The challenge presented by dissociations and synaesthesia for the neo-dualism of David Chalmers and Tim Bayne

A thesis submitted for the degree of Doctor of Philosophy

Department of Philosophy
Reading University

Robert Fletcher

May 2020

Abstract.

This thesis has, as its primary target, the neo-cartesianism, or property dualism of certain philosophers of mind: David Chalmers, Tim Bayne, and others. All begin with a pre-theoretic commitment to the view that all perceptual states are conscious. They define consciousness by saying that it is synonymous with having 'qualia' – a term directed at phenomenal properties which defy reduction to physical states. The thesis argues that this position is challenged by certain neurological conditions, - blindsight, visual form agnosia etc- which we can generalise as 'dissociations': conditions in which functions are separated from awareness. The thesis holds that these are examples of unconscious perception, which present the case for a different pre-theoretic position, and a redefinition of concepts used in the vicinity.

Insofar as the neo-cartesians recognise the empirical, their inclination to reject the dissociations is challenged with new cases and paradigms which have emerged in recent years. It is contended that the dissociations disrupt Chalmers' coherence between the phenomenal and the psychological, Bayne's commitment to the Unity of Consciousness, and Phillips' campaign against unconscious perception. In addition, synaesthesia is advanced as a problem case for Bayne's unity contention. I argue that the dissociations are not easily dismissed, and present a real challenge for neo-cartesianism.

In exploring this challenge, the thesis contributes to the case for a different philosophy of mind – a Higher Order Thought (HOT) approach. The notion of unconscious perception – it is contended - much more easily finds an accommodation in HOT theories, so this thesis contributes to existing arguments for that position. HOT theories also have an advantage in being receptive to reduction, in giving a developmental account of consciousness, and in being more open to the accommodation of empirical discoveries. In each respect, there is an advantage over the respective positions taken by Chalmers and Bayne.

Acknowledgements

I would like to record my gratitude to those people who have assisted me in my work for this thesis. Primarily, my thanks go to my supervisors – to James Stazicker, in the first instance, and to John Preston, and Severin Schroeder, who stepped in when James left for King's London. All of them have been patient, understanding and properly challenging. I owe a great deal to them, and it has been a privilege to have had them as my supervisors.

My thanks also go to Ian Phillips, who was most kind to meet me and discuss his work. Any remaining failures in understanding or representing his work are mine alone.

Some years ago, I invited Larry Weiskrantz to give a talk on blindsight at the school where I taught philosophy. I remember the occasion with fondness, for he was generous with his time and gave an enthralling talk. We passed a very pleasant evening, and the talk obviously planted a seed in me. Sadly, he died in February 2018, so all I can do here is acknowledge his influence and record my thanks to him.

In similar vein, I would like to mention Bob Kirk, who was teaching Philosophy of Mind at Nottingham University when I arrived to study for a Master's degree in 1980. I attended some of his third year classes, although I had had no background in Mind before then. I am not sure I understood much, and my focus was elsewhere in the department, but I do recollect enjoying the time I spent in those classes. The seed he planted took some time to germinate, but here it is.

These acknowledgements would not be complete without the recognition of my wife and family's encouragement from the beginning. I dedicate the final draft to them.

Declaration

I confirm that this is my own work, and the use of all material from other sources has been properly and fully acknowledged.

Robert Fletcher

Introduction

At the centre of this thesis and its arguments, we shall consider a variety of pathological and psychological conditions which we shall call, for brevity, *dissociations*. There are wide differences between them, making it hard to generalise, but there are also similarities. In each case, some preserved function is present – an ability to detect, or locate, or discriminate between stimuli, even though the subjects do not consciously recognise the stimuli. There is a dissociation between function and consciousness. The interpretation and treatment of these conditions by different philosophical traditions and schools is the issue explored by this thesis. That they exist is not in contention; how they are dealt with and described is. The first chapter of this thesis will be given to a description of the conditions. They are each of them, in their own way, deeply surprising and fascinating, and their discovery has often astonished experimenters. In the appropriate chapter we shall also consider another strange condition-synaesthesia – which, while not a dissociation nor a pathological condition, presents a further challenge to the target philosophers we wish to discuss.

In the next chapter we will identify all of these conditions more fully, but for the purposes of introduction here, we will define dissociations uncontroversially as: some preserved function in the absence of acknowledged awareness. This description is borrowed from Weiskrantz (1999). We shall find that the *full* range of functions in each condition is impaired, and in none of the cases will we find a complete preservation of discriminative capacities, but enough discrimination remains to say that some functional ability to distinguish between stimuli and to make judgements is preserved – abilities normally ascribed to perception.

Several of these dissociations will be referred to in the course of this study, but chief among these curious cases is the phenomenon of *blindsight*. (This is best introduced in Weiskrantz, 2009). A good deal of this present work will be about blindsight, but the case it makes will not be confined to it. To the interested lay observer, blindsight can be described as a case in which the function of perceptual discrimination is preserved (for example, a subject can distinguish between two stimuli,) all the while claiming still to be blind, claiming not to be able to see the stimuli concerned. We rule out the possibility, and associated counter-claim, that they are being insincere.

We can refine this definition according to the audience interested in it, whether psychological, neuro-scientific, or philosophical. In the terms used by each discipline it is:

- Some residual visual function in a field defect, in the absence of acknowledged awareness. This is a definition couched in psychological terms.
- To a neuroscientist, it is defined as: visual processing by sub-cortical pathways, in the absence or lesion of the primary visual cortex (V1), specifically either via the superior colliculus, or the lateral geniculate nucleus.
- A philosophical description might be inclined to characterise it as: perception, or at the very least, successful discrimination, or visual processing, all the while lacking consciousness. But of course, *how* it is characterised by different philosophers will be the point at issue. As we shall see, some philosophers would not accept this account.

Other dissociations can be defined in similar ways, for example, the case of optic agnosia, as studied by Goodale and Milner (2005) can be rendered as: successful navigation and spatial adjustments guided by visual systems, but with a control of action unconscious to the subject, and accompanied by verbal reports to that effect.

In another case -Marshall and Halligan's study of visuo-spatial neglect (1998) - they report that their patient, PS, failed overtly to process information in the visual field corresponding to that part of the brain which had suffered a lesion, a feature which was not mitigated by her ability to move her head and eyes. And yet, she was nevertheless able to make successful discriminations concerning the un-experienced part of her visual field. This would again appear to be a case of successful discriminations without an acknowledgement or recognition of such.

We will also include in this study, references to 'split-brain cases' which are the result of a commissurotomy to constrain the effect of epilepsy, an operation which severs the corpus callosum (which provides a communication link between the two hemispheres of the brain.) In these cases, and under certain experimental situations, it would appear that the subject is capable of making successful discriminations, without being aware of the accomplishment, on the basis of stimuli being processed in only one half of their brain.

What is common to these dissociations is the claim that they involve successful discriminations – or visually guided behavioural responses, but without awareness. They

present the possibility that they can be styled as cases of unconscious perception, where the function of perception comes apart from the phenomenology.

Some of these dissociations have been part of the academic landscape for some time. Blindsight first made the headlines in the psychological world in the late 60s and early 70s with the work of Weiskrantz and Nicholas Humphrey. The split-brain cases won a Nobel Prize for Sperry in the 1960s. However, evidence for these dissociations accumulates, and new cases and discoveries are being made all the time; how far they impact on philosophy, however, is very moot.

Synaesthesia has only recently been taken more seriously by neurologists. New scanning techniques and devices, and experimental studies have brought respectability to the condition, and philosophers have begun to make room for what it implies for philosophical positions. We shall argue that synaesthesia creates difficulties for the unity thesis held by Chalmers and Bayne later in the thesis.

How do these dissociations, with synaesthesia, impact on philosophy? I shall argue that they may require us to make changes to our conceptual repertoire, and radical changes to the way we think of human beings. I shall argue that they pose a challenge to certain philosophers and their systems, and ultimately, make a case for a quite different picture of mind in HOT theory.

Philosophical concepts.

In philosophical thinking in this area there are many concepts at play – most obviously, *consciousness*. There are others: *perception, qualia, what-it-is-like, awareness*, and so on, but let us deal first with consciousness. In the course of this thesis, we shall most commonly understand this term to mean *phenomenal consciousness*, as this is the way in which it is used by our target philosophers. In this sense we understand it to mean the possession of mental qualitative properties, or qualia. At the relevant points, we shall want to use consciousness in this narrow sense. We shall, however, want to create room for the idea of unconscious perception at many points throughout. We can think of consciousness in this sense as 'awareness[1]', such that successful acts of perception can be accomplished without awareness.

[1] Chalmers however, uses 'awareness' as a psychological or functional term, recognising that in everyday language it is often used as a synonym for consciousness (Chalmers 1996 p28). I wish to retain the term as a synonym.

In this sense, we shall want to speak of awareness, or consciousness of stimuli. We shall further argue that this may also be accompanied by mental qualitative properties, though this will not be a notion shared by neo-cartesians.

How these concepts are tied together has much to do with the general commitment of philosophers to a wider theory of mind. Our target philosophers may not share these definitions, and a large part of our task will be to make a case for our picture, with our study of dissociations. We shall find that the extent to which the dissociations create problems for these wider theories turns upon the definitions advanced. Some will claim that dissociations create no difficulties. Thus: the view of Jason Holt (2003): 'Blindsight alters nothing in our conception of consciousness, nor does it call into question its ontological status' (Holt, 2003, p.47). Diametrically opposite, and concluding a section on blindsight, Patricia Churchland (2000) claims: 'from the point of view of philosophy, it is important to see that this is an instance where empirical discoveries put pressure on us to make conceptual revisions' (pp 227-8). Similarly, Dennett asks: 'what is going on in blindsight? Is it visual perception without consciousness, when the functions of vision are all present, but all the good juice of consciousness has drained out?' (Dennett, 1991, p.325).

This study will ask whether the concepts we deploy in thinking about consciousness need to be revised in the light of recent empirical work on dissociations, and whether the Churchlands are right to think that blindsight (etc) should be conscripted as evidence for a wholesale conceptual revolution in the philosophy of mind.

Neo Cartesianism

The target selected as vulnerable to a reassessment of dissociations is the theories of mind in the philosophies of David Chalmers, Tim Bayne and others. Most of the positions targeted in this thesis can be described as neo-cartesian, insofar as they subscribe to certain positions which can be described as cartesian in character, with one major exception - they do not follow Descartes in predicating a *substance* dualism, but claim that, while there is but one substance in the world, that substance has a dualism of *properties* – both physical and mental. The mental properties in question, however, resist any reduction to the material. An entirely separate and non-reductive status is claimed for them and their relation to the physical world is variously described as emergent, or epiphenomenal. As we shall see, these properties are captured by the term 'qualia', or 'raw feels'.

In other respects, however, the views of these philosophers are similar enough to those of Descartes to merit the widely used term: neo-cartesian. To offer a few preliminary examples: they begin with a certain internalist approach to knowledge about the mind, and place a reliance on introspection. Descartes took his cogito argument to be self-evident; the neo-cartesians do not quite go as far as to claim incorrigible knowledge of the subjective realm, but they still regard the subject as the best authority for these inner states.

Within this family of views, however, there are many nuances. We shall find it convenient to attack Searle for his contribution to features of this position, but we acknowledge that he refuses to describe himself as a property dualist (Searle 1992, 2012). McGinn shares much of the position described, but argues that inspection can only yield a limited understanding of the mind-body problem (McGinn 1991). One of our targets in this thesis, Ian Phillips, is reluctant to take on any label.

Within the broad class of neo-cartesian philosophical views, however, there are some who are committed to a view which we shall take as the main target of this thesis. This is the view that perception is always conscious. For those philosophers, what it is to be conscious is identified with the possession of the non-reducible mental qualitative properties as described. David Chalmers' 1996 book *The Conscious Mind* contains the clearest and most celebrated version of neo-cartesianism. His views are broadly shared by Tim Bayne, who co-authored a paper with Chalmers, and who shares an opposition to HOT theories. Bayne will be another target of this thesis.

It will be convenient, throughout this thesis, to refer to them as neo-cartesian, and to bracket them with other philosophers sharing or inspiring many of the views in that stable (Nagel 1974, Jackson 1982, Levine 1983), but our main target will be the view outlined above: that perception is always conscious. We shall argue that neo-cartesianism has insecure conceptual foundations, is poorly consonant with the empirical discoveries involved in the dissociations we shall discuss, and lacks the explanatory width of other competing philosophies of mind. It will argue that a conceptual reconsideration is prompted by a careful consideration of the dissociations, and once this is granted, the way is paved for a review of a wider philosophy of mind; one which has a greater explanatory width over neo-cartesianism. The way this will be argued is as follows.

The first chapter of this thesis will serve to introduce the dissociations. They will be discussed in some detail, and some similarities between them will be drawn. It will be

necessary in later chapters to discuss some of them in further detail, but this introductory chapter will identify key elements. Even a sketch of the conditions may be enough to begin the case for conceptual revision, for they are often deeply surprising in their different ways. We may find that our conventional vocabulary fails to do justice to them, and the insights they present.

In chapter 2 it will then turn to examine the philosophical concepts and arguments as pre-theoretical commitments of our target philosophers. I will argue that introspection does not provide an incorrigible foundation for the convictions of neo-cartesians. In this part of the exercise we shall follow the lead provided by Stalnaker, Schwitzgebel, Dennett and others. I will argue that concepts such as qualia and terms such as 'what-it-is-like' are problematic. The approach will be to put pressure on the central citadel of neo-cartesianism – the concepts in which their account is framed. It is not expected that this work will make a substantial contribution to the attack made by others on their definitions of consciousness and associated terms. However, positions taken by the neo-cartesians might recognise the empirical challenge implicit in any serious consideration of the dissociations, and it is here that we expect to make some contribution. I will argue that the neo-cartesian philosophers under study do not have a full appreciation of the potency of the challenge presented by the dissociations – and new developments and paradigms have come to put pressure on their way of handling them; if they have set them to one side in their work, new empirical findings suggest this might have been premature.

In recent years, new scanning techniques and mechanisms have been developed, and used in many of the studies of the dissociations. To a degree much more fine-grained than ever before, it is now possible to investigate the neural underpinnings of perception, consciousness, and so on. In this, the dissociations are particularly useful, in that they allow the identification of specific areas of the brain responsible for aspects of mental functioning particularly in perception (see Lamme, 2001). Where one of the relevant areas of the brain is missing, or lesioned, otherwise malfunctioning, or where a more primitive processing pathway is triggered, bypassing those areas responsible for awareness, we find important aspects of perception may be preserved, without the usual awareness. In short, the dissociations seem to provide a case for saying that there is such a thing as unconscious perception, a position expressly denied by the neo-cartesians.

Many of these advances have been made since the publication of Chalmers' seminal work in 1996. While we appreciate that empirical findings alone cannot dent a philosophical position; our case will be that scientific advance can give reason to re-think concepts. At the very least it can call into question the concepts scientists want to work with. This thesis will argue that positions within neo-cartesianism are built on conceptual foundations that do not sit comfortably with the empirical findings we shall discuss. It may be that these findings prompt a re-think of the concepts, and if this is the case, the door to quite different theories of mind is opened.

The thesis will then turn, in chapter 3, to a close examination of the work of David Chalmers. We shall consider first his definition of the phenomenal, and how this locks out, by definition, a full consideration of the dissociations. For him, as for other philosophers in the neo-cartesian camp, there is a pre-theoretic assumption that perception is necessarily conscious. In the first part of Chalmers' work therefore, the conditions are not given much shrift. Chalmers only considers blindsight as a problem for the structural coherence of the phenomenal with the psychological, which is a notion figuring later in his work. In this section, Chalmers argues for a close parallel between the phenomenal and psychological, which allows him to claim that consciousness is non-reducible, but his account is still naturalistic. He claims that this structural coherence constitutes 'a central and systematic relationship, which ultimately can be cashed into a relationship between phenomenology and underlying physical processes' (Chalmers, 1996, p.225). This principle can help us to make progress in understanding consciousness, as an 'epistemic lever, allowing researchers to infer conclusions about experience from third-person data', and as 'a background principle in the search for physical correlates of consciousness' (Chalmers, 1996, p.234).

However, in an important section, he discounts blindsight as posing a difficulty for this parallel, or structural coherence. But, in the view of this work, more recent refinement of paradigms and empirical work has done much to weaken his objections to blindsight. We shall consider his claim that blindsight can simply be discounted as degraded but normal vision, his suggestion that we cannot really definitively say whether there is experience in blindsight, and other claims. If it turns out that, notwithstanding Chalmers' arguments, in blindsight we have a case of awareness, which satisfies the demand for the control of behaviour and availability for report, (the psychological) but which does not involve consciousness as defined with no associated qualia, or which may not be described in terms

of ‘what-it-is-like’(the phenomenal) then Chalmers’ principle of structural coherence is in jeopardy.

The next chapter (4) on Tim Bayne’s principle of the Unity of Consciousness begins with Bayne’s (and Chalmers’) claim that consciousness is necessarily unified. This claim is advanced as a constraint on other philosophies of mind, which, if conceded, specifically rules out consideration of higher order thought theories. They write in a joint paper in 2010: ‘The higher order thought thesis is incompatible with the unity thesis […] if the higher order though thesis is true, the unity thesis is false. And if the unity thesis is true, the higher order thought thesis is false.’ (Chalmers, 2010, p.533). If various disorders, or conditions can be shown to be a problem for the Unity Thesis, we can at the very least lift this constraint upon higher order thought theories. We may find that such theories could be an improvement on the way dissociations are handled.

We shall argue that split brain cases present a problem for the unity principle, in that they provide evidence of failures of integration and unity. We shall also contend that synaesthesia presents cases of phenomenological ‘danglers’- a superfluity or a super-unity. Synaesthesia is a condition taken much more seriously in recent years. Evidence for the condition is now no longer exclusively reliant on the subjects’ testimony, but is backed up by brain scanning methods. On the question of the reality of that condition, we shall follow Blake *et al*. We shall argue such cases pose difficulties for the unity principle, but present further problems, even for a reductive account of the phenomenal states they involve.

Chapter 5 deals with the work of Ian Phillips, who argued against blindsight as a case of unconscious perception during the early work on this thesis, picking up some themes introduced but not fully developed by Chalmers. Amongst other considerations, these include claim that light scatter within the retina, filtering into the sighted field, is responsible for the effect, returning us to the idea that blindsight is degraded but normal vision. We deal with Phillips’ claim that the successful discriminations within dissociations are not truly unconscious and are an effect of a degraded awareness coupled with a conservative response criterion; we shall deal with his question marks about whether the dissociations are truly cases of perception, using the definition presented by Tyler Burge (and endorsed by Phillips). I will argue that blindsight and the other dissociations survive these challenges to their status. His exchanges with Ned Block must be assessed.

Phillips is unwilling to be pigeon-holed as a partisan in the wider debate over consciousness, but his sympathies clearly lie with neo-cartesianism, as his exchanges with Block testify. He is robust in explaining the dissociations in ways that do not allow that they are a challenge for a neo-cartesian position. His take on the conditions is perhaps the most recent and the strongest challenge to the standard interpretation of them, and so this work must deal with his position. Throughout, we shall review the empirical evidence for the phenomena, against the claim that the experiments establishing them are, in different ways, flawed. In dealing with his arguments, however, we shall find that we will again raise questions about the conceptualisation of consciousness. We may find that Phillips is inclined to define 'conscious' and 'perception' in a highly restrictive way, making it hard to see what can qualify as 'unconscious perception'.

In chapter 6, I will argue that empirical findings, and the change in concepts intimated, help prompt a case for a higher order thought theory of mind. Amongst other things, such a theory would have the added advantage, on this view, of offering a more thorough-going reductive account of mind. Higher order thought theory proceeds from the following assumptions. Phenomenal mental states are not necessarily conscious; some certainly are, but others form a class of mental states which are unconscious, but in which some discriminations amounting to perception are made. In arguing this, higher order thought theorists begin with a different pre-theoretic position on phenomenal concepts. They argue further that what makes a mental state conscious is its being accompanied by a higher order thought in ways outlined in this chapter. In this way, an account of conscious mental states can be given in terms of unconscious mental states. A reductive account might then be completed with the help of neurological advances. The failures and preservations of mental functions involved in dissociations, and the identification of specific areas of the brain as responsible, provide additional and fine-grained reasons for giving an account of mental states in terms of brain states.

We recognise throughout that evidence pointing to specific areas of the brain responsible for different functions will not slay the dualist. Discovering a real role for the pineal gland did not – could not – undo the commitment to mental substance of the Cartesian prepared to die in a ditch. The modern dualist takes a similar stance. In Chalmers' case, (and in the accounts given by similarly minded philosophers such as Nagel) his pre-theoretic conceptual position is immune to the advances in neurological science. On their account, science, in being an objective discipline, has no route into the subjective character of

consciousness. Chalmers appears interested in the neurological evidence, but only as a means of establishing a theory of coherence between what he describes as the phenomenal, and the psychological. However, all that such a theory of coherence can do, in his view, is to establish how closely the phenomenal correlates with the neurological, or, to be more precise, the psychological evidence of the neurological.

It is the argument of this thesis that the dissociations allow us to work towards a theory of mind that treats the neurological and psychological evidence not simply as evidence for correlation, but for a closer account which will allow that the phenomenal can be reduced to the physical. We shall also argue that a revision of concepts in the vicinity also allows us to give an account of the evolutionary development of consciousness in its widest sense, and its development in children. These are areas of consideration to which our target philosophers seem unable to contribute. For them, consciousness – understood as the possession of qualia - leaps into existence with no preliminary. It is an all-or-nothing matter, and there seems to be no account of the way in which we reach this condition. Higher order thought theory has such an account.

Chapter 1

The Dissociations.

The purpose of this section is to give a detailed introductory account of the conditions we shall discuss. Still further details of the experiments studying them may be given in the appropriate section of the chapters to follow, especially where features of the conditions and paradigms have a specific bearing on the argument being used. Here, however, we shall identify the dissociations, and advance some similarities between them. Some of the dissociations are the result of neurological dysfunctions, or the product of brain damage, or the by-product of operations to correct other problems; others apply to neuro-typical people, or can be artificially induced in subjects.

Beyond this, however, how they are characterised and presented, is philosophically moot. This needs to be said, as the characterisation presented below is not without controversy.

I argue that there is something common to all – namely that these are situations and conditions in which some function is separated or dissociated from consciousness. Thus, the function of perceiving can be detached from the conscious awareness of doing so: subjects can have perceptual sensitivity ie be demonstrably sensitive to their environment, without being consciously aware of this sensitivity. Some of these sensitivities extend to recognising a face, making choices concerning the emotional valence of that face, distinguishing between line drawings of animals, navigating a route past obstacles – all without awareness of that ability; in fact, in most cases, lacking any confidence in being able to do so. The controversy in this characterisation lies in the fact that some philosophers, as we shall see, are committed to defending the belief that 'all seeing is conscious' (Chalmers 1996 (p.222: '[..] where there is consciousness, there is awareness, and where there is (the right kind of) awareness, there is consciousness), Phillips, 2015, p.445) and that, if this is the case, the idea of unconscious perception is an oxymoron. Some therefore define terms to exclude the possibility of dissociation, and others do not accept that there is empirical evidence to confirm this as an interpretation of the conditions below (see O'Shaughnessy, 2002, p.415 et seq). Shelving this for the moment, let us identify the conditions in question.

Blindsight

Perhaps the most famous of the dissociations we shall examine is ‘blindsight’. The condition owes its name to Larry Weiskrantz, who worked on it, and promoted its study from a beginning in the 1970s. The term was coined in a hurry, before a presentation, but its oxymoronic character has much in common with the term: ‘unconscious perception’.

Weiskrantz first worked on a monkey, called Helen, who was used as a subject for the study of the primary visual cortex (V1 area) in vision. Helen was operated upon to remove the V1 area of her brain. This cortex is the most important point in the brain in which signals from the optic nerve are processed. Helen was rendered cortically blind by the operation; that is, she had her primary visual cortex removed by operation. For 19 months after the operation, she showed no visual capacity.

However, over a period, Nicholas Humphreys, a graduate student of Weiskrantz, began to work with Helen, and found that, with time and persistence, Helen showed an ability to react accurately to moving targets, and, with some false starts, she could systematically anticipate and skirt obstacles in her path, and ultimately behaved indistinguishably from other monkeys. Despite being cortically blind, she was able to function surprisingly well[2] (Humphreys, 1974).

Later in the 1970s, work shifted to two human subjects in Oxford. These were: DB, who, at the age of 33 underwent surgery to remove a tangle of abnormal blood vessels in his brain, in the right medial occipital lobe – a process which involved the removal of part of his V1 area – the primary visual cortex. This left him with a scotoma in his left visual field – an area in which he was blind. A part of his operation included the insertion of a metal clip to close off some blood vessels; this ruled out the possibility of conducting any further tests with an fMRI scanner, to establish the areas of his brain involved in processing his visual information.

GY: At the age of 8, GY sustained a traumatic brain injury to his left medial occipital lobe, and which again caused damage, this time to the corresponding V1 area processing information from the right optic nerve, leaving him with a scotoma in his right visual field.

[2] See: https://www.youtube.com/watch?v=6ek2LBqM7dk

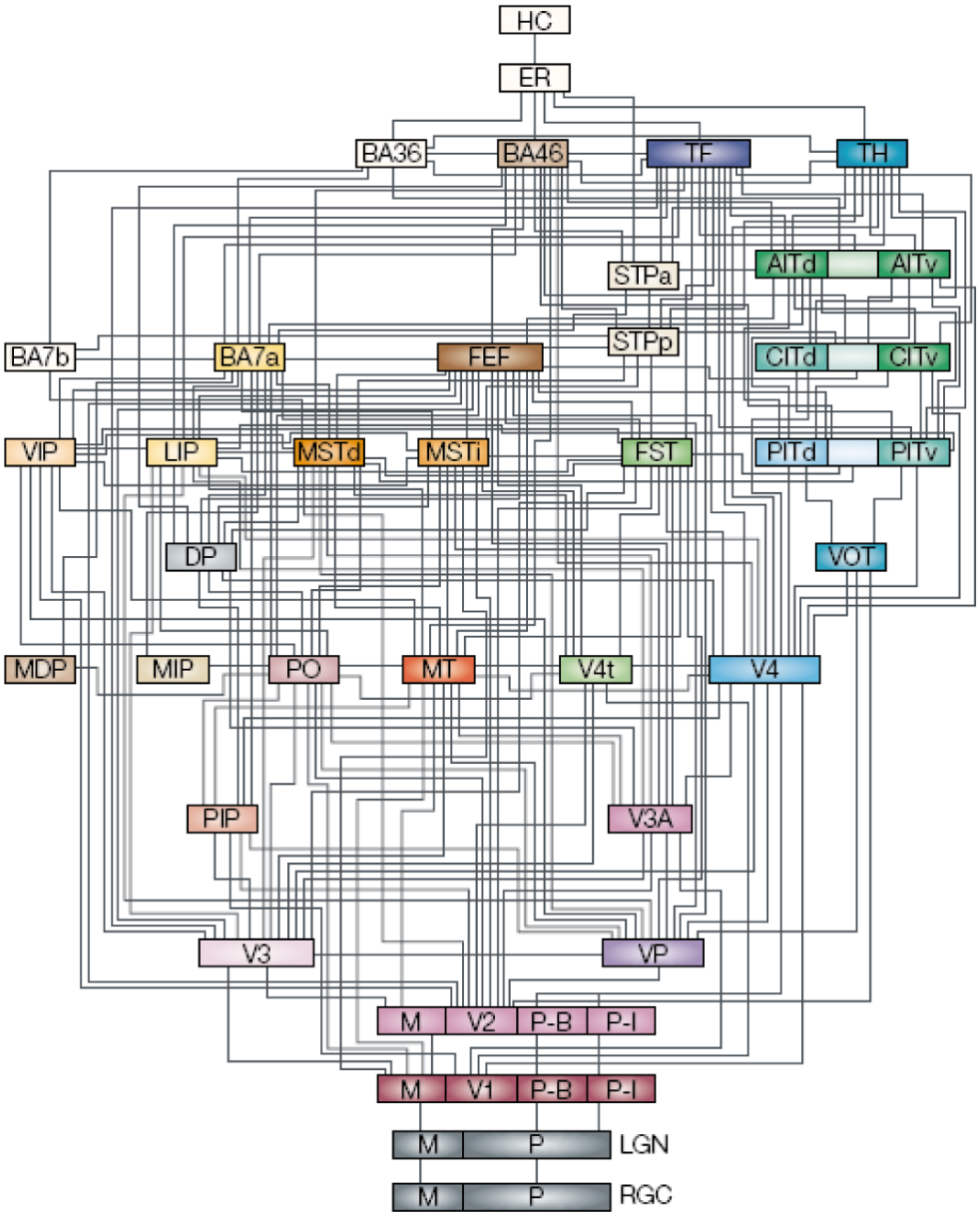
Both GY and DB have been extensively studied, and it was found that a degree of visual functioning has been preserved in their blind fields – they were able to make successful discriminations concerning various stimuli presented to their blind fields, way above chance.

Throughout, they protested that they could not see in that field, and were asked nevertheless to guess – at first in simple yes/no tests, and then in 2-alternative, forced choice (2afc) studies. Notwithstanding their protestations, their success rates in distinguishing between stimuli, and identifying features of stimuli, were 90-100%, where chance would be 50%.

