific characterisation of this 'depressive state of consciousness' and explain how this proposal can be rendered compatible with the plausible claims that a person who is depressed remains *awake* and retains the capacity to *dream.* I argue these observations can be accommodated if one views depression as an extended temporal process with depressive global states as temporal parts. Philosophers and psychiatrists are not only interested in understanding what depression is and what it involves, but also in a host of further questions which relate to its neurological basis and treatment. I conclude in **Section 6** by outlining two salient implications of the new hypothesis which shed light on these clinically pressing questions. This includes a suggestion for future empirical research which can shed light on the viability of the claim outlined in this paper and move research forward.

2. Depression in a Clinical Context

The obvious starting place when answering the question *What is depression?* is the clinical taxonomies in which the diagnostic category MDD is listed[130]. According to the latest version of the Diagnostic Statistical Manual (the DSM-5; 2013), Major Depressive Disorder is characterised by at least one discrete depressive episode lasting no less than two weeks. As exemplified by mental disorders generally, a depressive episode is diagnosed polythetically, on the basis of the presence of a group of common symptoms only some of which are necessary for a diagnosis (Fellowes [2021]). Thus, in addition to the presence of two core symptoms - a 'depressed mood' and a 'markedly diminished interest or pleasure in all, or almost all, activities' - the DSM states that at least four of the following symptoms must be present to warrant a diagnosis of Major Depressive Disorder[131]:

- Considerable weight loss or weight gain when not dieting, or decrease or increase in appetite,
- Sleep disturbances resulting in insomnia or hypersomnia,

[130] One might object to this on the basis of that mental disorders like MDD are too causally heterogenous to be of use to psychiatric research (See Hyman [2010], Lilienfeld [2014], Insel [2014], Hoffman and Zachar [2017], Tabb [2015], [2019]). While the hypothesis I motivate later may well have implications for this debate with respect to the category MDD and the Research Domain Criteria (RDoC), I set these issues aside here. Instead, I work with the mostly uncontroversial assumption that there *is* a distinctive phenomenon (or group of phenomena) tracked by the category MDD, although this may require substantial revision in accordance with the RDoCs aims at a later stage.

[131] In addition, it is stated that these symptoms must cause significant distress or impairment in socio-occupational contexts, and that symptoms better attributable to another medical condition (including other mental disorders or substance abuse) should not count towards a diagnosis. See also the ICD-11 [2020].

- Fatigue or loss or energy,
- Feelings of worthlessness or guilt, which may be delusional; that is, not merely self-reproach or guilt about being sick,
- Diminished ability to think or concentrate, or indecisiveness,
- Recurrent thoughts of death (not merely dying), recurrent suicidal ideation without a specific plan; the individual has made a suicide attempt or specific plan for committing suicide.

Reflection on these diagnostic criteria gives rise to two standard views of depression found in the empirical and philosophical literature (Radden [2003]). According to 'ontological descriptivism', associated with the DSM-5, depression as a mental disorder can be equated or identified with the characteristic signs and symptoms of MDD just listed (Kendler et al. [2010])[132]. This stands in contrast to a 'causal' view of depression which moves away from an understanding of depression in terms of its symptoms to the *causes* of these symptoms[133]. This is compatible with a number of different views about the appropriate level of explanation required here (Murphy [2010]). That is, it is compatible with both a biomedical view which stipulates that mental disorders must be characterised and explained at a biological level (i.e. in terms of genetic influence, molecular changes, brain dysfunction etc.) and with broader biopsychological models which leave room for a pluralistic, multi-levelled explanation[134]. On these causal views, MDD works in much the same way as a diagnosis of appendicitis: the diagnostic category refers to the hidden cause or explanation, at whatever level, of why an individual exhibits the range of characteristic symptoms[135].

3. The Phenomenology of Depression

Part of the appeal of ontological descriptivism is that when it comes to mental illnesses, the mental or psychological effects implicated seem to be *constitutively connected* to these disorders in a way which distinguishes them somewhat from somatic illnesses[136]. This idea is reflected in the fact that the majority of MDD's diagnostic criteria make reference to changes in a patient's conscious mental life - for example, *feelings* of worthlessness, a depressed *mood*, recurrent *thoughts* of death, and an inability to *think* and *concentrate.* This sort of claim suggests that in order to improve our understanding of depression psychiatrists ought to collect more detailed accounts of the changes to one's mental life which occur when one becomes depressed. This task has been taken up recently by Ratcliffe who, in

[132] This accommodates both a 'minimal descriptivist view' which identifies depression with its 'key symptoms' such as depressed mood (Ventham [2019], Wong [2016]), and a more demanding 'threshold' conception on which depression is identified with these core symptoms only when they are present over a long period of time or at a high level of severity (Maj [2011], discussed in Tully [2018]).
[133] This is concurrent with the way in which diagnostic categories are used in medicine more generally. Here, for example, a diagnosis of appendicitis *explains why* the patient has abdominal pain by conveying information about what is causing it (Haung [2016]).
[134] Murphy [2013], c.f. Davies, Savulescu, Roache and Loebel [2020] and Bolton and Gillett [2019)].
[135] For causal views of mental disorder, see (Beebee and Sabbarton-Leary [2010]; Kendler et al. [2010]; Murphy [2014]).
[136] Prima facie, the disturbances to one's mental life that are associated with MDD for example, don't seem to *mere* causal effects of the illness in the same way that the abdominal pain commonly associated with having appendicitis does.

his book *'Experiences of Depression'* ([2015]) offers a comprehensive analysis of what it is like to be depressed, with the aim of producing more fine-grained criteria which can be employed in clinical contexts to reduce false diagnoses. Motivated by a comprehensive analysis of first person reports of depression taken from both depression memoirs and an online questionnaire (SANE [2011]; Appendix [2015]) Ratcliffe observes that a diagnosis of MDD in fact tracks several salient (and often connected) changes to a subject's mental life, many of which are left out of, or brushed over, in the DSM-5's existing diagnostic criteria[137]:

- Changes to bodily resonance (feelings of fatigue, tiredness, lethargy, heaviness, sickness or nausea, loss of appetite, numbness).
- Feelings of hopelessness (in particular, the loss of an *ability to hope* simplicter ('existential hope') rather than feeling hopeless *about* something).
- Feelings of guilt and changes to one's self-narrative (feeling guilty about something specific, feeling guilty about something but not knowing what, feeling *that* one is guilty of something specific, feeling guilty simpliciter, feeling that one is essentially guilty or burdensome - that one could not have been otherwise or is to blame for one's depression - negative self-narrative, over-active rumination and repetitive thinking which reduces one's ability to concentrate).
- Disturbances to individual's sense of agency or free will (acts seem not only difficult but *impossible*, an impossibility which affects the way the world is experienced when one is depressed).
- Changes to the subject's experience of time (time seems to 'slow down' or 'stop' when one is depressed, a change which is attributable to changes in the overall structure of temporal experience as well as changes to the rate of 'temporal flow', the future appears 'closed' to depressed agents, - a feature which is connected to the way in which individuals find things salient or significant).
- Changes to the structure of interpersonal experience (feelings of estrangement, changes to how a person relates to another as a person, loss of interpersonal connection, being cut off from the world and others).

Ratcliffe's discussion of these phenomenological changes is extensive. The main idea which I wish to highlight here is the central claim or overarching view of depression which Ratcliffe claims is supported by a wide range of reports of depression. This is the idea that while MDD seems to reliably *track* the range of specific content-related changes just listed, individuals who suffer from depression frequently report that the experience of depression *cannot be reduced to* these types of changes[138]. That is, the report that any attempt to explain or describe depression solely in these terms misses something crucial; a feature of depression which is often connected to the assertion that depression is, even among mental disorders, particularly elusive ([2015] p.2; Ratcliffe et al. [2013])[139]. This pushes Ratcliffe to

[137] See (Radden and Varga [2013]) for an epistemological defence of the reliability of first-person reports of depression which this discussion relies upon.
[138] Where this is not to be understood as a general claim about phenomenal experiences but a specific claim about the experience associated with *depression*.
[139] Many conscious experiences are difficult to describe precisely. This is particularly true of what we might call 'extreme' experiences which plausibly reflect individual differences in mental traits of organisms (Machery [2017]) — the experiences associated with the absence or extremely vivid forms of imagination are a good

endorse a view on which depression, rather than involving a conjunctive set of changes to one's experience, involves a change to something more general and *fundamental* to experience. On his view, first-person descriptions of MDD support a view of depression as involving a radical and all-encompassing change to the total *structure* or fabric of a subjects' experience, one in which the robust subjective changes listed above take place within, or are to be made sense in terms of:

"Sufferers consistently indicate that depression i*s qualitatively different from what many of us regard as 'everyday 'experience*. The depressed person finds herself *in a different 'world'*, in an isolated, alien realm that is *cut off from the consensus reality where people have more mundane experiences of feeling 'more x 'or 'less y ' than usual*…Depression instead involves a disturbance of something that is *fundamental to our lives*, something that goes unnoticed when intact. What is eroded or lost is a 'sense 'or 'feeling 'of being *comfortably immersed in the world* ([2015] p.10, emphasis added).

According to Ratcliffe, this shift explains a number of salient features about depression as a phenomenon. For example, it is argued that this explains why individuals diagnosed with MDD often report that depression is elusive and hard to convey to others who have not experienced it. It is *because* depression is not reducible to a conjunction of particular changes to the content of one's experience (having a lower mood *than usual*, feeling more guilty *than I did before*) but involves a change to something more fundamental (but elusive), that it is difficult to convey to others what it is like to be depressed ([2015] p.2; p.158).

With this in mind, it is easy to appreciate how Ratcliffe's phenomenological analysis challenges the sufficiency of the two views of depression which emerge from the clinical characterisation of MDD. If we take Ratcliffe's claim about the mental changes which occur when one is depressed at face value, for example, current descriptivist views appear to be inadequate or incomplete because they make no reference to this change, nor many of the important aspects of it[140]. This is clinically problematic insofar as it implies that the current DSM-5 category MDD may include a variety of other predicaments which are distinguishable from depression (and thus should not be identified with it)[141]. Similarly, and derivatively, Ratcliffe's analysis raises problems for existing causal views which equate depression with a particular causal explanation for MDD's diagnostic symptoms. Here, the thought is that the causal view is doubly inaccurate because it not only overlooks the fact that there are subjective features essential to MDD (viz. the fundamental 'shift' described above) but also in virtue of the fact that it asserts that depression is to be identified with the causal explanation for these coarse-grained

example here (Zeman et al. [2020]). The claim Ratcliffe is highlighting with respect to depression is stronger. The idea is that individuals with depression lack an appropriate vocabulary or frame of reference with which to describe their experience; a claim exemplified in the reports that being depressed is to be 'cut off from the range of normal experience of feeling more x or less y' and is closer to being '*beyond* description', rather than merely being difficult to convey ([2015]; pp. 2-3).

[140] That is, even to the extent that they do attempt to account for these changes (in a 'depressed mood' for example) Ratcliffe's phenomenological analysis highlights a further limitation with this account viz. that these descriptions are provided at the *wrong level of explanation* — that is, are cashed in in terms of changes to particular contents of an individual's experience. If the phenomenological claim which emerges from (Ratcliffe [2015]) is correct, then any criteria described solely at this level will leave out something crucial.

[141] This seems to be Ratcliffe's point: 'a central theme of this book is that diagnostic categories such as Major Depressive Disorder are insufficiently discriminating and accommodate a variety of predicaments' ([2015] p.75).

symptoms. The causal view thus inherits the problems of ontological descriptivism and gains a few problems of its own — it searches for a causal explanation of a phenomenon in the *wrong place* because it has an incorrect, or ill-defined explanandum. In both cases, the central complaint here is the same: the phenomenological analysis of depression offered by Ratcliffe provides us with reason to think that these accounts miss out on or fail to acknowledge a crucial *level of explanation* necessary for understanding and diagnosing depression, namely, one which makes reference to the fundamental and structural features of an individual's conscious experience[142].

For Ratcliffe, these first-person reports of depression motivate a phenomenological theory of depression on which depression centrally involves an 'existential shift' to one's experience. Depression, on this view, involves a change in *'existential feeling'* which corresponds to the sense of 'world' examined by phenomenologists such as Husserl, Heidegger and Mearlau-Ponty[143]. Ratcliffe's phenomenological theory here is connected to his general endorsement of a methodological 'phenomenological stance' with respect to depression, "through which one comes to appreciate that certain questions cannot be satisfactorily addressed from the standpoint(s) of empirical science" ([2015] p.21). In my view, Ratcliffe's specific phenomenological theory about depression can and should be separated from the general observation that his work highlights viz. that depression involves a profound change or shift to the structure of one's experience. This latter claim ought to be regarded as an important source of data for psychiatric research on the nature of depression. One can, I think, accept both that first person reports are useful, that they point to the fact that something is missing in current attempts to characterise depression, without thinking that this is or *must* be something which is not amenable to empirical investigation.

From an empirical perspective however, the claim that depression involves a profound change to the structure of one's experience (that is, a change to something 'fundamental' which 'goes unnoticed when intact' ([2015] p.2) - particularly in light of the phenomenological theories of depression associated with it - appears at first glance to be something mysterious. What *is* this fundamental change that occurs, and how could this possibly be understood and studied in *scientific* terms? In the following section I outline an answer to this question, an answer which has only become available recently due to advances in consciousness science. This is the thesis that depression involves a change to one's global state of consciousness[144].

---

[142] For recent discussion of multi-levelled causal explanations in psychiatry see (Kendler and Gyngell [2020]; Cooper [2020]).

[143] See Ratcliffe (2015) Chapter 2 for a comprehensive statement of this idea, which is intimately connected to the kinds of possibility one can experience. Fernandez ([2014]) and Saarinen [(2018]) offer further elaboration of this phenomenological theme.

[144] The preceding discussion is suggestive of two distinct readings of Ratcliffe's position which are worth noting here. These arise from an ambiguity in the assertion that "certain questions cannot be satisfactorily addressed from the standpoint(s) of empirical science'' ([2015] p.21). On an initial reading, ''cannot be satisfactorily addressed'' can be taken to imply that there are certain questions one can ask about depression which are not amenable to empirical investigation in general or *at all,* given the nature of empirical enquiry. This sort of claim is in tension with the central proposal put forward in this paper, and is thus a claim I reject. However, there is a weaker reading of Ratcliffe's assertion here which complements the positive hypothesis I develop, on which Ratcliffe commits only to the idea that there are certain questions about depression that cannot be addressed satisfactorily by empirical science *alone* or to *begin with*. Here, for example, one might view the adoption of the

4. What is a Global State of Consciousness?

Changes to the content one experiences are not exhaustive of the changes to one's experience that are possible, or that even occur regularly. This fact becomes salient when we reflect on what happens to our experience when we fall asleep each day and enter a dream, or when we emerge from a dreamless sleep into waking consciousness. In these common and everyday cases, one's experience undergoes a distinctive change — the experiences of dreaming and wakefulness being often very different — which does not, prima facie, seem to be describable *solely* in terms of specific changes to the content one experiences[145]. In consciousness science and the philosophy of mind, such changes are referred to as a subject entering a different *global state of consciousness*, paradigmatic examples of which include wakefulness, lucid and non-lucid dreaming, drug-induced psychedelic states and post-comatose disorders such as the minimally conscious (MCS) and vegetative states (VS)[146].

As a central theoretical construct in consciousness science, a recent body of work has sought to understand and conceptualise what global states of consciousness consist in, such as to account for what distinguishes different global states (for example, dreaming and the psychedelic state) from each other[147]. While research on this is still on-going, there is an emerging and promising conceptual framework for understanding global states of consciousness which provides a firm foundation for my proposal. This is the claim that global states of consciousness are states of systems which regulate the range of cognitive, attentional, and content-related *capacities* that are online for a subject. I unpack this below[148].