A physiological explanation of blindsight starts with the observation that there are fully 10 different pathways from the retina to the brain – some by-passing the V1 primary processing area. The primary visual cortex is the recipient of 85% of information, in normal circumstances, arriving from the lateral geniculate, but between 10 and 15% of information arrives at different locations, making use of a number of neural fibres greater than the whole of the auditory system. These 'sub-cortical' or 'extra striate' routes have pathways to areas responsible for processing things like: motion, colour, shape, edge and some of which encode faces, and other familiar objects. Physiological studies provide evidence for no fewer than 20 different regions of the brain devoted to specific features of visual processing. Thus, seeing a cricket ball approach involves separate neurological processes, each of which is causally responsible for different things (the motion, the shape, the colour of the ball) each of which are then integrated, in the neuro-typical subject, into the complete item. The area known as V5 processes movement, V4 colour. Some features are processed faster than others.

The cases provided by DB and GY suggest that the contribution of V1 may not be necessary for the retention or recovery of some low level visual function. The claim is that, ordinarily, the V1 area has a role in creating visual awareness. It is contended by some (Weiskrantz, 1997, Lamme, 2001) that visual awareness is the product of feedback from the rest of the visual cortex to the V1 area, but when inactive or altogether absent, the subjects retain some visual capacity, yet the subject has no awareness of it.

(Weiskrantz reproduces Felleman and Van Essen's representation of the various connections between the primary visual cortex, V1 area, and the rest of the identifiable parts of the local cortex – a set of interconnections and feed-back structures so complex that it is has been called the 'oil refinery' Weiskrantz, 1997, p.205).



Since the first studies of DB and GY were conducted, other similarly affected subjects have been located and tested, with much the same results. The tasks set have moved on from simply asking subjects whether they are aware that an object is present in their blind field. Studies have looked into whether subjects can distinguish between horizontal and non-horizontal gratings, whether a light stimulus is coloured, what colour it is, in what direction it is moving, and so on. Other studies have looked into whether a word shown to the blindfield can prime the disambiguation of other words. (Thus the word: 'Water' might prime the subject to one meaning of the word 'Bank', and suggest a river bank, which might otherwise have been read as a branch of a high street bank, such as HSBC or Barclays.)

In all such cases, when asked to guess what they may have been shown, in their blindfields subjects have shown a degree of success in discriminations well above chance. In some cases, well above 90% correct. (See Trevethan *et al*, 2007 for particularly good illustration of successful form discrimination in DB).

Throughout, the response of subjects has always been to deny that they have any experience of the stimuli, and in the original trials, were forced to guess, 'could not see the point' and were astonished at their own results.

Superblindsight (TN)

Some opponents of the idea that blindsight can be regarded as unconscious perception have argued as follows: Blindsight does not count as unconscious perception, since, if perception involves access, and given that a thirsty blindsight patient (Marcel, 1986) would not reach for a glass of water in the blind field, it could not be considered that the blindsighted subject perceives the water.

Other arguments have been brought against the idea that, in blindsight responses are unspontaneous and must be prompted. Block however, raises the possibility of superblindsight, in which the response is not prompted, but spontaneous. In his paper on concepts of consciousness, (Block, 1995) he writes: 'A real blindsight patient can only guess when given a choice from a small set of alternatives ('X'/'O', horizontal/vertical etc). But suppose – interestingly, apparently contrary to fact – that a blindsight patient could be trained to prompt himself at will, guessing what is in the blind field without being told to guess' (Block in: Chalmers, 2002, p.211).

However, since Block wrote about his scepticism in finding a case of superblindsight, a case of bilateral damage to striate cortex has been investigated (de Gelder *et al,* 2008), in which a subject, TN, was found to retain sophisticated visuo-spatial skills in the absence of perceptual awareness. TN displayed quite a few of the more usual features of blindsight, including discrimination between rods presented horizontally and vertically to his blind field in forced choice tests, and a form of affective blindsight (see below) but the most interesting finding was that TN was able, unprompted, to navigate an obstacle course in a corridor

without colliding with randomly arranged boxes and chairs[3]. This case is as close to a case of superblindsight as has yet been found, since TN did not need to be prompted to make his decisions, nor was he made to choose between two available options.

Emotional contagion in blind fields: 'Affective Blindsight'.

DB and GY have been the subjects of many dissociation tests, but another from 2009 is sufficiently different to feature here (Tamietto *et al,* 2009). This study was not simply confined to the movement of a light source, but tested for the retention of a capacity for emotional contagion. This latter term refers to our tendency to synchronise our facial expressions with those of others. On encountering a shocked/sad/happy/angry/fearful face, there is a tendency for us to assume the same expression. Pictures of people with such faces, or bodily expressions were presented to the blind fields of DB and GY, and contractions and relaxations of their corresponding facial muscles were measured through electromyography – specifically two muscles, one exclusively involved in frowning, (Corrugator Supercilili) and the other involved exclusively in smiling (Zygomaticus Major). The study found that stimuli in the blind field nevertheless triggered the expected facial expressions; in fact, interestingly, it did so more quickly than the same stimuli shown to the unimpaired visual fields. DB and GY were also asked, in a 2AFC task, to guess the 'emotional valence' of the pictures shown to their blind fields, and were able to 'guess' accurately well above chance – in other words to discriminate correctly. The anatomical explanation of this is the suggestion that extra-striate areas, in particular, the amygdala, can still process the emotional content of the faces shown, possibly supplied by information via the superior colliculus, which remains intact after the destruction of v1 (Weiskrantz, 2009, p.10). It underscores the role of feedback to v1 as the means by which we become consciously aware of stimuli. This phenomenon has a great deal in common with prosopagnosia below.

Empirical developments in blindsight in recent years

Some recent empirical research has overtaken some of the early dismissals of dissociations. With this we can quickly dismiss one line of objection raised against blindsight. As recently as 2000, some philosophers were inclined to diminish the claims of blindsight studies on

[3] https://www.youtube.com/watch?v=ACkxe_5Ubq8

grounds of what was understood to be its limited nature. Thus: O'Shaughnessy: 'Several considerations ought to make us hesitate before accepting (the importance of blindsight). The first concerns the relatively low-key power of blindsight. (It is) restricted to 80% correct guesses, mostly unspontaneously elicited, concerning a few phenomena, like lights and their orientation from us' (O'Shaughnessy, 2000, p.417). And Holt: 'If blindsight is a form of perception, it is a quite rudimentary form. Only in simple tasks do patients perform better than chance. They cannot perform higher order tasks, like perceiving relationships between simultaneous blind field stimuli' (Holt, 2003, p.61).

In a relatively short period of time, new experiments and cases have conclusively seen off this first response to the condition. Success rates in simple tasks reach higher figures than 80% - in some cases, as much as 100%. Nor is the phenomenon confined to simple tasks. In his study of DB after some years had passed, Weiskrantz reports that in 2007, DB could identify low-contrast outline drawings of objects presented to his blind field, spontaneously, on first attempt with a success rate of 89% (25/28 items) (Trevethan *et al,* 2007). This success is worth some emphasis. His only failure was in not being able to distinguish between a horse and a stag, which, apart from its antlers was more or less identical. Furthermore, these drawings were presented with only 2% contrast, which made them very hard even for normally sighted people to see. Indeed, his blind field performance was significantly better than five normally-sighted control subjects.

DB was also able to report the presence of low-contrast 'Gabor' gratings with 8% contrast presented to his blind field – a sensitivity which exceeded the same test presented to his sighted field, where he was able only to detect a contrast of 12% (Trevethan *et al,* 2007). In another experiment, it was established that shapes and reactions can be primed in the blind field.

Many other discoveries have come to light in the last few years in a range of cases of dissociation, in which the processing of higher-order features is successfully accomplished, without conscious awareness[4]. Some of these developments will be explored more fully in later sections, but these recent empirical developments have a bearing on O'Shaughnessy's and Holt's claims above: that what is happening in these dissociative cases cannot be

[4] For example the unconscious processing of nude figures in the Jiang (2006) continuous flash study.

trivialised by claiming that it is only in simple tasks that the feature is found, or arguing that the discriminations in the tasks are rudimentary and do not involve higher order features.

'Numb-touch'

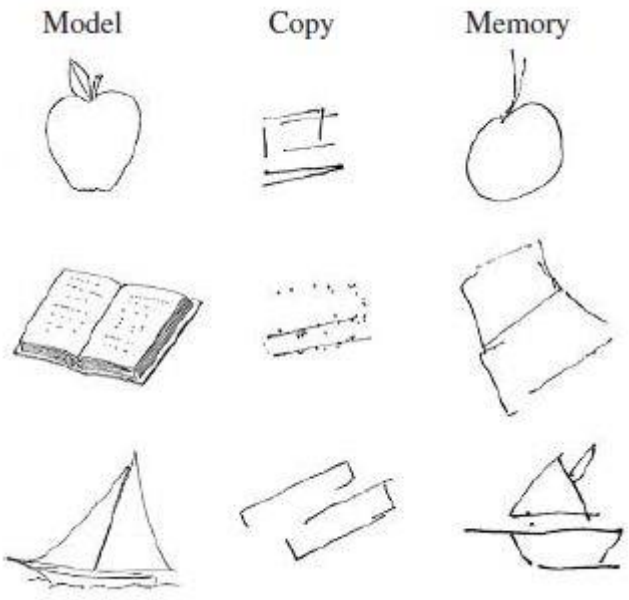
A case analogous to blindsight, but in a different sensory modality, was explored in France in 1982. The details are in: Paillard, Michel and Stelmach (1983). A woman, MSt, suffered damage to her left side parietal area, with a number of sequelae, among which was tactile deficit on her right side, so severe that she might cut or burn herself without noticing. During a series of tests, she was blindfolded and asked to respond if she felt pressure on her deafferented right hand. At first, she gave no response, even to the strongest pressure. However, when the test was repeated in subsequent sessions, the subject was progressively able to point, using her left hand, (and still blindfolded) to the location on her right hand where the pressure was being applied. This success was accompanied by her considerable surprise, and she commented: 'But I don't understand that. You put something there. I don't feel anything, and yet I go there with my finger. How does that happen?' She could also detect some gross differences in the size of objects being palpated by the right hand. Interestingly, the condition is different from the blindsight cases in one key respect – she did not need to be prompted to obtain a response, and so the case is immune to the criticisms of the two-alternative forced choice requirement in the blindsight tests.

Visual agnosia and 'vision for action'. Milner and Goodale's work with Dee Fletcher.

The lead work done on visual agnosia was done by Milner and Goodale, (2005) who researched the case of DF, a young woman who suffered the consequences of carbon monoxide poisoning, while taking a shower in her house in Italy. At first it was imagined she was cortically blind, like many of the blindsight patients above, but with no spared areas in her visual field.

Over time, she recovered some visual experience. She first experienced a vivid sensation of colour, but it became quickly obvious that she had lost an ability to recognise objects on the basis of their form alone. She had sustained some very regionally specific brain damage as a result of her hypoxia. In tests, she could be shown patterns of lines, even very faint in contrast, but could not tell if they were horizontal, vertical, or obliquely angled. She could not

identify objects from line drawings of them, nor could copy line drawings as given (see below). But she could draw objects from memory; thus it was established that her visual memory was unimpaired. The key finding in the studies of her was that she did not suffer at all from any motor disability. She had no problems at all in walking, picking things up, or angling her hands to receive or give things. In experiments where she was offered a pencil with either an horizontal or vertical orientation, she could not tell if it was one or the other, but she aligned her hand correctly to receive it.



She could walk unassisted, along a forest path, strewn with rocks and tree roots, with no difficulty (Milner and Goodale, 2005, p.19).

It was on the basis of these findings that Milner and Goodale, following work on primates done by Ungerlieder and Mishkin (1982), established that there are two separate, relatively independent visual processing systems. Both highways begin in the primary visual cortex V1, but then take different routes geographically: one along a ventral stream, ending at the bottom and sides of the hemispheres, in the inferior temporal region; the other takes a pathway higher in the cerebral cortex, ending in the posterior parietal cortex, along a so-called dorsal stream. The two highways have some interconnections, but would appear to be functionally specific. The ventral stream is responsible for the processing of retinal signals for the purpose of forming a perceptual representation, and it is this route that was most affected by DF's hypoxia. This stream creates a rich representation of the scene being perceived, but does not have any direct connection to the motor system. It is this stream that is involved in perceiving a football, but has no role in instructing the body to align itself in

such a way as to kick it into the goal. This function, which could be described as a more 'on-line' function, is played out by the dorsal stream, and is responsible for 'vision for action'. The evidence for this separation of visual processing would be stronger if a case where the dorsal stream were damaged, but not the ventral stream – the converse of DF's case. In fact the discovery of just such a condition was made much earlier by Rudolph Balint in 1909. Balint's subject was not agnosic like DF; he could recognise objects and people. He suffered though, from an inability to reach out and pick up objects. He would fumble and grope around, much like a blind man. He could still see the object of his attention, but he could not orient himself or his hand to pick it up.

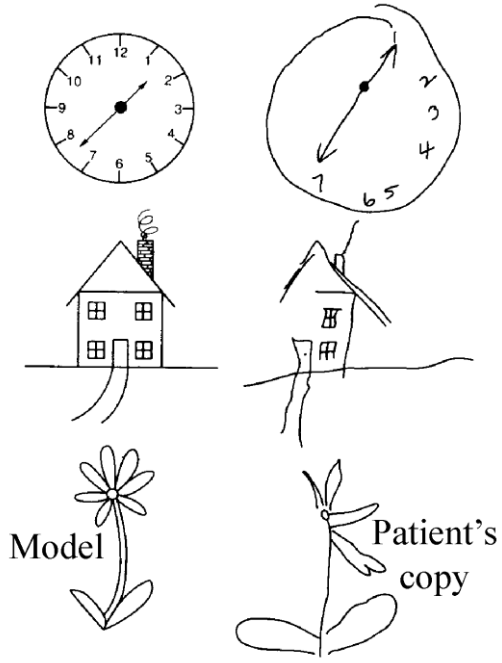
From our point of view, however, the important thing to acknowledge is the claim that conscious visual experience is a product of the ventral, not the dorsal stream. The dorsal stream is responsible for the processing of information to be used in the alignment of the hands and body in (say) returning a tennis serve. This conversion of visual information directly into action is done entirely without awareness. This remains intact in DF's case; however, her processing in the ventral stream is compromised – she finds it hard to tell where one object ends and another begins, especially if they are of similar colour. She cannot say what shape a presented object has. So, in DF's condition we have another case in which we have successful discrimination, (in the situations in which DF is orienting herself to pick up objects etc) but without any awareness that she is so doing.

This is not to say that visual information processed by the ventral stream always reaches awareness. Milner and Goodale concede that at present, they have 'no real idea what the critical difference is between neural activity that reaches awareness, and that which does not' (Milner and Goodale, 2004, p.114). But they would seem to take the view that the ventral stream is responsible for unconscious perception of the kind that is at work in subliminal stimuli (shown for very brief periods, say 50 microseconds), and which is at work in prompting behaviour.

Hemispheric neglect

In hemispheric neglect, we find a condition occasioned by damage in a different area of the brain - subjects have all of them suffered compromise to the right hemisphere to the brain – either through a stroke, or external trauma. These subjects are, many of them, hemianopic – that is to say, blind in their left visual fields. Damage to the right hemisphere means that visual information, processed normally in their right hemisphere, is extinguished. Subjects

with the condition do not seem to realise that the left hand side of their world even exists – women might apply make-up to one side of their face, men might only shave the right side of their face. When drawing pictures, subjects leave the left side of clocks, cats, flowers incomplete. See fig 1 below. Bisiach and Luzzatti (1978) present a study in which subjects who know Milan well, are invited to imagine being in the cathedral square, facing the Duomo, and to describe what they would see. The subjects who know the square well, typically describe all the buildings and facades that lie to the right of their orientation, but omit those to their left. If they are then told to imagine they turn through 180 degrees and to repeat their description of what they see, they describe the buildings they had previously left out.



There is a difference of opinion concerning whether the condition is a failure of attention, or a failure of the brain to represent space, but this may be a debate we can avoid commitment to, here. The condition is by no means rare (see the recent study in Guariglia *et al,* 2013) but let us focus on one study which is of particular interest to us. Marshall and Halligan (1988) reached the conclusion that stimuli in the neglected field, which are not therefore consciously seen, can still influence or prime responses. Their paper was the result of the study of a single patient, who was shown a series of line drawings featuring two pictures of a house, otherwise identical except that in some cases, smoke and flames are pictured as coming from the window of the second house on its left side (see below). The subject reported that the pictures

appeared identical throughout, since her neglect extinguished that part of the picture which featured the smoke and fire. When however, she was then asked which house she would prefer to live in, the subject always chose the house without the fire, though could not say why, and imagined she was guessing. This was the case, even though the subject was allowed free movement of her head and eyes, and so the effect was not a direct consequence of sensory loss in the left visual field. The subject simply failed to see the smoke on that side of the presented images.

Other studies have explored the condition still further, specifically on the question of whether unseen words presented in the neglected field can act as a prime for tasks given to the field showing no neglect (Berti and Rizzolatti 1992). Subjects were shown pictures in their left, neglected field as primes for pictures (of fruit, or animals) in their right, which they then had to classify correctly. It was found that the subjects could classify correctly very much faster if they were primed with a prime that was cognate with the picture shown to the side showing no neglect.

Berti and Rizzolatti conclude: 'patients with neglect are able to process stimuli presented to their neglected field to a categorical level of representation even when they deny the stimulus presence in the affected field'.

In his own review of dissociations, Weiskrantz describes the findings in this way: 'The subject may not "know" it, but some part of the brain does' (Weiskrantz, 1997, p.26). There are clear similarities with the phenomenon of blindsight. This account, which ascribes

psychological predicates to parts of the brain, is a controversial way of describing the situation, and we will address this issue later. For now, we might say that some processing on one side of the brain becomes available to the other side of the brain, but without the accompanying awareness of the stimulus in question.

Prosopagnosia

The characteristic features of this condition are these: subjects have suffered brain injury to the inferior posterior temporal lobe, mainly in the right hemisphere, but retain the ability to see. Their injury however, means that they are no longer able to recognise people from their faces alone, even the faces of family members. They are able to recognise people from non-visual clues, such as their voices. Their ability to see, and the ability to recognise people on the basis of their voices rules out blindness, or some form of intellectual impairment. Various studies have been conducted to see if such subjects nevertheless retain some form of preserved recognition. One used by Bauer (1984) made use of a test for galvanic skin response (not dissimilar to a lie detector test, in which a subject will involuntarily show some physiological response to stimuli related to a crime). Bauer worked with a subject LF, who was shown five faces for 90 seconds, accompanied by five names. For four of them, the name did not match the face; one of the five had the correct name attached. When LF was asked to choose which of the faces had the correct name, he chose at chance level (22%). However, when he was tested by using the machine measuring his skin conductance (rather than his verbal response) the responses as measured by the machine were 61% of trials – well above chance. Young and de Haan (‘*Face Recognition and Awareness after Brain Injury’* in Milner and Rugg, 1992) conclude that ‘it is inadequate to think (that LF has lost) recognition mechanisms. Instead, at least some degree of recognition does take place; what LF has lost is awareness of recognition’.

Young and de Haan worked on their own subject, PH, who seems to be doubly unfortunate – reporting: he is ‘profoundly impaired at recognising familiar faces. Most faces he simply says are unfamiliar. The only face we have noticed (from the hundreds he has been shown) being fairly consistently recognised is Mrs Thatcher’s.’ On formal tests, PH recognised 0/20 highly familiar faces, and was at chance in deciding whether the faces were familiar. In a forced-choice test in which he was asked to pick one of two simultaneously presented faces (one familiar, one unfamiliar) he was at chance in picking out the familiar. When, however, he was primed with names, in a parallel task, he was much more accurate

(118/128 correct). In this and other tasks, Young and de Haan establish that PH retains an ability covertly to recognise faces. They conclude that normally faces are stored away in recognition units in which the various features of faces are encoded, Even in prosopagnosia cases, this encoding persists. In the case of those with the condition, there is some impaired output from these units, and they remain disconnected from the other parts of the processing system. However, it is still possible, by unexpected indirect means, to access these units, and so some 'covert' recognition is preserved.

Cerebral achromatopsia

In this condition, subjects lose the perception of colour, - they will say that everything seems to them to be black, white or shades of grey, but yet studies indicate that they retain an 'implicit' or 'covert' ability to discriminate between colours. GW Humphreys *et al* report on their subject HJA ('*Covert Processing in Different Visual Recognition Systems'* in Milner and Rugg 1992). HJA suffered a bilateral lesion in the occipital cortex, and in consequence, a complete loss of colour vision[5]. When asked to name the colour of Mondrian patches offered to him, HJA would say he was guessing, but typically, he was 40-50% correct, where chance would have yielded a result of 6.3%. In another, slightly different, forced-choice study, HJA was asked to match colour patches. He was given a patch, and asked to choose from a collection of 20 others, which was the sample that matched his own. With widely separated colours, HJA got scores of 60% correct, when chance would be 5%.

HJA would sometimes claim that he was very confident of having got some tests right, but his confidence would have no bearing on whether he was right or not – there was no link between confidence and objective performance. As with blindsight, therefore, there would appear to be evidence of successful discrimination without awareness. Humphreys *et al* describe his case: '[In HJA] there is residual covert processing of colour, without colour identification […] he reports no conscious experience of colour' (Milner and Rugg, 1992, pp. 62-64).

[5] Coincidentally, he also was left with visual object agnosia, which we will not develop here, beyond saying that in his case, the agnosia was not accompanied with any covert recognition of objects. However, studying HJA's condition allowed Humphreys *et al* to conclude that the areas of the brain processing the recognition of objects and the recognition of colours were functionally separate.

Partially successful perception in post-commissurotomy patients.

Sperry's split-brain patient NG (Sperry, 1968) was shown images which one hemisphere alone (the right) could perceive (flashed briefly into the field of vision of the left eye). When asked if she saw images shown to the right hemisphere, she said 'No, nothing'. When the experiment was repeated for the other, left hemisphere, and NG asked if she saw images shown to that hemisphere, she correctly reported the image. However, when shown images to the right hemisphere, and then asked to reach under a screen and pick out, with her left hand, the object shown, she palpates a number of objects, and then correctly identifies the one shown in the image. On the face of it, it would seem that the right hemisphere is capable of making discriminations but without awareness, or at least without awareness being available to the hemisphere that controls speech.

The study went further. NG was flashed a picture of a nude woman to her right hemisphere. The effect is reported in Springer and Deutsch 2003: "NG's face blushes a little, and she begins to giggle. She is asked what she saw. She says: 'Nothing, just a flash of light', and giggles again, covering her mouth with her hand. 'Why are you laughing then?', the investigator inquires. 'Oh doctor, you have some machine!' she replies" (Springer and Deutsch, 2003, p.36). In this case, it would seem her right hemisphere processed the image to a point where a response of a certain kind is made (her embarrassment), though this was non-verbal, in view of the fact that the verbal centres of the brain are in the left hemisphere. The comment: 'Oh doctor, you have some machine' is a sign perhaps that the left hemisphere had registered the emotional and physiological responses triggered by the right hemisphere, and, in common with other cases, tried to make a verbal rationalisation of the situation, with incomplete information.

A similar kind of dissociation has been recognised in the field of amnesia. Amnesic patients have been shown to have good retention of certain types of events, even though the patients themselves say of themselves that they have no memory of them as such. (See: '*Consciousness and awareness in Memory and Amnesia'* by Daniel Schacter in Milner and Rugg, 1992).

Cases not involving brain damage.

Lest it be thought that unconscious perception could only be the result of deficit, or the by-product of trauma, we must acknowledge that there can be replicable cases of unconscious perception created in normal subjects. One of the advantages of being able to do so, is being able to side-step the worry that brain damage might also affect the cognitive processing underlying their reports, and so introducing uncertainties or biases to the paradigm. In his exchange with Phillips, Block (2015) adverts to Breitmeyer 2015 which is a review of 24 different ways in which unconscious perception can be induced in normal subjects by various means of 'blinding' them to stimuli. Some of these ways involve presentations to subjects of visual stimuli over very short time periods[6], and masked by other stimuli flashed to them. Other studies involve and exploit binocular rivalry. One example is styled 'continuous flash suppression' – a study in which a changing sequence of Mondrian colour patterns are presented to the perceptually dominant eye, while the perceptually subordinate eye is presented with the target stimulus of a word or picture. The sequence of changing colours suppresses conscious awareness of the target stimulus, for up to several seconds. When asked, subjects say they have no knowledge of what the stimulus is – they are just guessing – and there is no correlation between the confidence they have in their answers and what they have actually seen. They really are just guessing – but again, in their guesses, they are well above chance in giving the right answers.

More recently, cases of blindsight can be induced in subjects by application of trans-cranial magnetic stimulation such that, for example, they can perform fast corrective reaching movements without being aware of the presence of the visual signal that prompts their behaviour (see: Ro *et al,* 2004, Christensen *et al,* 2008).

Early Vision

Tyler Burge's definition of perception is pivotal in a later chapter. He writes: 'certain states in early vision (states in the first micro-seconds of visual processing) may count as perception by the individual, but fail to be conscious' (Burge, 2010, p.375). I illustrate this with the following experience. One early summer morning, some years ago, I walked to the bottom of

[6] Ie well below the time necessary for the stimuli to become consciously represented – typically around 40 ms

my garden. As I approached an unkempt area, the following events unfolded quickly and in this order: I felt uneasy; I leapt sideways, and it was only then I became aware of the 50 cm snake sunning itself on the ground. (It was a grass snake, and retreated quickly to cover). Gray writes: 'we become conscious of events only after we have had time to respond to them behaviourally, and often already have done so' (Gray, 2004, p.90). Estimates of how much time it takes for visual processing to become something of which we are conscious vary, but a common estimate is about 250-300 ms. This is not a trivial length of time, and might be the difference between surviving and not surviving an encounter with a predator. Detecting that predator at an unconscious level bestows an evolutionary advantage, and waiting for a conscious appreciation of the danger would be a disadvantage. An unconscious perception of such challenges is thus hardly a surprise, and we must have similar experiences often, even if we do not notice the sequence of events so vividly.

Perception under anaesthetic.

A particularly striking example of unconscious perception is given by a finding that patients under general anaesthetic can respond to questions and can be primed by auditory stimuli. In 1992, Jelicic *et al* reviewed several studies working on the presentation of auditory stimuli during anaesthesia, making the following finding: While no subject remembered any words or sounds played to them, it was clear they were successfully primed while under anaesthetic. Groups of colour words and fruit words were played to them, and after recovery, the primed group disproportionately responded to prompts asking them to name 4 colours, 4 fruits, with the colours and fruits suggested to them under anaesthetic. Jelicic *et al* conclude that there is therefore 'convincing evidence' of the retention of processing leading to the formation of an 'implicit memory' under anaesthetic. Subjects were unaware of having been primed, and have no explicit memory after the fact; The subjects were unconscious, but at some level, perceived the tapes played; that is, they successfully processed auditory information and encoded it in working, or 'implicit' memory.

Synaesthesia

As acknowledged, synaethesia is not strictly speaking a case wherein function is separated from awareness. If blindsight and the other dissociations are cases of awareness without the

qualitative feel, synaesthesia would seem to be a case of awareness with too much qualitative feel. However, we must not rush to the conclusion that the condition thereby validates the view of qualia held by qualia-supporters. On examination, other issues present themselves, and synaesthesia may present problems for those who cleave to the idea that consciousness is unitary. We shall present this argument in the chapter dealing with Tim Bayne.

Here are the basics of the condition: in otherwise normal individuals, experience in one sensory modality (sight, smell, hearing etc) reliably triggers the report of a sensation in another modality. In other words, sensory stimuli associated with one sense are accompanied, in appropriate circumstances, by qualitative 'feels' in another sense. It seems there are many different types of synaesthesia, that sensations in all modalities can elicit further sensations in all other modalities in different individuals, (thus: hearing/colour, touch/taste and so on) but some forms of synaesthesia are more common than others. Perhaps the most common version is 'coloured-grapheme' synaesthesia, in which letters are seen as coloured, but this would appear to be a case of two sensations in the same modality of vision. Another variant – colour/music, or colour/sound - is more clearly a case of triggered sensation in a different modality (a cross-modal variant). The condition is reported more commonly by women (on a ratio of 6:1), by people who are artistic, or left-handed, or musical (evidence suggests Liszt, Sibelius, and Messiaen had the condition). It can be brought on by brain injury, or epileptic or tumour-related seizures, but it may also, in other variants, be a condition which is inherited from birth. Indeed, some developmental theories have argued that everyone starts out as a synaesthete, in the first few weeks of life, and we learn to establish singular (and modally-appropriate) sensational responses to stimuli as pathways are laid down in the brain's structure.

It might also be noted that synaesthetes do not consider the condition to be a curse; in fact, quite the opposite. The condition confers upon the subject certain abilities which are welcomed by the subject, and some of those abilities, as we shall see, help diagnose the condition.

There are problems with generalising from the studies made of synaesthetes – we have noted already that there are different variants, however, within each variant there are idiosyncracies – different subjects have reported differences in vividness of experience, in the colours experienced with different words, or with musical notes, etc. There would appear to be differences in the location at which synaesthetic responses are generated. Many of the studies done on the condition collapse the different variants together, and there are obvious dangers in this. Some are classed as higher and lower synaesthetes, according to where they

report their concurrent experience as being (amongst other features), and some are classed differently to accommodate other features. It may be possible that no two subjects' sets of synaesthetic experiences/associations are the same. For example, within the class of colour/grapheme subjects, any individual letter may invoke quite different colour experiences – for one it may be that the letter A is accompanied by the experience of redness; for another, it might be yellow. In Sean Day's study (in Robertson and Sagiv, 2005, p.14) 43% of colour grapheme synaesthetes experience the letter A as red.

This of course, is only a sketch of the condition, and further detail will be introduced in the relevant section later. We shall see what contribution it has to make to our thinking about consciousness – I shall argue that it impacts on the suggestion that consciousness is unitary – a spin-off claim from the neo-dualist position, but one which puts a constraint upon accepting the higher order thought theories for which we shall argue.

What do these conditions have in common?

Neurologically, synaesthesia may turn out to have more in common with the deficits we will study than is immediately obvious, but for now I shall set it aside, and concentrate on the dissociations which preceded it in this chapter.

We should ask: what, if anything, is common to them? We can make an answer to this question on different levels; most simply, one common denominator is neurological. There has often been a trauma of some kind, and some parts of the brain usually responsible for processing information no longer function as before, and this has allowed, in the pathological cases, a study of the brain's functions. Some of the techniques which have the effect of creating a blindsight effect in neuro-typical subjects also interrupt the functioning of relevant parts of the brain – for example, trans-cranial magnetic resonance can be used to disable temporarily the primary visual cortex.

The neurological causes and the location of damaged neural tissue in these deficits will of course differ from condition to condition, but following Weiskrantz (1997, p.226) we can try to generalise:

- the lesion causing the deficit affects a 'major distribution point for a network capable of processing a variety of attributes and components' (usually the primary processing area of the brain for the function affected)

- a sub-part of this network can still be stimulated over a less direct route
- the normally important back-projection to the distribution region is now without effect because of the lesion
- this lack of feedback leads to an inability to a lack of awareness typical of such conditions, and (which may amount to the same thing)
- an output to what Weiskrantz calls the 'commentary stage' cannot be generated, by which we can understand as being an inability to report the features of the perception or discrimination. In consequence, the subject reports as 'blind' to the stimulus.

This might already suggest the beginnings of some neurological account of how consciousness arises, but there are other conclusions which can be drawn. From the dissociations we learn, amongst other things, that there are discrete areas of the brain responsible for different features of visual perception. The area for processing colour (V4-V8) for example is a different area from that processing motion (V5). The role of V1 – the primary visual cortex - is clearly important, since, when missing or damaged or interrupted in its function, consciousness is compromised.

The impairment by lesion or stroke of one of these areas, or sub-systems, while preserving others, provides a means by which the brain may be studied at a fine-grained level.

In passing, we should avoid thinking of this as a new kind of phrenology. In fact, Weiskrantz alerts us to the danger of thinking of these studies in those terms (Weiskrantz, 1997, p.234), since there are many inter-linkages or projections – connecting and feedback mechanisms which proliferate in the brain, and subjects differ - identifying discrete areas responsible for, or associated with the performance of different functions in all humans, while not impossible, is difficult. (Indeed, this is why subjects with highly specific damage to areas are so useful in brain research). We should also recognise that the plasticity of the brain allows for the development of new connections and pathways, to replace those lost, or interrupted by localised damage, so the functional architecture in one subject may not be replicated exactly in another.

There are further conclusions we can derive from the neurological study of dissociations. In many of the cases mentioned, there is a single dissociation between residual capacity and

acknowledged awareness, and these are useful insofar as they allow inferences about the hierarchy of control between different regions of the brain. This was the area of study, and conclusion, of Breitmeyer's 2015 work on the various psychophysical 'blinding' methods. He claims to have identified that a functional hierarchy of unconscious visual processing can be established. This in turn can begin to identify the neural correlates of unconscious visual processing, and to distinguish these from the neural correlates of conscious visual processing. Double dissociations, such the case of Dee Fletcher's visual agnosia, and its counterpart, optic ataxia, allow inferences to, or provide evidence of the independence of systems, or modules. In this case, the vision-for-action, and the vision-for-perception distinction.

Thus, much can be learnt from the study of the dissociations, especially since many of the psychological studies of them can now be supported by modern imaging techniques. But the scientific study of dissociations, and all of the results above, does not necessarily make the target philosophers against whom this thesis is aimed, sit up and take notice.

Some, such as Ian Phillips, remain committed to defending the view that all seeing is conscious, and will, as we shall see, reject the dissociations as cases of unconscious perception. He will contest them as being both unconscious and as cases in which genuine perception is involved. Others might define their terms in such a way that the dissociations cannot present a challenge – Chalmers and others define consciousness in terms of the possession of qualia, or awareness, or something that it is like to taste coffee, see the colour red, and so on, and as we shall see, this might mean that the dissociations create difficulties for a wider picture of mind (his 'structural coherence'), but they do not impact on what they consider consciousness to be.

However, if they are as we shall depict them, ie genuine cases of perception, or successful perceptive discriminations, in the absence of consciousness, the dissociations break apart what the neo-dualists would insist are linked by definition. To put it slightly differently, in these cases we seem to have situations in which successful discriminations are accomplished, but without the qualia of experience.

In the course of this study, we shall make the case for taking the dissociations more seriously. We shall argue that our target philosophers have a conceptual framework which allows the conditions to be recognised, but in ways that do not question the contention that seeing is necessarily conscious (e.g. O'Shaunessey, 2002, p.416). Others are inclined to question the empirical evidence for dissociation (Chalmers, Phillips). If, however, we adopt a

different conceptual framework, it may be that we can embrace the dissociations as unexceptional examples of mental states of which we are not conscious, and at the same time, pave the way for a consideration of a higher order thought theory. This would have the additional advantage, on this view, of being more open to the evidence of neuroscience, and would ultimately contribute to a reductive understanding of mind.

Chapter 2

On Phenomenal Concepts

Qualia, and 'what-it-is-like'.

The purpose of this section is to explore the phenomenal concepts at the centre of neo-cartesianism, and their role in an anti-physicalist theory. It will identify them and how they are defined; we shall see how they appeal to our intuitions; we shall show how they feature in a non-reductive account of mind. It will then go on to identify some of the shortcomings involved in the concepts, arguing that, as such, they shunt the philosophical consideration of mind into a cul-de-sac, and provide no basis for any account of the evolutionary development of consciousness in humans or other animals. The chapter will largely be a prima facie case against the understanding of phenomenal concepts as adopted by the neo cartesians. As used by them, these concepts disallow a treatment of the dissociations as unconscious perception, so a critical examination of those concepts is preparatory to any new theory with unconscious perception at its centre.

What are qualia; how are they defined and how is the concept of them acquired?

To be discussed is the concept of a quale and specific examples of qualia. In the conception presented by the dualists, these concepts are the product of having an experience of a certain sort, and which refer to the phenomenal character of such experiences. As we shall see, such concepts refer to the specific phenomenal properties which are created by an experience. They are invoked expressly to allow a distinctive way of thinking about subjective states, and are supposedly available only to those who have had the relevant experiences.

These will be our primary targets, though they belong to a wider family of terms which point at non-material phenomenal properties distinct from those that can be accessed or explained from a scientific perspective. Chalmers makes this clear: '[A number of alternative terms] include: "experience", "qualia", "phenomenology", "phenomenal", "subjective experience" and "what-it-is-like"'(Chalmers, 1996, p.6).It is important to note that Chalmers identifies them with consciousness from the start: '"To be conscious" […] is roughly synonymous with "to have qualia" (Chalmers, 1996, p.6).

Not all philosophers that use the terms, agree on what those terms should be confined to: for example, Dretske (1995) and Lycan (2001) think of qualia as being properties of objects external to us, while Chalmers thinks of them as characteristics of our experience. In this sense we can describe them as mental phenomenal properties. This account will restrict the discussion to that version which is central to the position taken by the non-reductive monists we have as our target – ie to a qualitative property or 'feel', or feature of experience. On this reading, a paradigmatic quale would be that felt quality of my experience which, in the presence of a certain scent, is what we may call *Patchouli*[7]. Similarly, *Redness*, *the taste of Coffee*, how *Middle C on a Piano sounds*, and the *feel of Rabbit Fur* would all be qualia.

It is also clear that Chalmers is committed to a belief in the felt quality of thought (at least of *some* thoughts) – that for example, there is something it is like to think of the Eiffel Tower, or to believe that it is tall (Chalmers, 1996, p.10). In other words, there is a qualitative feel to propositional attitudes as well. We shall however, confine our discussion to the qualitative character of phenomenal mental states.

However, before we look at these phenomenal concepts more closely, let us consider for a moment, what concepts are, and what role they play in the scheme of things. We can describe them as 'mental representations' or '[the] units of thought, the constituents of beliefs and theories' (Carey, 2009, p.5). They are mental categorisations of the world and its contents, and they figure in the internal models we make in the form of thoughts and theories. Taking this last point further, it will also be helpful to think of them as the components of a Kuhnian paradigm (Kuhn, 1996) for which the theory provides the architecture. In his account of scientific paradigms and progress, Kuhn makes it clear that it is common for concepts to be theory-specific. Thus the notions of superlunary and sublunary forces and spheres were conceptual features peculiar to Aristotelean cosmology; the concept of corpuscles appertains to a Cartesian theory of light. Such concepts are the product of observation, or inference from observation. It is clear from this historical perspective that sometimes the wrong inference can be drawn from observations.

At a different level, we should recognise that the concepts with which we apprehend the world are not fixed, but develop over time. Our concepts of number and agency become very much more sophisticated over time. Carey adds: 'Many representational capacities arise

[7] Italicised to indicate this being understood as a property of my experience, not of the liquid in the bottle itself.

from maturational processes. An example is stereoscopic representations of depth, which emerge in humans quite suddenly around six months of age' (Carey, 2009, p.12). Our basic concepts are ones we work on, as we develop. They are rarely static. We refine our concepts of justice, and fairness as we encounter complex problem cases. We add to our conceptual resources as we master a natural language, and enrich ourselves with an expanded vocabulary. Equally, we may jettison conceptual commitments and representational models that fail us. On that point, and returning to the Kuhnian account of scientific progress, one of the casualties of a transfer from one paradigm to another are the concepts specific to the rejected paradigm. One of the reasons for the transfer is a growing dissatisfaction, for one reason or another, with the existing conceptual scheme (see Kuhn, 1996, pp52 et seq). This might be occasioned by new scientific discoveries, or experimentation. Sometimes, the process of conceptual change can be prompted by a growing awareness of inconsistency, or internal contradiction.

Against this background, let us consider the theory with which we are currently concerned: neo-cartesianism. On this account, phenomenal concepts are concepts we can form as the result of an experience, and which refer to the felt character of that experience. Thus, when Frank Jackson's Mary (Jackson, 1982) exits the room and encounters the tomato for the first time, she has a visual experience as of seeing something red, and, we are told, forms the phenomenal concept of *Red*. Such a concept (we again are told) refers to non-material properties, distinct from any properties that can be accessed from a scientific perspective. Further, because Mary instantly knows what seeing red is like from her first exposure to it, the acquisition of the concept appears to be immediate. Mary comes to acquire it by reference to her inner experience, and nothing else would appear to be necessary. We can now make further progress in analysing them further, and in identifying what it is that the neo-cartesians appeal to, in positing them.

'What-it-is-like.'

At some point, early in any non-reductive monist's discussion of phenomenal concepts such as qualia, the term: 'what-it-is-like' features. The phrase first gained currency in the seminal work by Nagel: 'What is it like to be a bat?' (Nagel, 1974). Since its appearance in this paper, it has been advanced as a way of further characterising experience, or phenomenal consciousness, or specifically explaining the notion of 'qualia'.

Again, in a recent paper Chalmers and Bayne announce their own commitment to the idea: '[w]hen a state is phenomenally conscious, being in that state involves some sort of subjective experience. There is something it is like for me to see the red book' (Chalmers, 2010, p.503). Chalmers also makes the tie between qualia and the what-it-is-like formula explicit in Chalmers (1996): 'In general, a phenomenal feature of mind is characterised by what it is like for a subject to have that feature' (Chalmers, 1996, p.12).

In this way, (neo-dualists would say) there is something it is like to be me, but nothing it is like to be a brick, or a thermostat, or a zombie, since such items lack, or are incapable of qualitative experience. But as noted above, the idea can go beyond qualitative experience to embrace thoughts, beliefs, desires, such that the broader aspect of consciousness can be construed in terms of what-it-is-like for me to have such inner goings-on. There is nothing it is like to be the automatic doors that open on my approach since these cannot be said to have the belief that I am approaching and need access.

In his 1974 article, Nagel does not spend a lot of time on the specific meaning of the phrase, and only gives a short footnote on the question, saying: '[…] the analogical form of the English expression "what it is *like*" is misleading. It does not mean "what [in our experience] it *resembles*," but rather "how it is for the subject himself". This makes it reasonably clear that the term is not being used in any way other than pointing to the specific experience. Nagel wants to know: 'what it is like for a *bat* to be a bat.' And: 'An organism has consciousness if there is something it is like to *be* that organism – something it is like *for* the organism' (Nagel, 1974). In this sense, the term is not meant to be comparative. What it is like to smell a rose is not to be assimilated to, or differentiated from what it is like to smell a hyacinth.

Nagel does not offer an account of what it actually *is* like, to have an experience, in other words; what it would be like for someone to have the experience. Little is offered by any of those who would use the phrase. Chalmers has an unconvincing attempt – he suggests that what is *common* to visual sensations, perceptions in different modalities, bodily sensations, mental images conjured up internally, and the felt quality of emotions is that they are all states united by the idea that there is something it is like to be in them (Chalmers, 2010, p.5). This of course puts us no further forward; we are no closer to identifying the nature of 'what-it-is-like' by suggesting that several things have that character. We do not make much progress in closing in on the concept of loyalty by suggesting that x, y and z all

have that quality, and it is the one thing they have in common. Elsewhere, Chalmers has another attempt: while conceding, in a way rather instructively, that 'It is often hard to pin down just what the qualitative feel of an occurrent thought is [...]', he suggests that 'When I think of a lion, there seems to be a whiff of leonine quality to my phenomenology' (Chalmers, 1996, p.10). It is hard to say what progress this makes, if any.

It is the view of this work that the term: 'what-it-is-like' does not help to define qualia. At best, it is an attempt to enlarge upon the term by synonymy; at worse, it is circular. Others reach stronger conclusions. Lycan describes the phrase as 'worse than useless: it is positively pernicious and harmful, because nothing whatever is clarified or explained by reference to it' (Lycan, 1996, p.77). He is equally scathing about 'qualia' insofar as these terms are used in a deplorably general way 'as an umbrella term for whatever one finds puzzling about consciousness, subjectivity etc' (ibid).

How do our target philosophers arrive at the concepts?

To what do the neo-cartesians appeal, in their account of the acquisition of these phenomenal concepts? As suggested briefly above, the answer to this question is firstly that they are arrived at by a process of introspection, and secondly, derived from the experiences they relate to.

On the first point, Loar comments (in: '*Transparent Experience and the Availability of Qualia*' in (ed) Smith 2003, p.77) that 'all qualiaphiles think of qualia as introspectable', that they are discoverable by that process, and that nothing else is needed or is indeed possible. It is contended that the felt character of a colour experience is revealed to us when we have an experience of it. Chalmers subscribes to this in the opening sections of *The Conscious Mind*. He prefers to use the term: 'experience', but it is evident that he thinks that we know about qualia by dint of personal experience.

It might be argued that Russell himself bought into the characterisation: 'So far as it concerns knowledge of the colour itself, as opposed to knowledge of truths about it, I know the colour perfectly and completely when I see it, and no further knowledge of it is even theoretically possible. Thus the sense data [...] are things with which I have acquaintance, things immediately known to me just as they are' (Russell, 1912, p.25). The philosopher CI Lewis, who first coined the term in 1929, also took a similar view:

‘The quale is directly intuited, given, and is not the subject of any possible error because it is purely subjective’ (‘Mind and the World Order’ 1929).

As for the acquisition of the concepts, Chalmers in *The Character of Consciousness,* seems to adopt a Lockean account[8] of concept formation – ie that experience alone is sufficient to establish a concept: ‘The most important phenomenal concepts are those we acquire directly from having experiences with that sort of phenomenal character. For example, when one first learns what it is like to experience an orgasm, one acquires a phenomenal concept of the experience of an orgasm’ (Chalmers, 2010, p.251). And later: ‘I look at a red apple and visually experience its colour. This experience instantiates a phenomenal quality R which we might call phenomenal redness. It is natural to say I am having a red experience, even though experiences are not red in the same sense that apples are red’ (Chalmers, 2010, p.254).

In general philosophers who subscribe to the existence of qualia regard that existence as ‘just obvious’, and no further attention is given to it, until they encounter and want to deal with philosophers who are less convinced. But it is clear that our target philosophers consider that the mechanism for the acquisition of phenomenal concepts is personal experience, and they are knowable by introspection. Some neo-cartesians would also consider that this route is tied together with a sense of infallibilism. There is a long-standing tradition of philosophy that claims: I know all and only the contents of my mind infallibly and completely. (See Heil 1988 for a review of this position.)

Qualia as used in the attack on Physicalism

As we noted earlier, phenomenal concepts on the neo-cartesian account are central to critiques of physicalism. Qualia, once conjured and characterised, feature in a variety of thought experiments designed to torpedo forms of physicalism, portrayed as theories incapable of accommodating them – arguments concerning inverted qualia, Chalmers’ absent-altogether qualia (zombies), Block’s Chinese Mind; Jackson’s Mary and the Black and White Room, are all conjured as thought experiments to defeat functionalism, or any other form of physicalism. These arguments are well-known, and limitations on space forbid their

[8] cf: Locke: ‘An Essay concerning Human Understanding’ Book 2 chp 1.

elaboration here, but involved in each is some version of qualia, or 'what-it-is-like'. They are presented in Balog (2009) pp 8-9.

The conclusions the arguments draw are: Colour qualia are to be thought of as fundamental, subjective, non-physical emergent properties, not to be identified as functional or physical properties. For there to be qualia, God had to do extra work to ensure that they come with the physics: '[...] once god had fixed all the A-facts (physics), in order to fix the B-facts (qualia inter alia), he had more work to do. The B-facts are something over-and-above the A-facts' (Chalmers, 1996, p.41). As Frank Jackson more succinctly asserts: 'Nothing you could tell of a physical sort captures the smell of a rose' (Jackson, 1982).

Further, qualia are fundamentally irreducible features – they emerge from the physics of the brain, in a process which is hard for the neo-cartesians to spell out, partly because of the problems of accounting for emergence, but also because to do so means crossing the subjective/objective divide – the so-called explanatory gap. Again, there is a great deal of literature here, the exploration of which is forbidden by our limitations on space. This work, however, seeks to grapple with these questions by looking at the problem at a more fundamental level, that is, by asking hard questions of these phenomenal concepts themselves, and it is to this that we now turn.

How do dualists account for the acquisition and retention of phenomenal concepts?

In the *Blue Book*, Wittgenstein says one of the great sources of philosophical bewilderment is that a substantive makes us look for a thing that corresponds to it (1958 p1). This might apply to a number of candidate terms, but in each case, we might make the point that errors can ensue when one lets language determine one's ontology. Because there is a word in the language does not mean there is a something which that word singles out. We might be especially careful when a term of art is in question, which is what 'qualia' is.

One thing we might want to be sure is that there is indeed something in the world which the term singles out. Neo-cartesians, or qualia enthusiasts, should show that there are such things as qualia as defined, and the mere invocation of them is not sufficient to establish the term, let alone, with them, demolish the case for physicalism. Some philosophers argue that qualia, in the sense that they have been conjured by neo-cartesians, do not exist at all.

Dan Dennett tells a story of a concept that has got off to a bad start namely 'fatigues', in his book: *Brainstorms* (1978). Dennett cautions against the assumption that a term identifies something substantial, and that very often, conceptually speaking, people can get off on the wrong foot. In the case explored, we imagine an Amazonian tribe not previously exposed to advanced biology, which has come to think that tiredness is the possession of '*fatigues*': the more tired, the more *fatigues* someone has. The word is embedded in the language of the tribe. The tribe has not been able to isolate a *fatigue* (unsurprisingly) and ask the visiting explorers if, with the advances of modern science they have discovered what nature *fatigues* have, what connection to they have to bodies, whether they have a location in time and space. Our response, says Dennett, is to point out that the tribe has got off to a bad start with forming the concept – *fatigues* are not good theoretical entities as a foundation in the tribe's biology. The same may be true of qualia. In fact, Dennett does argue elsewhere (Dennett 1994) that qualia should be 'quined' out of existence.

However, we should allow the neo-cartesians to make their case for their understanding of the concepts. Let us remind ourselves of the main features of the neo-dualist argument which features phenomenal concepts of this kind.

Firstly, the possession of qualia by a being, ie that being's consciousness, is defined as being *something it is like* to be that being (Chalmers, 1996, p.4). We need to ask if anything is gained by such a definition.

Secondly, on their account, we acquire phenomenal concepts simply by experience. Mary comes to know what it is like to see red by seeing a tomato for the first time. She acquires the concept of *Red* when she exits the black and white room. The process by which Mary comes to know the concept of what it is like to see red as the result of a process of introspection – the internal contemplation of her own inner experience. According to Chalmers, the redness in question is a property of her experience, not of the tomato. The claim is that we know about this more directly than we know about anything else (Chalmers, 1996, p.xii-xiii). In other words, what we are forming a concept of, is the character of our experience. We shall ask if a neo-cartesian account of concept acquisition and retention is one that works.

Thirdly, the experience of phenomena is identified with consciousness. What it is for me to be conscious is for me to have the experiences which take this phenomenal character. ("'To be conscious' is roughly synonymous with: 'to have qualia'" (Chalmers, 1996, p.6).

We will ask what this implies for the wider theory of which the phenomenal concepts are a part.

Qualia as 'what-it-is-like'.

We begin with the claim that phenomenal concepts can be defined in terms of 'what-it-is-like'. To recall: Chalmers writes: 'a mental state is conscious if it has a qualitative feel – these are also known as phenomenal qualities, or qualia for short. To put it another way, a being is conscious if there is something it is like to be in that mental state' (Chalmers, 1996, p.4). We began this chapter with a few critical remarks about this phrase. It is time to press further. What are we saying when we define qualia in terms of 'what it is like'? Are we saying anything useful; are we saying anything at all which can be of use in defining terms? On the first question – is the phrase at all ampliative? Several philosophers have had occasion to pronounce that no – there is nothing that this phrase adds to the discussion. Collectively they argue that 'what-it-is-like' is as ineffable as 'qualia'.

Nicholas Humphrey puts the problem well: 'The difficulty is one which has haunted our discussion of sensation from early on – the difficulty of pinning down that elusive extra ingredient, the X factor. We have several times had occasion to stress that when S (has a sensation), S –even as the subject of it – does not entirely know "what it's like" is like, or at any rate he certainly cannot say' (Humphrey, 2006, p.80).

Peter Hacker takes the same view, approaching the question from an ordinary language point of view (Bennett and Hacker, 2003, p.277 et seq). He asks: what could count as an answer to the question: 'What IS it like to' (have a sensation)?. Standard answers would be: 'pleasurable', 'agonising', 'disturbing', 'delicious', 'overpowering', 'exquisite'. Not only are these the standard answers, but they are the only answers. And sometimes the answer is: 'nothing at all' (this latter would be the case when seeing a lamppost, or a table, for example.) These characterisations however, do not yield a 'specific feel' to any qualitative experience. Hacker writes: what individuates the experiences are the things they are experiences of. As he puts it: 'The experiences differ in so far as their objects differ', and later: 'pleasures are individuated by their objects' (Bennett and Hacker, 2003, pp.274-5).

Hacker concedes that the experience of hearing Beethoven's 9th Symphony is different from smelling a rose, and that both are different from seeing a sunset, but to make the further claim that there is some unique distinctive feel to these is an altogether more

questionable claim. Just as Chalmers suggested there was a leonine 'whiff' to seeing a lion, he would have to be committed to saying there is some Germanic-Romantic 'lilt' to the sensation of hearing the 9th, or some Hesperrhodos 'smack' to the experience of smelling a rose, or a Columbian 'tang' to the taste of coffee, at which point it becomes obvious that we are no further forward in characterising qualia.

Hacker also attacks an associated position taken by Chalmers, when he, Chalmers, asks (in sketching out the Hard Problem): 'Why do conscious experiences have their specific character' (..) 'why is seeing red like *this,* rather than like *that*'? Hacker makes the point that these terms: 'like *this'* and 'like *that*' are used by Chalmers as ways of referring to the specific qualities experiences are alleged to have, attempting to eff the ineffable, to pin down the distinctive with indexical expressions. Invoking Wittgenstein, Hacker comments that this is confused.

> The content of experience. One would like to say 'I see red *thus*'. 'I hear the note you strike *thus*', 'I feel sorrow *thus*' (…etc) One would like to people a world analogous to the physical one, with these *this*es and *thus*es. But this only makes sense where there is a picture of *what is experienced,* to which one can point as one makes these statements. (Remarks on the Philosophy of Psychology 1 §896)

There is, as Hacker would contend, nothing to which the experiencer can point in these cases, and we are left with the suspicion that one vague and poorly established concept is being used to support another; perhaps both stand in need of independent justification; perhaps, on examination, neither is grounded convincingly.

A suspicion creeps into the picture. We might ask whether, in talking about 'what-it-is-like' we are saying anything at all. It may be that invoking the description really adds nothing at all to further the definition. Ned Block was clear about this in his paper making the distinction between p-consciousness and a-consciousness (Block 1995). In attempting a definition of phenomenal consciousness: 'Phenomenal consciousness is experience; what makes a state phenomenally conscious is that there is something it is like to be in that state. Let me acknowledge from the outset that I cannot define p-consciousness in any remotely non-circular way' (repr in (ed) Chalmers, 2002, p.206).

Let us now turn to the question whether phenomenal concepts can be fixed or acquired in the way described by the neo-cartesians. We need to answer two questions: *Do*

we form concepts with phenomena as the raw material? And: *can* we form concepts with phenomena as the raw material?

Can qualia be understood as intrinsic features of experience?

As we have noted, neo-cartesians think of qualia as appertaining to the qualitative character of the experience of, say, seeing the red tomato for the first time. The phenomenal concepts derived are ones formed as the result of an introspective examination of one's experiences. They are the result of a contemplation of one's experience. But is this a true picture of the process? We do have concepts of *redness* and *birdsong*, but are they the product of a contemplation of experience itself?

According to neo-cartesians, qualia are to be understood as the qualitative character of a feature of experience, and not of objects themselves. Their claim is that experience has a specific qualitative character, and this is what is discoverable by a process of internal examination – the 'feel' associated with seeing green, or the taste of coffee, etc. Yet, we can ask – what actually is it that is attended to in the moment that Mary sees the tomato for the first time?

Tye (1999) claims that there are several problems associated with the idea of phenomenal consciousness, some of which feature elsewhere in this work, in some form or other, but relevant here is the question of transparency. The problem has antecedence in Moore (1922) and Harman (1990). Tye invites us to focus on the character of our *experience* when considering (say) a blue patch and to try to become aware of the experience itself, rather than the patch itself. He concludes: 'The task seems impossible; one's awareness seems always to slip through the experience to the blueness [..] instantiated in the external object' (Tye, 1999, p.30).

His counter-claim is that we do not find any intrinsic feature of the experience, from which we form our concepts. We 'slip through the experience, and end up concentrating on what is outside again, on external features or properties.' What this is identifying is the problem of transparency. Tye suggests it may be the sort of thing 'GE Moore had in mind when he remarked that the sensation of blue is diaphanous'[9] (Dennett makes similar points in

[9] GE Moore: 'The Refutation of Idealism' quoted in: M Tye: 'Visual Qualia and Visual Content revisited' 1993

Dennett, 1991). Tye again: 'Introspection does not seem to reveal any further distinctive features of the experience over and above *what* one experiences [..]'.

The simple point is that the concept of qualia as the neo-cartesians would have it – as applying to the intrinsic features of experience, is contestable. There may be no such thing as qualia as imagined in this way. It may be that we do not form the relevant concepts as the result of a contemplation of our phenomena. If, as Tye writes: 'I cannot make myself aware of any features of my experience over and above, or apart from, what I am experiencing' (Tye, 1995, p.31) then it is hard to see how concepts can be derived from the character of the experience itself. But let it pass – we should now turn to a different question: *can* phenomenal concepts be derived from an introspective contemplation of phenomena?

Can introspection of experience give us a secure grip on phenomenal concepts?

I first want to raise an epistemological point about the reliability of introspection. In recent years, a number of philosophers have put pressure on whether introspection can provide us with the concepts as described. They question what neo-cartesians claim that introspection can deliver, and have begun to ask if we are especially authoritative about our inner mental states. In this category we find: Stalnaker (2008), Schwitzgebel (2011), Noe (2002), Kornblith (1998). We find that most people are poor introspectors of their own ongoing experience. We make errors in judgement about our inner states, when we are not fully attentive, in conditions dependent on our mood and for other reasons. This may be particularly true of our emotional states. The neo-cartesian claim that we are secure in our internal knowledge, and that we cannot be wrong about our phenomenology (infallibilism) no longer goes unchallenged (Dennett 1991, and see also Noe 2002, and Heil 1988).