A view on which global states of consciousness are states of multi-dimensional capacitation takes as its starting point the rejection of a traditional conception of global states which understands these in terms of 'conscious levels'. On this traditional view, born out of the clinical literature on post-comatose

---

phenomenological stance as a necessary but initial step in a longer process of naturalisation; one which seeks to subsequently *integrate* phenomenological observations and insights into an empirical account of depression (for example, by identifying the proximate neurobiological causes of these phenomenological changes, or by translating them into a terminology which *is* amenable to empirical investigation). Thank you to an anonymous referee for prompting me to detail this alternative reading of Ratcliffe's position.

[145] See for example Windt's ([2015]) influential argument that dreaming cannot be characterised in the standard psychological terms used to characterise changes in wakeful consciousness. This serves to distinguish global states from other sorts of mental phenomena which occur *within* a particular global state such as emotions, thoughts, or moods.

[146] These are also sometimes called "background states" (Chalmers [2000]), "levels" (Boly et al. [2013]; Laureys, [2005]), "modes" (Bayne & Hohwy, [2016]), "global modes" (Fortier-Davy & Millière [2020]), and "aspects" (Klein, [2017]) of consciousness (McKilliam [2020]).

[147] Within consciousness science, the concept of a global state is typically contrasted with the notion of 'local states 'of consciousness or 'conscious contents '(occurrent experiences such as perceptual experiences of various kinds, imaginative and affective experiences, experiences of agency, and occurrent thoughts) which together form the basis of the Neural Correlates of Consciousness research programme (Koch et al. [2016]; Boly et al. [2017], Klein et al. [2020]). In addition, the notion of global states has been gaining prominence as a result of work on the nature and constitution of specific global states such as dreaming and wakefulness. See (Windt [2015]; Voss and Hobson [2014]; Crowther [2018], Soteriou [2020]) for a short, non-comprehensive list of such work.

[148] The account I will sketch here draws mainly on ideas found in (Bayne et al. [2016]); Bayne and Carter [2018], Crowther [2018], Fortier-Davy and Milliere [2020], Mckilliam [2020].

disorders, consciousness (in the global sense) is a *gradable* property which admits of a linear ordering from states which are 'least conscious' to those which 'most conscious' [Figure 1]. For example, this view characterises Minimally Conscious State patients as having a higher level of consciousness than Vegetative State patients do, and so on with respect to the global states associated with dreaming, wakefulness, and psychedelics[149].

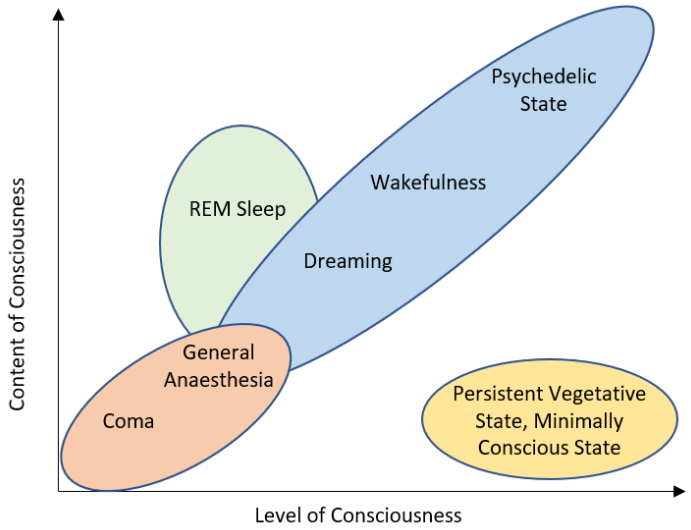

Figure 1: The Levelled Conception of Global States (adapted from (Bayne et al. [2016] p.406)

This levels-based view has been objected to on several grounds. One central concern, found in (Bayne et al. [2016]), comes in the observation that while there may be an *intuitive* sense in which some global states of consciousness are 'lower' than others — for example, when comparing the PVS with wakeful consciousness — further conceptual and empirical reflection reveals that global states of consciousness in fact differ from each other in *multiple respects* and are not measurable along a single dimension[Figure 2][150]. For example, while there is an intuitive sense that the psychedelic state one enters after having taken psychedelic drugs is a 'higher' state of consciousness than the wakeful state we typically occupy, on closer inspection we find that a strict ordering is not possible here: there will be some aspects in which the psychedelic state is improved or enhanced relative to wakefulness - such as the vividness and intensity of the sensory contents one perceives - but many aspects of a subject's experience are impaired in such states — for example, an agent's decision making and attentional capacities (Bayne and Carter [2018]). Similarly, some aspects of the psychedelic state will be *neither*

[149] See (Laureys [2005]; Koch et al. [2016]; Boly et al. [2013]; Overgaard and Overgaard [2010].
[150] Bayne et al. ([2016], p.407) also argue against the levelled view of conceptual grounds, on the basis that our concept of consciousness does not admit of degrees.

enhanced nor impaired but different or novel in a way which is not strictly comparable to wakefulness (for example, the experiences of ego dissolution frequently reported at high dosages of psychedelics)[151]. Thus, while a partial ordering may be possible, the linear ranking implicated in a levels-based conception of global states appears unfeasible.

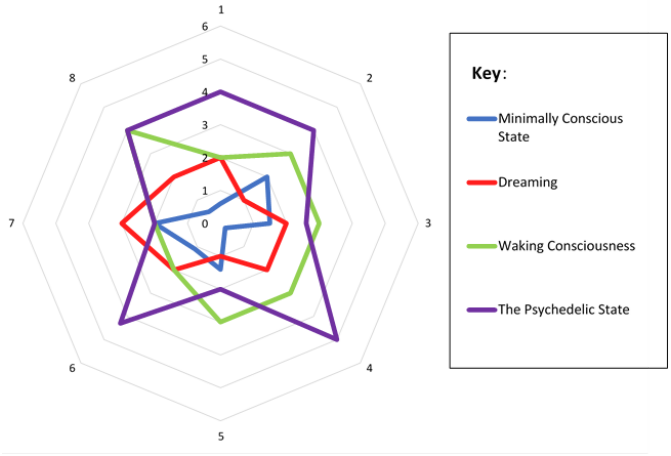

Figure 2: The Multi-Dimensional View of Global States (adapted from (Bayne et al. [2016]). This generalises to all global states of consciousness. As in Bayne et al., the number of dimensions represented here, as well as the placement of particular global states along them, is only intended to be illustrative and does not commit to a claim about the actual or ideal number of dimensions of the global state space. This remains an open empirical question. For now, the reader can view the eight dimensions depicted here as corresponding to a range of dimensions taken from the four dimension families described below.

In light of this, Bayne et al. ([2016]) propose an alternative conceptual framework for global states of consciousness according to which global states of consciousness are *multi-dimensional constructs* measurable across several different dimensions. The search for a comprehensive list of the dimensions which frame global states of consciousness is still on-going. However, recent literature suggests that there are nonetheless several *families* of dimensions which characterise the global state space. These dimensions, which are drawn from a contrastive examination of different global states of consciousness, are listed below:

Dimensions of the Global State Space:

(1) **Sensory content** (corresponding to the intensity and volume of contents experienced (Bayne et al. [2016]) and the quality of conscious contents (Fazekas and Overgaard [2016]).

[151] In this sense, Bayne et al. ([2016] p.409) write that global states of consciousness are similar to economies. Just like one cannot compare two economies on a single dimension, but along several different dimensions, they argue that global states of consciousness similarly resist comparison in absolute terms.

(2) **Cognitive content** (corresponding to mental control and manipulation, attentional capacities and concentration, (enhanced or impaired) capacity to formulate and distinguish novel thoughts (Bayne and Carter [2018]).
(3) **Subjective dimensions relating to the subject's experience of unity, self and time** (corresponding to the experience of time as stopping or slowing, an openness to past and present, altered experiences of self, the boundary between self and environment, experiences of 'ego-dissolution 'and experiences of disembodiment (Bayne and Carter [2018]).
(4) **Functional dimensions** (corresponding to the extent to which information can be globally broadcast, both in terms of the ability to which conscious contents *can* be broadcast, and the range and number of consumer systems it can be broadcast to (Bayne et al. [2016]; McKilliam [2020])[152].

This work raises an obvious question regarding how these dimensions ought to be understood and characterised precisely. According to a recent proposal, broadly continuous with Bayne et al.'s multi-dimensional framework, the dimensions which frame the global state space ought to be analysed in terms of *capacities* which are currently online for a subject (Crowther [2018]; Soteriou [2017]; [2020]; Fortier-Davy and Millière [2020]; McKilliam [2020]). One way of motivating this claim, drawn from the philosophy of mind, is via an examination of the nature of the dreaming state — in particular, the claim that when one is dreaming when one is in a general state of *incapacitation* (Crowther [2018]). This reflects the general idea that there are capacities that individuals can only exercise while awake which the individual is unable to exercise whilst asleep and dreaming; when the individual is asleep, these capacities remain but are 'fettered' in some way (Soteriou [2020])[153]. The 'capacities account' outlined by McKilliam ([2020]) generalises this idea to all global states of consciousness. This gives the proposal in this paper clear positive content:

Global states of consciousness: states of creatures (systems) that regulate (i) the range and quality of conscious contents the creature is capable of experiencing while in that state, (ii) the range of cognitive systems into which those contents can be mobilised while in that state, and (iii) the range of attentional capacities the creature has while in that state. (Mckilliam [2020] p.14)[154].

According to this capacitation account then, global states of consciousness are characterised by a profound shift in a content of an subject's experience which is constitutively connected to the fact that

[152] The reader who is familiar with consciousness science will have recognised that as listing some of these features as dimensions, as opposed to constitutive features of consciousness, the authors' claims seem to be in tension with some major theories of consciousness. For discussion see (Bayne and Carter [2018]).
[153] One might interpret this as the claim that when an individual is asleep a range of capacities she is able to exercise while awake are *masked* (Fara [2008]). In this sense, global states are markedly distinct from other mental phenomena such as moods which (while they may dispose creatures to experience certain emotional content (Textor and Grzankowski *unpublished*)) do not regulate which *capacities* for certain experiences are online for a creature.
[154] The notion of regulation utilised by McKilliam is ambiguous between a strong view of global states as causally acting or *regulating* these capacities in order to keep them on or offline (one might think that sleep is a good example of a state of this kind) and a weaker view on which one views global states as a cluster of on/offline capacities which regularly co-occur in something like a homeostatic property cluster sense (Boyd [1999]). The weaker claim is all that I commit to here.

when one's global state of consciousness changes, the range of content, cognitive and attentional capacities that are online for the subject changes. Thus, what is central to occupying, for example, a psychedelic state as opposed to a state of wakefulness is a simultaneous structural change to perceptual, cognitive, self and temporal-related experiences which is reflected in a profound and holistic change to one's conscious experience (Bayne and Carter [2018])[155].

Replacement of the levelled based view of global states with the multi-dimensional account has opened up numerous exciting research avenues within consciousness science. One such possibility, which has not yet been discussed, is that the multi-dimensional view, in contrast to the previous levelled conception, prompts us to ask what *other* psychological phenomena may fall under - or be best explained in terms of - a change to one's global state of consciousness. I explore this in the next section, where I develop the hypothesis that depression is one such phenomenon which is apt for explanation and elucidation in these terms.

5. The Depressive State of Consciousness

In Section 2 I argued that an application of Ratcliffe's analysis, which seeks to challenge existing accounts of depression on the basis of their inability to account for the phenomenology of depression, faces a significant challenge with regard to reconciling this idea with a naturalistic approach to psychiatry. This paper's central claim is that the recent advances in consciousness science just described provide a way of meeting this challenge. Here, I'll develop the hypothesis that when an individual is depressed, she departs from a state of wakefulness to a *distinctive depressive state of consciousness*, a change which is reflected in an experience of an 'existential shift' as described by Ratcliffe. As stated, the aim here is to provide conceptual motivation for this hypothesis via the claim that this can account for the phenomenology of depression. The real test will come in its empirical applicability, which must come later[156].

On the view I am proposing, the changes an individual undergoes when she becomes depressed resemble - and are changes of the same *kind* as - the changes to her experience that occur when she departs from waking consciousness and starts dreaming, or when she enters a psychedelic state. In the context of the phenomenological features of depression outlined in Section 2, the central motivation for this claim is twofold. First, an understanding of depression as a global state of consciousness

[155] A crucial question for the capacitation account just sketched which has yet to receive significant attention concerns the *relation* between the capacities which characterise global states and the specific states e.g., 'wakefulness' 'dreaming' themselves. On the view just sketched, the correct relation between these is a reductive one: on this view, to be in a global state of consciousness such as dreaming *just is* to realise a set of values which correspond to a range of consciousness-related capacities. However, this view may be objected to. One line of resistance is the observation that such an account necessarily fails to recognise that the capacity related dimensions which frame the multi-dimensional space *interact* in ways specific to a given global state of consciousness. Relatedly, one might reject this view on the basis that such an account does not allow for the possibility of a non-reductive view on which a given state of consciousness ('wakeful consciousness') plays explanatory role(s) over and above the explanatory roles of online capacities (for example, with respect to knowledge (Soteriou [2020]; O'Shaughnessy [2002])). While I am sympathetic to a non-reductive view, I leave this issue open here as the hypothesis I offer is not dependent upon its acceptance.

[156] I include a brief suggestion as to how this might be implemented in Section 6.1.

captures the central phenomenological claims highlighted by Ratcliffe, namely, that depression, while often characterised by numerous specific changes to the content of experience (relating to time, mental agency and so on), resists explanation in these terms in a way which distinguishes it from other sorts of experiences. That is, that depression appears - to those experiencing it - to involve a disturbance or shift to a more *fundamental* aspect of experience. The hypothesis about depression under consideration here offers an explanation as to why depression has these features. The 'shift' it denotes marks a change from one global state of consciousness to another (an elusive process which one is often not explicitly aware of) while the former claim is captured by the fact that global states also *distinctively* resist analysis at the level of conscious content. In both depression and global states of consciousness, there is a *level of explanation* - that is, one at the level of global changes to experience - required to account for the phenomenon; a level of explanation which disappears, or is left out, when one attempts to describe the phenomenon wholly at the level of conscious content.

The second line of motivation reflects the multi-dimensional capacitation account of global states outlined above. This is the striking observation that many of the changes to experience reported by individuals diagnosed with MDD - those documented extensively by Ratcliffe — are experiential changes of the *same kind* as those reported by Bayne and Carter ([2018]) with respect to the psychedelic state. In both cases, individuals report robust phenomenological changes or alterations to their experiences of time, their sense of self, bodily experience, mental agency, concentration, and attention (Ratcliffe [2015]; Bayne and Carter [2018]). In both cases, these changes do not tend to happen disparately but appear, from the reports, to come into effect all at once. The hypothesis here is well placed to make sense of and *explain* the phenomenological similarities here. The thought is that both of these phenomena involve a change to an individual's global state of consciousness, and changes to one's global states of consciousness are *just the sorts of thing* which involve simultaneous experiential changes of this sort. In addition, and perhaps yet more telling, are the numerous observations that for many individuals who are depressed it is the *capacity* for certain sorts of experiences — what Ratcliffe refers to as 'existential' forms of experience, like the capacity to hope — and not merely the degradation or enhancement of ordinary forms of experience that go missing[157]. Here again, the hypothesis that depression involves a change to one's global state is well placed to explain this feature, via the assertion that global states are precisely the sort of thing which regulate the range of mental and content related *capacities* which are online for a creature.