Schwitzgebel (2011) mounts a particularly formidable attack on neo-cartesian claims for what introspection can deliver. In: 'The Unreliability of Naïve Introspection' (2011, chp7) he advances the view that 'Most people are poor introspectors of their own on-going conscious experience' (Schwitzgebel, 2011, p.118). He offers a number of everyday situations in which we claim to have a secure grasp of our inner phenomenology, only to find ourselves mistaken. Despite our confidence claim to know ourselves better than others, we find that any individual, and for a variety of reasons - inattention, self-deception, as well as pathological delusion, can be wrong or untrustworthy about his inner states. Introspection

may not be a reliable epistemological basis, and grounding for what-it-is-like, and phenomenal concepts. Yet, this is the starting point for the neo-cartesian. In the Introduction to *The Conscious Mind*, Chalmers claims: 'We know about consciousness more directly than we know about anything else [..]' (Chalmers, 1996, p.xiii).

Our discussion remains focussed on whether the phenomenal concepts at the heart if the neo-cartesian argument are securely reached. Here we offer the view that: the neo-cartesian grounds his argument in introspection; introspection does not deliver a secure epistemology and an insecure epistemology can be at best a poor ground for the kind of phenomenal concepts at issue. There is, however, an even more fundamental objection concerning the genesis of these phenomenal concepts, to which we now turn.

Phenomenal concepts and the private language argument.

In this section we will consider whether Chalmers' account of the genesis of phenomenal concepts is immune to an argument inspired by Wittgenstein's private language argument. Let us first re-state and examine his position. The argument is laid out fully in his later work (Chalmers 2010), but some features make an appearance in Chalmers 1996.

Phenomenal concepts are advanced as concepts derived from experience, with which we frame our beliefs about phenomena. He writes: 'When one believes that one is having a red experience, one deploys a phenomenal concept of a red experience. The most important phenomenal concepts are those that we acquire from having experiences with that sort of phenomenal character' (Chalmers, 2010, p.251). We may suppose that phenomenal redness, where that term refers to the concept involved and generated by red-seeing, is one such phenomenal concept, as is the general term: 'qualia' insofar as it operates as a concept.

In Chalmers' account, 'the reference of expressions such as 'phenomenal redness' is fixed via a relation to certain paradigmatic red things in the external world that are ostended in learning the public-language term 'red'. (..) the phenomenal concept involved here is relational in that it has its reference fixed by a relation to external objects' (Chalmers, 2010, p.255). As Balog puts it: according to this view, 'we can apply phenomenal concepts directly to the phenomenal states we are currently aware of, without the mediation of any functional, representational, behavioural, or physical definition, or physical criteria' (Balog, 2009, p 11).

Several consequences emerge: recall that one of the functionalism-defeating arguments used by Chalmers is the possibility that two individuals could undetectably have an inverted spectrum in their experience – imagine the situation in which we both encounter the same cricket ball, we both describe it as red, but in considering it, I have the phenomenal property of my experience is of red, yet my inverted twin has an experience with the phenomenal property of green. How can this be accommodated in Chalmers' concept fixing story?

Chalmers puts it thus: The phenomenal concept is fixed by relation to the external object, but 'The property that is referred to need not be relational, however […] there are counterfactual worlds in which red experiences are never caused by red things' (Chalmers, 2010, p.255). The property invoked in me and my inverted twin might be different, as it is not fixed by the object in the way our concept of red is.

Our question is: can the phenomenal concept of red be fixed independently and irrespective of the property invoked? What problems are associated with this picture? Chalmers exposes the flank himself. If a concept is to be fixed independently of the qualitative property invoked, a private language may be required. The qualitative property becomes a kind of 'beetle in a box'.

He compares the example of Mary and her inverted twin: 'When Mary has a red experience for the first time, she learns something *different* from what is learned by her inverted twin, who has green experiences where Mary has red' (Chalmers, 1996, p.206). She has a concept of this quality – call it "*R"* that is distinct from the corresponding concept – call it "*G*" in the inverted counterpart. He goes on: 'My qualitative concept "*R*" plays little direct role in communicative practices. In that way, he acknowledges, it resembles Wittgenstein's 'beetle in a box' (Chalmers, 1996, p.206).

He concedes: 'Of course, the view I have set out here is just the sort of view that Wittgenstein directed his "private language" argument against. The nature of the private language argument is contested, so in response I can say only that I have seen no reconstruction of it that provides a strong case against the view I have laid out' (Chalmers, 1996, fn 13 p.206, and 2010 p.273).

One powerful question which Chalmers must deal with is – how can such concepts be fixed, if they are so irrelevant, and float so free of the usual processes of sensation-naming. In Chalmers 2010, p.272 he discusses the process by which phenomenal concepts are fixed.

He criticises those constructions of the private language argument that have, at their centre, the belief that concept possession requires 'a strong sort of "repeatability"', and 'an exceptionally strong sort at that, requiring the recognisability of correct repeated application' (Chalmers, 2010, p.273).

He argues that this kind of reading of the private language argument sets too high a bar for concept formation – he gives an account of concept formation thus:

- The lifetime of a direct phenomenal concept is limited to the lifetime of the experience that constitutes it. (Though it may persist of there is a sufficiently vivid memory of it).
- The concepts in question are predicable of any number of entities during their limited lifetimes, and these predications can be true or false.
- This sort of predicability, with assessibility for truth or falsehood, seems sufficient for concepthood. Clarifying: 'Merely the "hypothetical repeatability" involved in present predicability of the concept to actual and hypothetical cases, is required for concept possession' (Chalmers, 2010, pp.272-273).

On the face of it, this points to a process of concept-fixing which sets the bar way too low. To begin with, it seems insupportably rapid. Right at the beginning of the chapter in Chalmers 2010, he writes: 'When one first learns what is it like to experience an orgasm, one acquires a phenomenal concept of the experience of an orgasm'. He seems to be suggesting a single experience is sufficient to fix a concept. He rejects the idea that concepthood requires persistence (Chalmers, 2010, p.272). All that is sufficient is for the qualitative properties to be present to the mind. Concept fixing is a product of the special non-physical nature of qualia themselves (Balog 2009, p12).

A further point is this: he does not spell out why he takes the view that the private language argument has no bearing on his position, so let us give a reading of it that creates difficulty for him.

Chalmers asks: what justifies phenomenal judgements (in which phenomenal concepts are deployed)? (Chalmers, 1996, p.196). He gives the response: 'I think the answer to this is clear: it is *having* the experiences that justifies the beliefs'. He concedes that having the experience may not be the only factor relevant, but he writes approvingly of a 'first person

warrant' for our beliefs about experiences (Chalmers, 1996, p.197). This would line up with other commitments the neo-cartesian would make.

However, we can ask whether the process of fixing a phenomenal concept, and justifying phenomenal judgements using them can be as 'first person' as he claims.

The private language argument in Wittgenstein's 1953 *Philosophical Investigations* (paragraphs 258 et seq) gives an account of how the meaning of terms or concepts becomes established, and the role of experience in this. In brief, Wittgenstein argues that a private language is incoherent. What is a private language in this regard? It is a language, the subject matter of which is exclusively the subjective experience of the speaker. It is private in the radical sense that it cannot, even in principle, be shared by another person. Wittgenstein was opposing the position that Descartes, Locke and other philosophers took in this respect.

The neo-cartesian understanding of qualia is that they are essentially private and, as Dennett puts it, according to tradition, there are no ways we can say to each other, 'no matter how eloquent one is, and no matter how co-operative one's audience is, exactly what way one is currently seeing, tasting etc. [..] Qualia are ineffable – in fact the paradigm cases of ineffable items' (D Dennett: 'Quining Qualia'). In short: how can the concept of qualia be grounded given that the neo-cartesian must start from his or her own case, trying to attach a name to an intrinsically unshareable feature of experience, with no way of communicating the qualitative nature of that experience? The kinds of item that Chalmers is advancing here, concepts of which he is trying to fix, would seem to be paradigmatic examples of the unshareably private experiences against which Wittgenstein pitched his argument. The intrinsically unshareable character of experience as described by Chalmers is essential if it is to be possible for Mary to have the inner experience of Red when looking at a tomato, and for her inverted twin undetectably to have the experience of Green when doing the same. Chalmers posits this, and also requires that there be a process of fixing the phenomenal concepts to the phenomenal character of the experience.

The central claims of Wittgenstein's argument are that merely associating a name with an inner sensation does not suffice to endow the sign with a use, that there is no private equivalent of public ostensive definition, and that there is no such thing as a rule of use which is in principle incommunicable to anyone else. Something of this argument is spelt out below.

Associating a name with the inner sensation is only going to work if the 'process brings it about that I remember the connection *right* in future. But in the present case I have

no criterion of correctness' (Wittgenstein, 1953, para 258). Impressing a connection (between sign and sensation) upon myself is insufficient to establish it for future reference. It might *seem* right, but it is only a semblance of right. We cannot do what we might do with external objects – ostensively define it by pointing to it, since it is internal, and has no public aspect. Might one use one's memory to confirm the association? No – for we cannot be certain that this process produces a memory which is actually correct, and known to be correct (Wittgenstein, 1953, para 265). Justification in using the sign consists in appealing to something independent. The rule for the use of the concept is not established by association or appeal to memory.

I cannot give myself the meaning of a word in this way, in just the way that I cannot give myself money. I can transfer the money from one pocket to another, and write myself a receipt, but none of this ceremony has the practical consequence of a gift (Wittgenstein 1953 para 268). In just the same way, I cannot arrive at private definitions of words, insofar as I do not learn a rule for use in this way.

Over subsequent paragraphs (272- 278), Wittgenstein addresses exactly the kind of phenomenal concept that Chalmers has in mind – one derived from the intrinsically unshareable private experience of blueness. He adds the thought - - when one considers the blueness of the sky, 'the idea never crosses your mind that this impression of colour belongs only to you'. A private language built upon such notions is a fiction since genuine acts of naming, or concept forming take place in complex social settings, governed by conventions, which fix the subsequent use of that term. A private ceremony cannot do this, and an intrinsically unshareable experience, such as the one posited by Chalmers, would act like the beetle in the box (Wittgenstein, 1953, para 293), in not being able to feature in any kind of language game, and thus could not fix a concept.

Even supposing our putative private concept former could fix a concept such as this, they would be in the position of the customer who walked into a confectioners, and asked the shopkeeper: 'I want a bag of sweets that taste just like this' (pointing to their mouth).

On this view of this thesis then, Chalmers' account struggles to meet the challenge of Wittgenstein's private language argument. Of course, if it is indeed impossible to arrive at a private language, then any further comments about Chalmers' account is strictly unnecessary, but let us nevertheless add other criticism of his suggestions for how phenomenal concepts

are fixed. Recall that he argued that they do not require persistence, and can be fixed by a single experience alone.

This is not easy to maintain. Concept formation and fixing, as any parent knows, is grounded in a far longer process of repetition and correction, extension and boundary-limitation. Sometimes, the process can be accelerated, - once 'red' has been grasped, 'yellow' might be grounded more quickly - but I submit the claim that no phenomenal concept can be established on the basis of one experience alone, no matter how vivid that experience (which is what seems to be claimed by Chalmers). No phenomenal concept may be established without the correction and boundary-limitation involved in a community's policing of the use of the concept, that requires repeated testing to confirm the truth and falsity of its application. As we saw above, it is not simply repeatability (buying multiple copies of the morning newspaper) that fixes the concept; the morning newspaper must be checked against the BBC news, the internet, and other sources. And no phenomenal concept can be established without some prior stage-setting in which it can be determined that a specific definiendum is what we are directing our definiens at. Fixing 'Red' as a concept requires the determination that it is the colour of an item at which I am directing my attention. Further, if there is a temptation to think there is a 'wow-factor' when Mary leaves her study in the black and white room and sees the tomato for the first time, it can be claimed that Mary's task in acquiring the concept of 'Red' would be made far easier by the fact that she already has the concept of 'coloured item' from her hours of study in that room.

At this point we acknowledge that there are other accounts of phenomenal concepts which do not need to invoke phenomenal consciousness itself. This line of enquiry has become known as the Phenomenal Concept Strategy. Variations on this are offered by Loar, Tye and Carruthers, but exploration of these would take this thesis too far from its original focus, but we can offer the beginnings of solution, borrowing on a point made by Chalmers himself. He offers the view that the phenomenal is in a relation of structural coherence with the psychological (1996). Our psychological concepts characterise mental states functionally, and phenomenal concepts characterise the phenomenal aspects of mental states. It seems at least plausible that our acquisition of phenomenal concepts is on the model of our acquisition of psychological concepts. This lines up with another Wittgensteinian philosopher's position – that of P Strawson (1959) who wrote: 'One can ascribe states of consciousness to oneself only if one can ascribe states of consciousness to others' (p100). We can easily imagine that concepts are acquired and fixed by the same token, such that it is on the basis of having

attributed states of consciousness to others, having learnt and applied the relevant concepts, that we can attribute states to ourselves, and thereby securing our own concepts. This would be a process subject to public referencing, and meet the Wittgensteinian requirement.

Other philosophers have taken a different approach to this question (Papineau 2011, Prinz and Pris 2013). Papineau claims that one way of circumventing the problems posed for phenomenal concepts by the private language argument is to argue that we have a 'shelf-supply' of potential concepts for types of experience, [and that when] we first have a new kind of experience, we take a potential experience concept from the appropriate shelf, and lock it onto the type of experience at hand' (Papineau, 2011, p.9). For example, when I am confronted with a horse, I might take a potential *animal species* concept from the shelf and use it to form a concept that is locked onto the species: horse. He continues: 'My ability to refer to horses does not involve rules of any kind'. This would appear to be a form of nativism as far as the origin of concepts is concerned. That in itself raises many questions.

Prinz and Pris counter this move, claiming that there is no such 'shelf-supply' of phenomenal categories for potential phenomenal concepts. In any event, the application of a rule would still be a necessary part of the process of fixing the concept. Mary would need some sort of rule correctly to apply the shelf-supply concept, if one such did exist, to the new experience. Prinz and Pris claim that 'the role of the rule is played by the "ready-made concepts" (or "shelf-supply" of categories for potential concepts) (Prinz and Pris, 2013, p.13). So that, on this view, Papineau's ability to refer to horses without involving any rules, is false.

So, to conclude this section – can concepts of *blueness*, of *qualia*, which have built into them the kind of privacy which allows the inverted spectrum condition, be fixed? This thesis takes the view that there is much that is problematic about the neo-cartesian strategy. It is not easy to say what exactly we are trying to define, given the ineffability of qualia, and our suggestion that nothing is added to the target by adding the locution 'what-it-is-like'. On this view, the private language argument has a lot more purchase on the process of defining phenomenal concepts than Chalmers allows, or can allow. For him and neo-cartesianism generally, the experience must be self-evident to the experiencer and the experiencer can easily form a concept on the basis of his own experience. We argue that the kind of privacy which allows inverted qualia possibilities is the very privacy that the private language argument tilted at. Chalmers recognises the problem posed by the argument, but does not

meet it. Our claim is that, at the very least, phenomenal concepts are not well grounded. Some are confident in taking the conclusion further.

The identification of possession-of-qualia with consciousness.

Finally we turn to the consequences of identifying consciousness with the possession of qualia, linking the two concepts together. As we have hinted already, this has the effect of confirming the neo-cartesian in his non-reductive position. On this account, mental qualitative properties cannot occur unconsciously by definition - there is nothing it is like to be a driver of a car on a long journey, zoned out for a while, or a sleepwalker, while sleepwalking, or someone with blindsight when stimuli are shown in her blind field. The neo-cartesian observes that it is interesting that blindsight patients can respond appropriately to visual inputs, but this does not deflect them from their position concerning consciousness. Perception always occurs consciously; the concept of qualia has tied into it, the idea that it is something of which we are consciously aware.

But then, what account can we give to the behaviour of the sleepwalker, or the blindsight patient? Can we really deny that the sleepwalker has genuine qualitative states? Sleepwalkers and blindseers are capable of really quite sophisticated discriminations: performing actions without conscious control, negotiating obstacles, and so on. The same is true of the drivers who fall into an autonomic mode when driving long distances.

It may be that, if we abandon the pre-theoretic commitment to the idea of identifying consciousness with the possession of qualia, and uncouple the two, a more compelling account of the dissociations becomes possible. This might be reason enough to proceed, but if at the same time, our new conceptual infrastructure gives better purchase to a more reductive account of mind, on this view this would be progress. This argument will feature and be developed in later chapters.

There is a final feature of the conceptualisation of consciousness and relevant items in the field. According to Chalmers, to be conscious is to have qualia, or phenomenology. This is, in a sense, an all-or-nothing question. If you have qualia, there is something it is like to be that creature, and not if not. Chalmers deduces that certain lower forms of life have a phenomenology in virtue of their perceptual psychology, but at some point it will ‘wink out’ (Chalmers, 1996, p.295). Dogs, mice are assumed to be conscious: ‘it is entirely plausible

that there is something it is like to be a mouse' (Chalmers, 1996, p.294). But in these concessions, there is not much sense of gradation – it is still down to the possession of phenomenology. You either have it, or you don't. Despite Chalmers' descent into the phylogenetic scale all the way to thermostats, (Chalmers, 1996, p.295) what is lacking in the conceptual story is any sense of developmental increase.

This point brings us back to the Kuhnian points made at the outset of this chapter. If a conceptual scheme is static in this sense, it lacks an ability to explain how consciousness has developed. As we shall see, Chalmers allows that there is a close coherence or correspondence between the phenomenological and the psychological. In the realm of the psychological, there can be increments in the perceptual discriminations and psychology of evolving animals, and it is possible that these will correspond to an increasingly rich phenomenal manifold. But there is no explicit causal link between the two. It may be that there are ways of conceptualising the picture in such a way as to allow a developmental account of consciousness that in turn allows us to tell a story of the growth of consciousness over evolutionary time. It will be the contention of this thesis that just such a conceptualisation exists, and for the reasons advanced is to be preferred over the neo-cartesian account.

Conclusion

The purpose of this section was to establish a *prima facie* case against qualia, the attendant term 'what-it-is-like', the way they are conceived, and what role they play in the non-reductive aspect of non-reductive monism. We have seen that dualists think of them as being a 'just obvious' feature of our consciousness, built upon an incontrovertible ability to introspect and identify them.

It is clear that the dualists form their phenomenal concepts in a Lockean way, by dint of experience – when Mary exits the black and white room, she has a 'wow' moment and forms the phenomenal concept '*Red*' when she sees the tomato. Having taken the step of introducing qualia as a defining feature of consciousness, which is itself something over and above the physical features of the world, dualists must then resort to exotic accounts of how the (non-reductive) qualitative features of experience are generated (emergence, panprotopsychism etc) and how their relation to the physical world can be explained (or not: epiphenomenalism).

We have argued that qualia as a concept is not grounded well by introspection and that this term, along with ‘what-it-is-like’, is at the very least poorly formed. There is a tendency for dualists to define the one in terms of the other, when both stand in need of justification. We have argued that the private language argument poses a significant challenge to a dualist phenomenal concept strategy. Finally, we argue that the pre-theoretic identification of consciousness with the possession of qualia is at the heart of the dualist position, but we do not have to be committed to this. Other conceptual foundations are available; ones that can provide the possibility of an account of experience in reductive terms, avoiding the problems adverted to above. The same conceptual foundation may allow us to give a better account of the development of consciousness over time, or the differences in phenomenal sophistication between different species today. In the next chapters we shall consider how our target philosophers handle the dissociations we have introduced. It is the contention of this work that the dissociations will help to prise open the dualist view of consciousness, and make the case for a different conceptual foundation.

Chapter 3

The implications of dissociations for Chalmers and the neo-Cartesian position.

In this chapter we shall look at the foundational conception at the heart of Chalmers's work in *The Conscious Mind* - the concept of phenomenal consciousness. We shall argue that as he establishes and builds upon this foundation, he subscribes to more than one aspect of what can be described as a 'neo-Cartesian position'. We argue that phenomenal consciousness as he understands it, rules out by definition, the possibility of consciousness lacking qualitative properties. As we shall see, there are steps in Chalmers' argument. The first introduces a stipulative understanding of experience as necessarily conscious. If perception as we ordinarily understand it is experience, this would mean that there can be no such thing as unconscious experience. The second step in Chalmers' position is more empirical, in which he argues for a coherence between the phenomenal concepts he advances, and corresponding psychological terms.

We shall argue for a different stipulation or pre-theoretical conceptualisation against the first step, and present more empirical argments against the second. We shall claim that the stipulation as advanced puts obstacles in the way of work towards naturalistic theories which ground consciousness in states of the brain. We shall argue that the dissociations that are the centrepiece of this work are more of a problem for Chalmers' position than he is ready to concede. Were he more ready properly to accommodate the dissociations, it might occasion a re-think of the conceptual foundations in his work.

This then is the line of advance:

i) We shall develop and examine Chalmers's understanding of phenomenal consciousness, within the family of neo-Cartesian positions, and what that concept commits him to.

ii) We shall argue that that this leads him to a blind alley as far as any naturalistic explanation is possible, and indicate the inadequacy of the position in dealing with the empirical evidence that dissociation provides.

iii) We shall identify the advantages that a different conception might have in accommodating the empirical evidence, and in making progress towards a more plausible account of mind.

Neo Cartesianism in Chalmers

Firstly, however, some comments on the neo-Cartesianism in his work. Philosophers in this tradition differ to the extent to which they can be described as such. In doing so, they present more superstructure above the waterline to be aimed at. Chalmers sometimes swings close to positions that are more easily attacked, but he is for the most part very careful. As we set out his position, we shall make it clear what target he presents.

It is possible for philosophers uncontroversially to agree on some of the basics: mentality is comprised of the *phenomenal* – perceptions, sensations, proprioception - and the *intentional*, or propositional attitudes – beliefs, desires. It has become a commonplace to think that the latter can be held unconsciously, until we call them to mind. Nevertheless, our behaviour can be shaped by them. Non-occurrent beliefs can play a part in our causal nexus. But let us set them apart for the moment; it is the phenomenal that interests us here.

According to the conception presented by Chalmers, phenomenal consciousness has a qualitative character which is essential to its nature. All that I perceive and sense has such a qualitative character.

Supporting this is an appeal to the character of our epistemic access to phenomenal consciousness – to what Chalmers argues is more intimately known than anything else. While this is not a claim for an *infallible* access through introspection to the phenomenal, there is a clear appeal to our intuitions. The claim of infallibility would be more true to the original Cartesian argument based on what is self-evident, but it is not a target presented by Chalmers.

Chalmers' position also has the consequence of making any reductive explanation of consciousness more difficult. Attaching an essential qualitative character to my sensory states raises the question of why they should be like this, and why they should have any such character at all – the so-called 'hard problem.' In the event that there are sensory states that do not have such a qualitative character, if we find examples of successful sensory discriminations of the same fundamental kind as we find in standard cases but stripped of their qualitative overlay, we may find that process more amenable to a reductive account.

How can Chalmers' position be described as neo-Cartesian? The term clusters several philosophers who might be said to owe their inspiration to assumptions found in Descartes'

work, but they by no means subscribe to the same assumptions as each other, and none of them buy into the substance dualism of the *Meditations*.

Having said that, there is a number of contentions which Descartes makes, which find echoes in the works of recent philosophers. The first is this. Descartes, in a reply to Arnaud said: 'As to the fact that there can be nothing in the mind, insofar as it is a thinking thing, of which it is not aware; this seems to me to be self-evident' (Cottingham *et al*, 1985, p.171). One reading of this is that Descartes seems to be committed to the view that all mental states are conscious.

Some philosophers writing recently, have come close to this position. Galen Strawson (1994) for example, argues that all our occurrent, experiential phenomena are necessarily conscious. Our other mental states are ones of which we are necessarily aware. Chalmers does not offer quite the same target, although there are hints that he is tempted to take a similar line. As determined above, one of our tasks will be to identify exactly where Chalmers stands on this question.

The other positions borrowed from Descartes might include doctrines about our epistemic access to consciousness – the claims, for example, that it is through introspection that we arrive at knowledge of our mental states; or that we cannot be wrong in the judgements we make (if I think I am in pain, then I am). Again, different philosophers are divided over the extent to which we should cleave to such doctrines. As we have already indicated, in the case of Chalmers, his commitment to the first is implicit – he begins *The Conscious Mind* with the line: 'There is nothing we know about more directly than consciousness' (Chalmers, 1996, p.3), and it may be claimed that the appeal to 'what-it-is-like' talk is an appeal to subjective experience and what we know of our own case from within. Most philosophers though, even the usual suspects here, have tended to recoil from a full endorsement of the doctrine of infallibility.

The final doctrine of Cartesianism is the claim that mind is not reducible to physics. On one, very common, reading of Descartes, mind is composed of a substance entirely different to material things[10]. Most neo-Cartesians do not propose a *substance* dualism; most subscribe to some form of monism, but hold to some form of non-reduction – explanatory, ontological. Expanding on the first, they hold that there is an *explanatory gap* (Levine 1983)

[10] Revisionist philosophers include Seager (1999, p.12), Strawson (2006). Here is Seager: 'I am sometimes tempted to doubt that Descartes was really a Cartesian Dualist'.

between functional, or physical accounts of mental phenomena and consciousness. Chalmers builds his case on the failure of logical supervenience. 'Conscious experience is almost unique in its failure to supervene logically [on the physical]' (Chalmers, 1996, p.71). That 'consciousness accompanies given physical processes is a further fact, not explainable simply by telling the story about the physical facts' (Chalmers, 1996, p.107). Nevertheless, while Chalmers builds his case on the 'irreducibility of consciousness' he still regards his position as a *naturalistic* one. Let us turn now to the specific claims made in Chalmers' work.

The Conscious Mind.

In his seminal 1996 book *The Conscious Mind*, Chalmers has three broad projects.

1. Firstly, he establishes and defines his understanding of phenomenal consciousness, in a section in which he then goes on to conclude that that consciousness is irreducible to the physical – arguing that it fails to supervene logically upon the physical. He famously introduces the 'hard problem' of consciousness: why should there be any phenomenal properties at all, accompanying neural processing; why are sensations like this rather than like that.
2. Secondly, he proposes a working framework that could act as the foundation for a science of consciousness – as a way of working towards a solution of the easy problems of consciousness, such as: 'How does the brain integrate information?' or 'How does the brain process information about its environment?' It is clear however that this will not have traction on the hard problem as defined.
3. Thirdly, he offers some thoughts on how consciousness might be explained: controversially arguing for the idea that experience should perhaps be included as a fundamental feature in our ontology – that our ontology needs to be expanded to include experience alongside mass, spin etc as fundamental features of the world. This leads him to propose panprotopsychism as a possibility. This is a product of the constraints on theories of mind and consciousness introduced by his definition of phenomenal consciousness and the failure of experience, as he has it, to supervene logically upon the physical.

This anatomy more or less coincides with Chalmers' own contents description, spelt out in the Introduction (Chalmers, 1996, p. xv).

This chapter will not look in detail at 3. Instead it will develop the first and second themes. If it can be shown that Chalmers' project in those themes is flawed, or otherwise faces difficulties with the possibility of unconscious perception, or unconscious mental qualitative states, in the dissociations we are examining we may want to shift our pre-theoretic commitment to a different picture. We may then find that the constraints leading to 3 will be lifted, and we can, as planned, offer alternative theories of consciousness rather than resort to panprotopsychism, especially in the event this involves a costly counter-intuitive expansion of ontology.

Phenomenal Consciousness

Let us look carefully at the first claim, and in particular, the understanding that Chalmers and others in a similar stable have of 'phenomenal consciousness'. We should remark first of all that there is much conceptual competition between philosophers in this vicinity. Terms used by philosophers and other interested parties are often used in strikingly different ways by others. As careful as they are individually, there is, as Platchias notes, 'terminological chaos in this area. "Awareness" for instance, a key term […], is used in different contexts to mean different things' (Platchias, 2011, p.66). We shall need to proceed with care. 'Consciousness', 'experience', 'awareness', 'phenomenal quality' and other terms are all in the mix, and distinctions and differentiations may need to be drawn.

Although Chalmers is at pains to make careful distinctions between the phenomenal and the psychological, he does add to the terminological mix by leaning towards regarding perception as a psychological term, when most other philosophers would regard it as having a phenomenal flavour. In an instructive section (Chalmers, 1996 p18) he weighs up the arguments. If we regard 'perception' as a phenomenal term, 'the possibility of subliminal perception counts against this construal, though some would argue that it only qualifies as perception in a weakened sense of the term' (loc cit). He argues, correspondingly, for adopting 'sensation' as the correct term in this vicinity on the grounds that 'unconscious perception' makes more sense than 'unconscious sensation'.

A couple of points must be made. Firstly, for all the care Chalmers is taking, it is possible he is looking for a term involving mental qualitative properties, of which it would be more obviously oxymoronic to prefix the term: 'unconscious'. This is the true position, regardless of whether we prefer 'perception' over 'sensation' or vice versa: his view is that

mental qualitative properties do not and perhaps cannot occur unconsciously. Throughout this thesis we shall ourselves prefer the term 'perception' over 'sensation' to refer to states involving phenomenal properties, as this is the more common usage.

Bracketing 'perception' as a psychological term allows him here and elsewhere to accept that blindsight is a case of 'subliminal perception' (Chalmers 2010, pp 94 and 503), but we must remember that this is not to concede anything to the argument here.

Chalmers makes it clear he links consciousness with mental qualitative properties when he begins his work with the following definition: '[..] a mental state is conscious if it has qualitative feel – an associated quality of experience. These qualitative feels are also known as phenomenal qualities, or qualia for short. The problem of explaining these phenomenal qualities is just the problem of explaining consciousness. This is the really hard part of the mind-body problem' (Chalmers, 1996, p.4).

The existence of qualia as defined, and the commitment that phenomenal consciousness is to be understood as having this qualitative character, is a controversial position. It is shared broadly among neo-cartesian philosophers, but is not confined to those of a non-reductive persuasion. CI Lewis, first coined 'qualia' as a term in 1929: 'There are recognizable qualitative characters of the given, which may be repeated in different experiences, [...] I call these "qualia." [...] The quale is directly intuited, given, and is not the subject of any possible error because it is purely subjective' (Lewis 1929).

In a similar vein is John Searle: 'Every conscious state has a certain qualitative feel to it, and you can see this if you consider examples. The experience of tasting beer is very different from hearing Beethoven's 9th Symphony, and both of those have a different qualitative character from smelling a rose, or seeing a sunset' (Searle, 2010, p.39).

Frank Jackson, in announcing himself a 'qualia freak' in '*Epiphenomenal Qualia'* (Jackson 1982), says: 'I think there are certain features of the bodily sensations especially [...] which no amount of purely physical information includes. Tell me everything physical there is to tell [...] and you won't have told me about the hurtfulness of pains, the itchiness of itches, pangs of jealousy (etc)'. Jackson goes further, to claim: 'Nothing you could tell of a physical sort captures the smell of a rose' and that this is sufficient to establish the falsity of Physicalism. Other philosophers have also borrowed the term, or its synonyms, for example Robert Kirk in '*Raw Feeling*' in 1994 (see especially pp 2-5)

Qualia feature in a variety of arguments and thought experiments designed to torpedo forms of physicalism and functionalism. See for example, Jackson's Mary and the Black and White Room, Block's Chinese Mind. The common feature to these arguments is the point that while physics can tell

us much about the process of seeing or experiencing, there are other facts concerning the qualitative character of experience that physics cannot capture. Therefore, a purely physical account is incomplete.

The concept of phenomenal consciousness, however, is closely identified with another term: 'what-it-is-like'. This term was first popularised by Nagel[11] (1974), and taken up by several philosophers as a way of amplifying the concept. Thus Chalmers: 'We can say that a being is conscious if there is something it is like to be that being […] similarly, a mental state is conscious if there is something it is like to be in that mental state' (Chalmers, 1996, p.4).

As we explore the idea, we must ask the following connected questions: what is meant by 'what it is like', is the phrase illuminating, and what, if anything, does the phrase do to explicate the idea of phenomenal consciousness?

What is meant by: 'What-it-is-like.'

Interestingly and as we have noted, Nagel does not spend a lot of time on the specific meaning of the phrase, and only gives a short footnote on the question, saying: '[..] the analogical form of the English expression "what it is *like*" is misleading. It does not mean "what (in our experience) it *resembles,*" but rather "how it is for the subject himself". This makes it reasonably clear that the term is not being used in any comparative way. Nagel is quite clear: he wants to know: 'what it is like for a *bat* to be a bat.' And: 'An organism has consciousness if there is something it is like to *be* that organism – something it is like *for* the organism'. The point, perhaps, of using the notion is to make an appeal to the perspectival or first-person aspect of consciousness. This is in keeping with a neo-Cartesian inclination to introspection as a way of pinning down their understanding of phenomenal consciousness. It also draws attention to the qualitative character of experience.

Is 'What-it-is-like' illuminating?

Concerning the second question, whether the term succeeds in characterising, or elaborating upon qualia, we can offer the following thought. If there was a hope that the phrase might point to something particular, it would need some elaboration, but Nagel does not offer an account of what it actually *is* like to have an experience. In fact, little is offered by any of those who would use the phrase. Chalmers makes the following attempt: he suggests that what is *common* to visual sensations,

[11] Though he concedes he was not the first to use it. He attributes, in Nagel 1986 (p.15), first use to Timothy Sprigge.

perceptions in different modalities, bodily sensations, mental images conjured up internally, and the felt quality of emotions is that they are all states united by the idea that there is something it is like to be in them (Chalmers, 2010, p.5). We have already observed that this does not seem to take us further forward. As we have recorded, Chalmers has another attempt: while conceding, in a way rather instructively, that 'It is often hard to pin down just what the qualitative feel of an occurrent thought is [...]', he suggests that 'When I think of a lion, there seems to be a whiff of leonine quality to my phenomenology' (Chalmers, 1996, p.10). It is hard to say what this pins down; indeed, it looks very circular. Chalmers suggests that the character of a thought about a lion is 'subtly different from what it is like to think of the Eiffel tower'. It is not certain what the nature of this difference could be. As I summon the image of the Tower, and then summon a lion, either on its own, or as part of the image of the Tower, I can detect no qualitative change in the thoughts.

We can make the following additional comments, concerning whether the phrase is illuminating. As we note above, Nagel always intended that his phrase not be understood to involve comparison; he was not asking what is it like for me to be a bat, or for a librarian to be a carpenter. He was asking: what is it like for a bat to be a bat, or a human being to be a human being. But to repeat the term adds nothing to whatever is being specified. If we eliminate the repeated term we are then left with: 'what is it like to be a human being?' – a question no more interesting than an enquiry into the attitudinal features of the life of a human being, and can be answered by reference to them. Such a question has a reply in acquainting ourselves with the biographical details of the human being in question. Nagel claims that there is something 'precise that it is like [for us] to be us' and that 'while we do not possess the vocabulary to describe it adequately, its subjective character is highly specific' (Nagel 1974 in Chalmers, 2002, p.221), but his locution does nothing to identify this specific character.

A still more critical contribution comes from William Lycan (1996) who writes to condemn the use of the phrase by others, but his comments apply quite as much to Chalmers. Lycan observes that there is a tendency for the neo-cartesian tradition to use the term in a way to lump together many different phenomena and issues indiscriminately together (Lycan, p.70). 'What it is like' is now ambiguous between phenomenal character (that is, a quale in the strictest sense) and the conscious experience of such a quale, and perhaps also what one knows, introspectively, in virtue of having such an experience. These different uses of the term are to be found in *The Conscious Mind*: thus, Chalmers writes (Chalmers, 1996, p.10) 'The qualitative feel of an occurrent thought [is cashed in as] there is something it is like to be having such thoughts'. This equates the phrase with qualitative feel, or phenomenal character. However, in a later section (Chalmers, 1996, p.103-104) in which Chalmers discusses Jackson's Mary, Chalmers defends the argument that Mary gains knowledge of a new fact when she comes to know what it is like to see red for the first time. This is to use the phrase in a way to refer to what one comes to know by having the experience of a quale for the first time. Lycan is

inclined to think that equivocation between the different uses of the term is 'deplorable', and that 'nothing is clarified or explained by reference to it' (Lycan, 1996, fn 3, p.176). He goes on to urge that both psychologists and philosophers should separate questions about qualia from questions about awareness and introspective consciousness, as this leads to confusion in research about consciousness.

In a more recent contribution Jonathan Farrell (2018) also argues that 'What-it-is-like' talk is at least unclear. It is moreover not technical in the sense that it introduces a specifically philosophical meaning, either by indicating that the words used in the locution are being used in a different or more precise way, or given meaning by some non-explicit way, such as ostension. Farrell concludes that we should not accept that talk of this kind sheds any light upon the object of our investigation: consciousness. Our view, which does not conflict with what Farrell concludes, is that What-it-is-like talk has its appeal precisely because it is not technical, and is easily understood by those in the trade and those who are not. It makes an appeal to a shared experience – it is perhaps like a masonic handshake – you are part of the club if you accept its meaning.

Taken together with previous comments made in chapter 2, these attacks on the central concepts at the heart of neo-cartesianism should make us wonder whether 'what it is like' talk adds anything useful to the discussion.

What does 'what-it-is-like' add to the concept of phenomenal consciousness?

The attacks above recognise that the notion of phenomenal consciousness and qualia and their reliance on what-it-is-like as a characterisation have a number of critics; the comments hint at the problem of a settled unimpeachable notion of phenomenal consciousness.

To begin with, there would appear to be a circularity in definition; a point admitted by Block, in his contribution to the *Companion to the Philosophy of Mind* (Guttenberg, 1994, p.210-211): '[Concerning phenomenal consciousness] there is no non-circular definition to be offered; the best that can be done is the offering of synonyms, examples, and one or another type of pointing to the phenomenon. For example, I used as synonyms 'subjective experience' and 'what it is like to be us'. In explaining phenomenal consciousness, one can also appeal to conscious properties or qualities eg the way things seem to us or immediate phenomenological qualities. Or one can appeal to examples, the way things look or sound, the way pain feels, and more generally the experiential properties of sensations, feelings, and perceptual experiences.'[12]

[12] This article is worth quoting a little more extensively: in a later section, (p 216) he argues for an equivocal usage of the terms in this vicinity: 'Am I saying that the word 'conscious' is ambiguous? I don't think that the different concepts of consciousness I have mentioned indicate a straightforward ambiguity […] I would rather say that 'conscious' (together with 'aware', 'experience', and other

It will be noted that one of the synonyms here is '*subjective* experience'. So making it clear that tied together with the concept of phenomenal consciousness is subjectivity, with an implication of objective inaccessibility and privileged subjective access. Indeed, across the several definitional accounts we can see a subjective or perspectival aspect being built together with the notion of phenomenal consciousness; consciousness is locked together with the qualitative properties of experience, such that such properties are *essentially* conscious, and we know this by virtue of our immediate introspective access.

The implications of the concept as constructed.

As we have seen, these features are all part of a piece, used in an interdefinitional way. They are locked-together characteristics of mental states that appeal to what intuitively seems obvious, a subjective, how-it-feels nature, known by introspection. The recourse to phenomenal qualities, or qualia, or to say we are the possessors of mental states with a specific qualitative nature, are all unhelpful in taking the analysis further. It may be argued that this strategy is actively misleading, since reference to qualities or properties would seem to give the mental states in question an intrinsic nature. This commits us to a way of thinking about qualia which makes them a something experienced, or a characteristic of that experience, rather than a feature of a something which is itself experienced.

It becomes clear that the concepts constructed here carry within them important commitments. This is firstly that qualitative character is a property essential to phenomenal consciousness. Secondly (or rather, what amounts to much the same thing) that consciousness is locked into first order verbs of perception as noted by Platchias (2012, p.71), such that all qualitative, sensory states are understood as being conscious. There cannot be states with qualitative character in the sense intended that lack consciousness. Apart from anything else, this has important consequences for the way in which the kind of dissociations we are interested in, are to be understood.

Of course, the concept as constructed then allows Chalmers to proceed with his arguments against reduction. Phenomenal consciousness defined in this way, becomes the centre of the hard problem. It allows him to argue that consciousness does not supervene upon the physical. He bases one argument, for example, on the logical possibility of phenomenal zombies, that is, creatures that

words similarly used) *should* be ambiguous.' He clarifies this, saying that 'consciousness' is a cluster concept, like religion. And that the various types of consciousness he would identify (access-consciousness, self consciousness etc) are indeed so different from one another that a failure to differentiate can lead to serious confusion. It might be claimed that one response to this multiplication of concepts is to suggest that there is one fundamental kind of consciousness, and that these are all wrinkles in that one fundamental kind.

are physically and functionally identical to human counterparts, but lack a qualitative character to their experience or a sense of what-it-is-like. The very conceivability of such postulated creatures is sufficient to undermine the case for reduction. Chalmers argues that a reductive account based on functionalism fails to capture the character of what-it-is-like, since a zombie is, by definition, a functional equivalent of me, but lacking internal qualitative states, or phenomenology. His claim is that a reductive account requires that consciousness be logically supervenient upon the physical facts. The conceivability of zombies (he claims) shows that consciousness is at most naturally supervenient upon the physical facts, but not logically supervenient.

The main thrust of the argument is on Chalmers 1996 p.38: 'If B-Properties supervene logically on A-Properties, then once God (hypothetically) creates a world with certain A-Facts, the B-Facts come along for free as an automatic consequence. If B-Properties merely supervene naturally on A-Properties, however, then after making sure of the A-Facts, God has to do more work to in order to make sure of the B-Facts: he has to make sure there is a law relating the A-Facts and the B-Facts.'

As noted, his argument is that the conceivability of zombies shows that the qualitative properties of consciousness do not supervene logically upon the physical facts. At most, they are naturally supervenient upon the physical facts. God has to do more work to make sure that certain B-facts of consciousness (the what-it-is-like character) come along with the A-facts of my physical constitution. Physicalism would require that A-facts alone would be sufficient to account for all the properties; that it does not demonstrates that physicalism is insufficient.

As we commented above, a second argument is built upon Mary emerging from the black and white room, possessed of all the physical facts about colour experience, and who nevertheless learns something new when she encounters the tomato for the first time – what-it-is-like to see red for the first time. Physicalism (the possession of all the physical facts) is incomplete – it might be sufficient for Mary to recognise the redness for what it is, and to provide a neurological account of what happens inside her as she sees it for the first time, but it fails to furnish us with an account of what happens when Mary emerges from the room after her confinement – ie the qualitative character of her experience - what-it-is-like to see red – the specific quality of experience that is differentiable from (say) the greenness of the tomato's stalk.

It is unnecessary to go on – the arguments are well-known. Much ink has been spilt in trying to rebut these arguments, but much of their appeal comes from a conceptual foundation which we see as flawed. The adoption of a different conceptual foundation might make the debate around them unnecessary.

In each of the arguments as presented, however, we need only note the pre-theoretical understanding of phenomenal consciousness. The key thing about this understanding is that, if the notion of what-it-is-like is locked into the concept of conscious mental states from the beginning, by

definition, there cannot be any cases of unconscious qualitative states. Rosenthal makes the same point: '[according to the Cartesian conception] the idea that a mental state could have a sensory character and yet not be conscious is simply unintelligible. For it may seem that the very idea of a nonconscious state with sensory qualities is, in effect, a contradiction in terms' (Rosenthal, 2005, p.33).

Accordingly, the various candidate exemplars of unconscious states with qualitative character such as we find in blindsight, the dissociations herein, etc receive no debate in the first section of Chalmers book. They make no appearance because of the way mental states are defined. The dissociations present no challenge to Chalmers' viewpoint because of the way that viewpoint is conceptually constructed. However, it should now be clear that this way of looking at mental states and phenomenal consciousness is stipulative and pre-theoretical. There are of course, compelling reasons to adopt it – there are appeals to what is considered obvious, intuitive and common sense, and there is recourse to our epistemic access to mental states – ie introspection. But there are arguments to be made against these motivations. The next section will explore these.

Criticisms of Chalmers' conception of phenomenal consciousness.

In order to make the case for a reconsideration of this concept, we should first of all consider its weaknesses. We can make a quick and rather cheap point to counter any argument built on what is considered obvious. It was once thought incontrovertibly obvious that the sun went round the earth, but this did not secure the truth of the proposition. The curvature of space-time by gravity is not obvious, but this does not make it any less true. These may be empirical matters, but they make the point that an appeal to the obvious (as Chalmers does, on p. xiii), is an appeal to intuitions, and as such needs careful examination. The argument from introspection needs more attention, as it is largely from this that the point above proceeds.

Rosenthal sets the scene: 'Most of our knowledge of mind derives from introspection […] This close fit may tempt some to hold that introspection is a privileged source of knowledge that is somehow immune from error' (Rosenthal, 2005, p.43). As we have noted, Chalmers hints that this is his position when he says we know about consciousness more intimately and directly than anything else. However, this view today is not uncritically absorbed.

It is now a commonplace to argue that we do not have authoritative access, via introspection, to propositional attitudes; we do not necessarily reach definitive judgements about our desires and feelings via introspective means. Freud and Schopenhauer have both made the point that we have motivations of which we are unconscious, and to which we do not have direct access. We may easily not know we are depressed; not know we need to talk about something. We may easily not be aware

we are anxious, or angry. Other propositional attitudes might be suppressed, others unconsciously held.

As for qualitative experiences, perceptions, Shoemaker comments: '[…] Cartesian views about our access to our own minds, including claims about infallibility and self-intimation have fallen out of favour' (in Guttenberg, 1998, p.396), and it may not be necessary to make too many further comments here.

In experimental situations studying phenomena such as change blindness, inattention blindness etc, there is real doubt whether introspection delivers in a reliable way and whether this therefore can sustain the kind of concept construction at the heart of the neo-Cartesian project. There is a further point, which can be briefly made, which is to argue that with all introspective judgements, there is a problem of criteria of correctness. We have already launched a Wittgenstein-inspired criticism of concepts built upon such judgements, and the impossibility of private rule-following, which need not be repeated.

Our view is confident in endorsing the broad externalist position taken by this tradition, and Schwitzgebel's conclusion: 'I see no reason to accept […] introspective foundationalism. Indeed I suspect that the opposite is nearer the truth: our judgements about the world tend to drive our judgements about our experience. Properly so since the former are more secure' (Schwitzgebel, 2011, p.137).

Let us summarise: there is a clear sense in which Chalmers and others adopt their conception on the strength of what seems intuitively obvious and indubitable, and known by introspection. These grounds, however, are not compelling, and the motivation for accepting the position has to be set alongside those for an alternative conception.

Explanatory inadequacy I: The 'absence of analysis'.

In one of the arguments he makes rejecting the possibility of reduction, Chalmers maintains that any functionalist account fails to provide further analysis of what is essential to consciousness. He claims that functionalism must be rejected on grounds of inadequacy. The section is worth quoting at length.

> For consciousness to be entailed by a set of physical facts, one would need some kind of analysis of the notion of consciousness – the kind of analysis whose satisfaction physical facts. [Functional analyses] simply miss what it means to be a conscious experience. Although conscious states may play various causal roles, they are not defined by their causal roles. Rather what makes them conscious is that they have a certain phenomenal feel, and this feel is not something that can be functionally defined away. […]

> [According to functionalism] all we have to do to explain consciousness is to explain our ability to make certain reports, or to perform certain sorts of discrimination […] But on the face of it, it is entirely conceivable that one could explain all these things without explaining the *experience* that accompanies the report or the discrimination. […]
> [In dealing with questions such as: 'Do mice have beliefs?' answers] depend on how we draw the boundaries in concepts […] But compare: Does a mouse have conscious experience? Does a virus? These are not matters for stipulation. Either there is something it is like to be a mouse or there is not (Chalmers, 1996, pp.104-105).

In this section we can see all the elements of the criticism we have levelled so far. According to the text, what it is for something to be conscious is for it have phenomenal feel, and what we are to understand by phenomenal feel is that there is something it is like (to be conscious). The target concept is circularly defined, and wrapped up in the concept of consciousness of a mental state is our being introspectively aware of it. We can only know what it is like by reference to our own case, a situation which has led to defenders of this position saying: (In reference to Louis Armstrong's reply to the question: What is jazz?) 'If you gotta ask, you ain't ever gonna know'. Third person approaches to consciousness 'miss the point' or 'change the subject', or 'define away the problem'.
But note that, in contradistinction to the title of the section given by Chalmers, it is *this* approach which fails to provide analysis. We are not allowed, by this definition, to make any progress in analysing further. Consciousness is primitive, is what it is, and there is nothing further to be added. Chalmers' analysis extends to what he defines as the psychological categories only.
But note: 'Does a mouse have conscious experience? Does a virus? These are not matters for stipulation'. This makes the question an empirical one, and is not decided by prior definition. As we have argued here, however, the way consciousness is construed in *The Conscious Mind* **is** decided by prior definition.

It may be argued that a more productive analysis is possible in the event that Chalmers's conception of consciousness is jettisoned. There are two powerful reasons why. The first is this. Recall that the concept of phenomenal consciousness deployed by Chalmers involves the commitment to the idea that the qualitative character of experience is essentially conscious. Now, in addition to the various dissociations explored, there are other, well-grounded cases of sleep-walking and automatism in the literature. In such cases the people concerned show very complex behaviour, such as eating a snack, or driving a car, all asleep, or at least lacking awareness of their activity. This neo-Cartesian view is to deny them sensory states, but what reason could we possibly have to deny that that the

sleepwalkers are genuinely in qualitative states? We lack any convincing account of their behaviour, and cannot explain why or how they navigate their environment successfully.

We might for example refer to the case of Kenneth Parks, who, in 1987, made Canadian legal history when he drove 23 km while asleep, and not only killed his mother-in-law, but also sustained serious knife wounds to his hands. He is one of a number of cases of automatism which have made legal precedent (see June Callwood: *The Sleepwalker* 1990). Again, Parks would have made discriminations which had a qualitative character during his 23km drive, and yet, as the account has it, he was completely unaware of them. It is not clear what account Chalmers can make in these situations, other than to put them down as cases in which some phenomenal consciousness is present, or as being under-determined. The better account however, would say that sensory properties are registered without awareness, and these properties played a role in determining behaviour.

To work with another example, and considering the case of the long distance driver who comes back into the moment, after 40 miles of driving, in which all kinds of motor functions, sensory registrations, judgements, and discriminations, have taken place, but to which she has had no access during those 40 miles. Chalmers takes the view that the qualitative properties of our mental states are accompanied by a subjective feel, and adds that this can also be expressed as: 'what it is like to be a cognitive agent' (Chalmers, 1996, p.6). In the case of the driver, some of her mental states are qualitative: during her period 'off-line' she may have had a mental state associated with the visual information of an indicator light flashing ahead. She will have manoeuvred appropriately. On this view, the driver acted appropriately to visual information without there being an associated 'feel'. She acted as an agent, but without there being anything-it-was-like in the sense of a *qualitative feel*.

It might be claimed that in these two examples of automatism, Chalmers could deny that qualitative properties explain the behaviour of the subjects, and that an adequate explanation can be found in an account using representational states. In reply we might argue with Platchias (2011, p 80) that 'mental qualitative properties are the properties in virtue of which we distinguish between sensations'. Both Parks and the long-distance driver engage in a rich sensory interaction with their environment, involving the successful discrimination between the colour of traffic lights, for example. Without some qualitative properties to make the necessary distinctions between such sensations, we lack any convincing or realistic explanation of their behaviour. More is involved in the account of their behaviour than successful sensory registration. It requires at least discrimination between colours and this suggests the subjects are in genuine qualitative states. We shall make similar points concerning Chalmers' philosophical zombies later.

The second sense in which the neo-Cartesian conception is explanatorily impoverished is given in a short, provocative paragraph in Rosenthal (Rosenthal, 2005, p.40). Thus: '[W]e typically come to make more fine-grained discriminations as we master more subtle concepts pertaining to

various distinct sensory qualities. Experiences from wine-tasting to hearing music illustrate this process vividly. […] If consciousness is intrinsic to sensory states, the relevance of concepts remains mysterious. The neo-Cartesian [denies] that sensory differences exist when we are unaware of them. But it will be even more difficult to explain how learning new concepts can actually cause sensory qualities to arise that previously did not exist'.

In anticipation of a later section, in which we shall look at this in more detail, let us try to say how this indicates a weakness in the Chalmers account. In short, it struggles to provide a convincing account of learning, or the development of consciousness. Recall what the neo-Cartesian is committed to: there are no unconscious sensory differences; all qualitative sensory states are conscious. All the features of an experience are already present to us: there is nothing in an experience that is not already present. How then do we develop the richness of experience that we enjoy as an adult compared to when we were very much younger? How have we developed the richness of our conscious experience over evolutionary time, from a point thousands of years ago, when, we can imagine, we were at the level of sophistication the higher apes are now.

Chalmers must be committed to the view that the learning of new concepts is alone sufficient for the creation of new sensory qualities, since these cannot occur unconsciously. According to this thinking, the acquisition of new concepts results in our sensory states coming to have properties or qualities they did not previously have. But this expects a great deal from the learning of new concepts: it is hard to see how learning a new concept can create a new qualitative property of an existing perception.

There is, on the other hand, a different account - as we come to learn further conceptual discriminations in various sensory experiences, we come to be conscious of further properties *latent* in our experience. Rosenthal gives us the relevant examples. Studying musical theory and appreciation gives us conceptual distinctions such as the difference between augmented/diminished chords, major/minor chords, suspensions, syncopation, harmonic progression and so on. The study of wine and the acquisition of conceptual discriminations such as 'full-bodied', 'tannin-rich' and sensory descriptors such as 'tobacco', 'leathery' and so on add to our conceptual resources concerning wine-tasting. The claim then is, that armed with these additions to our conceptual toolkit, our perceptual qualities are enhanced – and we can identify and amplify qualities already latent in the experience. This is true in adult life, but also in early language learning. We were once *un*conscious of the smoky quality of Shiraz, but armed with more sophisticated conceptual resources, we become aware of it. This seems a more plausible account than suggesting that the addition of a new concept singlehandedly generates a new sensory quality. We shall develop this argument in the concluding chapter, but Chalmers' model would seem to be explanatorily inferior to Rosenthal's in this respect.

The more plausible account to give here is to argue that new concepts make us more aware of the sensory qualities already immanent in the mental states.

Explanatory inadequacy II

Chalmers' neo-Cartesian conception has another major shortcoming in its contribution to the understanding of Mind; this is in its treatment of many of the dissociated cases we have looked at. It is instructive that he does not make clear reference to dissociations in the first half of his book. There is only one brief reference to subliminal perception (Chalmers, 1996, p.18). This is perhaps unsurprising, since, if the model rules out unconscious perception by definition, there would not be a reason to consider it. However, as we have trailed, our argument will be that the dissociations make a difficulty for his conception, that Chalmers' view of phenomenal consciousness rules out a fuller consideration of the dissociations.

In developing this, we need to take some care in identifying the target presented. On one view, the answer lies in Chalmers' distinction between phenomenal and psychological mental terms. He accepts that terms like perception, sensation, awareness have a dual life – they can be understood as phenomenal, and also as psychological terms. It is his view that perception, understood as a psychological term, indicates the processes in which cognitive systems are sensitive to environmental stimulations. But the term can also be understood purely phenomenally. 'Perception', taken as a phenomenal term, involving conscious experience of what is perceived, has, against it, 'the possibility of subliminal perception' but this is quickly qualified: 'but some would argue that this qualifies as perception only in the weakened sense of the term' (Chalmers, 1996, p.18).

For the moment, let us go on. Chalmers suggests that the phenomenal component implied in these terms can be stronger in some than in others. It is stronger in 'sensation' than 'perception' – 'as witnessed by the fact that the idea of unconscious perception seems to make more sense than unconscious sensation'. He suggests reserving 'perception' as the psychological counterpart to the phenomenal term: 'sensation'.

Out of all this qualification, it becomes clear that, on this understanding of perception, where 'perception' is understood as a phenomenal term, the idea of unconscious perception is an oxymoron. Where, however, it is understood as a psychological term, the idea of unconscious perception is given a hearing, but as we shall see, objections are dismissed.

So, apart from the very brief reference to subliminal perception, there is no consideration of blindsight or other dissociations in this first part of the book. It is our view that the reason for this is because his definition of phenomenal consciousness is yoked together with qualitative character. The

dissociations do not feature in this section, since the very idea of sensory registration with qualitative character cannot happen unconsciously.

But then we can ask: What account can Chalmers' neo-Cartesian conception give of these cases – cases for which a great deal of empirical support and evidence has now accumulated? There would appear to be no reason why we cannot shift the burden of explanation. If the response is to deny the phenomenon, he would be open to the same criticism neo-Cartesians often make of their opponents. This seems to be the kind of response Chalmers wants to make, however; in dismissing subliminal perception as perception in a weakened sense, and in his suggestion that one reason to explain the successes in discrimination in blindsight is on the grounds that they are the product of a weak sort of conscious experience (Chalmers, 1996, p.227). In the brief discussion of Helen and a possible candidate filling the super-blindsight bill, (Chalmers, 1996, p.228) he again accounts for their performative success by saying: 'there is reason to believe that phenomenal consciousness is actually present.' Both comments suggest he is quick to discount the cases, and the brevity with which they are dealt is instructive.

Secondly, can the neo-Cartesian conception allow us to make progress on a more comprehensive account of the workings of the mind? Chalmers clearly wants to be able to contribute to further understanding on the (easier) questions already identified: he sets up a research programme in the second section of the book to do so, and remains predisposed to physicalism. ('Temperamentally, I am strongly inclined toward materialist reductive explanation' (Chalmers, 1996, p. xiv).) We might therefore anticipate that he would endorse any conceptual reconfiguration to permit a more reductive approach. Many of the dissociations are neurologically and psychologically investigable: specific areas of the brain have been compromised by strokes, surgery or other damage, and can be identified, and linked with functions. If we can avoid conceptual stipulations which might commit us to regarding 'the qualitative character of phenomenal consciousness' as pleonastic, if we allow that there can be mental qualitative properties outside consciousness, then we have a class of mental states that are outside consciousness, and amenable to a more reductive account. This in turn can begin to make sense of those studies of the various neuronal processes which comprise the components of an experience. In a very important sense we have got off to a bad start with the conceptual foundations as they stand, in *The Conscious Mind*, and only by finding new foundations can we work towards a better connection with empirical findings in neuroscience. It is to this different conception that we might now turn.

An alternative conception of consciousness.

What might a non-Cartesian re-conception look like? This section will identify the main features of such a re-conception and leave a fuller exposition until a later chapter.

Chalmers and others begin with a pre-theoretical stipulative definition, which owes something to common usage, but which yokes together mental states with consciousness. In such an account, 'conscious' would seem to be characteristic intrinsic to the state in question, a characteristic which the neo-Cartesians universally attempt to amplify with locutions such as 'qualitative feel', or 'what-it-is-like'. Side-stepping this, we might uncouple consciousness from mental states, specifying each notion independently. This may be a different stipulation, departing from common usage, but it is one that can be defended.