The general hypothesis that depression involves an altered state of consciousness can be fleshed out by drawing on the multi-dimensional framework outlined in the previous section. This provides the beginnings of a more specific interpretation of the thesis that depression involves a change to one's global state of consciousness according to which when one is depressed one enters a state that regulates changes to the range and quality of conscious contents one is capable of experiencing, the range of cognitive systems into which those contents can be mobilised, and the range of attentional capacities

[157] Indeed, this phenomenological distinction is one of the central themes of the book. See, for example, ([2015] Chapter 4 'Loss of Hope').

one has. But what are these capacities more specifically? The material in Section 3 allows for a more detailed characterisation of the depressive state as follows[158]:

Dimensions of the Depressive State: depression is a state of consciousness whose key capacity related dimensions include: (**1) dimensions relating to bodily experience** (feelings of fatigue, tiredness, lethargy, heaviness, sickness or nausea, loss of appetite, numbness and sense of agency) **(2) dimensions relating to cognition** (impairments with respect to mental control, impairments to an agent's attentional capacities and concentration, the capacity to experience hopeful thoughts, an increased capacity for a range of guilt related thoughts, and impairments to self-narrative), **(3) dimensions relating to the subject's experience time** (changes in the overall structure of temporal experience, changes to the rate of 'temporal flow', changes in the way the future appears to subjects) and **(4) dimensions relating to the subject's experience of others** (feelings of estrangement, changes to how a person relates to another as a person, loss of interpersonal connection, being cut off from the world and others).

Like existing characterisations of specific global states such as the characterisation of the psychedelic state offered in (Bayne and Carter [2018]), this multi-dimensional characterisation of depression is provided with the expectation that this will require further development and revision as further work on depression and global states of consciousness emerges. Nonetheless, it is still informative enough to do explanatory work, which I return to in the next section. Before I do so however, further clarification of my hypothesis is required. In particular, more needs to be said in order to render this proposal consistent with a number of claims which are obviously true with respect to depression. This is that when one is depressed one is (i) still awake, (ii) one retains the capacity to dream, (iii) one may return to an ordinary state of wakeful consciousness for intermittent periods while depressed, and finally (iv), that we should leave open the possibility that there may be different sub-kinds of depression, yet to be distinguished, which are lumped together on the current characterisation I have provided.

There is a straightforward way of accounting for (i) which readers who are familiar with the literature on post-comatose disorders will recognise. The strategy here appeals to the distinction between a subject's being awake (e.g., a state associated with various levels of arousal) and that same individual being in a state of *wakeful consciousness*[159]. The sense in which depression constitutes a departure from waking consciousness would then, on this view, resemble the departure from waking consciousness that occurs in the standard post-comatose disorders as well as in individuals who are in psychedelic states. This is straightforward. The points implicated in (ii-iv) however require further conceptual scaffolding. I propose that we can account for these features of depression if we take a step back from consideration of individual depressive *episodes* and account for their occurrence within a broader understanding of 'depression' as an extended temporal process [Figure 3].

[158] This is taken to be analogous to the broad characterisation of the psychedelic state recently offered by Bayne and Carter ([2018]).
[159] Soteriou [2020]; Owen [2009]; Bayne and Shea [2010]; Laureys [2005].

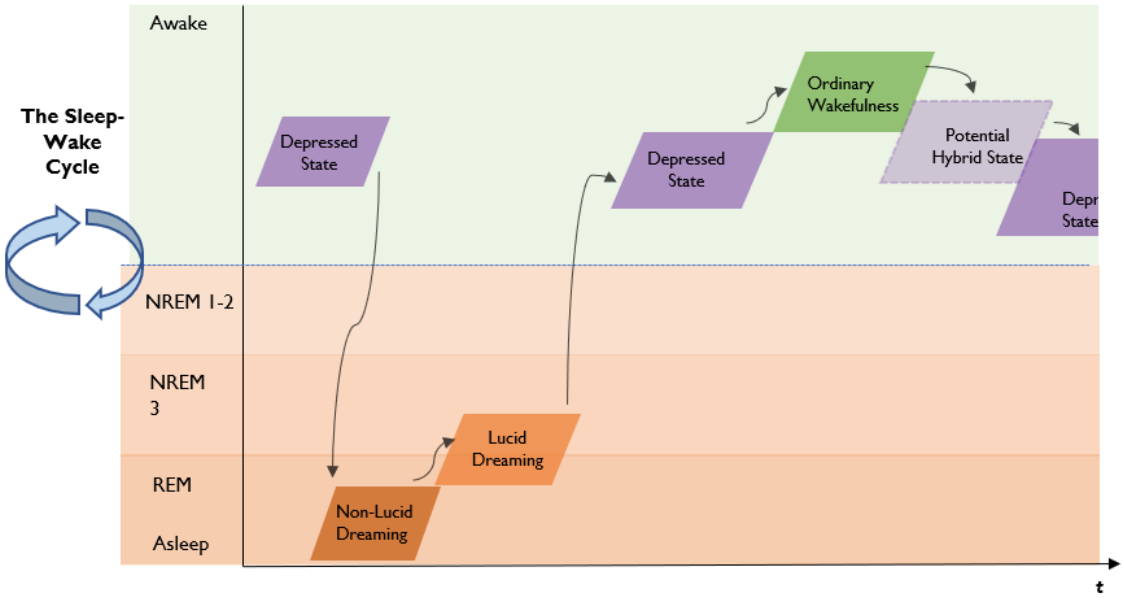

Figure 3: Depression as a temporal process with the depressive state as temporal parts. This is an over simplistic representation of the actual process in several respects. For example, the changes between entering one state of consciousness from another, particularly in the wakeful-dreaming case are often non-instantaneous, and the recourse back to occupying a state of non-pathological wakefulness is likely to occur for periods of days or months rather than hours. Similarly, the placement of dreaming in relation to sleep stages in overly simplistic insofar as REM sleep and dreaming are now understood to be doubly dissociable.

On the view I have in mind, an individual who is *depressed* is so characterised in virtue of the fact that this individual frequently (but not invariantly) occupies a depressive state of consciousness where one would typically occupy a state of ordinary wakeful consciousness. Thus characterised, a diagnosis of depression - and the hypothesis that depression involves a departure from wakefulness to a distinctive *depressive state* of consciousness - is consistent with that same individual sometimes occupying healthy wakeful states of consciousness for periods of time (as per the common observation that depressive episodes often obtain intermittently and not continuously)[160]. This is also consistent with that same individual frequently reporting experience of dreams[161]. Here again, the idea would be the same: if one understands depression as a broader temporal process, it is the frequent recourse *back* from dreaming and other sleep states *to* the depressive (non-wakeful) state that is important for capturing the sense in which that individual is depressed. Finally, this view also accommodates the possibility that depression could admit of various sub-kinds which may correspond to different levels

[160] This periodicity is a common feature of diseases. A good example in the mental realm is individuals diagnosed with Tinnitus. Here, an individual can warrant a diagnosis of Tinnitus despite not having the central feature of this disorder continuously, either in virtue of the fact that the individual only 'hears' the distinct sounds intermittently (in the case of 'Intermittent Tinnitus') or in virtue of the fact that these are not perceived whilst the individual is dreaming (Henry et al. [2014]).

[161] A further explanatory virtue of this hypothesis is that it is well placed to account for the strong empirical connections made between sleep, dreams, and depression (Thase [2006]; McNamara et al. [2010]; Aviram and Soffer-Dudek [2018]).

of severity. On the view proposed here, these sub-kinds would themselves amount to distinct global states of consciousness which would constitute hybrid states between the depressive state and wakefulness[162]. This allows for the further possibility and that an individual may - over the course of their depression - occupy several of these states over a long period of time[163].

In closing, I want to consider where this leaves us vis-a-vis the view of depression endorsed by ontological descriptivists and causalists. The hypothesis just outlined is, I think, consistent with a revised version of either of these views. This follows from the observation that the claim that I am making here is one of a *necessary condition* on a diagnosis of depression. This can be accommodated both in terms of a descriptivist view which acknowledges the depressive state (or aspects of it) as central diagnostic symptoms of depression, and with a biomedical, causal view if the depressive state is taken as a central explanandum of the constitutive explanation provided. A distinctive global state of consciousness will certainly have a constitutive mechanistic explanation, in accordance with a biomedical or biopsychosocial view (more on this below) which renders it compatible with a causal account. There is however a third view of depression to which this hypothesis gives rise, which I wish to mention briefly in closing, without endorsing. This is one on which depression just *is* the sort of extended temporal process involving depressive states sketched above. On this view, which would constitute a sort of middle ground between descriptivism and causalism about depression, depression *qua* depressive states is *both* an explanans and explanandum. That is, the depressive state would itself have a constitutive mechanistic causal explanation *and* also work to explain a number of causally down-wind effects or diagnostic symptoms of depression (such as loss of appetite, or sleep disturbances). Yet, crucially, on this view, it would be inappropriate to identify depression with either of these.

6. Empirical Evidence and Application

The argument so far has attempted to show that there is a viable analysis of the 'phenomenological shift' evidenced in first-person reports of depression which is explicable in naturalistic terms. This view has thus far been motivated by phenomenological considerations: it is attractive, I have argued, because it alone has the resources to *explain why* depression feels - from the inside - the distinctive way it does. This should prompt philosophers and psychiatrists to take it seriously. Depression, however, has other distinctive features which reveal themselves in the empirical literature on MDD. These are often expressed in the form of questions or puzzles: *Why* is there currently no single mechanistic or neurological explanation for depression which can explain various aspects of the disorder? How are we to best intervene and *treat* depression, when such a neurobiological explanation is not forthcoming? In this last section, I will outline two important empirical implications of this

[162] This would thus amount to a waking analog of lucid dreaming which has been proposed as hybrid state of consciousness between wakefulness and dreaming (Valat and Ruby [2021]; Voss et al. [2009]).
[163] Ratcliffe ([2015]) suggests something like this hypothesis and it is implicated in much of the clinical literature on depression (Otte et al. [2016]). One might connect this idea to the observation that MDD is frequently co-morbid with other disorders such as anxiety (see Richards and O'Hara [2014], along with the 'specifiers' of MDD listed in the DSM-5).

hypothesis which shed light on these questions. The aim here is not to offer anything like the final word on these, but rather to point to several exciting lines of research which can be taken up in the future.

6.1. Global State and MDD Research: A Reciprocal Relationship

A central line of criticism posed against MDD as a diagnostic category is grounded in the fact that despite significant advances in the understanding of the neurology of many mental disorders, there is currently no single neural mechanism identified which can explain all aspects of depression[164]. One explanation for this is that there is no unified kind here at all: while at first glance MDD appears phenomenologically unified, upon empirical investigation, we find that there is no neurological unity to these reported cases of depression. It thus may follow that we ought to eliminate the category from our scientific discourse and replace it with a series of more fine-grained categories which correspond to the neurologically unified mechanisms we do find[165].

This eventuality should not be ruled out. However, if the hypothesis I am proposing here is plausible, such a conclusion may be premature. Here, the thought is that the hypothesis that depression involves a distinctive depressive state of consciousness may help explain *why* a mechanistic explanation for depression has not yet been forthcoming. This is due to the fact that there is currently no established account of the neural mechanism(s) underlying global states of consciousness[166]. This is to say, on the view that depression involves a change to one's global state of consciousness, the fact that a mechanistic explanation has not been found for depression is *unsurprising*, for this would amount to a mechanistic explanation for a global state of consciousness.

Thus interpreted, the hypothesis here interestingly offers us reason to be cautiously optimistic about the discovery of a mechanistic account of depression within the next few decades. In particular, the hypothesis here opens up the possibility that, as the work in consciousness science continues to progress, this can be used to inform and motivate new lines of research into the mechanistic basis of depression. In time, this work may allow for the production and implementation of empirical and objective *tests* for depression in patients using fMRI and other neuroimaging techniques employed for the purposes of detecting changes to global states of consciousness. While much of this will depend on future developments within the NCC research programme, work along these lines may already be possible. I have in mind here the potential application of what is known as the 'perturbational complexity index (PCI)', which utilises transcranial magnetic stimulation (TMS) to engage distributed interactions in the brain and to measure their algorithmic complexity (Casali et al. [2013]; Sitt et al.

[164] However, a number of common biomarkers have been identified (Otte et al. [2016]). Time constraints permit me from discussing these here. For a recent meta-analysis of this research which challenges the predictive success of these biological markers see (Kennis et al. [2020]).
[165] This is the line of thought often driving discussion of the RDoC as an organising framework for psychiatric research (Insel [2014]).
[166] Although this research is well underway. See (Howhy and Bayne [2016] and Koch el al. [2016]) for overviews of current research and main predictions made NCC research, and (Revonsuo [2006] and Owen and Guta [2019]) for arguments that this ought to be understood in terms of neural mechanisms.

[2014]; Schartner et al. [2015]; Hudetz et al. [2016]). PCI has been shown to reliably discriminate global states of consciousness in individuals during wakefulness, sleep, and anesthesia, as well as in patients in the MCS, with high PCI values associated with wakefulness and psychedelics (Schartner et al. [2017]) and low values in impaired conscious states like the clinically diagnosed VS and MCS patients. Accepting that the PCI can be used as a correlate or *marker* of (as opposed to constitutive of) an individual's global state of consciousness opens up a potential empirical line of research which can begin to shed light on the viability of the hypothesis outlined in this paper. That is, accepting the correlational utility of the PCI, the hypothesis in this paper suggests that individuals diagnosed with MDD would, as a whole, exhibit statistically significant different PCI values relative to healthy controls subjects[167]. Such a study would thus be clinically worthwhile.

On the other hand, the hypothesis defended in this paper also has general applicability and utility for furthering research into the nature and neural bases of global states of consciousness. This can be done both at a personal and sub-personal level. On the subjective level, for example, the characterisation of the depressive state outlined in the previous section provides a new resource for identifying the subjective dimensions which characterise the general global state space (Bayne et al. [2016]). While this works to reinforce many of the dimensions outlined by Bayne et al. and Bayne and Carter, the dimensions drawn from Ratcliffe's phenomenological analysis also highlight new families of dimensions - particularly, those relating to bodily experience and intersubjectivity - which have not yet been properly considered in the on-going research on global states. The hypothesis here encourages the investigation of these dimensions in the contexts of other global states of consciousness. In addition, on a sub-personal or neurological level, the hypothesis here encourages the application of the various biomarkers of depression (Otte et al. [2016]) to NCC research, in order to facilitate new lines of research on the neural basis of global states which may currently be overlooked.

6.2. Psychedelic Psychiatry

The search for a mechanistic basis for depression is important not only for the purpose of vindicating MDD as a diagnostic category, but for the purposes of facilitating improved treatment options for depression. Currently, two main treatment options for MDD are available, including psychotherapy and pharmacotherapy[168]. As is well known, both psychotherapy and pharmacotherapy have limited efficacy as treatment options. Both have been shown, for example, to be effective in only around 50% of cases, a statistic which continues to spur the need for development of new therapeutic interventions[169]. Within this vein, a promising avenue of research examines the therapeutic potential of psychedelics (such as Ketamine, Mescaline and Psilocybin, the active ingredient in magic

[167] While there is early evidence that depression is associated with higher levels of brain complexity (Li et al. [2008]), a comprehensive study using the PCI is lacking.
[168] Which is offered will typically depend on how severe the depressive episode is, with the majority of depressive episodes treated with a combination of medication and psychotherapy (Cleare et al. [2015]; NICE [2016]; Otte et al. [2016].
[169] Recent studies suggest there is no meaningful difference between the effects of psychotherapy and pharmacology, although combining these is significantly more effective than one alone (Cuijpers et al. [2009]).

mushrooms) as a treatment for depression (Rucker et al. [2018]; Nutt [2019]; Nutt et al. [2020])[170]. While research on psychedelic psychiatry is still in its early days, initial results have been promising. For example, four separate trials have reported improvements in depressive symptoms after psilocybin-assisted psychotherapy (Griffiths et al. [2016]; Ross et al. [2016]; Grob et al. [2011]; Carhart- Harris et al. [2016]), including one in which 'treatment-resistant depression' was the primary criterion for inclusion (Carhart- Harris et al. [2016])[171]. As a result, major research centres have now been set up at Imperial College London and John Hopkins University to study the therapeutic potential of psychedelics for depression and related disorders.

A central question for this research, and one to which answers are currently speculative, concerns the explanatory *rationale* for psychedelic psychiatry. The key question is as follows: what explains the current positive results of psychedelic therapy, and what ought to motivate the funding of further research in this field? (Letheby, [forthcoming]). Current answers to this question offered by leading researchers in this area, locate this rationale within the therapeutic features of psychedelics (Carhart-Harris and Nutt [2017]). This is done both at cognitive, neurological, and psychological levels. For example, it has been hypothesised that the positive effects of psychedelics on mental disorders like depression is due to the promotion of neural plasticity in the key circuits relevant to treating neuropsychiatric disorders (Ly et al. [2018]), that psychedelics relax the precision weighting of the predictive models encoded in the brain (Cahart-Harris and Friston [2019]) and, on a psychological level, that psychedelics play a crucial role in limiting an agent's belief and revising cognitive biases implicated in mental disorders like depression. This latter claim is often expressed in terms of the idea that psychedelics, in contrast to typical treatment responses, directly address the psychosocial causes of depression, and thus involve a "reset" or "opening up" of a therapeutic window which facilitates insight and emotional release and a revision of a subject's general outlook and lifestyle ([Nutt et al. 2020] p. 34; Cahart-Harris and Nutt [2017]).

However, if the hypothesis in this paper is true, then there may be further explanatory rationale for this research in the case of depression which is grounded more firmly in the *nature of depression* itself. That is, if depression *does* involve a change to one's global state of consciousness, then one can explain the efficacy above via the thought that the ingestion of psychedelics — which presumably, given their primary role in changing one's global state of consciousness — target the very same mechanisms which underly and maintain a central aspect of depression. That is, the fact that depression involves an altered state of consciousness may play a significant role in *explaining why* psychedelic treatment is particularly effective in the case of treatment resistant and other forms of depression. Namely, in virtue of the fact that psychedelics act on the *same mechanisms* implicated in the maintenance of depression[172]. Their ingestion thus may prompt a repair of the general neural circuitry which modulates changes to an individual's global state of consciousness such that an ordinary state of

[170] These are also being explored as potential treatment options for other mental disorders such as anxiety, addiction, and OCD. See (Carhart-Harris and Goodwin [2017]) for a recent review.
[171] Similar results were reported in studied with a different psychedelic (ayahuasca) in patients with 'recurrent depression '(Osorio et al. [2015]; Sanches et al. [2016]).
[172] In this sense, the efficacy of psychedelic therapy on depression can be viewed as offering indirect empirical support for my hypothesis.

wakeful consciousness is, after a psychedelic 'trip', reinstated or more reliably maintained[173]. The 'reset' often discussed in the context of psychedelic therapy here would thus be grounded in something *shared* by depression and the psychedelic state which is used to treat it - namely, a resetting of one's global state of consciousness.

7. Conclusion

Existing attempts to utilise phenomenological work on depression in a psychiatric context are challenged by the fact that the experiential 'shift 'characteristic of depression appears to resist analysis in naturalistic terms. The primary aim of this paper has been to establish that there is a viable empirical analysis of the phenomenological shift implicated in depression - on which depression involves a change to one's global state of consciousness - which is explicable in scientific terms. In addition to its ability to explain and account for the distinctive phenomenology of depression, I have suggested that the view on which depression is a disorder of consciousness also has rich explanatory potential in empirical and psychiatric contexts, opening up new conceptual and empirical connections between recent research on global states and the neurobiology of depression and psychedelic psychiatry. The upshot is, in my view, an exciting and conceptually well-motivated hypothesis about the nature of depression which is apt for empirical development.

[173] Scott and Cahart-Harris ([2019]) have recently suggested that psychedelics ought to be developed also as a treatment for post-comatose disorders of consciousness like the PVS. The therapeutic value of this research, given the risks posed by psychedelics for patients, has recently been challenged (Peterson et al. [2019]). The hypothesis here has additional application in this context, suggesting a missing link between psychedelic therapy for depression and the therapeutic potential of psychedelic therapy for treating other disorders of consciousness.

0 | Conclusion

In the introduction, I suggested that although it is not essential to the success of the arguments presented, the four chapters which comprise this thesis can be viewed as falling within a broader philosophical research programme -- one which approaches the philosophical and scientific study of the mind via an examination of the natural categories to which paradigmatic mental phenomena belong. The application of this research strategy to the case of consciousness explored in this thesis has given rise to the idea that, in addition to re-conceptualizing consciousness as a natural kind, a mature scientific and philosophical understanding of subjective experience requires researchers to recognize the reality and importance of a particular category of conscious phenomena known as 'global states of consciousness'. This in turn has given rise to the idea, and empirical hypothesis, that we can utilize *this* concept to revolutionise the understanding of Major Depressive Disorder in psychiatry. Given the scope and ambition of these ideas and broader research programme, the material presented thus far leaves a number of questions regarding the nature and significance of natural kinds and global states of consciousness unanswered. In leaving these issues open, I intend that this is beneficial insofar as the material defended in this thesis remains compatible with different answers to these questions which might be articulated and defended in future (that is, it doesn't hinge on any one answer to these questions being correct). However, it will be useful to highlight these issues briefly in closing. This will serve to situate this work within a broader context and significance. As such, I conclude here with a discussion of what I see as the main research questions to be taken up and explored going forward. These fall into two groups: (1) questions relating to a comprehensive natural approach to the mind and (2), those relating to global states of consciousness and their theoretical significance for clinical psychiatry.

Natural Kinds in the Mind

In Chapter 2, I argued that one way of facilitating a mature science of consciousness is to re-engineer our existing concept (or concepts) of consciousness to a natural kind concept. This was said to reflect the general observation that some concepts are better suited to scientific theorizing about the world than others, and thus that progress in scientific and philosophical domains – like consciousness science -- can be (and indeed, has often been) made by working to refine our

existing conceptual schemes to reflect this. I suggested that a useful new concept of consciousness was one on which 'consciousness' is said to refer to the homeostatic property cluster associated with our initial samples of consciousness, a cluster of properties which underwrite a privileged epistemic role for consciousness in this scientific domain. This idea was explored further in Chapter 3, where I suggested that one way in which consciousness, exemplified in the case of 'dreaming', could fail to refer in neuroscientific domains is if it is discovered that there is not a single natural kind or homeostatic property cluster associated with our initial samples of consciousness, but several.

For someone examining what general lessons can be drawn from these ideas – such that we can see these arguments as falling within a broader 'natural kind approach' to the *mind* – there are at least two central foundational questions which need to be explored going forward, so as to lend this programme greater support. The first asks whether there is a broader psychological theory of the nature of our philosophical and scientific concepts which can be utilized in support of a general research programme of this type. That is, the arguments above invite the question, what are the nature of our philosophical concepts more generally, and their relation to the world, such that these particular projects of engineering are possible? An answer to this question would plausibly draw on empirical work on the nature of concepts to provide a detailed story as to how this conceptual change occurs in practice, and how projects of conceptual change achieve the sort of epistemic progress that I and others typically attribute to them. Second, and relatedly, one can ask here how this general proposal regarding conceptual engineering relates to naturalism as a philosophical project. In the introduction, I noted that the more general research project I have in mind aims to situate mental phenomena 'firmly, but not reductively in the natural world'. In future, I think a great deal more can be said with respect to how exactly projects of conceptual engineering of the sort developed in this thesis work to achieve this goal. This in turn will require us to think more about how 'naturalism' can and ought to be understood as a philosophical project, as well as the ontological commitments of the pragmatic naturalist view of natural kinds I have drawn on, and how, if at all, these fit together.

In addition to examining foundational questions that arise when one attempts to generalize from the work in this thesis to the broader project concerning natural kinds and the mind, there is also important work to be done in applying the general natural kind framework for consciousness science developed here to new debates and methodological issues in this field. For example, in Chapter 2, I discussed the idea that conceptualizing consciousness as a natural kind commits one to several counterintuitive claims or underexplored consequences with respect to consciousness, such as admitting the possibility of genuinely vague or borderline cases of consciousness. The literature on the natural kind approach to consciousness has thus far been silent on, and has yet to recognize, a further upshot of a natural kind conception of consciousness which I think could be integral to a mature scientific understanding of consciousness. This concerns the observation that natural kinds in the special sciences tend to exhibit robust *individual differences* among individuals of a given species, in a manner which can be *explained* by appeal to the ontological structure of natural kinds as homeostatic property clusters. An important theme I think is worth

exploring in future work then is the way in which this feature of natural kinds might be put to work as a methodological principle in consciousness science and philosophy – serving as a working hypothesis which can provide an alternative explanatory hypothesis for trends in empirical data which has yet to be properly considered.

Global States of Consciousness in Psychiatry

In Chapter 4, I motivated the hypothesis that depression involves a change to a subject's global state of consciousness akin to the changes which occur when a subject enters a dream or psychedelic state. The aim here was primarily to demonstrate that this hypothesis has explanatory purchase insofar as it alone can explain what depression *feels like* from the inside – that is, it can account for the salient phenomenological features of depression which have been brushed over or ignored in existing psychiatric research. This is a promising start. However, going forward, there is crucial work to be done in further spelling out the empirical aspects of this model to allow for greater comparability with existing empirical models of depression. In particular, it will be important to specify a set of concrete empirical predictions made by this model in order to test its empirical and explanatory purchase. In Chapter 4 I made one suggestion of this kind with respect to the idea that different global states of consciousness are correlated with changes in Perturbational Complexity Index values, however identifying further methods of testing this model will be crucial going forward.

While Chapter 4 examines the idea that we can use the concept of global states of consciousness to analyze and elucidate the nature of depression, there is an understanding of the work in Chapter 4 which views this idea as also laying out the beginnings of a broader research programme. This examines how the concept of a global state can work to transform the conceptual toolkit of psychiatry more generally. The idea here would be that the concept of a global state of consciousness could serve as the basis of a 'paradigm shift' within psychiatry -- revolutionising our understanding, diagnosis, and treatment of mental disorders. However, for this to have real applicability, I think a more precise and detailed account of global states of consciousness and their explanatory significance is required. This will involve, in the first instance, the development of a more rigorous and detailed account of the metaphysics and epistemology of global states of consciousness which can serve as the foundation for this type of research programme. As discussed in Chapters 3 and 4, the current view of global states of consciousness in the literature suggests these are states which regulate the range and number of *consciousness-related capacities* which are currently online for a creature [Bayne et al. (2016), Bayne and Carter (2018), McKilliam (2020)]. This initial view however leaves many important questions unanswered. For example, how exactly should the concept of a 'consciousness-related capacity' be defined and understood in this context, and how can these capacities be measured, for the purposes of detecting a change to a creature's global state of consciousness? What is the ontological relation between these capacities and the global states of consciousness themselves? And how can an analysis of global states of consciousness explain why some global states of consciousness – like some drug-induced states, or depressive states – *feel bad*? The material presented and discussed in this thesis remains

silent on these questions, but they are nonetheless crucial and need to be answered if this broader programme in psychiatry is to succeed. Answering these questions is thus the next job for philosophers and consciousness scientists interested in testing this model of depression and moving research in consciousness science forward.

Bibliography

Allen, C. (2004). Animal pain. *Noûs* 38 (4):617–643.

Allen, K. (2015). Hallucination And Imagination. *Australasian Journal of Philosophy* 93 (2):287-302.

Allport, A. (1988). What concept of consciousness? In Anthony J. Marcel & E. Bisiach (eds.), *Consciousness in Contemporary Science*. Oxford University Press.

American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub, 2013.

Andrillon T, Windt J, Silk T, Drummond SPA, Bellgrove MA and Tsuchiya N (2019) Does the Mind Wander When the Brain Takes a Break? Local Sleep in Wakefulness, Attentional Lapses and Mind-Wandering. *Front. Neurosci.* 13:949. doi: 10.3389/fnins.2019.00949

Antony, M. V. (2008). Are our concepts CONSCIOUS STATE and CONSCIOUS CREATURE vague? *Erkenntnis* 68 (2):239 - 263.

Arnulf I. (2011). The 'scanning hypothesis' of rapid eye movements during REM sleep: a review of the evidence. *Archives italiennes de biologie*, *149*(4), 367–382. https://doi.org/10.4449/aib.v149i4.1246

Aserinsky, E & Kleitman, N. (1953). Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science (New York, N.Y.)*, *118*(3062), 273–274. https://doi.org/10.1126/science.118.3062.273

Aviram, L and Soffer-Dudek, N. "Lucid dreaming: intensity, but not frequency, is inversely related to psychopathology." *Frontiers in psychology* 9 (2018): 384

Baird, B., LaBerge, S., & Tononi, G. (2021). Two-Way Communication in Lucid REM Sleep Dreaming. *Trends in cognitive sciences*, *25*(6), 427–428. https://doi.org/10.1016/j.tics.2021.04.004

Baird, B., Mota-Rolim, S. A., & Dresler, M. (2019). The cognitive neuroscience of lucid dreaming. *Neuroscience and biobehavioral reviews*, *100*, 305–323. https://doi.org/10.1016/j.neubiorev.2019.03.008

Balog, K. (2009). Phenomenal Concepts. In Brian McLaughlin, Ansgar Beckermann & Sven Walter (eds.), *Oxford Handbook in the Philosophy of Mind*. Oxford University Press. pp. 292--312.

Balog, K. (2012). In Defence of the Phenomenal Concept Strategy1. *Philosophy and Phenomenological Research* 84 (1):1-23.

Bayne, T.; Hohwy, J. & Owen, Adrian M. (2016). Are There Levels of Consciousness? Trends in Cognitive Sciences 20 (6):405-413.

Bayne, T. & Shea, N. (2020). Consciousness, Concepts and Natural Kinds. *Philosophical Topics* 48 (1):65-83.

Bayne, T., & Carter, O. (2018). Dimensions of consciousness and the psychedelic state. Neuroscience of Consciousness, 2018(1), niy008. https://doi.org/10.1093/nc/niy008.

Beaney, M., "Analysis", *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/sum2021/entries/analysis/.

Beebee, H. & Sabbarton-Leary, N. (2010). Are psychiatric kinds real? European Journal of Analytic Philosophy 6 (1):11-27.

Bernardi, G., & Siclari, F. (2019). Local patterns of sleep and wakefulness. In Handbook *of behavioral neuroscience* (Vol. 30, pp. 33– 47). London, San Diego, Cambridge and Oxford: Elsevier.

Birch, J. ; Ginsburg, S. & Jablonka, E. (2021). The Learning-Consciousness Connection. *Biology and Philosophy* 36 (5):1-14.

Birch, J. (2020). The Search for Invertebrate Consciousness. Noûs.

Bird, A. & Tobin, E. (2008). Natural kinds. In Edward N. Zalta (ed.), Standford Encyclopedia of Philosophy.

Block, Ned (1995). On a confusion about a function of consciousness. *Brain and Behavioral Sciences* 18 (2):227-–247.

Block, N.J. (2007). *Consciousness, Function, and Representation: Collected Papers*. Bradford.

Bolton, D. & Gillett, G. (2019). The Biopsychosocial Model of Health and Disease: New Philosophical and Scientific Developments. Springer Verlag.

Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., & Tononi, G. (2017). Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. Journal of Neuroscience, 37 (40), 9603–9613. https://doi.org/10.1523/jneurosci.3218-16.2017

Boly, M., Seth, A. K., Wilke, M., Ingmundson, P., Baars, B., Laureys, S., Edelman, D., & Tsuchiya, N. (2013). Consciousness in humans and non-human animals: Recent advances and future directions. Frontiers in Psychology, 4, 625. https: //doi.org/10.3389/fpsyg.2013.00625%20

Boyd, Richard (1991). Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds. *Philosophical Studies* 61 (1):127-148.

Boyd, Richard (1999). Homeostasis, species, and higher taxa. In R. A. Wilson (ed.), *Species: New Interdisciplinary Essays*. MIT Press. pp. 141-85.

Boyd, Richard (1999). Kinds, Complexity and Multiple realization. *Philosophical Studies* 95 (1):67-98.

Boyle, A. (2019). Mapping the Minds of Others. *Review of Philosophy and Psychology* 10 (4):747-767.

Brun, G. (2016). Explication as a Method of Conceptual Re-engineering. *Erkenntnis* 81 (6):1211-1241.

Burge, T. (1979). Semantical paradox. *Journal of Philosophy* 76 (4):169-198.