Following Rosenthal, and Platchias and others, there is an alternative starting position defined as follows: '[on this conception, by contrast to the Cartesian conception above] consciousness is not essential to mental states, and thus consciousness may well be an extrinsic characteristic of whatever mental states have it' (Rosenthal, 2005, p.33) or, to put it differently, 'mental qualitative properties are not invariably conscious; they may occur unconsciously and so are not essentially conscious' (Platchias, 2012, p.64).

On such a reading, a mental state can be sensory, such as a perception, or it can be a propositional attitude, such as a belief, or a desire. What makes either of these conscious, is some higher order mental state: a thought or perhaps the process of attending to the lower order state, but in any event, something construed independently of the original mental state itself. In other words, a conscious mental state is simply a mental state we are conscious of, in virtue of the Higher Order Thought. Rosenthal argues that the content of the higher order thought in question must be, roughly, that one is in that very state.

As this thesis contends, if we construe mental states, even sensory states, in this sense, we can allow that they can occur unconsciously. Mental states possessing the so-called qualitative character, or qualia, can occur unconsciously. This sunders consciousness as a necessary feature from sensory states, and it is able to accommodate the mounting empirical evidence connected with the dissociations we are here studying.

So the first advantage to this new conception is that it does not deny the phenomenon. It properly understands blindsight and the other dissociations as strange cases of experience, though not of the usual kind. The second gain from the reconception is that it lends hope to those who seek a reductive analysis of consciousness. As we have seen, and as Rosenthal puts it (Rosenthal, 2005, p.47) 'if mental states are all conscious, no informative, or nontrivial explanation of consciousness is possible'. Consciousness on the neo-cartesian model is presented as basic, and no further definition,

beyond a circular appeal to ill-formed notions such as 'what it is like' is possible. Philosophers who resist this notion of consciousness, and who might seek a more reductive explanatory account have a potentially productive line of enquiry. One of the intuitions to which the neo-Cartesian approach appeals is the perceived difficulty in reducing consciousness directly to physical states. On this Higher Order Thought model, the claim is that consciousness can be reduced, in the first instance, to mental items, rather than directly to brute matter. A conscious mental state is simply a mental state of which we are conscious. As Seager puts it: 'Mental states which do not intrinsically involve consciousness can be seen as closer to the natural world' (Seager, 1999, p.61). This kind of reduction helps to bridge the explanatory gap, by introducing intermediate mental structures which can be linked on the one side to our intuitions about consciousness, and on the other, eventually, to the physical world, with relative ease.

On one side of this attempt to bridge the gap, a putative reductive account can begin to locate unconscious mental states such as those in blindsight, in the processing of signals in the V4/V8 areas of the brain. It can explain the orientation of DF's hand for the purposes of posting letters etc as the result of processing, in the dorsal stream, of signals received, and so on. None of this is mysterious, and the results of these empirical studies can be given their due weight. Much of the scientific study of consciousness is framed in talk of 'the neural correlates of consciousness', which is a formula designed to satisfy everyone, since it allows that consciousness is only correlated to brain activity, and cannot be reduced to it. With the re-conception here offered, we can begin to replace this phrase with a search for 'the neural realisers (producers, generators or some such) of mental states', but this is the work of neuroscience, and is, as yet, far from completion.

On the other side of the gap, there is also work to be done. There are several versions of the Higher Order claim, suggesting variously that it is in virtue of a Higher Order Thought, or Higher Order Perception, that a mental state is conscious. Other versions might speak of attention paid or devoted to a lower order mental state that brings it into consciousness. This section will not attempt to decide between them, but merely acknowledge that what they share is all that this part of the thesis requires – a recognition of the shortcomings in the neo-Cartesian approach, and a common acceptance that the dissociations are better explained by the HOT model.

There is a third advantage to this step. Chalmers advances the possibility of zombies as part of his argument against reduction in *The Conscious Mind* (p.94 et seq). The point is to defeat functional explanations of consciousness by imagining a situation in which a functional copy of myself could exist but which has no internal phenomenology. And so therefore a functional account fails, since it cannot capture the 'what-it-is-like' aspect of experience. Much has been said about the question of whether such zombies are possible, whether conceiving such a situation can imply their possibility, and so on, but it is worth commenting that the higher order approach need not fear zombies, and one

of Chalmers' broadsides against reduction goes wide of the mark in this case. The Higher Order approach is quite happy to embrace, indeed makes its case on the idea that there are all kinds of unconscious mental states, and cognitive processes existing beneath our conscious mental states. It contends that only a fraction of our mental lives, at any one time, is subject to the 'what-it-is-like' test. Of course, it may be Chalmers can find other complaints against the Higher Order approach, but the possibility of zombies would seem to have no traction. For this form of physicalism, the possibility of zombies is not fatal.

Finally, there is a fourth recommendation to the conception considered. It will have a more convincing story to tell about the evolution of consciousness. We have already hinted at a similar point in the section concerning learning above. We may imagine that, at some point in our evolutionary history, hominids lacked the sophistication of the conscious awareness we have today. To use the phrase we have been keen to caution against, there would not have been much that it was like to be one of the earlier hominid variants. This is to move backwards from such developments as wine tasting, and music appreciation, to more fundamental sensory states, and propositional attitudes, to understanding signals and evaluating a situation for its dangers. (Let us be careful; there will always have been an evolutionary and selective advantage to feeling pain, and pain is something our early ancestors and other animals undoubtedly felt and feel. As a sensory state, it is arguably the best candidate for one of which we are necessarily aware. As such, pain creates particular difficulties for the Higher Order position).

However, pain aside, we have argued that the process of acquiring new concepts is what generates new conscious sensory qualities. Following Platchias (2012, p.84), the Chalmers position is to claim that learning new concepts results in sensory states having qualitative properties they did not previously have. The argument is that this is asking a lot of the acquisition of new concepts alone. We have yet to make a full case for the HOT alternative, but to anticipate it: on this model, any new concepts yield new experiential properties in virtue of making us conscious of such properties already inherent in our sensory states. (But which were, to that point, still unconscious). This brings us closer to the spirit of Kant's synthesis in making the creation of new sensory qualities in the evolution of consciousness dependent on bringing together concepts and (hitherto unconscious) experience. Several writers tie together the acquisition of language with the development of consciousness (Jaynes (1976), Humphrey (1992), McPhail (1998), who work on the assumption that language gives voice to unconscious sensations, or states previously held, and to which concepts come to be attached, as a way in which our consciousness is enlarged, over evolutionary time.

Let us make this point more vivid by comparing two approaches. Firstly, that taken by Feinberg and Mallett (2017). This work is a largely scientific attempt to locate the origins of consciousness, and coincidentally to address some philosophical problems, such as the 'hard

problem[13]'. It starts out with an endorsement of Chalmers' view on qualia, and what-it-is-like, and finds appeal in Searle's emergentism. Such a view seems to be wedded to a neo-Cartesian view of phenomenal consciousness – they are led to pitch the beginnings of consciousness at a point in Earth's history over half a billion years ago, describing arthropod trilobites in the Cambrian ocean as conscious. They are inclined to resist Chalmers' view that experience is fundamental (see Feinberg and Mallett, 2017, p.10) and everywhere, but by endorsing the neo-Cartesian picture, they nevertheless conclude that it is widespread, much more so than previously suspected. (Feinberg and Mallett, 2017, p.206). Revealingly, and undoubtedly as a product of the same commitment, they find themselves troubled by the idea of 'partial consciousness'. 'At first we thought the idea of partial consciousness sounds absurd, like being "half pregnant"'. (Feinberg and Mallett, p.285 fn 43)[14] (Also revealingly, they dismiss the dissociations which are the primary subject of this thesis in a short paragraph on p.207: 'The problem with [this kind of objection] is that the types that have been detected in humans and monkeys are so weak and incomplete that they effectively amount to inattention. [They] cannot be distinguished from a degraded type of consciousness'.)

Let us compare this position with that taken by Julian Jaynes and Nicholas Humphrey. Jaynes, writing in 1976 and before the development of Higher Order Thought theories, nevertheless has a view of consciousness which is more dynamic and developmental than the neo-cartesian picture above. It has concepts at its centre, and links consciousness to learning and words. Jaynes gives an account of consciousness as a far more recent phenomenon, and finds a literary record of its development in the Iliad, as the earliest written poem. He observes: 'There is, in general, no consciousness in the Iliad […] no words for consciousness or mental acts. The words in the Iliad that in a later age, come to mean mental things (*thumos, psyche, etor, kradie*) have different meanings, all of them more concrete.' There is no concept for will, or word for it. Jaynes draws the conclusion that 'Iliadic men have no free will of their own, and certainly no notion of free will' (Jaynes, 1976, p.69-70) The evidence is: 'Iliadic men did not have subjectivity as do we, *he had no awareness of his awareness of the world, no internal mind-space to introspect upon*' (Jaynes, 1976, p.75 my emphasis).

[13] It would, however, fail to satisfy Chalmers or Searle that the problems (the explanatory gap, the case against reduction etc) are solved. Their basic claim is that, these problems are dissolved by an approach that 'integrates the neurobiological, the neuro-evolutionary and the neuro-philosophical domains.' See p.226-7. By some magic in the way these levels are integrated, problems like those associated with qualia can be dealt with. They are analysed as 'the result of a unique, multifactorial neurobiological substrate and recursive interactions between higher and lower neuro-hierarchical levels.' (p.225). But they are 'related closely to the features of auto- and allo-ontological irreducibility. (..) If our scientific explanation of (these) irreducibilities is correct, then these two features could bring qualia into the realm of science.' Sic.

[14] They persuade themselves, that it must exist, however, since 'there is a period between a (human) fertilised egg (unconscious) and full consciousness at birth, and this period must be one of partial consciousness'.

He holds that the relationship between language and consciousness is complex, but: 'The central assertion of this view […] is that each new stage of words literally created new perceptions and attentions, and such new perceptions and attentions resulted in important cultural changes which are reflected in the archaeological record' (Jaynes, 1976, p.132). A final quote: 'Let no-one think that these are *just* word changes. Word changes are concept changes, and concept changes are behaviour changes' (Jaynes, 1976, p.292).

From these quotes, it is clear that Jaynes situates the development of consciousness as we know it in parallel with conceptual changes and the emergence of linguistic sophistication. The Iliad records some of these changes. It is an old text, but the oral version of the written text, is surely much older. We can perhaps see in it the beginnings of a consciousness of existing mental states, an awareness of our awareness. In it is a record of how the addition of new words – concepts – reflects and assists the sophistication of consciousness. The evidence might be dismissed as circumstantial, but as an account, it has a sense of development rather more than the view that the same consciousness, defined as the possession of qualia, is unchangingly present over time.

Jaynes' view finds some support in Humphrey (1998), in which Humphrey likens the imagery in the cave paintings at Lascaux (made 30,000 years ago) to highly naturalistic pictures of animals drawn by Nadia, an autistic girl from Nottingham. The similarities are indeed striking, and Humphrey suggests that the similarity might tell us something about the minds of the cave artists – that they had 'pre-modern' minds perhaps limited in the same way Nadia is, by undeveloped language, and other cognitive impoverishments. On this basis, he argues, we can imagine that 'language was absent in the general population of human beings living in Europe 30,000 years ago', and that they perhaps saw the world in the way that Nadia did. The naturalistic style died out over time, and so did Nadia's interest in drawing in the same way, when she began to acquire a modicum of language. Humphrey admits that the similarities hardly 'add up to a solid deductive argument' but there may be something in the way in which the acquisition of concepts enhances our experience and representation of it, and it is certainly sits more comfortably with the Higher Order Thought understanding of consciousness as mental states we are, or become, conscious *of*. The cave artists saw the world in a certain way, before language; they had mental states of which their consciousness could be said to be limited, in much the way an autistic girl might be described. On the acquisition of improved conceptual tools, they saw (and represented) the world differently.

This again may not be conclusive, but it is illuminating insofar as in it we can see some of the shortcomings of the neo-Cartesian position as it relates to language and concept learning, and the greater appeal of the account of some its opponents.

Summary

The first section of Chalmers' book establishes his understanding of phenomenal consciousness, which is then used to conclude that that consciousness is irreducible to the physical. We conclude that a conception in which qualitative character is an integral feature of consciousness is an obstacle to analysis, and cannot acknowledge dissociations and other empirical phenomena as evidence of unconscious qualitative states. Chalmers' conceptual starting point might sit with common intuitions but is nevertheless stipulative. The dissociations present, as yet, no direct challenge to Chalmers, but this position only obtains because of the conceptual foundations with which he begins.

However, a different conceptual starting point is in a position to take such phenomena more seriously. At the beginning of this chapter it was noted that blindsight features later in *The Conscious Mind*, as a possible difficulty for his claim that consciousness and cognition cohere. It is to this that we shall now turn our attention.

The challenge of dissociations for Chalmers' structural coherence

In the midsection of *The Conscious Mind*, Chalmers turns his attention to a science of consciousness. The consequence of his approach in the first section, was to suggest that consciousness may not be reductively explained, or rendered in terms of something more basic. The hard problem is an insoluble barrier to this. But according to his line of thinking, this should not deter us from finding an answer to the easy (easier) problems of consciousness: making use of subjective reports of experience, and finding neural correlates of consciousness etc. He concedes that, although consciousness does not supervene logically upon the physical, it may nevertheless be imagined that it supervenes naturally. In this event, this relationship may be underwritten by psychophysical laws which are 'on a par with the laws of physics as part of the basic furniture of the universe' (Chalmers, 1996, p.213).

The most that Chalmers will allow is a systematic relation between the phenomenology and the psychology of the mind, which are bound together by many lawful relations. The relationship is spelt out:

> 'Awareness is the psychological correlate of consciousness, roughly explicable as a state wherein some information is directly accessible and available for the deliberate control of behaviour and for verbal report. […] Where there is consciousness, there is awareness. My visual experience […] is accompanied by a functional perception'. But the arrow goes both ways. 'With awareness appropriately defined [as involving direct access with some reportable content directing behaviour] there is a corresponding conscious experience' (Chalmers, 1996, pp.220-2).

This is elevated into a principle: the 'principle of structural coherence'. This principle underlies the science of consciousness as Chalmers sees it, and is the working assumption behind the resolution of the easy problems. 'This is a central and systematic relation between phenomenology and psychology, and ultimately can be cashed out into a relation between phenomenology and underlying physical processes' (Chalmers, 1996, p.225).

Awareness, as a category, receives further development wherein 'accessibility', and 'availability for control of behaviour', and: 'for verbal report' are given more attention. It is clear the availability for control and for verbal report are third person litmus tests for the presence and extent of awareness. They are again stipulative, and up for debate, and there are reasons for being careful here. For example, we should not want to rule out the accreditation

of consciousness to creatures without language, or pre-linguistic children. It is on this basis that Chalmers is ready to shift his account to speak of 'report*ability*' and 'access*ibility* for verbal report' rather than any criterion more demanding. The reason for doing so is to 'refine the notion so as to deal with problem cases', but the point we might want to make here is that the notion may be refined so as to side-step problem cases too, if they turn out to be too inconvenient. Note however that there is a limit here: 'refinement can only go so far, as we lack an experience meter with which to confirm and refine these hypotheses empirically' (Chalmers, 1996, p.226).

There would appear to be no compulsion to accept Chalmers' criteria as he defines them, and when assessing for awareness, a case may be made for adding other tests, such as success in sensory discrimination, neurological support (where this is available), and the explanatory power of the account[15]. We do not have to limit ourselves to the tests that Chalmers imposes. It is true, there is the possibility of error (as recognised early in *The Conscious Mind*: 'We can coherently imagine a situation in which phenomenal property P occurs without a psychological property and vice versa' (Chalmers, 1996, p.23)). But we should not be too worried about this as falsifiability is an important test of scientific postulates and the construction of theories. What we do *not* want to do is to rule out candidates by a prior definition which is too narrow.

Chalmers on Blindsight.

With this in mind, let us now turn to Chalmers' position on blindsight. He admits that his coherence principle may be put in difficulty if the claims for it are upheld. On the face of it, it would be a case of where the phenomenal and the psychological come apart; a case in which one might be said to be aware, but lack the experience as Chalmers defines it. Indeed any case of unconscious experience, or unconscious mental qualitative states (such as we might

[15] We may consider for a moment, what such additional criteria would bring to the table. Chalmers might object that neurological evidence could certainly be advanced as support for claims about awareness, where this is construed as a psychological term; but it cannot be used as evidence for the presence of mental qualitative states. There are several ways of expressing this point, but according to Chalmers, no amount of fine-grained neurological evidence (eg areas v4-8 firing) explains the what-it-is-like subjective aspect of seeing a yellow banana. We can establish a regular correlation, but not a causal explanatory link. However, we can make an argument to the best explanation well enough. If a neurological area fires invariably when a yellow banana is experienced, then the best explanation for a neurological event of that kind, is the experience itself, and the argument that the experience just is the neurological event is stronger.

find in *numb-touch*, or DF's visual form agnosia) would pull the two apart so a way to deal with them needs to be found. Not only do they challenge the coherence principle, but it also presents difficulties for the concepts which he claims are combined by the principle.

We might make this a little clearer. In *The Conscious Mind* ch 6, in which the discussion of blindsight figures, Chalmers is arguing for a systematic relationship between phenomenal consciousness as he understands it, and the psychological manifestations of it on the other (and ultimately the physical processes that underpin the psychological). Phenomenal consciousness is irreducible, but naturally supervenient upon the physical, in what he expects to be a regular, law-like relationship. Establishing such relationships, he contends, would allow a science of consciousness to be developed. While there is a limit to what we can know, on a third person basis, about consciousness, the coherence between the phenomenal and the psychological would constitute an epistemic lever, leading from knowledge about physical processes to knowledge about experience (Chalmers, 1996, p.237).

All of this is something which a naturalist about mind would have some sympathy with, and it does not seem as if a critic should have enthusiasm for dismantling the principle of coherence between the phenomenal and the psychological here. But it is important to remind ourselves the wider picture being promoted by Chalmers. He is promoting a principle bridging the phenomenal with the psychological, with the phenomenal seen as otherwise unfathomable. The position favoured is one which ultimately sees no need for a bridge, as such, between the phenomenal world with a fundamentally irreducible character, and psychology and ultimately a science of physical processes. We shall argue that the dissociations make a case for a closer relationship between these than the metaphor of a bridge would contend.

But let us return to Chalmers' treatment of the conditions. He makes only a brief reference to blindsight and similar dissociations, and the case he makes against them can be summarised as follows.

A. Firstly, he makes the point that blindsight may not be what those interested in it, claim for it: 'it is not obvious that there is no experience in these cases […] perhaps there is faint experience'. Again: 'It is possible that blindsight subjects have a weak sort of experience, [which corresponds to] a weak sort of awareness'. He returns to this theme in a discussion of Block's p-conscious/a-conscious distinction: [In cases of superblindsight] 'there is reason to believe that phenomenal consciousness is actually

present'. These are claims that experience is after all present in blindsight. The main thrust of them is to contend that the dissociations somehow involve *degraded* forms of experience. These are empirical questions which it ought to be possible to resolve.

B. Blindsight cases are 'far from being standard cases of awareness'. Subjects with the condition seem to lack the usual sorts of access to the information at hand: 'their access is curiously indirect – not straightforwardly available for verbal report, and the deliberate control of behaviour. [Information] is only made available by unusual methods such as prompting or forced choice.' So the condition is one which does not meet the criteria of awareness he has set up for judgement: information is not directly available for global control, that is, for a range of different behaviours.
C. Thirdly, 'the description of the situation is somewhat underdetermined, given our lack of access to the facts of the matter.' In other words, there is a number of interpretations which can be put on the cases, and nothing definitive can be said, since we do not have ways of pursuing the matter further.
D. Such cases can never 'damage the principle of coherence; they can only bolster and refine it' since 'any conclusions about the presence or absence of consciousness in these cases are drawn precisely on functional grounds. […] the evidence for unusual states of consciousness usually relies entirely on evidence for unusual states of awareness'. This is a claim that the empirical investigation of dissociations cannot establish that conscious experience dissociates from awareness – the principle of coherence would therefore be immune to any evidence accrued. But it goes further – the empirical study of dissociations (whatever is found) will always have the effect of supporting the position taken by Chalmers (All references: Chalmers, 1996, pp.227-8).

To evaluate the case here, we need to go further into the blindsight cases (and other dissociations). We shall look at each claim in turn.

A. The claim that there is experience, in Blindsight.

In a sense, this is a claim that Blindsight is in a sense not real, and this must be dealt with first. Has blindsight been over-described – might the effect be caused by a badly degraded form of perception in which there is perhaps some experience? (This is a line of attack which

would have to embrace a collection of dissociation cases, necessarily different in each case). For blindsight, one way of interpreting this claim is as follows:

i) It may be that islands of striate tissue somehow survive in the cortices of DB, GY and others, and we cannot be sure of the precise extent of damage until a full autopsy is possible (as far as I am aware, the 2 most studied cases are still alive). It is not clear, given the surprising plasticity of the brain over time, that neighbouring tissue could not take on some of the role of striate cortex, and be a means by which some normal function could be responsible for the effect.

ii) There is a claim that light scatter within the retina could also produce a blindsight response. Recall that the blind field is not total in DB and GY, and light, bouncing around within the retina could hit light receptor cells that feed normally operating areas of V1, and so some kind of shadowy image is created, allowing high success rates of discrimination.

iii) Blindsight may be challenged on grounds that response bias has crept into the study, and that therefore, the reports of the subjects themselves are problematic.

Similar objections can be made against other dissociations, (similar in the sense that they question the 'reality' of the phenomenon,) but the grounds for thinking so, in each case may be ad hoc, and there may be no one single defeating objection. For the most part, the claim will be that there is some experience in each of the cases, which defeats the idea of dissociation.

However, here we will focus on blindsight as the primary, and most comprehensively studied and vigorously defended example; if the reality of this case, and its description as a form of unconscious perception can be defended, there is no need to pursue others.

i) Blindsight as the result of degraded but normal vision?

To explore this, we will borrow and reflect upon the following papers and debates: the collection of papers written by and in reply to: '*Is blindsight an effect of scattered light, spared cortex, and near threshold vision?'* (Campion, Latto and Smith, 1983). This pre-dates *The Conscious Mind*, but it explores the claim made by Chalmers in more detail, and may in any case have been part of his reasoning in writing what he does. It makes a clear case for rejecting blindsight as a dissociation in the relevant sense.

Campion *et al* argue (p.446) that: 'We cannot prove that blindsight as defined as non-striate[16] vision, does not exist, but parsimony should lead us to reject it on the present evidence', claiming that the experimental identification of blindsight is more simply accounted for by other means. They argue that not enough work has been done to eliminate the possibility that what has been interpreted as blindsight might just as likely have been produced by spared islands of striate cortex: that is, not lesioned by whatever accident befell the subjects in the first instance. Equally, it might be that 'near threshold vision' might be responsible, that success in the tests is explained by peripheral vision at the edge of an inaccurately defined scotoma, or that light scattering or spill-over within the retina and then reaching unimpaired parts of the visual process as a whole might be involved. In other words, nothing particularly special is happening in blindsight, except some form of degraded but otherwise normal visual function.

To say that the paper stirred up the debate around blindsight would be an understatement. The peer reviews and commentaries included in the paper are extensive and many. Let us try and separate out the light scatter question and deal first with their point about degraded vision, that is, the idea hinted at by Chalmers, that there is still experience in the cases.

Campion *et al* make a number of claims, but we might firstly acknowledge their general position regarding the scientific investigation of blindsight. They begin with a very general point: 'there is no way of verifying the fidelity of the report of personal experiences' (Campion *et al,* p.425). And: 'From the theoretical point of view, the unconscious aspect of blindsight is […] trivial, and from the practical point of view it is impossible to treat scientifically' (Campion *et al,* p.427). This is odd, since, if true, the rest of their paper is unnecessary. We shall return to these claims as the first part of our rebuttal. Campion *et al* go on to argue that the effect in blindsight might be accounted for by the product of degraded striate mechanisms by reference to comments made by the patients themselves, suggesting some form of experience.

[16] Striate vision is neuro-typical – visual signals are processed in the primary visual cortex (V1) before being referred out to other areas responsible for processing the colour (V4-8) or the movement (V5), or the shape of the object perceived. The role of V1 is not clear, and it might be that the conscious registration of stimuli is mediated by feedback from secondary sources to V1, rather than having been referred there in the first instance. Either way, in cases of Blindsight, the V1 area has been lost (either surgically removed or damaged), and this has the effect of inducing cortical blindness. But there are many neural pathways (up to 15) by which information reaches these areas, without first having been registered in striate cortex V1, so non-striate vision is understood here as being successful visual experience of a (unconscious) kind, without the involvement of striate cortex.

At this point we can make an important distinction. There are some subjects who report no experience whatsoever. Subsequent researchers have called this condition Type 1 blindsight. Then there is a claim that some subjects do have some kind of experience accompanying their successful discriminations. These have been bracketed off as Type 2 blindsight.

Weiskrantz' most famous and most studied subject, DB falls into the first category. Weiskrantz deals with the question of whether blindsight can be degraded normal vision in Weiskrantz (2009 p.208 *et seq*) by focussing on this strongest case. DB's response was always emphatic: systematically, throughout all the studies done on DB, for all stimuli shown to his blind field, he reported: 'Nothing there; all guesswork', (and yet scored successful performances between 90 and 100%).

To pursue the matter further, DB was asked to compare the periphery of his intact field, and to say whether stimuli therein at all resembled those in his impaired field. Weiskrantz records: 'The answer was always quite firmly negative, and he remarked: "No, even way out there in that part of my field, I *see*". This suggests DB was quite capable of acknowledging experience in the intact but peripheral field, and its absence in the blind field.

Indeed, at no stage does DB report any sensation accompanying the stimuli in his blindfield. He is quite emphatic that what is happening to him is not like the kind of experiences we would imagine someone with degraded but normal vision would have, and there are further experimental conclusions which support the claim that blindsight in DB's case is real.

Campion *et al* concede this, but say of DB: while 'his most common response was that he saw nothing at all, if *pressed*, he might say that he perhaps had a feeling that the stimulus was either pointing this or that way, or was 'smooth' (the O) or 'jagged' (the X)' (Campion *et al,* p.435, their emphasis). It may be argued that this amounts to an experience, but he was highly reluctant to think of it as such, and has the hallmarks of something like an intuition. Indeed, on the occasions DB had such intuitions, Weiskrantz noted that they were sometimes distracting, and misled him into making choices. In one paradigm, testing for DB's ability to detect rapidly moving stimuli, DB reported 'seeing' some peculiar waves. Weiskrantz adjusted the experiment, changing the luminance contrast of the stimulus, so as to eliminate the 'feelings', or intuitions. DB's discrimination ability remained as high as before. Type 1 cases then would appear to be (on the basis of the subjects' own reports) the least

easily dismissed as some sort of degraded striate vision - they would appear to be genuine cases.

It may be complained that we cannot rule out the survival of some small parts of striate cortex, - at the point of writing, we still lack a case of demonstrable blindsight, and with an autopsy showing complete destruction of the striate cortex, and complete preservation of extrastriate connections. In the case of DB, we are limited by the circumstances of his injury. The correction of his neurological problem (which created his blindsight in the first place) necessitated the insertion of metal clips which has subsequently meant that the extent of the destruction of his striate cortex could not be confirmed by modern MRI scanning.

But we need not rest the case on DB's condition alone. A further study of TN and affective blindsight in his case adds to the argument here. TN was tested for affective blindsight in Andino *et al,* 2008. It was established that TN has total cortical blindness, and describes his visual experience as complete darkness. He was unable to detect strong sunlight coming through his window on a clear and bright day. Nevertheless, he was significantly above chance in discriminating between angry/happy, sad/happy and fearful/happy faces in tests for his affective blindsight. In TN's case it has been possible to confirm by MRI scan the absence of both his left and right visual cortices, and also identify, using an EEG, some of the higher brain areas correlating to the processing of facial images. These correspond to the same areas active in visual processing in non-blindsighted people.

We can argue, therefore, that there is some neurological support for the condition, although a neo-cartesian would argue that this confirms only that there are quite specific neural correlates of consciousness.

Let us turn to type 2 cases.

Type 2 Blindsight cases.

Here, on the face of it, the evidence is more helpful for Campion *et al'*s (and, of course, Chalmers') case. In some cases of blindsight, the subject offers a commentary suggesting a limited, or degraded form of awareness alongside their successful discriminations in their blindfields. These comments are typically something like 'seeing a black object moving across a black background', or a 'feeling' that something was presented.

Similarly, in another study (Richards 1973 collected in the Campion paper) three subjects 'did report occasional visual sensations such as a "pin-prick", a "prickling", or "gunfire at a distance"' (Richards p.436)[17]. In Campion *et al*'s own study, they record that a subject NG felt that his blindfield stimuli were similar to "a car's headlights passing behind me at night" (Campion *et al,* p.440).

This kind of response find some support in later studies, in work by Stoerig and Barth (2001) who looked into the 'feelings' which Weiskrantz' patient GY described when studied, to see if they were perhaps perceptual in some diminished way. GY had described his case as being 'similar to that of a normally sighted man, who, with his eyes shut against sunlight, can perceive the direction of motion of a hand waved in front of him' (originally in Beckers and Zeki, 1995). (GY commented: 'You are kind of aware that something has happened, but you don't quite see it' (Weiskrantz, 1997, p.145). He goes on: 'I mean; I can't describe something I don't understand myself'.) Stoerig and Barth conclude that GY's vision is degraded and his experiences are basically visual in nature.