Cappelen, H. (2018). Fixing Language: An Essay on Conceptual Engineering. Oxford: Oxford University Press.

Carhart-Harris, R. L., and D. J. Nutt. (2017): "Serotonin and brain function: a tale of two receptors." *Journal of Psychopharmacology* 31.9 (2017): 1091-1120.

Carhart-Harris, Robin L., and Guy M. Goodwin (2017):. "The therapeutic potential of psychedelic drugs: past, present, and future." *Neuropsychopharmacology* 42.11. 2105-2113.

Carhart-Harris, Robin L., and K. J. Friston (2019). "REBUS and the anarchic brain: toward a unified model of the brain action of psychedelics." *Pharmacological reviews* 71.3:316-344.

Carhart-Harris, Robin L., et al. (2018). "Psilocybin with psychological support for treatment-resistant depression: six-month follow-up." *Psychopharmacology* 235.2: 399-408.

Carnap, R. (1947). *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago, IL, USA: University of Chicago Press.

Carnap, R. (1947). On the application of inductive logic. *Philosophy and Phenomenological Research* 8 (1):133-148.

Carruthers, P. (2018). Comparative psychology without consciousness. *Consciousness and Cognition* 63:47-60.

Carruthers, P. (2019). *Human and Animal Minds: The Consciousness Questions Laid to Rest*. Oxford University Press: Oxford.

Carruthers, P. & Veillet, B. (2017). Consciousness operationalized, a debate realigned. *Consciousness and Cognition* 55:79-90.

Casali, A. G., et al. (2013). "A theoretically based index of consciousness independent of sensory processing and behavior." *Science translational medicine* 5.198: 198ra105-198ra105.

Cavallero, C., Foulkes, D., Hollifield, M., & Terry, R. (1990). Memory sources of REM and NREM dreams. *Sleep: Journal of Sleep Research & Sleep Medicine, 13*(5), 449–455.

Chakravartty, A. (2007). *A Metaphysics for Scientific Realism: Knowing the Unobservable*. Cambridge University Press: Cambridge.

Chalmers, D.J. (1996). The Conscious Mind: In Search of a Fundamental Theory. Oxford University Press.

Chalmers, D. J. (2000). What is a neural correlate of consciousness? In T. Metzinger (Ed.), Neural correlates of consciousness (pp. 17–39). MIT Press.

Chalmers, D.J. (2006). Phenomenal concepts and the explanatory gap. In Torin Alter & Sven Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.

Chalmers, D.J. (2010). *The Character of Consciousness*. Oxford University Press: Oxford.

Chalmers, D.J. (2011). Verbal Disputes. Philosophical Review 120 (4):515-566.

Chalmers, D.J. (forthcoming). What is Conceptual Engineering and What Should it Be? *Inquiry: An Interdisciplinary Journal of Philosophy*.

Charcot, J. M. (1889). Clinical Lectures on Diseases of the Nervous System. (vols. 3). London: The New Sydenham Society.

Churchland, P.S. (1988). Reduction and the neurobiological basis of consciousness. In Anthony J. Marcel & E. Bisiach (eds.), *Consciousness in Contemporary Science*. Oxford University Press.

Cipolli, C., Mazzetti, M., & Plazzi, G. (2013). Sleep-dependent memory consolidation in patients with sleep disorders. *Sleep medicine reviews*, *17*(2), 91–103. https://doi.org/10.1016/j.smrv.2012.01.004

Cipolli, C., Ferrara, M., De Gennaro, L., & Plazzi, G. (2017). Beyond the neuropsychology of dreaming: Insights into the neural basis of dreaming with new techniques of sleep recording and analysis. *Sleep medicine reviews*, *35*, 8–20. https://doi.org/10.1016/j.smrv.2016.07.005

Clark R.E, Manns J.R, Squire L.R. Classical Conditioning, Awareness, and Brain Systems. *Trends in Cognitive Sciences.* 2002;6:524–31

Cleare, A. et al. Evidence-based guidelines for treating depressive disorders with antidepressants: a revision of the 2008 British Association for Psychopharmacology guidelines.

Conee, E. (1994). Phenomenal knowledge. *Australasian Journal of Philosophy* 72 (2):136-150.

Cooper. R (2020) 'When answers are hard to find, change the question: Asking different causal questions can enable progress, in Davies, Will; Savulescu, Julian & Roache, Rebecca (eds.) (2020). Psychiatry Reborn: Biopsychosocial Psychiatry in Modern Medicine. Oxford University Press.

Corns, J. (2016). Pain eliminativism: scientific and traditional. *Synthese* 193 (9).

Craver, C. F. (2009). Mechanisms and natural kinds. *Philosophical Psychology* 22 (5):575-594.

Crowther, T .(2018). Experience, dreaming, and the phenomenology of wakeful consciousness. In F. Dorsch, F. MacPherson, & M. Nida-Rumelin(Eds.) *Phenomenal presence*. Oxford: Oxford University Press.

Cuijpers, P., van Straten, A., Warmerdam, L.& Andersson, G. Psychotherapy versus the combination of psychotherapy and pharmacotherapyin the treatment of depression: a meta-analysis.

D'Atri, Aurora, et al. "Cortical activation during sleep predicts dream experience in narcolepsy." *Annals of Clinical and Translational Neurology* 6.3 (2019): 445-455.

Davies, W. & Roache, R. (2017). Reassessing Biopsychosocial Psychiatry. British Journal of Psychiatry 210 (1):3-5.

Davies,W.; Savulescu, J. & Roache, R. (eds.) (2020). Psychiatry Reborn: Biopsychosocial Psychiatry in Modern Medicine. Oxford University Press: Oxford.

Dawes, A.J., Keogh, R., Andrillon, T. *et al.* (2020) A cognitive profile of multi-sensory imagery, memory and dreaming in aphantasia. *Sci Rep* 10, 10022.

Dawkins, M.S. (2017) 'Animal Welfare with and without consciousness'. Journal of Zoology (301;1-10). https://doi.org/10.1111/jzo.12434

Debner J.A, Jacoby L.L. Unconscious Perception: Attention, Awareness, and Control. *Journal of Experimental Psychology: Learning, Memory, & Cognition.*

Dement, W., & Kleitman, N. (1957). The relation of eye movements during sleep to dream activity: An objective method for the study of dreaming. *Journal of Experimental Psychology*, 53(5), 339.

Dement, W. C., & Vaughan, C. (1999). *The promise of sleep: A pioneer in sleep medicine explores the vital connection between health, happiness, and a good night's sleep.* Dell Publishing Co.

Dennett, D. C. (1976). Are Dreams Experiences? *Philosophical Review.* 85 (2):151.

Dennett, D.C. (2001). 'Are we explaining consciousness yet?' Cognition 79 (1):221-37.

Descartes, R.(1984). Philosophical writings of Descartes, Volumes I–III, (Cottingham, J. Stoothoff, R. & Murdoch, D. Trans.). In J. Cottingham, R. Stoothoff, D. Murdoch,& A. Kenny (Vol. IIIed), Cambridge: Cambridge University Press.

Diaz-Leon E., (2016). Phenomenal concepts: Neither circular nor opaque. *Philosophical Psychology* 29 (8):1186-1199.

Díaz, R. (2021). Do People Think Consciousness Poses a Hard Problem?: Empirical Evidence on the Meta-Problem of Consciousness. *Journal of Consciousness Studies* 28 (3-4):55-75.

Dinets, V. (2016) No cortex, no cry. *Animal Sentience* 3(7).

Domhoff, G. W. (2003). *The scientific study of dreams: Neural networks, cognitive development, and content analysis.* American Psychological Association. https://doi.org/10.1037/10463-000

Dorsch, F. (2012). *The Unity of Imagining*. De Gruyter: Berlin, Boston.

Dorsch, F. (2015). Focused Daydreaming and Mind-Wandering. *Review of Philosophy and Psychology* 6 (4):791-813.

Dresler, M., Wehrle, R., Spoormaker, V. I., Koch, S. P., Holsboer, F., Steiger, A., Obrig, H., Sämann, P. G. & Czisch, M. (2012). Neural correlates of dream lucidity obtained from contrasting lucid versus non-lucid REM sleep: a combined EEG/fMRI case study. Sleep, 35 (7), 1017-1020

Edelman, D. B., & Seth, A. K. (2009). Animal consciousness: a synthetic approach. *Trends in neurosciences*, *32*(9), 476–484. https://doi.org/10.1016/j.tins.2009.05.008

Edelman, D. B. (2016) Leaving the door open for fish pain: Evolutionary convergence and the utility of 'just-so stories'. *Animal Sentience* 3(36)

Eilan, N. (2014). Intelligible Realism about Consciousness: A Response to Nagel's Paradox. Ratio 27 (1):32-52.

Eklund, M. (2002). Inconsistent Languages. *Philosophy and Phenomenological Research* 64 (2):251-275.

Eklund, M. (2015). Intuitions, Conceptual Engineering, and Conceptual Fixed Points. In Christopher Daly (ed.), *The Palgrave Handbook of Philosophical Methods*.

Eklund, M. (2020). Reply to critics. *Inquiry: An Interdisciplinary Journal of Philosophy* 63 (5):535-561.

Elwood, R. W. (2016) A single strand of argument with unfounded conclusion. *Animal Sentience* 3(19)

Ereshefsky, M. & Matthen, M. (2005). Taxonomy, Polymorphism, and History: An Introduction to Population Structure Theory. Philosophy of Science 72 (1):1-21.

Ereshefsky, M. & Reydon, T. A. C. (2015). Scientific kinds. *Philosophical Studies* 172 (4):969-986.

Erlacher, D., and Schredl, M. "Practicing a motor task in a lucid dream enhances subsequent performance: A pilot study." *The Sport Psychologist* 24.2 (2010): 157-167.

Farah, M. J. (1984). The neurological basis of mental imagery: A componential analysis. *Cognition.* 18, 245-272.

Faw, B. (2009). Conflicting intuitions may be based on differing abilities - evidence from mental imaging research. *Journal of Consciousness Studies,* 16, 45-68.

Fazekas, P., & Nemeth, G. (2018). Dream experiences and the neural correlates of perceptual consciousness and cognitive access. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *373*(1755), 20170356. https://doi.org/10.1098/rstb.2017.0356

Fazekas, P., Nemeth G., & Overgaard, M. (2019). White dreams are made of colours: What studying contentlessdreams can teach about the neural basis of dreaming and conscious experiences. Sleep Medicine Reviews, 43,84–91

Fazekas, P., and Overgaard, M. "Multidimensional models of degrees and levels of consciousness." *Trends in Cognitive Sciences* 20.10 (2016): 715-716.

Fellowes, S. (2021) "The value of categorical Polythetic diagnoses in psychiatry." British Journal for the Philosophy of Science.

Fenwick, P., Schatzman, M., Worsley, A., Adams, J., Stone, S., & Baker, A. (1984). Lucid dreaming: Correspondence between dreamed and actual events in one subject during REM sleep. *Biological Psychology, 18*(4), 243–267. https://doi.org/10.1016/0301-0511(84)90056-5

Fernandez, A.V. (2014). Depression as existential feeling or de-situatedness? Distinguishing structure from mode in psychopathology. Phenomenology and the Cognitive Sciences 13 (4):595-612.

Fiala, B. & Nichols, S. (2019). Generating Explanatory Gaps. Journal of Consciousness Studies 26 (9-10):71-82.

Flanagan, O. J. (2000). *Dreaming Souls: Sleep, Dreams, and the Evolution of the Conscious Mind*. Oxford University Press.

Fortier-Davy, M., & Millière, R. (2020). The multi-dimensional approach to drug-induced states: A commentary on Bayne and Carter's "Dimensions of Consciousness and the Psychedelic State". Neuroscience of Consciousness, 2020(1), niaa004. https://doi.org/10.1093/nc/niaa004%20

Foulkes, D. (1982) *Children's Dreams: Longitudinal Studies.* Wiley: New York.

Foulkes, D (1990) "Dreaming and Consciousness," *European Journal of Cognitive Psychology* 2: 39–55. DOI: 10.1080/09541449008406196.

Foulkes, D (1996) "Dream Research: 1953–1993," Sleep 19: 609–24

Foulkes, D. (1999) *Children's Dreaming and the Development of Consciousness.* Cambridge, MA: Harvard University Press.

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

Fulford, J., et al., (2018) The neural correlates of visual imagery vividness e An fMRI study and literature review. *Cortex* [online]. https://doi.org/10.1016/j.cortex.2017.09.014.

Galton, F. (1880). Statistics of mental imagery. *Mind.* 5, 301-318.

Garfield, J. L. (2016). Illusionism and Givenness. *Journal of Consciousness Studies* 23 (11-12):73-82.

Godfrey-Smith, P. (2016) Pain in parallel. *Animal Sentience* 3(21)

Goff, p. (2011). A posteriori physicalists get our phenomenal concepts wrong. *Australasian Journal of Philosophy* 89 (2):191 - 209.

Goff, P. (2017). *Consciousness and Fundamental Reality*. New York, USA: Oup Usa.

Goodman, Nelson (2000). The New Riddle of Induction. In Sven Bernecker & Fred I. Dretske (eds.), *Knowledge: Readings in Contemporary Epistemology*. Oxford University Press.

Griffiths R. R, Johnson M.W, Carducci M.A et al (2016) Psilocybin produces substantial and sustained decreases in depression and anxiety in patients with life-threatening cancer: a randomized double-blind trial. J Psychopharmacol 30:1181–1197

Griffiths, P. E. (1999). Squaring the Circle: Natural Kinds with Historical Essences. In Robert A. Wilson (ed.), *Species: New Interdisciplinary Essays*. MIT Press. pp. 209-228.

Grob CS, Danforth AL, Chopra GS et al (2011) Pilot study of psilocybin treatment for anxiety in patients with advanced-stage cancer. Arch Gen Psychiatry 68:71–78

Häggqvist, S. & Wikforss, A.M. (2018). Natural Kinds and Natural Kind Terms: Myth and Reality. British Journal for the Philosophy of Science 69 (4):911-933.

Haslanger, S. (2000). Gender and race: (What) are they? (What) do we want them to be? *Noûs* 34 (1):31–55.

Hearne, K. (1978). Lucid dreams: An electro-physiological and psychological study. Liverpool, UK: University of Liverpool, England.

Henry, J. A., et al. "Underlying mechanisms of tinnitus: review and clinical implications." *Journal of the American Academy of Audiology* 25.1 (2014): 5-22.

Hobson J. A. (2009). REM sleep and dreaming: towards a theory of protoconsciousness. *Nature reviews. Neuroscience*, *10*(11), 803–813. https://doi.org/10.1038/nrn2716

Hobson, J. A, Pace-Schott, E. F. & Stickgold, R. (2000). Dreaming and the brain: Toward a cognitive neuroscience of conscious states. *Behavioural and Brain Sciences* 23 (6):793-842; 904-1018; 1083-1121.

Hobson, J. A. (1988). *The dreaming brain.* Basic Books.

Hobson, J. A. (1988). *The dreaming brain*. New York, NY, US: Basic Books.

Hobson, J., Pace-Schott, E. F. & Stickgold, R. (2003). Dreaming and the brain: Toward a cognitive neuroscience of conscious states. In Pace-Schott, E.F, Solms. M, Blagrove M. & Harnad, S. (eds.), *Sleep and Dreaming: Scientific Advances and Reconsiderations*. Cambridge University Press. pp. 793-842.

Hoerl, C. (2015). Writing on the page of consciousness. Proceedings of the Aristotelian Society 115 (3pt3):187-209.

Hoffman, G. A., and Zachar, P. "RDoC's metaphysical assumptions: Problems and promises." *Extraordinary science and psychiatry: Responses to the crisis in mental health research* (2017): 59-86.

Holman, Emmett L. (2013). Phenomenal concepts as bare recognitional concepts: harder to debunk than you thought, …but still possible. *Philosophical Studies* 164 (3):807-827.

Holzinger & Mayer (2020).Lucid Dreaming Brain Network Based
on Tholey's 7 Klartraum Criteria. *Frontiers in Psychology*.

Horikawa, T., Tamaki, M., Miyawaki, Y., & Kamitani, Y. (2013). Neural decoding of visual imagery during sleep. *Science*, 340(6132), 639– 642.

Horikawa, T., Tamaki, M., Miyawaki, Y., and Kamitani, Y. (2013). Neural decoding of visual imagery during sleep. *Science* 340, 639–642. doi: 10.1126/science.1234330

Horwitz, A.V., and Wakefield, J.C. The loss of sadness: How psychiatry transformed normal sorrow into depressive disorder. Oxford University Press, 2007.

Howell, M. J., & Schenck, C. H. (2015). Rapid Eye Movement Sleep Behavior Disorder and Neurodegenerative Disease. *JAMA neurology*, *72*(6), 707–712. https://doi.org/10.1001/jamaneurol.2014.4563

Hu Y, Goodale MA. Grasping after a Delay Shifts Size-Scaling from Absolute to Relative Metrics. *Journal of Cognitive Neuroscience.* 2000;12:856–68

Hudetz A.G., Liu X, Pillay S. , et al. Propofol anaesthesia reduces Lempel-Ziv complexity of spontaneous brain activity in rats. Neurosci Lett 2016;628:132–5.

Humphrey, N. (2011). *Soul Dust: The Magic of Consciousness*. Princeton University Press.

Hyman, S.E. "The diagnosis of mental disorders: the problem of reification." *Annual review of clinical psychology* 6 (2010): 155-179.

Ichikawa, J. (2008). Scepticism and the imagination model of dreaming. *The Philosophical Quarterly,* 58(232), 519–527.

Ichikawa, J. (2009). Dreaming and imagination. *Mind and Language.* 24 (1):103-121.

Ichikawa, J. (2016). Imagination, dreaming, and hallucination. In Amy Kind (Ed.), *Routledge Handbook of the Philosophy of Imagination* (pp. 149–162). London: Routledge.

Insel, Thomas R. "The NIMH research domain criteria (RDoC) project: precision medicine for psychiatry." *American Journal of Psychiatry* 171.4 (2014): 395-397.

Irvine, E. (2012). *Consciousness as a scientific concept: a philosophy of science perspective*. Springer.

Irvine, E. (2019). Developing Dark Pessimism Towards the Justificatory Role of Introspective Reports. *Annalen der Philosophie* 86 (6):1319-1344.

Irving, Z. C. (2018). Psychology Off Tasks: Self-Report in the Science of Dreaming and Mind-Wandering. *Journal of Consciousness Studies* 25 (5-6):63-84.

Jackson, F. (1986). What Mary Didn't Know. *Journal of Philosophy* 83 (5):291-295.

Jackson, F. (2007). A priori physicalism. In Brian P. McLaughlin & Jonathan D. Cohen (eds.), Contemporary Debates in Philosophy of Mind. Blackwell.

Jorem, S. (2022). The good, the bad and the insignificant—assessing concept functions for conceptual engineering. *Synthese* 200 (2):1-20.

Jouvet, M. (2000). *The paradox of sleep: The story of dreaming*. Cambridge (MA) and London: MIT press.

Justus, J. (2012). Carnap on concept determination: methodology for philosophy of science. European Journal for Philosophy of Science 2 (2):161-179.

Kahane, G. & Savulescu, J. (2009). Brain damage and the moral significance of consciousness. *Journal of Medicine and Philosophy* 34 (1):6-26.

Kammerer, F. (2016). The hardest aspect of the illusion problem - and how to solve it. *Journal of Consciousness Studies* 23 (11-12):124-139.

Kendler, K. S., Zachar, P., & Craver, C. (2010). What kinds of things are psychiatric disorders? *Psychological Medicine*, *41*, 1143–1150.

Kendler, K. S. & Parnas, J. (eds.) (2012). *Philosophical Issues in Psychiatry II: Nosology*. Oxford University Press: UK.

Kendler, K.S & Gyngell. C (2020). 'Multi-level Interactions and the Dappled Causal World of Psychiatric Disorders. in Davies, Will; Savulescu, Julian & Roache, Rebecca (eds.) (2020). Psychiatry Reborn: Biopsychosocial Psychiatry in Modern Medicine. Oxford University Press.

Kennis, M., et al. "Prospective biomarkers of major depressive disorder: a systematic review and meta-analysis." *Molecular psychiatry* 25.2 (2020): 321-338..

Keogh, R & Pearson, J. (2018). The blind mind: No sensory visual imagery in aphantasia. *Cortex*. [online]. Available from https://doi.org/10.1016/j.cortex.2017.10.012.

Kerr N. H. (1983). The role of vision in "visual imagery" experiments: evidence from the congenitally blind. *Journal of experimental psychology. General*, *112*(2), 265–277. https://doi.org/10.1037//0096-3445.112.2.265

Key, B. (2016) Why fish do not feel pain. *Animal Sentience* 3(1).

Khalidi, M.A. (2013). Natural Categories and Human Kinds: Classification in the Natural and Social Sciences. Cambridge University Press.

Khalidi, M.A. (2018). Natural kinds as nodes in causal networks. *Synthese* 195 (4):1379-1396.

Kincaid, H. & Sullivan, J.A. (2014). Classifying Psychopathology: Mental Kinds and Natural Kinds. In Harold Kincaid & Jacqueline Anne Sullivan (eds.), Classifying Psychopathology: Mental Kinds and Natural Kinds. MIT Press. pp. 1-10.

Kind, A. (ed.) (2016). *The Routledge Handbook of Philosophy of Imagination*. Routledge.

Kingma, E. (2013). Naturalist accounts of mental disorder. In K. . W. . M. Fulford (ed.), The Oxford Handbook of Philosophy and Psychiatry. Oxford University Press. pp. 363.

Klein, C. (2017). Consciousness, Intention, and Command-Following in the Vegetative State. *British Journal for the Philosophy of Science* 68 (1):27-54.

Klein, C., Hohwy, J., and Bayne, T. "Explanation in the science of consciousness: From the neural correlates of consciousness (NCCs) to the difference makers of consciousness (DMCs)." *Philosophy and the Mind Sciences* 1.II (2020).

Koch, C., et al. "Posterior and anterior cortex—where is the difference that makes the difference?." *Nature Reviews Neuroscience* 17.10 (2016): 666-666.

Konkoly, K. R., Appel, K., Chabani, E., Mangiaruga, A., Gott, J., Mallett, R., Caughran, B., Witkowski, S., Whitmore, N. W., Mazurek, C. Y., Berent, J. B., Weber, F. D., Türker, B., Leu-Semenescu, S., Maranci, J. B., Pipa, G., Arnulf, I., Oudiette, D., Dresler, M., & Paller, K. A. (2021). Real-time dialogue between experimenters and dreamers during REM sleep. *Current biology : CB*, *31*(7), 1417–1427.e6. https://doi.org/10.1016/j.cub.2021.01.026

Kornblith, H. (1993). *Inductive Inference and its Natural Ground*. MIT Press.

Kosslyn, S.M., Thompson W.L., and Ganis, G. (2006) *The Case for Mental Imagery.* Oxford: Oxford University Press.

Kouider, S., Stahlhut, C., Gelskov, S. V., Barbosa, L. S., Dutat, M., de Gardelle, V., Christophe, A., Dehaene, S., & Dehaene-Lambertz, G. (2013). A neural marker of perceptual consciousness in infants. *Science (New York, N.Y.)*, *340*(6130), 376–380. https://doi.org/10.1126/science.1232509

Kriegel, U. (2015). Perception and Imagination. In S. Miguens, G. Preyer & C. Bravo Morando (eds.), *Pre-reflective Consciousness: Sartre and Contemporary Philosophy of Mind*. Routledge. pp. 245-276.

Kripke, S. 1980. *Naming and Necessity*. Oxford: Blackwell.

LaBerge, S. P. (1980). Lucid dreaming as a learnable skill – a Case Study. Perceptual and Motor Skills, 51 (3), 1039-1042. 10.2466/pms.1980.51.3f.1039

LaBerge (1985). *Lucid Dreaming.* New York, NY: Ballantine Books.

LaBerge, S. Levitan, L. and Dement, W.C. (1986): "Lucid dreaming: Physiological correlates of consciousness during REM sleep." *The journal of mind and behavior* 251-258.

LaBerge, S. & Gackenbach, J. (2000). Lucid dreaming. In E. Cardeña, S. J. Lynn & S. Krippner (Eds.) Varieties of anomalous experience: Examining the scientific evidence (pp. 151-183). Washington, DC: American Psychological Association.

La Berge, S.(2007). Lucid dreaming. In D. Barrett& P. McNamara (Eds.), *The new science of dreaming, Volume 1: Biological aspects* (pp. 307–328). Westport: Praeger Perspectives.

Ladyman, J. (2012). Science, metaphysics and method. *Philosophical Studies* 160 (1):31-51.

Lau, H. & Michel, M. (2019). A Socio-Historical Take on the Meta-Problem of Consciousness. *Journal of Consciousness Studies* 26 (9-10):136-147.

Laureys, S. (2005). The neural correlate of (un)awareness: Lessons from the vegetative state. Trends in Cognitive Sciences 9 (12):556-559.

LeDoux, J. ; Michel, M. & Lau, H. (2020). A little history goes a long way toward understanding why we study consciousness the way we do today. Proceedings of the National Academy of Sciences of the United States of America 1.

Lee, A. Y. (2018). Is consciousness intrinsically valuable? *Philosophical Studies* 175 (1):1–17.

Lee, A. Y. (forthcoming). Degrees of Consciousness. *Noûs*.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly* 64 (October):354-61.

Levy, N. (2014). *Consciousness and Moral Responsibility*. Oxford University Press.

Li, Yingjie, et al. "Abnormal EEG complexity in patients with schizophrenia and depression." *Clinical Neurophysiology* 119.6 (2008): 1232-1241.

Lilienfeld, S. O. "The Research Domain Criteria (RDoC): An analysis of methodological and conceptual challenges." *Behaviour research and therapy* 62 (2014): 129-139.

Lipton, P (2003). *Inference to the Best Explanation*. Routledge: UK.

Lipton, P. (1994). Truth, existence, and the best explanation. In A. A. Derksen (ed.), The Scientific Realism of Rom Harré. Tilburg University Press.

Loar, B. (1997). Phenomenal states (second version). In (N. Block, O. Flanagan, & G. Güzeldere, eds). In Owen J. Flanagan, Ned Block & Guven Guzeldere (eds.), *The Nature of Consciousness*. MIT Press.

Loar, B. (1999). Should the explanatory gap perplex us? In *The Proceedings of the Twentieth World Congress of Philosophy*. Philosophy Documentation Center. pp. 99-104.

Loar, B. (2003). Qualia, Properties, Modality. *Philosophical Issues* 13 (1):113-129.

Lovibond, P., & Shanks, D. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. Journal of Experimental Psychology,28(1), 3–26.

Ly, Calvin, et al. "Psychedelics promote structural and functional neural plasticity." *Cell reports* 23.11 (2018): 3170-3182.

Dawkins, M.S. (2012). *Why Animals Matter.* Oxford University Press. Oxford: UK.

Machery, E. (2005). Concepts are not a natural kind. Philosophy of Science 72 (3):444-467.

Machery, E. (2006). How to Split Concepts: A Reply to Piccinini and Scott. Philosophy of Science 73 (4):410-418.

Machery, E (2017) "Kinds or Tails?." *Extraordinary Science and Psychiatry*: 15-36.

Magnus, P. D. (2011). Drakes, seadevils, and similarity fetishism. *Biology and Philosophy* 26 (6):857-870.

Magnus, P. D. (2012). Scientific Enquiry and Natural Kinds: From Planets to Mallards. Palgrave-Macmillan.

Magnus, P. D. (2014). NK≠HPC. *Philosophical Quarterly* 64 (256):471-477.

Magnus, P. D. (2015). John Stuart Mill on Taxonomy and Natural Kinds. *Hopos: The Journal of the International Society for the History of Philosophy of Science* 5 (2):269-280.

Magnus, P. D. (2018). How to be a Realist about Natural Kinds. *Disputatio* 7 (8).

Maher, P. (2007). Explication Defended. Studia Logica 86 (2):331-341.

Maj, M. (2011). Differentiating depression from ordinary sadness: Contextual, qualitative and pragmatic approaches. World Psychiatry, 11(1), 43–47

Malcolm, N. (1956). Dreaming and skepticism. *Philosophical Review.* 65 (January):14-37.

Mandik, P. (2016). Meta-Illusionism and Qualia Quietism. *Journal of Consciousness Studies* 23 (11-12):140-148.

Margolis, E. and Laurence, S. "Concepts", *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2021/entries/concepts/>.

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, 64, 17-24.

Marques, T. & Wikforss, A.M. (eds.) (2020). *Shifting Concepts: The Philosophy and Psychology of Conceptual Variability*. Oxford: Oxford University Press.

Martin, M.G.F.(2001). Out of the Past: Episodic Recall as Retained Acquaintance, in C. Hoerl & T. McCormack (eds.), *Time and Memory*. Oxford University Press. pp.257-284.

Martin, M.G.F. (2002). The transparency of experience. *Mind and Language*, 17, 376–425.

Martin, M.G.F. (2004). 'The Limits of Self-Awareness'. *Philosophical Studies*, vol. 120, pp. 37-89.

Martin, M.G.F (2006). 'On Being Alienated' , in: T. S. Gendler & J. Hawthorne (eds.), *Perceptual Experience* . Oxford: Oxford University Press.

McGinn, C. (2004). *Consciousness and its Objects*. Oxford University Press University Press.

Mckelvie, S. (1995). The VVIQ as a psychometric test of individual differences in visual imagery vividness: A critical quantitative review and plea for direction. *Journal of Mental Imagery.* 19, 1-106.

Mckilliam, A.K. (2020). What is a global state of consciousness? Philosophy and the Mind Sciences: 1 (II).

McNamara, P. et al. "Impact of REM sleep on distortions of self-concept, mood and memory in depressed/anxious participants." *Journal of affective disorders* 122.3 (2010): 198-207.

Merikle P.M., Smilek D., Eastwood J.D. Perception without Awareness: Perspectives from Cognitive Psychology. *Cognition.* 2001; 79:115–34

Merritt, J. M., Stickgold, R., Pace-Schott, E., Williams, J., & Hobson, J. A. (1994). Emotion profiles in the dreams of men and women. *Consciousness and Cognition: An International Journal, 3*(1), 46–60. https://doi.org/10.1006/ccog.1994.1004

Metzinger, T. (2013). Why are dreams interesting for philosophers? The example of minimal phenomenal selfhood, plus an agenda for future research. *Frontiers in Psychology.* 4:746.

Metzinger, T. & Michelle Windt, J. (2007). Dreams. In D. Barrett & P. McNamara (eds.), *The New Science of Dreaming*. Praeger Publishers.

Michel, M. (2019). Consciousness Science Underdetermined: A short history of endless debates. *Ergo: An Open Access Journal of Philosophy* 6.

Michel, M. (2019). Fish and microchips: on fish pain and multiple realization. Philosophical Studies 176 (9):2411-2428.

Michel, M. (forthcoming). Calibration in Consciousness Science. *Erkenntnis*:1-22.

Morales, J. & Lau, H. (2020). The Neural Correlates of Consciousness. In Uriah Kriegel (ed.), The Oxford Handbook of the Philosophy of Consciousness. Oxford University Press. pp. 233-260.

Moruzzi, G., & Magoun, H. W. (1995). Brain stem reticular formation and activation of the EEG. 1949. *The Journal of neuropsychiatry and clinical neurosciences*, *7*(2), 251–267. https://doi.org/10.1176/jnp.7.2.251

Murphy, D. (2010). "Explanation in psychiatry." *Philosophy Compass* 5.7. 602-610.

Murphy, D. (2013) "The medical model and the philosophy of science." n K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Z. Sadler, G. Stanghellini, & T. Thornton (Eds.), *International perspectives in philosophy and psychiatry. The Oxford handbook of philosophy and psychiatry* (p. 966–986). Oxford University Press.