Chalmers might say that all these comments are an endorsement of his claim that; 'Blindsight patients have a weak sort of [phenomenal] experience, in which case one might also want to say they have a weak sort of awareness' (Chalmers, 1996, p.227). In which case, blindsight studies confirm his position, that the phenomenon is 'compatible with the coherence between consciousness and awareness' as Chalmers would define them. Of course, Chalmers does not provide a case for his argument that blindsight patients have a weak kind of experience, (or specify exactly what he means with the term: 'weakened experience') so we can attack only those philosophers and other investigators that do.

It is to this that we now turn. There are powerful counters to this conclusion, and to the case made largely by Campion *et al* and others.

Resisting the idea that Blindsight is degraded vision.

Let us deal first with the stance that Campion *et al* take on the philosophy at the centre of the blindsight issue. Their caution against subjects' subjective reports, their claims (Campion *et*

[17] Sic: One might observe *en passant* that gunfire and pin-pricks are hardly *visual* sensations.

al p.425, quoting Corso 1967, p.9) that 'there is no way of verifying the fidelity of the report of personal experience', and that the scientific investigation of blindsight is impossible, would seem to be partly methodological, and partly an over-enthusiastic endorsement of the explanatory gap. The comments sit oddly with the rest of the paper, which does deal with subjective reports, and indeed the case above is largely made on the basis of the reports made by subjects. To make a rather obvious point: they may not cleave to an equivocal position on the subjects' reports, while also making strong claims about what the subjects say of their experience.

There is an additional point that, if we cannot rely on subjects' reports, we are left with no ability to define the extent and location of scotoma. We shall give it no further attention, except to say that it takes an unwarrantedly sceptical view of the first-person accounts of subjects, when others, even Chalmers, would concede that the best authority for the inner states of subjects are the subjects themselves.

In what follows, we shall argue that Blindsight is more than degraded vision. There are several routes to follow. We shall first consider a paradigm offered by Weiskrantz, to rule out the interpretation. In Weiskrantz (1999) he describes a study in which he compared DB's blindsight performance in his blind field, with a part of DB's sighted field, - an area that was poor in quality, and produced very fuzzy images – a spared amblyopic crescent in the left half-field. He establishes that: 'Measured visual acuity (properly: discriminative capacity) was in fact poorer in that region of the half field than it was in the scotoma'. In Weiskrantz' own words: 'The intact field seems to be biased towards object identification, and the blindsight field towards stimulus identification, and it is hard to explain this on a simple degradation basis' (Weiskrantz, 1999, p.40). This is worth some emphasis. When comparing the performance of the intact field with that of the blind field, as measured by forced choice guessing, the intact field was better at discriminating visual shape (a triangle over a square, say), but there were conditions when the blind field was better at detecting a stimulus (say a light shone very briefly). The idea that blindsight is perhaps some sort of degraded vision might explain the success in many paradigms, but it is a stretch to think it can explain a better performance in stimulus detection in the blind field over the sighted field.

A second route invites us to turn to the reports made by Type 2 blindsight cases. The question we need to explore here is whether the kind of experience is of the kind which

would support Chalmers' position. Do the comments that subjects make actually support the view that there is experience in these cases? Kentridge (2015) observes that one route available is a broad definition: 'One approach to answering this question is simply to assert that any phenomenal experience elicited by a visual stimulus will be, by definition, a visual experience.' Adopting this rule would appear to settle the matter by definition. A report of an experience of any kind, even a vague feeling, such as we find in cases of Type 2, would diminish the challenge which blindsight poses to Chalmers' coherence thesis.

Most investigators do not take this line, but Overgaard and Grunbaum (2011) come close: 'A visual process is one in which a subject at some level reacts to something visual. […] It should follow that if there is any kind of preserved conscious experience in blindsight subjects caused by visual stimuli […] those experiences should be conceived of as visual. Even in the case that a subject has experiences usually associated with tactile or other kind of perception' (Overgaard and Grunbaum, 2011). They do concede: 'we have not argued that perception in the absence of conscious experience cannot, as a matter of principle take place. It remains an empirical question […]', so there is no principled road-block to the question, and one we can explore.

However, our preferred standard here is, at the very least, that the kind of experience elicited shares the same modality as the modality in which the stimulus was received, or we would be faced with an odd situation when considering synaesthesia. A colour-sound synaesthete might, on seeing the colour blue, have a concomitant experience of F#; If we followed Kentridge's let-out for the blindsight-sceptic, the phenomenal experience of F#, elicited by the colour blue would be, by the definition given, a visual experience. But how could hearing F# be a visual experience?

Our position then is this: for it to be the case that there is, in blindsight, some experience, consistent with the idea that it is a case of degraded vision, the experience in question must be visual in character. The same is true *mutatis mutandis* of other dissociations involving other senses. With this in mind, let us revisit the reports of the subjects.

Helpfully for Chalmers' position, Brogaard claims 'there is a growing consensus that Type 2 blindsight is a kind of veridical visual experience' (Brogaard, 2014) on grounds that subjects 'typically report that they are aware of 'something' or have the feeling that 'something is happening' but deny being directly aware of the external object that triggered the experience.' In this respect, type 2 blindsight might provide a reason for Chalmers' block

to succeed – Brogaard argues 'type 2 blindsight subjects […] have a form of residual awareness that is positively correlated with their residual visual abilities'. The reports of subjects seem to tie together their perceptual success with some awareness – not properly visual, not standard, not easily describable, but perhaps sufficient for Chalmers to claim that there is weak experience. Can this be made to stand up?

We might note first that many of the reports speak of additional experience, insofar as they do, in non-visual modalities – 'gunfire', 'prickling', 'feeling', or a 'feeling of knowing'. In these cases it is clearly harder to sustain the view that some visual experience is involved. Even in the cases where some visual imagery seems to be involved: 'a hand moved across my face while my eyes were closed' – these sound more like metaphors, attempts to capture something not easily expressible, something not easily ascribed to any sensory modality. In fact, we could easily read the reports as suggesting that the subjects are unwilling to describe their experiences as being visual experiences. We might be more convinced that there was experience in these cases if the subjects were prone to describe some of their other, entirely normal, experiences as being accompanied by gunfire, or shadows. We might then be inclined to think of their case as one not unlike synaesthesia, and then to imagine that their blindsight gunfire is of a piece with intact-field gunfire. But of course, they do not report this.

Fiona Macpherson has also addressed the question of whether Type 2 blindsight is a case in which some kind of degraded experience is present in subjects, and dealing with much of Brogaard's position. (Macpherson 2015) Ultimately, Macpherson concludes that the available evidence comes down against the degraded-experience argument, but the route she takes provides us with another line of reasoning.

She proceeds from preliminaries in which she discusses candidate structural features of perceptual experiences. She advances several, but begins with one she advances as being held by many philosophers:

'Necessarily, perceptual experiences are conscious' (Macpherson, 2015, p.105).

It would be legitimate to suppose that that the philosophers she has in mind are neo-cartesian, and indeed this does look like the position we are attacking. She does argue that it is extraordinarily difficult to establish the existence of structural features of experience, but this is one that must be endorsed by Chalmers. Her article mostly explores the question whether

colour is a structural feature of visual experience, and how this fits with the neo-cartesian assumption that consciousness is necessarily locked into perceptual experience. However, picking up Macpherson's line of thinking, I want to advance a further structural feature of experience. This is:

> A perceptual experience of movement necessarily involves a perceptual experience of a thing moving.

We can agree with Macpherson that the neo-cartesians set up the contention that type 2 blindsight is a situation where 'limited consciousness of the stimulus in the blindfield exists'. The neo-cartesians would also argue that 'A subject might report having some form of consciousness of the left to right movement [of a stimulus] but deny having consciousness of anything else.'

The neo-cartesian case would seem to be reliant on the kind of reports included in Zeki and fftyche (1998) – examples include:

Patient 1, who 'declared he could distinguish no object, … but knew that something had moved through his blind field'.

Patient 2 reported: 'The 'moving something' had neither form nor colour. It gave him the impression of a shadow.'

Patient 3: 'The 'moving things' have no distinct shape, and the nearest approach to colour that can be attributed to them is shadowy-grey'

In each of these cases it is clear that movement is being reported. GY gives other examples, in which he describes his experience as being 'aware of 'something moving', but it appears as 'black on black', or, as 'similar to a normally sighted man who, with his eyes shut against sunlight, might perceive the direction of the motion of a hand waved in front of him'. While such reports include 'shadow' and 'hand' as things which might be considered as things which move in GY's blind field, in full context, these comments do look like metaphors for movement as such, rather than a thing moving.

The question this thesis raises is this: the neo-cartesian case would seem to be built upon the following argument:

> Necessarily, perceptual experiences are conscious.

> In the case of blindsight, the subject is conscious of movement, but not of the thing being moved. (We should surely want to say of the subject that if s/he was conscious of movement and of the thing being moved, this would hardly count as a case of blindsight at all.)
>
> Subjects' reports seem to confirm the idea that movement (and movement alone) is experienced.
>
> Therefore, the blindsight subject is conscious of a fraction of what is normally experienced by an ordinarily sighted person; s/he has a degraded experience.

This is the conclusion that neo-cartesians are driven to by the assumption of their first premise. We can however proceed from a different preliminary structure of experience, and arrive at quite different conclusions:

> The qualitative character of experience of movement necessarily involves the qualitative character of experiencing a thing moving – they are inseparable.
>
> In strange blindsight cases, the subject acknowledges the registration of some movement, but it is not possible for the subject to have the *conscious experience* of movement without the experience of the thing that is doing the moving.
>
> The blindsight performances in reporting movement are not therefore the result of degraded experience in the way required by the neo-cartesians; reports made by subjects are metaphorical and not meant to be taken reports of qualitative character.

What account should we then give of the subjects? In blindsight subjects, in the absence of V1 striate cortex, no conscious experience of their processing of movement in other visual processing areas is possible. This processing is nevertheless happening (in V5 or the Middle Temporal), and is somehow made available to areas of the brain accessible when choice-guessing is forced upon the subject. It never becomes conscious in the absence of V1, and this accounts for the subjects' reports of blindness throughout. Something must be happening to produce the 'black on black' responses, but the explanation is best left to an account of neurological function rather than the result of degraded experience necessarily conscious, as no differentiation within this is possible.

There is a further possibility: ie that the neuro-anatomical basis of Type2 may be different to that in pure, or Type1 blindsight, and so the question of degraded awareness (ie

experience, for Chalmers) in the condition (one way or the other) may be orthogonal to the issue under discussion. Type2 could be the result of some differentiation in neural damage, compared to Type1. Type2 blindsight (or Riddoch's syndrome: Brogaard 2015, Zeki and fftyche 1998) on this reading might yet be another provocative example of the kind of dissociations we are concerned with, and one which gives us a more fine-grained insight into the neurological basis of aspects of consciousness. It may be that certain connections in GY, and GR's cases are preserved, that have been lost in DB, connections which may (or may not) be responsible for the black-on-black reports given by GY.

Other studies have borne out a view of this kind. Azzopardi and Cowey (1998) used a signal-detection treatment of a study on GY to reach a conclusion that his blindsight is qualitatively distinct from near-threshold vision. 'Our results explain why patients with blindsight are apparently *more* "aware" of moving stimuli than of static stimuli. […] they also imply that blindsight is unlike normal vision near threshold." Their results would seem to support the view that there is a second visual system at work in blindsight – a conclusion reached by many of the correspondents with Campion *et al* (Puccetti, Bridgeman, Pasik and Pasik in the Campion *et al* paper)

We should be careful not to over-generalise - there is no escaping the fact that what blindsight subjects report does appear to be highly variable. All kinds of factors may be involved in this, and it is at least possible that the subjects are influenced by the expectations of experimenters, and what the subjects themselves may already have heard about their condition. However, the claim that there is a weak kind of conscious experience of the kind needed by Chalmers, is not borne out by the reports as given. Forced metaphors and quasi - visual imagery do not endorse the claim that there is conscious experience involved in the condition. Indeed, over several paradigms, GY withdrew his first intimation of awareness of movement. In 1980, he described his condition as being one in which he had a visual experience; in 1994, he described it as 'just having a feeling'. Finally conceding in 1998: 'a shadow (moving) is the nearest I can get to putting it into words so that people can understand'. When the definition of 'qualia' was given to him in Persaud and Lau (2008), and he was then asked if he ever had qualia of moving stimuli in his blind field, GY was emphatic: 'No; never'.

A final, and potent point. There is, in GY's case, an argument for saying that his type 2 blindsight is superimposed on an underlying type 1. GY gave his comments when being tested with fast moving stimuli. When he was tested for his awareness of more slowly moving stimuli, he this time reported seeing nothing at all, (during a 2AFC test concerning the movement of a light pen). He was still nevertheless above chance (90-100%) in his guesswork (Weiskrantz, 1997, p.64). In another paradigm, he hinted at some awareness when being asked to detect Gabor patches. When Weiskrantz diminished the contrast used in the patches shown, GY's responses became consistent with the reports given in Type 1 blindsight cases. Weiskrantz was clearly determined to arrange his paradigms so as to eliminate any stray experience, and was inclined to rule out experimental results where there was any suggestion of it in his subjects.

What is our local conclusion here? We have explored the suggestion that there is experience in blindsight, along the lines of the claim in Chalmers (*The Conscious Mind* p.227), and that the effect is somehow the result of spared islands of striate cortex. But on the basis of the subjects' reports, we conclude that neither type1 nor type2 blindsight count as involving experience in the sense that Chalmers needs to preserve his principle of structural coherence. But the claim that blindsight provides a lock-down case against that principle is not yet out of the woods, and we need to deal with other claims which undercut its potency

ii) Awareness, and success in discrimination - an artifact of light scatter?

There is a further sense in which there may be experience in blindsight, in the sense meant by Chalmers. That is, if there is found to be 'light scatter' in the retina during testing. This would also account for the success in discriminations in subjects. Light scatter in this sense is stray light from stimuli shown to the blindfield, finding its way to parts of the retina that correspond to the unimpaired parts of the visual field. Such light scatter could then account for the 'feeling' that something had moved or a stimulus had been presented, and create a degraded awareness on the strength of some experience.

Campion *et al* argue that none of the scientific investigations into blindsight provided adequate controls for light scatter, and they explicitly make the claim that it is 'clearly inadequate to rely upon subjects' subjective reports' (Campion *et al*, p.437). They argue that it is not enough to perform procedures in which the intact field is flooded with light (to

overwhelm the scattered light effect, without 'the accompanying psychophysics to demonstrate that such procedures are effective.'

They make use of a series of experiments conducted with three subjects. Two of their subjects, NG and BW gave descriptions of their experiences which seemed to suggest some limited form of awareness. NG was the candidate who had described his awareness accompanying the stimulus as being similar to 'a car's headlights passing behind me at night', while BW described his experience accompanying the stimulus as 'a sort of halo of light'. A third subject was dismissed as not really showing any signs of blindsight awareness or success in discrimination. Before the specific experiment testing for a light scatter effect, however, NG died, leaving only one subject. It is hard to build too much of a case on one subject, and there are obvious risks in drawing too many conclusions. However, their general point is that experimental procedures should have been taken to control for the possibility.

One test to establish the possibility of light scatter, even in normally sighted people was to make use of the blind spot, where the optic nerve leaves the retina in all cases. If stimuli presented to the blind spot can be detected, in a similar way to blindsight cases, then the case for thinking that light scatter is responsible in blindsight cases becomes stronger (See Campion *et al*, 1983, p.443). They manage to get some good results with the test – with a test of 30 presentations to the blind spot, the subjects scored 100% correctly on the question of whether a stimulus was shown or not. (The target presented was one of sufficient size to create the effect - 1° 40´ of arc, but fell to chance if the target stimulus was only 1°of arc).

However, there are problems with drawing too many conclusions from this test. Firstly, the blindspot is very much smaller than the scotomas with which blindsight subjects present, so that the possibility of light scatter into a good field, is greater - a fact admitted by Campion *et al*. Put plainly, if the diameter of the blindspot is tiny (which it is: 5°x7°) then light spill-over into a visually functional field has but a short distance to travel, and is a more likely effect. Campion et al themselves concede this: 'one cannot extrapolate directly from this paradigm to a patient with a scotoma since the blind spot covers a much smaller area, and so light scatter would be more significant'. More importantly, showing that there is a light scatter effect with the blind spot does not establish that this is the effect which creates the experience and hence awareness in blindsight (if that is what the subjects' reports attest to).

Nevertheless, their conclusions are arguably over-reaching: although they argue ‘’we cannot show that all blindsight is attributable to light scatter’, ‘we have shown the extent to which light scatter *could* be responsible for “blindsight” within the paradigms described’.

Resisting light scatter as a cause of awareness and perceptual success.

Barbur and Ruddock (peer reviewers in the Campion *et al* paper, p.448) were critical of Campion *et al*’s light scatter tests. Campion *et al* had claimed that their patient BW could derive information about the orientation of a stimulus from scattered light. Barbur and Ruddock’s patient , G, could do no such thing. There were other differences, too, and this led Barbur and Ruddock to claim that ‘the fact that there is so much variability between “blindsight” in different subjects is an indicator that it cannot be attributed to a single factor such as light scatter’ (Campion *et al*, p.449).

Weiskrantz (2008) also makes the following defence, using his patient, DB. On the question of awareness, he says of DB, that he was punctilious, and always reported any spread of light into the good field, and in tests to identify the blind field. Could he, DB, ever ‘see’ the targets? ‘No’, he replied, ‘That would have been cheating *myself*’.

On the issue of perceptual success as a result of light scatter, Weiskrantz also did a test putting results from blindfield alongside those from the natural blindspot.(Weiskrantz and Warrington in: Campion *et al* p.466). In this study, DB was shown a target stimulus of 0.5° of arc in four different places in his scotoma, together with the same stimulus shown to his natural blindspot. The results were well above chance in the scotoma, ranging between 77%-97%, and at chance in the blindspot (43%). Further, using DB’s reports, in him there is a genuine case of complete lack of awareness, accompanied by perceptual success in the scotoma, where none exists in the natural blindspot. The blindspot result tends to rule out light scatter as a possible artefact in creating the perceptual success.

We might make two further observations. One from Weiskrantz himself: in more complex tests conducted with Warrington (cited in Campion *et al,* p.465), DB was able to make discriminations between sine-wave gratings presented at different angles to his blind field, (and actually to a surprising level of acuity). The results, as Weiszkrantz suggests: ‘make any interpretation in terms of diffusion or stray light quite impossible.’ It is hard to see how light from such stimuli could stray over into sighted fields, together with information

concerning the target's eccentricity. We can add to this argument. In Weiskrantz (2009), he records a re-test of DB's blindsight after some 17 years of absence. In the re-test, DB was able to make some impressive shape-recognition discriminations – very low contrast, outline drawings of animals were shown to his blindfield, and he was able to identify them with 89% accuracy[18]. (Trevethan *et al,* 2007). DB was given no information concerning the category from which these drawings were taken. It is very hard to see how this result can be accounted for by the light-scatter theory. Complete images would have to be scattered, intact, around the retina for this to work.[19]

Finally light scatter cannot be the explicans of other forms of unconscious perception. In De Gelder *et al* (2002) we found that GY's and DB's forced-choice guesses of the emotional valence (fear/happy) of projected faces to the blind field was 87% correct. Slightly differently, Marcel (1998) was able to establish that subjects could be primed by pictures and words shown to their blind field, which then influenced their semantic interpretation of ambiguous words shown to their sighted field. These results are extremely hard to explain by the light scatter theory; complex shapes (pictures, words) are being projected, and for a successful discrimination, the shapes, faces in their entirety would have to make their way to the sighted field within the retina, and this fails to persuade.

Conclusion

We began with the suggestion that some experience accompanies the perceptual success in blindsight, and this accounts for some of the subjects' reports of awareness and their success in discriminations. The claim we are exploring here is that this awareness and success can be accounted for by light scatter into and within the retina during testing, reaching those parts of the retina corresponding to the sighted field. We acknowledge that this might also account for the reports given by subjects, attesting to some form of experience. However, we have found that the light scatter effect can be experimentally ruled out, and the effects persist. We have found that more sophisticated versions of blindsight success cannot be accounted for by light scatter, since such scatter would necessarily be diffuse, and it does not seem credible to think that whole and complex images could make such a transition to the sighted field.

[18] This represented three incorrect guesses out of 28. His response of 'horse' when an antlered deer was presented was regarded as wrong, even though there are obvious similarities in shape.
[19] Curiously, when the same images were shown to his sighted field, and to 5 control subjects with no impairment, DB's blindsight performance was far better.

iii) Blindsight and Response Bias

In addressing this question, we intrude upon a very live discussion which we will need to give more attention in a later section. At this point we can offer an outline of the issue, and some preliminary points. To put the point simply, the claim is, in the experimental paradigms investigating blindsight, safeguards have not been taken to eliminate response bias which might have crept into the study. It is suggested that the reports of the subjects may be mis-representing the situation, and subjects may be under-reporting awareness, or influenced by what they know the researchers are looking for.

Here is some of the detail of the issue, raised in the first instance by the Campion paper, and then followed up by Ian Phillips. Blindsight (and with the necessary changes, the other conditions in question here) can be understood as a dissociation between a function established as a product of a high success rate in discrimination tasks, and any acknowledgement of the conscious awareness of a stimulus given. The latter is typically established by the subject's response to the question: 'Did you see the stimulus?'(answer y/n) The former is often (though not always) established by a forced choice, or forced guess -was the stimulus coloured (or not), was it angled to the horizontal (or not), was the face happy (or not), in what is known as a 2 alternative forced choice test (2afc). Only those responses scoring 'no' with the first question are followed up with the 2afc question after. As we have already noted, the success rate in the second part of the test can be high (>90%), even though subjects began by saying 'no' to the first question. Where this is the case, we have a case of dissociation.

In this kind of paradigm, (it is claimed) a subject's judgements are the product of her perceptual sensitivity and her response criterion (signal detection theory Green and Swets 1966). In other words, whether a subject responds 'yes' on any given trial is a function not only of whether a subject is sensitive to a signal, but also her readiness and confidence in admitting it, a variable threshold which must be crossed in order for the subject to be confident of saying 'yes'.

The principal question here is known as the 'criterion problem' (Megan Peters *et al*, 2017) – just because a subject says she did not see a stimulus, or have confidence of seeing it, 'does not mean she has zero subjective experience of it. It may be that her experience fell

beneath some (potentially very) arbitrary threshold for reporting "seen". All experiments that ask participants to rate stimulus visibility […] are potentially prey to this problem.'

The criterion adopted by the subject may vary over the course of an experimental paradigm, according to a number of factors – the subject may become tired (some experiments require hundreds of responses) or they may be influenced by experimental instruction, or the expectations of investigators. They may be naturally under-confident, or that day be in a low-mood state. There may be implicit or explicit pay-offs – (quoting from Phillips 2016 on-line paper) 'to take an extreme example: observers will have a different awareness threshold when they are paid $100 each time they consciously detect a stimulus than when they are penalised $100 each time they fail to detect a stimulus'.

In consequence, subjects may adopt a conservative response criterion which might be more inclined to produce one result to the y/n preliminary – one which underestimates the extent of the awareness involved. Phillips applies this to the case of neglect already mentioned in Marshall and Halligan, in which a subject unconsciously favours a house which does not have smoke and fire coming from a window on its left side, even though she claimed not to see the feature because of left-side neglect. Phillips argues: '[The questions faced by the subject, PS] are prone to response bias and so do not convincingly establish that PS was not (dimly) conscious of the flames.'

Resisting response bias claims.

More attention to this question will be given in the chapter concerning Ian Phillips' contribution to the debate, but our response here is twofold – firstly, to defend the subjective responses of individuals, against the problem of the criterion, and secondly, to count them as only one standard of judgement, and so side-step the criterion problem. It may be that there are other ways of testing for some sensory registration that do not involve the reports of subjects, and so we factor out any claim there is response bias.

On the first question, at the heart of the problem of the criterion is the wider question of how far we should trust the subjective responses of individuals in these dissociation studies. While there might be very good reason to exercise some judiciousness in trusting the responses, the idea that we should not trust them (all of them) at all is not justified. To

mistrust all, on the basis that some might be untrustworthy, is itself a Cartesian conceit, straight from the First Meditation.

In general, such a view loads up the burden of proof too heavily against the subjects themselves, and is too dismissive of their reports. Phillips says of Weiskrantz, when he 'bluntly insists: 'I trust my subjects' reports' that this is foot-stamping, and (quoting Kentridge 2015) 'it is ultimately dangerous to rely on introspection'. But it seems to me that the blindsight cases are being asked to meet impossible standards of proof, and taking the line that: for all that people may protest that they have no experience, *they may be wrong* is not a good bedside manner. If it were indeed the case that subjects' reports are unreliable, or too subjective, then we would of course lose the ability to define scotoma and establish their limits in the first place. Underwood, in the Campion *et al* paper, (p.463) goes further: 'If we were not to ask humans about their beliefs about their internal states, then a great deal of activity in psychology and medicine is called into question'.

The second response is to look for a wider support for the absence of experience. Our position here has been that the subject must be the best judge of whether he/she is conscious of a stimulus or not. But there are ways of grounding this. Verbal reports should be understood as simply part of, or the main feature of a wider criterion, embracing objective evidence and agentive responses of other kinds. Those who feel that objective methods need to be taken into account – typically the psychologists and neuroscientists, but some philosophers too (eg Flanagan 1992) – will argue that the evidence of fMRI scans, PET scans of brain activation, percentages of correct responses in recognition or stimulus tests, reaction times to stimuli or targets are valuable in making judgements about the ascription of consciousness. We can call this the 'outer criterion'. This need not collapse into behaviourism; those who contend that outer criteria are necessary, or helpful, need only claim that these kinds of methods and evidence have an appropriate place alongside subjective criteria.

In the light of this, the scepticism shown by Phillips *et al* is especially hard to defend if there is substantial neurological evidence lined up to support the claims made by the subjects. Taking for a moment the case of blindsight in TN (de Gelder *et al*) – the study established through a variety of brain imaging assessments that TN completely lacked any functional striate cortex, (through bilateral stroke damage) and yet still showed remarkable blindsight capabilities. Here, the evidence of neurological deficit is incontestable, and

supports the reports made by TN. Yet here is Phillips' reply: 'evidence of neurological damage, even taken together with behavioural and report-based evidence, does not offer a direct route to claims about consciousness'. At which point we have to wonder what evidence would be sufficient to bring Phillips round. Only his final comment ('that more empirical and theoretical work is required before the traditional view that all seeing is conscious is taken off the table' Phillips 2016) gives grounds for thinking his position is not a Cartesian redoubt, impervious to any kind of attack. Still, the burden must surely be on him to say what work/evidence *would* provide the progress he hints may be possible.

The commitment to outer criteria in studying blindsight is consistent with other studies in neglect and agnosias, which understand the conditions as the product of differential degradation of parts of the brain, and which themselves act as evidence for modularity in brain processing. Blindsight itself is probably best understood as the product of visual information processing by other pathways and systems in the brain, without feedback loops to the (absent) V1 area – which may be an essential, but not sufficient condition of consciousness. Empirical evidence is building to support the view that, in the absence of V1, other routes to various processing areas are in play: Superior colliculus to Middle Temporal (MT) for discriminations in orientation of movement, (Barbur *et al* 1993 working with GY), the Lateral Geniculate Nucleus (LGN) to V4 for discriminations concerning colour (Stoerig and Cowey 1992), and the colliculo-pulvino-amygdalar pathway to Superior Temporal Sulcus for processing emotional stimuli such as the emotional valence of faces (Gonzales Andino 2009 working on TN).

In the light of such evidence, the problem for many working in the field is no longer to prove blindsight, but rather to use it as a tool to understand vision. The physiological results contribute to an understanding of mid-brain vision and help us to formulate the differences between cortical and sub-cortical vision. Such results however are predicated upon the absence of conscious experience which we may suppose is a product of the loss of V1 as a location for feedback information from the wider vision system.

Neurological evidence as part of the wider criterion for judging is one route, but there may be others too. Here, we can conscript one of our targets, Tim Bayne. Bayne proposes (Bayne 2010) that we consider, as evidence of perceptual success, not just their verbal reports, but also the fact that the non-verbal responses of the subjects in blindsight trials were intentional. 'It is utterly commonplace to suppose that the non-verbal behaviour of an

organism can give us evidence about its experiential life' (Bayne, 2010, p.97-8). Bayne calls this the 'agentive criterion' – the idea that reports are merely one tool among many when it comes to specifying that creature's experience.

The agentive criterion would allow that the subjects have successful access to the information received – in the case of GY, there is a quite deliberate and specific control of his arm in mimicking the movement of the stimulus shown to his blind field. His explanation of this – suggesting that he 'knows' that there has been a stimulus shown, but cannot explain it fully -looks like the deliberate control of behaviour, and the formation of some sort of belief.

Similarly, Helen (Humphrey's monkey) acted without hesitation and automatically, when a nut was thrown into her environment[20]. She did not give a subjective report, and could not offer a view of her experience, but she did spontaneously pick up nuts thrown to her, and avoid obstacles in her environment – in other words acted as an agent.

Finally, there is a relatively new line of enquiry and experimental paradigm that seems to establish that subjects unconsciously perceive stimuli, and can confirm this in a way which does not involve the problem of the criterion at all. As we have seen, this is in experiments that explore 'affective blindsight'. (Weiskrantz, 2009, p.12, Tamietto *et al*, 2009, Gonzales Andino *et al*, 2009). Recall that subjects DB and GY were tested for their ability to discriminate between fearful/happy faces presented to their blind fields, and, as before, they could do so on 2AFC tests, significantly above chance – in GY's case, especially if the faces shown were moving, rather than still photos. While this feature of the experiment is still subject to the criterion problem, it is doubtful whether the next is.