Murphy, D. (2014). Natural kinds in folk psychology and in psychiatry. In H. Kincaid & J. A. Sullivan (Eds.), *Classifying psychopathology: Mental kinds and natural kinds* (pp. 105–122). Cambridge: MIT Press.

Mutz, J. and Javadi, A.H.(2017). "Exploring the neural correlates of dream phenomenology and altered states of consciousness during sleep." *Neuroscience of Consciousness* n.

Nado, J. (2019). Conceptual engineering, truth, and efficacy. *Synthese* 198 (Suppl 7):1507-1527.

Nanay, B. (2016). Imagination and Perception. In Amy Kind (ed.), *Routledge Handbook of the Philosophy of Imagination*. Routledge.

National Institute for Health and Care Excellence.Depression in adults: recognition and management. NICE https://www.nice.org.uk/guidance/cg90 (2016).

Nemirow, L. (1990). Physicalism and the cognitive role of acquaintance. In William G. Lycan (ed.), *Mind and Cognition*. Blackwell.

Neuhäusler A., Schredl M., Göritz A. S. (2018). General knowledge about lucid dreaming and lucid dream induction techniques: an online study. *Int. J. Dream Res.* 11, 179–185. 10.11588/ijodr.2018.2.50491

Nielsen, Tore A. (2000). A review of mentation in Rem and NRem sleep: “Covert” Rem sleep as a possible reconciliation of two opposing models. *Behavioral and Brain Sciences* 23 (6):851-866.

Nir, Yuval & Tononi, Giulio (2010). Dreaming and the brain: from phenomenology to neurophysiology. *Trends in Cognitive Sciences* 14 (2):88-100.

Noordhof, P. (2002). Imagining objects and imagining experiences. *Mind and Language.* 17 (4):426-455.

Nutt, D. , Erritzoe, D., and Carhart-Harris, R.C. (2020). "Psychedelic psychiatry’s brave new world." *Cell* 181.1 24-28.

Nutt, D. (2019): "Psychedelic drugs—a new era in psychiatry?." *Dialogues in clinical neuroscience* 21.2 139.

O'Shaughnessy, B. (2000). *Consciousness and the World.* Oxford University Press.

O'Shaughnessy, B. (2002). Dreaming. *Inquiry: An Interdisciplinary Journal of Philosophy.* 45 (4):399-432.

O'Shaughnessy, B. (2008). *The Will: A Dual Aspect Theory*. Cambridge University Press.

Occhionero, M. & Cicogna, P. (2016). Phenomenal consciousness in dreams and in mind wandering. *Philosophical Psychology* 29 (7):958-966.

Osório, F. L., et al. "Antidepressant effects of a single dose of ayahuasca in patients with recurrent depression: a preliminary report." *Brazilian Journal of Psychiatry* 37.1 (2015): 13-20.

Otte, C.. et al. "Major depressive disorder." *Nature reviews Disease primers* 2.1 (2016): 1-20.

Oudiette D, Constantinescu I., Leclair-Visonneau L., Vidailhet M., Schwartz S., et al. (2011) Evidence for the Re-Enactment of a Recently Learned Behavior during Sleepwalking. PLOS ONE 6(3): e18056. https://doi.org/10.1371/journal.pone.0018056

Oudiette, D., Dodet, P., Ledard, N., Artru, E., Rachidi, I., Similowski, T., & Arnulf, I. (2018). REM sleep respiratory behaviours mental content in narcoleptic lucid dreamers. *Scientific reports*, *8*(1), 2636. https://doi.org/10.1038/s41598-018-21067-9

Overgaard, M., & Overgaard, R. (2010). Neural correlates of contents and levels of consciousness. Frontiers in Psychology, 1, 164. https://doi.org/10.3389/fpsyg.2010.00164%20

Overgaard, M. (ed.) (2015). *Behavioral Methods in Consciousness Research*. Oxford University Press.

Owen, M. , and Mihretu P.G (2019). "Physically sufficient neural mechanisms of consciousness." *Frontiers in systems neuroscience* 13: 24.

Ozdemir, E. (2021). Empirical Evidence Against Phenomenal Theses. PhD Dissertation, Victoria University of Wellington

Pace-Schott, E. F. & Hobson, J. A. (1998). The Neuropsychology of Dreams: A Clinico-Anatomical Study. *Trends in Cognitive Sciences.* 2 (5):199-200.

Pagel, J. & Blagrove, M. & Levin, R. & States, B. & Stickgold, B. & White, S. (2001). Definitions of Dream: A Paradigm for Comparing Field Descriptive Specific Studies of Dream. Dreaming. 11. 195-202. 10.1023/A:1012240307661.

Papineau, D. (1994). Mental disorder, illness and biological disfunction. Royal Institute of

Papineau, D.(2002). *Thinking About Consciousness*. Oxford University Press UK.

Paul, L.A. (2012). Metaphysics as modelling: the handmaiden's tale. *Philosophical Studies* 160(1):1-29.

Peacocke, C. (1985). Imagination, experience and possibility. In J. Foster & H. Robinson (Eds.), *Essays on Berkeley* (pp. 19–35). Oxford: Clarendon Press.

Pereboom, D. (2011). *Consciousness and the Prospects of Physicalism*. Oxford University Press.

Peterson A, Tagliazucchi E, Weijer C. The ethics of psychedelic research in disorders of consciousness. *Neurosci Conscious*. 2019;2019(1):niz013. Published 2019 Oct 9. doi:10.1093/nc/niz013

Phillips, I. (2018) The methodological puzzle of phenomenal consciousness *Phil.Trans. R. Soc. B*373201703472017034 7http://doi.org/10.1098/rstb.2017.0347

Pinder, M. (2020). On Strawson's critique of explication as a method in philosophy. *Synthese* 197 (3):955-981.

Prinz, J. (2016). Against Illusionism. *Journal of Consciousness Studies* 23 (11-12):186-196.

Prinzing, M. (2018). The Revisionist's Rubric: Conceptual Engineering and the Discontinuity Objection. *Inquiry: An Interdisciplinary Journal of Philosophy* 61 (8):854-880.

Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*. Routledge.

Putnam, H. (1975). *Mind, Language and Reality: Philosophical Papers*. Cambridge, England: Cambridge University Press.

Raatikainen, P. (2021). Natural kind terms again. European Journal for Philosophy of Science. 11 (1):1-17.

Radden, J. (2003). Is This Dame Melancholy?: Equating Today's Depression and Past Melancholia. Philosophy, Psychiatry, and Psychology 10 (1):37-52.

Radden, J. & Varga, S. (2013). The epistemological value of depression memoirsi a meta-analysis. In K. W. M. Fulford (ed.), The Oxford Handbook of Philosophy and Psychiatry. Oxford University Press. pp. 99.

Ratcliffe, M. (2013). A Bad Case Of The Flu?: The Comparative Phenomenology of Depression and Somatic Illness. Journal of Consciousness Studies 20 (7-8):198-218.

Ratcliffe, M. (2015). *Experiences of Depression: A Study in Phenomenology*. Oxford University Press.

Remulla, A., & Guilleminault, C. (2004). Somnambulism (sleepwalking). *Expert opinion on pharmacotherapy*, *5*(10), 2069–2074. https://doi.org/10.1517/14656566.5.10.2069

Revonsuo, A. (2005). *The self in dreams* (pp. 206– 219). The Lost Self: Pathologies of the Brain and Identity.

Revonsuo, A. (2006). *Inner presence: Consciousness as a biological phenomenon*. MITPress.

Revonsuo, A., Tuominen, J., & Valli, K. (2015). The avatars in the machine: Dreaming as a simulation of social reality. In T. Metzinger, & J. Windt (Eds.), *Open MIND*. Frankfurt am Main: MIND Group (optional: commentary and reply). https://open-mind.net/papers

Revonuso, A., Tuominen, J and Valli, K (2015). 'The Simulation theories of dreaming: how to make theoretical progress in dream science'. *OPEN MIND.*

Richards, C.S. and O'Hara M.W. (2014). *The Oxford handbook of depression and comorbidity*. Oxford University Press: UK.

Rosen, M. (2012). *Dream pluralism: a philosophy of the dreaming mind*. Thesis (PhD) Macquarie University, Faculty of Human Sciences, Department of Cognitive Science 2012. Available from: http://hdl.handle.net/1959.14/284236.

Rosen, M. (2013). What I make up when I wake up: anti-experience views and narrative fabrication of dreams. *Frontiers in Psychology,* 4:514. doi: 10.3389/fpsyg.2013.00514.

Rosen, M. (2015). I'm thinking your thoughts while I sleep: sense of agency and ownership over dream thought. *Psychology of Consciousness: Theory, Research, and Practice.* 2 (3):326-339.

Rosen, M. G. (2018). Enactive or inactive? Cranially envatted dream experience and the extended conscious mind. *Philosophical Explorations*, 21(2), 295– 318.

Rosen, M.G. (2018). How bizarre? A pluralist approach to dream content. *Consciousness and Cognition* 62:148-162.

Rosen, M., & Sutton, J. (2013). Self-representation and perspectives in dreams. *Philosophy Compass, 8*(11), 1041–1053.

Ross S., Bossis A., Guss J. et al (2016) Rapid and sustained symptom reduction following psilocybin treatment for anxiety and depression in patients with life-threatening cancer: a randomized controlled trial. J Psychopharmacol 30:1165–1180

Ruby P.M. (2020) The Neural Correlates of Dreaming Have Not Been Identified Yet. Commentary on "The Neural Correlates of Dreaming. Nat Neurosci. 2017". *Front. Neurosci.* 14:585470. doi: 10.3389/fnins.2020.585470

Rucker, J. JH, Iliff, J., and Nutt, D.J. "Psychiatry & the psychedelic drugs. Past, present & future." *Neuropharmacology* 142 (2018): 200-218.

Saarinen, J. (2018). A critical examination of existential feeling. Phenomenology and the Cognitive Sciences 17 (2):363-374.

Sanches, R.F., et al. "Antidepressant effects of a single dose of ayahuasca in patients with recurrent depression: a SPECT study." *Journal of clinical psychopharmacology* 36.1 (2016): 77-81.

Sankey, H. (1997). Induction and Natural Kinds. *Principia: An International Journal of Epistemology* 1 (2):239-254.

Sankey, H. (2001). Scientific realism: An elaboration and a defence. *Theoria A Journal of Social and Political Theory* 98 (98):35-54.

Sartre, J.P. (1940, 2012). *The Imagination*. Routledge.

Saunders, D. T., Roe, C. A., Smith, G., & Clegg, H. (2016). Lucid dreaming incidence: A quality effects meta-analysis of 50years of research. *Consciousness and cognition*, *43*, 197–215. https://doi.org/10.1016/j.concog.2016.06.002

Sawyer, S. (2020). Truth and objectivity in conceptual engineering. Inquiry: An Interdisciplinary Journal of Philosophy 63 (9-10):1001-1022.

Scarpelli, S., and De Gennaro, L. (2017). Electrophysiological pattern of dream experience. *J. Public Health Emergency*. 1:7. doi: 10.21037/jphe.2017.07.03

Scarpelli, S., D'Atri, A., Bartolacci, C., Mangiaruga, A., Gorgoni, M., and De Gennaro, L. (2019). Oscillatory EEG activity during REM sleep in elderly people predicts subsequent dream recall after awakenings. *Front. Neurol.* 10:985. doi: 10.3389/fneur.2019.00985

Scarpelli, S. (2019)., et al. "The functional role of dreaming in emotional processes." *Frontiers in psychology* 10 459.

Scharp, K. (2013). *Replacing Truth*. Oxford, England: Oxford University Press UK.

Schartner M.M., Carhart-Harris R.L., Barrett A.B. et al (2017).. Increased spontaneous MEG signal diversity for psychoactive doses of ketamine, LSD and psilocybin. Sci Rep 2017; 7:46421.

Schenck C.H., Mahowald M.W. (2002). REM sleep behavior disorder: clinical, developmental, and neuroscience perspectives 16 years after its formal identification in *Sleep*. *Sleep.* 2002; 25:120–30

Schomacker, A. (2020). *Bridging the Gap: How Conceptual Change Can Remove the Obstacles to a Scientific Explanation of Consciousness.* Schwabe Verlag Gmbh: Berlin, Germany.

Schredl M., Erlacher D. (2004). Lucid dreaming frequency and personality. *Pers. Individ. Dif.* 37, 1463–1473. 10.1016/j.paid.2004.02.003

Schwartz, S., & Maquet, P. (2002). Sleep imaging and the neuro-psychological assessment of dreams. *Trends in Cognitive Sciences, 6*(1), 23–30. https://doi.org/10.1016/S1364-6613(00)01818-0

Schwitzgebel, E. (2016). Phenomenal Consciousness, Defined and Defended as Innocently as I Can Manage. *Journal of Consciousness Studies* 23 (11-12):224-235.

Scott, G., and Carhart-Harris, R.L. (2019) "Psychedelics as a treatment for disorders of consciousness." *Neuroscience of consciousness* 2019.1 (2019): niz003.

Seth, A.K. (2016) 'Why fish pain cannot and should not be ruled out'. *Animal Sentience* 3(14).

Seth, A.K., Bayne, T. Theories of consciousness. *Nat Rev Neurosci* **23**, 439–452 (2022). https://doi.org/10.1038/s41583-022-00587-4

Shea, N. (2012). Methodological Encounters with the Phenomenal Kind. Philosophy and Phenomenological Research 84 (2):307-344.

Shea, N. & Bayne, T. (2010). The Vegetative State and the Science of Consciousness. British Journal for the Philosophy of Science 61 (3):459.

Shepherd, J. (2018). *Consciousness and Moral Status*. New York: Routledge.

Shepherd, J. & Levy, N. (forthcoming). Consciousness and morality. In Uriah Kriegel (ed.), *Oxford Handbook of the Philosophy of Consciousness*. Oxford University Press.

Siclari (2021). 'Neurobiology of dreams'. Reference module in neuroscience and biobehavioural psychology. https://doi.org/10.1016/B978-0-12-822963-7.00119-5

Siclari, F., & Tononi, G. (2017). Local aspects of sleep and wakefulness. *Current Opinion in Neurobiology, 44*, 222–227. https://doi.org/10.1016/j.conb.2017.05.008

Siclari, F., & Tononi, G. (2017). Local aspects of sleep and wakefulness. *Current Opinion in Neurobiology*, 44, 222– 227.

Siclari, F., Baird, B., Perogamvros, L., Bernardi, G., LaRocque, J. J., Riedner, B., … Tononi, G. (2017). The neural correlates of dreaming. *Nature Neuroscience*, 20(6), 872– 878

Siclari, F., LaRocque, J. J., Postle, B. R., & Tononi, G. (2013). Assessing sleep consciousness within subjects using a serial awakening paradigm. *Frontiers in Psychology*, 4, 542

Siewert, C.P. (1998). *The Significance of Consciousness*. Princeton University Press.

Sikka, P., Valli, K., Virta, T., & Revonsuo, A. (2014). I know how you felt last night, or do I? Self- and external ratings of emotions in REM sleep dreams. *Consciousness and Cognition: An International Journal, 25,* 51–66. https://doi.org/10.1016/j.concog.2014.01.011

Simion, M. (2018). The 'should 'in conceptual engineering. Inquiry: An Interdisciplinary Journal of Philosophy 61 (8):914-928.

Simion, M. & Kelp, C. (2020). Conceptual Innovation, Function First. *Noûs* 54 (4):985-1002.

Simon, J.A. (2017). Vagueness and zombies: why 'phenomenally conscious' has no borderline cases. *Philosophical Studies* 174 (8):2105-2123.

Sinhababu, N. (2016). Imagination and Belief. In Amy Kind (ed.), *The Routledge Handbook of Philosophy of Imagination.* pp. 111-123.

Sitt JD, King JR, El Karoui I, et al. Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. Brain 2014; 137:2258–70.

Slater, M. H. (2013). Are Species Real? An Essay on the Metaphysics of Species. Palgrave-Macmillan.

Slater, M. H. (2015). Natural Kindness. *British Journal for the Philosophy of Science* 66 (2):375-411.

Sneddon, L.U. and Leach, M. C. (2016) Anthropomorphic denial of fish pain. *Animal Sentience* 3(28)

Snodgrass, M. (2002). Disambiguating conscious and unconscious influences: Do exclusion paradigms demon-strate unconscious perception?American Journal of Psychology,115(4), 545–579

Soffer-Dudek N (2020) Are Lucid Dreams Good for Us? Are We Asking the Right Question? A Call for Caution in Lucid Dream Research. *Front. Neurosci.* 13:1423. doi: 10.3389/fnins.2019.01423

Solms, M. (1995). New Findings on the Neurological Organisation of Dreams: Implications for Psychoanalysis. *Psychoanalytic quarterly.* 64(1):43-67.

Solms, M. (1997). The Neuropsychology of Dreams: A Clinico-Anatomical Study, Hillsdale, NJ: Erlbaum.

Solms, M. (1999). The New Neuropsychology of Sleep: Commentary by Mark Solms. *Neuro-Psychoanalysis* 1: 183–95.

Solms, M. (2000). Dreaming and REM Sleep Are Controlled by Different Brain Mechanisms. *Behavioural and Brain Sciences* 23: 843–50.

Solms, M. and Turnbull, O. (2002). *The Brain and the Inner World*. New York: Other Press, LLC.

Sosa, E. (2005). Dreams and philosophy. Proceedings and Addresses of the American Philosophical Association 79(2): 7-18. Available from https://www.jstor.org/stable/30046211.

Sosa, E. (2007). *A virtue epistemology: Apt belief and reflective knowledge* (Vol. I). Oxford: Clarendon Press.

Sosa, E. (2009). *Reflective Knowledge*. Oxford University Press.

Soteriou, M. (2013). *The Mind's Construction: The Ontology of Mind and Mental Action*. Oxford University Press

Soteriou, M. (2017). Dreams, agency, and judgement. *Synthese*: 1-16. Available from https://doi.org/10.1007/s11229-017-1496-7.

Soteriou, M. (2019). I—Waking Up and Being Conscious. Aristotelian Society Supplementary Volume 93 (1):111-136.

Soteriou, M. (2020). Dreams, agency, and judgement. Synthese 197 (12):5319-5334.

Spener, M. (2015). Calibrating Introspection. *Philosophical Issues* 25 (1):300-321.

Speth, J., Frenzel, C., & Voss, U. (2013). A differentiating empirical linguistic analysis of dreamer activity in reports of EEG-controlled REM-dreams and hypnagogic hallucinations. *Consciousness and Cognition: An International Journal, 22*(3), 1013–1021. https://doi.org/10.1016/j.concog.2013.07.003

Steriade, M., Timofeev, I., & Grenier, F. (2001). Natural waking and sleep states: a view from inside neocortical neurons. *Journal of neurophysiology*, *85*(5), 1969–1985. https://doi.org/10.1152/jn.2001.85.5.1969

Strauch, I. and Meier, B. (1996) *In Search of Dreams: Results of Experimental Dream Research*. SUNY Press, New York.

Strawson, P. F. (1963). Carnap's Views on Conceptual Systems versus Natural Languages in Analytic Philosophy. In Paul Arthur Schilpp (ed.), The Philosophy of Rudolf Carnap. Open Court: La Salle. pp. 503–518.

Strevens, M. (2019). *Thinking Off Your Feet: How Empirical Psychology Vindicates Armchair Philosophy*. Harvard University Press.

Stumbrys T., Erlacher D., Johnson M., Schredl M. (2014). The phenomenology of lucid dreaming: an online survey. *Am. J. Psychol.* 127, 191–204. 10.5406/amerjpsyc.127.2.019

Stumbrys T., Erlacher D., Schädlich M., Schredl M. (2012). Induction of lucid dreams: a systematic review of evidence. *Conscious. Cogn.* 21, 1456–1475. 10.1016/j.concog.2012.07.003

Sturgeon, S. (2000). *Matters of Mind: Consciousness, Reason and Nature*. Routledge.

Sutton, J. (2009). Dreaming. In Symons J. & Paco Calvo (eds.). *Routledge Companion to the Philosophy of Psychology*. Routledge.

Sytsma, J. & Ozdemir, E. (2019). No Problem: Evidence that the Concept of Phenomenal Consciousness is Not Widespread. *Journal of Consciousness Studies* 26 (9-10):241-256.

Sytsma, J. & Machery, E. (2010). Two conceptions of subjective experience. *Philosophical Studies* 151 (2):299-327.

Szasz, T. S. (1961). The myth of mental illness: Foundations of a theory of personal conduct. HarpPeren.

Tabb, K. (2019). Philosophy of psychiatry after diagnostic kinds. Synthese 196 (6):2177-2195.

Tabb, K. (2015). "Psychiatric progress and the assumption of diagnostic discrimination." *Philosophy of Science* 82.5: 1047-1058.

Tahko, T. E. (2015). *An Introduction to Metametaphysics*. Cambridge University Press: Cambridge.

Takeuchi, T. (2005). 'Dream mechanisms: is REM sleep indispensable for dreaming?. Sleep and Biological Rhythms. https://doi.org/10.1111/j.1479-8425.2005.00165.x.

Tart, C. (*1965*). Toward the experimental control of dreaming: A review of the literature. Psychological Bulletin, 64, 81–91.

Tartaglia, J. (2016). What is at Stake in Illusionism? *Journal of Consciousness Studies* 23 (11-12):236-255.

Taylor, J.H. (2018). Powerful Qualities, Phenomenal Concepts, and the New Challenge to Physicalism. *Australasian Journal of Philosophy* 96 (1):53-66.

Taylor, J.H. (forthcoming). Consciousness as a natural kind and the methodological puzzle of consciousness. *Mind and Language*.

Tekin, Ş. (2016). Are Mental Disorders Natural Kinds? A Plea for a New Approach to Intervention in Psychiatry. *Philosophy, Psychiatry, & Psychology*, *23*(2), 147-163.

Thase, M. E. "Depression and sleep: pathophysiology and treatment." *Dialogues in clinical neuroscience* 8.2 (2006): 217.

Thompson, E. (2015). *Dreamless sleep, the embodied mind, and consciousness*. In T. Metzinger & J. Windt (Eds.), *Open MIND*. Frankfurt am Main: MIND Group (optional: commentary and reply). https://open-mind.net/papers

Thompson, E. (2015). Steps toward a neurophenomenology of conscious sleep. In Open *mind*. Frankfurt, Germany: MIND Group.

Thompson, E. & Batchelor, S. (2014). *Waking, Dreaming, Being: Self and Consciousness in Neuroscience, Meditation, and Philosophy*. Cambridge University Press.

Tore N., Microdream neurophenomenology, *Neuroscience of Consciousness*, Volume 2017, Issue 1, 2017, nix001, https://doi.org/10.1093/nc/nix001

Tsou, J. Y. (2013). Depression and Suicide are Natural Kinds: Implications for Physician-Assisted Suicide. International Journal of Law and Psychiatry 36 (5-6):461-470.

Tsou, J.Y. (2016). Natural Kinds, Psychiatric Classification and the History of the DSM. History of Psychiatry 27 (4):406-424.

Tully, I. (2019). Demarcating depression. Ratio 32 (2):114-121.

Tuominen, J. ; Stenberg, T. ; Revonsuo, A. & Valli, K. (2019). Social contents in dreams: An empirical test of the Social Simulation Theory. *Consciousness and Cognition* 69:133-145.

Tye, M. (2003). A theory of phenomenal concepts. In Anthony O'Hear (ed.), Royal Institute of Philosophy Supplement. Cambridge University Press. pp. 91-105.

Vallat, R., and Ruby, P. M. (2019). Is it a good idea to cultivate lucid dreaming? *Front. Psychol.* 10:2585. doi: 10.3389/fpsyg.2019.02585

Vallat, R., and Ruby, P.M. "Is it a good idea to cultivate lucid dreaming?." *Frontiers in psychology* 10 (2019): 2585.

Valli, K. (2011). Dreaming in the multilevel framework. *Conscious Cogn.* 20, 1084–1090. doi: 10.1016/j.concog.2011.04.004

Valli, K. & Revonsuo, A. (2006). Recurrent dreams: Recurring threat simulations? *Consciousness and Cognition.* 15 (2):464-469.

Valli, K., and Revonsuo, A. (2009). The threat simulation theory in light of recent empirical evidence: a review. *Am. J. Psychol.* 122, 17–38.

Valli, K. & Revonsuo, A. (2006). Recurrent dreams: Recurring threat simulations? *Consciousness and Cognition* 15 (2):464-469.

Veillet, B. (2015). The cognitive significance of phenomenal knowledge. *Philosophic*

Ventham, E. (2019). Reflective blindness, depression, and unpleasant experiences. Analysis 79 (4):684-693.

Voss, U., & Hobson, A. (2015). What is the state-of-the-art on lucid dreaming? In T. Metzinger, & J. Windt (Eds.), Open MIND. Frankfurt am Main: MIND Group (optional: commentary and reply). https://open-mind.net/papers

Voss, U., Holzmann, R., Tuin, I., & Hobson, A. J. (2009). Lucid dreaming: A state of consciousness with features of both waking and non-lucid dreaming. *Sleep*, 32(9), 1191– 1200.

Voss, U., Schermelleh-Engel, K., Windt, J., Frenzel, C. & Hobson, J. A. (2013). Measuring consciousness in dreams: The lucidity and consciousness in dreams scale. Consciousness and Cognition, 22(1), 8-21. 10.1016/j.concog.2012.11.001

Voss, U.; Schermelleh-Engel, K. ; Windt, J. ; Frenzel, C. & Hobson, A. (2013). Measuring consciousness in dreams: The lucidity and consciousness in dreams scale. *Consciousness and Cognition* 22 (1):8-21.

Voss, U, and Hobson, A.. What is the state-of-the-art on lucid dreaming? -Recent advances and questions for future research. Open MIND. Frankfurt am Main: MIND Group, 2014.

Voss, U, et al. "Lucid dreaming: a state of consciousness with features of both waking and non-lucid dreaming." *Sleep* 32.9 (2009): 1191-1200.

W. James. (1980). *The principles of psychology.* Vol. 2. Dover Publications.

Walton, K. L. (1990). *Mimesis as Make-Believe: On the Foundations of the Representational Arts.* Harvard University Press.

Wamsley, E.J. and Stickgold, R. (2021) 'The psychology of dreams'. Reference module in neuroscience and biobehavioural psychology. https://doi.org/10.1016/B978-0-12-822963-7.00226-7

Wikforss, A. (2010). Are natural kind terms special? In Helen Beebee & Nigel Sabbarton-Leary (eds.), The Semantics and Metaphysics of Natural Kinds. Routledge.

Wilson, J. M. (2021). *Metaphysical Emergence*. Oxford: Oxford University Press.

Wilson, R. A. (2005). *Genes and the Agents of Life: The Individual in the Fragile Sciences Biology*. New York, NY, USA: Cambridge University Press.

Wilson, R.A. (1999). *Species: New Interdisciplinary Essays*. MIT Press.

Windt, J. M. (2010). The immersive spatiotemporal hallucination model of dreaming. *Phenomenology and the Cognitive Sciences.* 9 (2): 295-316.

Windt, J.M (2013). Minding the dream self: Perspectives from the analysis of self-experience in dreams. *Behavioral and Brain Sciences* 36 (6):633-633.

Windt, J. M. (2013). Reporting dream experience: Why (not) to be sceptical about dream reports. *Frontiers in Human Neuroscience*, 7, 708.

Windt, J.M. (2015). Just in time - dreamless sleep experience as pure subjective temporality: A commentary on Evan Thompson. Open MIND. Frankfurt am Main: MIND Group.

Windt, J. M. (2016). Dreaming: A Conceptual Framework for Philosophy of Mind and Empirical Research.

Windt, J. M. (2018). Consciousness and dreams: From self-simulation to the simulation of a social world. In R. J. Gennaro (Ed.), The *Routledge handbook of consciousness* (pp. 420– 435). New York: Routledge.

Windt, J. M. (2019). Can a microdynamic approach to sleep-onset imagery solve the overabundance problem of dreaming? Commentary on Tore Nielsen's "microdream neurophenomenology". *Neuroscience of Consciousness*, 2019(1), niz005.

Windt, J. (2020). Consciousness in sleep: How findings from sleep and dream research challenge our understanding of sleep, waking, and consciousness. *Philosophy Compass* 15 (4).

Windt, J. M., & Voss, U. (2018). Spontaneous thought, insight, and control in lucid dreams. In The *Oxford handbook of spontaneous thought*. Oxford: Oxford University Press

Windt, J.M. & Metzinger, T. (2007). The philosophy of dreaming and self-consciousness: What happens to the experiential subject during the dream state? In Deirdre Barrett & Patrick McNamara (eds.). *The New Science of Dreaming.* Vol 3: Cultural and Theoretical Perspectives. Greenwood Publishing Group. pp. 193-247.

Windt, J.M & Metzinger, T. (2012). The self in dreams. In Encyclopedia of Sleep and Dreams: The evolution, function, nature, and mysteries of slumber(pp-). Greenwood Press.

Windt, J.M; Nielsen, Tore & Thompson, Evan (2016). Does Consciousness Disappear in Dreamless Sleep? *Trends in Cognitive Sciences* 20 (12):871-882.

Windt, J.M & Noreika, V. (2011). How to integrate dreaming into a general theory of consciousness—A critical review of existing positions and suggestions for future research. *Consciousness and Cognition* 20 (4):1091-1107.

Winlove, C.P. Milton, F. Ranson, J. Fulford, J. MacKisack, M. Macpherson, F. Zeman, A. (2018). The neural correlates of visual imagery: A co-ordinate-based meta-analysis. *Cortex.* Volume 105 (4-25).

Wittgenstein, L. 1967: *Zettel*. Berkeley: University of California Press.

Wong, M. (2016). The mood-emotion loop. Philosophical Studies 173 (11):3061-3080.

Wong, W., Noreika, V., Móró, L., Revonsuo, A., Windt, J., Valli, K., et al. (2019). The Dream Catcher experiment: blinded analyses disconfirm markers of dreaming consciousness in EEG spectral power. *BioRxiv Neurosci. Consciousness*. doi: 10.1101/643593

World Health Organization. International Statistical Classification of Diseases and Related Health Problems (11th Revision). 2020. https://icd.who.int/en. Accessed 2021.

Zadra, A. and Stickgold, R (2021). "When brains dream: exploring the science and mystery of sleep." New York, NY: W. W.Norton & Company, 2021.

Zeman, A.Z.J. Della Salla, S. Torrens, L.A. Gountouna V.E. McGonigle, D.J. Logie, R.H. . (2010) Loss of imagery phenomenology with intact visuo-spatial task performance: A case of 'blind imagination'. *Neuropsychologia*. Volume 48, 1 (145-155). doi: 10.1016/j.neuropsychologia.2009.08.024.

Zeman, A.Z.J. Dewar, M. Della Sala, S. (2015) Lives without imagery - Congenital Aphantasia. *Cortex* (378-30). https://doi.org/10.1016/j.cortex.2015.05.019.

Zeman, A. Z.J. Dewar, M. Della Sala, S. (2016) Reflections on aphantasia. *Cortex.* Volume 74 (336-337). doi: 10.1016/j.cortex.2015.08.015.

Zeman, A.Z.J. Mackiasack, M. Onians, J. (2018). The Eye's mind - Visual imagination, neuroscience, and the humanities. *Cortex.* 105.(1-3). doi: 10.1016/j.cortex.2018.06.012.

Zeman, A. et al. (2020). "Phantasia–the psychological significance of lifelong visual imagery vividness extremes." *Cortex* 130.: 426-440.

Zimmer, C. (2010). The brain. Discover, 28-29.