DB was also tested for his capacity for emotional contagion. This term refers to our tendency to synchronise our facial expressions with those of others. On encountering a shocked/sad/happy/angry/fearful face, there is a tendency for us all to assume the same expression. Pictures of people with such faces, or bodily expressions were presented to DB's blindfield, and contractions and relaxations of their corresponding facial muscles were measured through electromyography – specifically two muscles, one exclusively involved in frowning, (Corrugator Supercilii) and the other involved exclusively in smiling (Zygomaticus major) The study found that stimuli in the blind field nevertheless triggered contractions of the muscles to produce a mirror of the faces shown in the pictures; (interestingly, it did so

[20] https://www.youtube.com/watch?v=rDIsxwQHwt8

more quickly than the same stimuli shown to unimpaired visual fields.) The muscle contractions married up with the emotional valence of the picture and the verbal responses of the subject.

What does this show? These responses establish that at some level, neurological sub-systems processed the visual stimulus and made the information available to visuo-motor areas, and the muscles contracted or relaxed appropriately. But the responses were automatic and involuntary, and so not subject to any choice the subject could make concerning whether to respond or not, conservatively, or liberally.

In summary, and in rejecting Chalmers' and Phillips' contention that there might after all be experience behind the discrimination successes of blindsight and dissociation cases, their case seems to rest on a rejection of subjects' testimony that they have no experience in the test situations, and to claim that there may after all be some degraded experience present, guiding the subjects' responses. This contradicts the subjects themselves – an odd recourse, since Chalmers and others have been quick enough to appeal to first person reports in the grounding of qualia and phenomenal consciousness, when it suits them. For them now to be sceptical of their reports, seems selective. Our response is that the starting position with subjects is that they should be taken at their word. Not all positive results are likely to be the product of a conservative response bias, even if some are. It is illegitimate to make the leap from some to all in such cases. Lastly, we should factor in the neurological evidence, and other criteria for perceptual success and absence of awareness. Neurological studies can give us fine-grained evidence for the role of specific brain areas, and allow us to make progress with understanding what area is responsible for different functions. If we were always to claim that the testimony of subjects is to be doubted, irrespective of the evidence presented, we could make little progress in understanding the neurological basis of consciousness.

Conclusion.

Recall that Chalmers' first argument against blindsight was that: 'it is not obvious that there is no experience in these cases […] perhaps there is faint experience' (Chalmers, 1996, p.227). We have explored the possibility that there could be a degraded form of experience in the dissociation cases, that blindsight in particular could be an artefact of light scatter in the eyeball, and that the self-reports of subjects might be the product of a conservative response

criterion, and under-reporting of experience. As we remarked at the outset of this chapter, these are largely empirical questions which should not be beyond investigation. Our first conclusion was that neither Type 2 and, with stronger reason, Type 1 blindsight can be regarded as degraded vision if any credibility is attached to the accounts of the subjects themselves. This, however, is the theme which runs through this whole section – to what extent should we ground our conclusions on the testimony of the subjects themselves? At the one end (the sceptical end) of the scale, we find Campion *et al* arguing 'there is no way of verifying the fidelity of the report of personal experiences', and that therefore it is impossible to treat scientifically (Campion *et al*, 1983, pp.425 and 427). Alongside this we have Chalmers who says 'we lack an experience meter [to confirm the link between reportability and experience]' (Chalmers, 1996, pp.112, 226, 239). These positions caution against a reliance on subjects' testimony at all. We were to do so, we then find Phillips arguing that the testimony may be flawed with response bias. At the same time, other critics of the blindsight phenomenon put a great deal of emphasis on the accounts of subjects, especially where these seem to hint at some experience (as in the commentaries of those with type 2 blindsight). On this view, it is odd to claim that 'it is not obvious that there is no experience in these cases' while at the same time lamenting the 'lack of an experience meter with which to confirm or refute these hypotheses' (Chalmers, 1996, p.226-7). Equally, it seems odd, in a section of *The Conscious Mind* which is dedicated to creating a science of consciousness, to find reasons to discredit the very considerable corpus of scientific findings which have accumulated on blindsight.

A more measured approach would be to exercise care in interpreting and making use of the accounts of the subjects themselves, and in creating experimental paradigms to investigate the conditions, but ultimately to trust the subjects and to put their accounts alongside other evidence, such as non-verbal reports and neurological evidence. Where neurological studies establish the absence or destruction of specific areas in multiple cases, and even more so, where there is a double dissociation[21], we have compelling evidence of what might be the medical underpinnings of the conditions, - a causal account of the neurological deficits creating the dissociations. Resisting this approach on a priori grounds would have little to commend it.

[21] An example would be visual agnosia and optic ataxia – the ventral/dorsal stream malfunctions.

B Chalmers' second point: Blindsight cases are 'far from being standard cases of awareness'

Let us now consider the second case that Chalmers makes against blindsight: Blindsight cases are 'far from being standard cases of awareness'. Subjects with the condition seem to lack the usual sorts of access to the information at hand: 'their access is curiously indirect – not straightforwardly available for verbal report, and the deliberate control of behaviour. [Information] is only made available by unusual methods such as prompting or forced choice' (Chalmers, 1996, p.227).

How should we respond to this case against blindsight? First a prefatory comment about the terms being used. Recall that 'awareness' here is advanced by Chalmers as a psychological term, one wing of a structurally coherent relationship with phenomenal terms, for: 'where there is consciousness, there is awareness' (Chalmers, 1996, p.220). At the same point he remarks: 'My visual experience of a book is accompanied by a functional perception of a book'. It is clear from these comments that Chalmers is content to put perception and awareness together, such that, to be successful perceptually is to have awareness of one's environment. The challenge Chalmers wants to head off with his treatment of blindsight is the suggestion its supporters make that it constitutes a case of unconscious perception – awareness *without* the accompanying phenomenal character.

More specifically, and in this context, Chalmers advances his definition of awareness as something which is 'directly available for global control'. When the subject is aware, 'that information is available to bring to bear in the direction of a range of behavioural processes' (Chalmers, 1996, p.221). In this way, Chalmers defines awareness (or perception) in terms of its function in the organism's life.

This definition or something like it, first saw light of day in Block's distinction between P-consciousness and A-consciousness, (Block 1995) in which, we may imagine, Chalmers saw a correspondence between the phenomenal on the one hand and the psychological, on the other. In the article, Block himself dismisses blindsight as 'not available as a premise in reasoning, or for rational control of action, and thus saw it as rejecting the idea of unconscious perception.

More recently, as Chalmers' hand has been picked up by Phillips and others, Block has come to defend the idea of unconscious perception, and whether or not he has revised his

P-conscious/A-conscious distinction, he nevertheless does not regard blindsight and other dissociations as being easily dismissed as case of unconscious perception. In later contributions to the debate, the definition more frequently used at the outset of argument is one borrowed from Tyler Burge. We shall return to this, but for now let us make our case on Chalmers' original dismissal.

Recall that Chalmers says: it is 'only by prompting and forced choice' that information is available to control processes, and in any case, these are fewer, in the case of blindsight. 'Access is curiously indirect and not straightforwardly available for verbal report and deliberate control of behaviour'.

The first charge to make is this: Chalmers sets a very high bar with his criticism that information is available to fewer control processes. Consider what would satisfy him – suppose that information in these cases is made available to a full range of control processes. In this event, the subject concerned would be functionally like a zombie, with information reaching all control processes, and reacting in no wise differently from a neurotypical human being, but lacking the relevant inner life. All of the cases we study here, are cases in which less than the full range of control processes are involved, but just because we have a case in which information is available to a limited number of control processes, does not make the condition less interesting, or do anything to blunt the charge that what is happening in these cases is perception of a stimulus without the associated phenomenality.

We can see this clearly in the case of visual agnosia presented by Dee Fletcher, as studied by Goodale and Milner (2004). Though not strictly a standard case of blindsight, it is a dissociation not dissimilar. The visual impairment in this case stemmed from damage to the ventral stream of visual processing. DF lost the ability to distinguish the shape, or orientation of an object before her, and the ability to distinguish it from its background (or: the ability to identify edge). On the other hand, she retained an ability to see colour and surface detail, and did not lose her sensorimotor functions. This meant that while she could not identify objects before her, she nevertheless had no difficulty in positioning her hand to grasp objects presented in different orientations. She could not *tell* if a pencil was upright, but she could, when asked to take it, orient her hand to take hold of it correctly. She could not say it was a pencil, but she could say what colour it was. In this case, it would be natural to say DF lacks awareness of (some of) her visual mental states.

DF's case allowed Goodale and Milner to claim that there are two visual processing streams – the ventral, responsible for processing form and orientation, and also the conscious processing of visual information, on the one hand, and the dorsal, responsible for the unconscious control of action. It is this latter stream that is preserved in DF, which allows her deliberately to control behaviour, even though her perception of objects is absent.

Goodale and Milner record that DF's deficit does not prevent her from walking without stumble along a steep and uneven path to reach a picnic site, or making tea in a kitchen – all of which do not need prompting or involve forcing a choice upon her. Several control mechanisms would seem to be at play, and she has a deliberate control of behaviour throughout, and yet she has no awareness of specific kinds in her visual processing.

Sceptics could argue that DF is simply suffering from degraded vision, just as similar sceptics have argued that blindsight is simply degraded vision. However, support for the idea that there are two visual processing streams may be found in a mirror image case to that of DF. Rudolph Balint described a syndrome in 1909 in which a man presented with bilateral damage to his parietal lobe, or dorsal processing areas, which had no effect on him recognising objects and people, but left him with a failure to reach out and pick up the objects properly. He would often miss the object by several inches (Goodale and Milner, 2005, pp.31-2). Other cases have come to light.

Double dissociations of this kind start to make a very strong case for different, quasi-independent processing modules, handling the sets of abilities dissociated. In DF's case, brain imaging reveals that her pattern of intact and impaired visual abilities maps accurately onto the areas of brain activation in her dorsal and ventral systems. While we concede that we cannot prove that the one is causally related to the other, we do have a neurological account with considerable explanatory power; we do have a conceptual framework which can accept the notion of unconscious perception, and both stand as improvements over a reluctance to take the phenomenon seriously. Dee's access to information – the correct way to grasp the pencil - is indeed indirect: it arrives as the product of an unconscious process, but this is a reason not to dismiss it, but to take it more seriously.

Resisting Chalmers' claim with Superblindsight.

A second case, that of TN, (de Gelder *et al*, 2008) provides us with a distinct case, this time a blindsight case to meet Chalmers' points directly. It would seem to be strongly suggestive of an instance of superblindsight – 'a situation where subjects are trained to have much better access to the information in the blindfield' (as per Chalmers, 1996, p.228). Chalmers goes on: 'There are clearly conceivable cases of awareness without consciousness in the vicinity, but Block himself notes that there is no reason to believe that such cases are actual. [In those examples which come closest] there is reason to believe that phenomenal consciousness is actually present.'

Here, Chalmers hides behind Block in the claim that 'there is no reason to believe such cases are actual', but it is his defence nevertheless. We shall explore TN's case as a possible example of superblindsight. On the last claim (that phenomenal consciousness may be present), Chalmers offers no evidence at all, and moves on without further development.

The paper presented by de Gelder *et al* 2008 is the product of a short study of TN, who experienced two strokes in succession, suffering lesions which damaged his visual cortex in both hemispheres, and causing clinical blindness over his entire visual field. The neurological damage was confirmed by brain imaging techniques, and no active visual cortex could be found when TN was shown visual stimuli while undergoing imaging sessions. The neurological damage was confined to the occipital cortex in both hemispheres. The examining team record: 'Behaviourally, TN was blind across the whole visual field, [but] he did show some evidence of visuo-motor integration in spatially guided reaching', and was able to discriminate the difference between a long rod placed in horizontal or non-horizontal positions. This puts TN into a similar category to that occupied by DF, and alongside other blindsight subjects, but the really interesting feature of the case came when the team 'constructed a complex obstacle course, consisting of boxes, chairs and so on, arranged randomly along a long corridor, without any person to guide him, […] Astonishingly, he negotiated it perfectly, and never once collided with any obstacle' (de Gelder *et al*, 2008).

This would seem to be a case where TN's access to information is clearly better than those subjects who have to be prompted, and no part of TN's passage down the corridor is forced choice. It puts TN alongside Helen, Humphrey's monkey as a possible superblindsight candidate, and as such presents a more significant problem for Chalmers than any so far.

C The claim that the situation in Blindsight is under-determined.

Recall the point was:

'the description of the situation is somewhat underdetermined, given our lack of access to the facts of the matter.'

Expanded, this suggests that there is a number of ways in which we can categorise blindsight and other dissociations; further progress is likely to be limited since we do not have the means to close in on any of them since we lack any way of progressing with a third person approach. We lack the 'experience meter' to which Chalmers refers 'This process of refinement can only go so far, as we lack an experience meter with which to confirm and refute these hypotheses empirically' (1996 p.226). This remark is close in spirit to those made by Campion *et al* (1983, p.425-6) '[..] experience lies outside the field of scientific psychology'. And: 'From the practical point of view [blindsight] is impossible to treat scientifically.'

These are claims that there is no objective, scientific way that the inner subjective can be investigated, a claim which has roots in Nagel's 1974 paper.

The absence of an experience meter.

Let us discuss Chalmers' claim that we lack access to the facts of the matter. If we can establish that we are not blocked off from access to consciousness by some principled commitment, then the condition that is blindsight might be more definitively described. In spelling out his position, we find ourselves returning to the what-it-is-like formulation. In *The Conscious Mind*, Chalmers clearly fences off the phenomenological into what is felt, first hand, first person, and in appealing to the fact that we have no 'experience meter' is arguing that we have no third person access to phenomenology. We have first person access, but only to our own experience. In making these claims, Chalmers shares the same commitment to the explanatory gap arguments of Nagel (1974), Block (1978), Jackson (1982) and McGinn (1989) – arguments suggesting that we cannot get from the facts of science to the facts of experience.

Chalmers is quite clear about how far subjective reports can be used to contribute to the understandings of science in pages 215-216. While he makes a case for using subjects' reports as a useful constraint on a theory of consciousness, and arguing for plausibility in rejecting the wilder theories that can be drawn from first person accounts alone, he nevertheless maintains that a science of consciousness will always retain a speculative character not shared by other theories in most scientific domains, precisely because we cannot poke around in others' minds to measure their conscious experience.

Despite all these careful, generous concessions to plausibility etc, there is something of the doctrine of privileged access in setting consciousness outside the limits of scientific investigation in this way. Chalmers expresses interest in what science can contribute, and recognises that much can be learnt from the studying the brain's computational and representational capacities, all this comes with a reminder that the qualitative properties of the subject's experience remain untouched. The 'very feel' of experience is left out by science, no matter how deep the analysis goes.

Flanagan's third person phenomenology

However, following Flanagan (1992) we can narrow the gap between the first person and third person. In promoting his own, naturalistic theory of consciousness, Flanagan refers to Hume's remark that it was possible to have knowledge or conception of a missing shade of blue, even though we do not experience it; in a similar way, we have knowledge or conception of others' inner phenomenology, even though we do not experience it. The right kind of approach would embrace a rich first-person phenomenology, combine it with a theory of how the phenomenology connects up with behaviour, and at the same time provide a neuro-scientific realisation theory about how all kinds of mental states – conscious and non-conscious – are realised in neurology. In time, and with the addition of detail, we would expect these elements smoothly to connect together, and explain anomalies like the dissociations we are studying. In fact, the combined theory would welcome information taken from dissociations for what it can tell us about the inner structures and components of mental life, and how these are themselves connected to neurological structures and processes.

Flanagan argues for a 'third-person approach to phenomenology' in the following ways: Firstly he warns against the approach which privileges the first person, saying there are

several mistakes to avoid. The first is the belief that, 'by virtue of being the sole subject of your experiences, no-one else can imaginatively come to see the missing shade that is you.' In coming to grasp others' experiences, I need to show sensitivity in detecting what others say, do, and intend; I must deploy knowledge of a shared form of life, and become a 'mental detector'. Just as a metal detector detects metal beneath the ground, 'so too mental detectors detect invisible mental states. Good mental detectors understand a great deal about the form of life in which they live, and about behavioural regularities.' (Flanagan, 1992, p.102-3).

The second mistake to avoid is the assumption that because there is a limit here to physicalist explanation, that that means it is false. Physicalism can still be true even when there are limits to what it can explain. Heisenberg identified one limit in the 1920s – the momentum and position of fundamental particles cannot simultaneously be determined – the more we know of the first, the less we know of the second. Similar observations might be made of M theory, and the question of multiple universes. Despite the limits involved in empirical investigation in these domains, Physicalism as a general theory of the world is not undermined. Similarly, there may be limits to what we can determine of others' experience, but this does not defeat the contention that what there is, and all there is, is physical stuff, and the relations combining it.

Frank Jackson's Mary in the black and white room argument works with a similar intuition. Mary, a colour perception scientist in the well-known thought experiment is granted complete physical knowledge of what happens physically when the colour red is perceived, but according to the thought experiment, learns something new on exiting the black and white room, - she learns what-it-is-like to see red – something new, not captured by the physical description of red-perception.

For the sake of clarity, here is the argument:

- Mary (before her release from her room) knows everything physical there is to know about colour perception.
- On exiting the room, and seeing a tomato for the first time, she learns something new. She learns what it is like to see red for the first time (we can imagine her sense of surprise, perhaps.)
- If she has learned something new, then all of the physical facts are not all of the facts about colour experience.
- Therefore physicalism is false in contending that physical facts alone are sufficient.

Flanagan contends that the argument is easily defeasible, since it opens with the contention that Mary has complete physical knowledge of colour perception. Flanagan argues that actually she does not. He insists that the experience of seeing red is a physical event, and she does not have this knowledge in the situation described. A full physicalist account of red-seeing would contain information about the sensations Mary would have on seeing the tomato. 'Jackson has made [Mary] extremely smart. [But] he can't have Mary both so smart she knows all that he says she knows, and so stupid that she will be so surprised upon seeing that first tomato' (Flanagan, 1992, p.100).

We can amplify and add to Flanagan's point. One of the reasons the story has such an appeal is that, because the room is monochromatic, there are so few points from which an imagined red-experience can be extrapolated. The story loses a great deal of its appeal if we allow that the room is populated by colours but not including red, and so contains yellow and oranges, and purples and blue. Mary can advance to a third person phenomenology – a sense of what-it-is-like to have a red-experience without actually doing so – much more easily in such circumstances. She can extrapolate to red from the shades approximating to it. In this way we diminish the sense of surprise when she sees the tomato. Our own insights into other people's phenomenology come with such extrapolations, if we know someone well, and share a form of life. In some circumstances it might even be possible to know the inner states of a person better than they do themselves. Ordinarily, people have special authority about their inner states, but are not infallible, and outsiders with good knowledge of their friends, a sense of detachment, not distracted by the stresses and preoccupations of the other, can see the truth of a situation better than their friend. 'You're right; I just don't feel the same way about him anymore.' 'You're right; I am not valued or supported by my colleagues' (Flanagan, 1992, p.92).

The reply could be: 'Ah yes, you can tell me *what* my inner state really is, but you can't know what the *character* of that state is – what it is like. The specific sense upon realising that work colleagues have not been supportive is not something that can be captured naturalistically, but we must allow that we could approximate to it by referring to similar experiences of our own.

The third person approach to what it is like to be a bat.

When a form of life is not shared, the exercise becomes more difficult, but it is not impossible. We can go a long way to extrapolating to the inner life of a bat for example, even though we employ different sensory mechanisms. This was addressed by Akins in her contribution to a book on Dennett. (ed Dahlbom, 1993) Nagel posed his question: 'what is it like to be a bat?' anticipating the conclusion that we cannot know, not only the bat's inner life, but that of others in general. Akins, however, argues that we can make a good approximation of the inner life of a bat, and *a fortiori*, those of others like ourselves.

Her richly provocative articles have more to say about the neurophysiology and behaviour of bats than we have space for here, but a brief summary of her position is that we can work back from behaviour and physiology and imagine how things would be for you, if you were in that subject's situation – you can empathetically project yourself into that subject's point of view. Let us see how this can work:

If humans are lucky in having (during daylight) an independent source of light – the sun – which makes it possible for objects to be readily located by our eyes, bats lack this permanent possibility of detection, and must produce its own signals, and have them bounce back from objects in its world. The advantages of light signals are many, and include the possibility of detecting small objects, because light has short wavelengths, and signals may be focussed by lenses and small apertures. Sound, however has problems, including a high absorption rate, and some materials and surfaces scatter sound (in the way stealth fighters avoid detection by radar by having surfaces that disperse outgoing signals). It is not a medium that makes it easy to focus; distance and size of object may be detectable, and signals can be processed to establish the fact that something the size of a moth is at twelve o'clock high, but otherwise an object is indeterminate. The one important advantage of course, is that echolocation works just as well in the dark, so the bat has exclusive access to prey not available to competitors at night.

How does this help to answer the question about the experience of the bat? The conclusions that can be drawn might be limited, but in keeping with Flanagan's natural method, the experience of the bat is not impenetrable. The processing properties of the bat's auditory system, and the nature of sound suggests that what the bat hears is a much poorer representation compared to that constructed on the basis of our visual mechanisms. The auditory picture it constructs on the basis of signals received back from objects is, in all

likelihood, poorly focussed, highly partial in the sense that the bat will probably be less interested in creating a reconstruction and permanent memory of the bigger features of its environment, and more interested in the moth-sized signals that move rapidly, and flutter. The 'picture' will be rebuilt repeatedly and rapidly especially as the bat approaches its prey. There are many details of the bat's cortical arrangements and their lack of complexity which suggest that the bat does 'not possess the necessary building blocks for complex spatial representations' (Akins 1993 (a), p.150), but it may not need to. The bat is purely interested in its prey, and avoiding objects in its flight path. All of this would suggest that the complexity of the bat's experience would be much reduced compared to ours, but there is here nevertheless an attempt to describe in representational terms, what it is like to be a bat. The bat's representational mechanism would lack the definition of reconstructions put together by the processing of visual information; its auditory 'picture' would be very different from ours, but it is doing what humans do for the same reasons – building a model of its environment.

As Akins concedes, this story is one which does not address the central question that Nagel (and we may suppose, Chalmers) raises – the phenomenal aspect of experience – the 'what-it-is-like' nature of experience, which is the point at issue. It is projecting an account of the inner life of the bat in representational terms, but not phenomenal. Chalmers and Nagel can reply: As good as we can become at empathetically projecting ourselves into the inner life of others, it might still be possible that you, my neurological equivalent, still see the Mediterranean Sea as red where I see it as blue, or have nothing phenomenologically happening where I have a whiff of leonine quality when we both imagine the lion. As good as we can become at using the various branches of science to arrive at a representational account of what is happening in perception, this still misses its qualitative aspect.

Akins, however, asks a telling question: 'In order to justify the essential mysteriousness of consciousness, the theory (put forward by Nagel) must show that the mystery is, in principle, one that science could never diffuse.' (1993 (a), p.127) It must give us good reasons to believe that there are some facts about conscious experience that will remain inaccessible to scientific explanation. The relevant sciences: neurophysiology, neuroanatomy etc are still in their infancy (a position not a lot further advanced as it was in 1993) and so cannot yet be expected to demonstrate their own explanatory inadequacy. In the absence of any obvious shortcomings, the matter could be settled *a priori*, but here again we can be sceptical. We can make no progress in discriminating between the representational and

phenomenal properties of experience on the basis of introspection, and in consequence, we cannot know, *a priori*, what insights will result from empirical investigation.

In the light of this, we can return to the suspicion that science has something to contribute – in developing an objective phenomenology - allowing conclusions to be drawn from neuro-science, and on the basis of this, informing our empathetic projections. If it turns out that that 'what-it-is-like' boils away to very little or nothing, then the neurophysiological account above, aimed at capturing the representational picture-building may be what it means to give an account of experience.

Of course, this returns us to the spirit of Flanagan's third person phenomenology, which aims to give science its due.

Conclusion

Recall that, according to Chalmers, the absence of an Experience Meter prevents an access to the facts of the matter, and allows that blindsight and the dissociations may have multiple interpretations. The question of access to the facts of experience depends very much on whether we subscribe to the explanatory gap position first introduced by Nagel, and whether this commits us to a principled divide between those facts about consciousness that are accessible to science, and those that are not. Akins demands that Nagel's and Chalmers' thinking on this identifies the reasons why there are some facts about conscious experience that are inaccessible to science. The jury is out on the question, and until and unless it can be shown that science has no tractability, there must be a presumption that empathetic projection, observation of behaviour in psychophysical tests and the accumulating detail from neuroscience is a route into others' experience. The claims (in Campion *et al*, for example, with which we have imagined Chalmers would have sympathy) that 'experience lies outside the field of scientific psychology' must be countered by the reply made by Judith Economos in the peer review section: 'What in the world cannot be investigated scientifically? Only something supernatural, by definition'. (Campion *et al* 1983, p.452) Going on: 'I trust it was only laziness or contempt for philosophy that led them to take such a heavy philosophical position.'

Equally, Chalmers must not so easily dismiss the interpretations of blindsight, and other dissociations on the grounds that we do not have the kind of access which his own conception of experience in any case makes impossible.

D Blindsight can never damage the principle of coherence.

Chalmers' final case against blindsight in his dismissal is as follows:

> Such cases (of dissociation) can never 'damage the principle of coherence; they can only bolster and refine it' (Chalmers, 1996, p.227) since 'any conclusions about the presence or absence of consciousness in these cases are drawn precisely on functional grounds. […] the evidence for unusual states of consciousness usually relies entirely on evidence for unusual states of awareness'.

Let us unpack this claim and what it amounts to. It reads like a claim that the empirical investigations of dissociations cannot have any contribution to our understanding of consciousness – the principle of coherence which he has identified, is immune to any evidence accrued. It seems also that the empirical evidence (whatever is found) will have the effect of supporting his position.

The first point is this. We remind ourselves that Chalmers is arguing for a coherence between phenomenology and psychology, such that, in normal cases, entirely uncomplicated cases of phenomenology (say- the qualitative state of tasting coffee) correspond to uncomplicated cases of psychology – the associated awareness involved. He repudiates a reading of blindsight and dissociations as cases in which consciousness/phenomenology comes apart from the associated psychological terms identified by their functional role. It is not possible to have awareness (of a stimulus) without an associated qualitative state. Hence the repeated claim that (some degraded) phenomenal consciousness is present '[…] in these cases, perhaps there is a faint experience that bears an unusual relation to verbal report' (Chalmers, 1996, p.227). Or, (Discussing Helen and DB on page 228) 'there is reason to believe that phenomenal consciousness is actually present.'

As we see in these quotes, Chalmers maintains that unusual states of awareness seem to go hand in hand with unusual states of consciousness, and so the principle of coherence is not damaged. There can be no dissociations.

However, it is our claim that dissociations are genuine cases in which the principle of coherence breaks down: they are the coming-apart of phenomenality and awareness. The repeated use of the term: 'unusual' superficially suggests coherence, but in point of fact, what is unusual is quite different in each case. In Blindsight, consciousness of the stimulus is unusual in the sense that it appears to be absent, but awareness is only unusual in the sense that its registration may need to be prompted somehow, or comes as unexpected to the individual concerned, or that the functional role is preserved without accompanying phenomenality.

Secondly, the claim: 'any conclusions about the presence or absence of consciousness in these cases are drawn precisely on functional grounds' is again suspect. (On the face of it, this is an odd thing for Chalmers to maintain, since he has held that the route into an individual's consciousness is privileged to the first person). We hold, however, that we can certainly draw conclusions about the absence of consciousness on the basis of the testimony of subjects. It may be that Chalmers is arguing that (say) we identify the extent and position of a subject's blindfield by functional means, and so claims about blindsight are open to the criticism that they are not definitively or objectively established, but this seems to have a feeling of desperation about it. We certainly cannot allow this argument to rule out the possibility of dissociations in principle. If the functional deficiency corresponds to a precise area of neurological damage, in a repeated number of cases, we have further evidence to go on, and an argument to best explanation becomes compelling.

Once again, we can conclude that Chalmers has not given due weight to blindsight, and the other dissociations. Far from bolstering and supporting the coherence principle, dissociations are just that: in being cases where function comes apart from the phenomenal, they present a real challenge to the principles to which Chalmers is committed.

Chapter 4

Tim Bayne and The Unity Contention

Prefatory note: on the Mereological principle.

Let us turn now to a different claim made by philosophers in the neo-cartesian stable. It is important to deal with this claim, as it constitutes an obstacle to the kind of alternative theory we wish to advocate. Before we do so, a specific problem needs to be dealt with.

The following chapter will deal with the challenge posed by the dissociations to the Unity of Consciousness. In it we shall make reference to brains split by the division of the connecting tissue between them – the corpus callosum. We shall also have something to say about alien hand syndrome and other conditions. In the course of this chapter, we shall find that the description of these conditions by some neurologists and experimenters may involve the attribution of psychological predicates (such as 'has a belief', or 'has the thought that') to the brain, and indeed to parts of it – hemispheres, in the case of the split brain cases, or even to smaller sub-systems involved in the processing of information (Zeki and Micro-consciousnesses).

It is appropriate, before we begin, to anticipate and confront an issue which has a bearing here. Hacker and Bennett have directed a substantial part of their book, *Philosophical Foundations of Neuroscience,* (2003) against what they describe as the mereological fallacy – ie the fallacy of attributing psychological predicates to something other than a person, or whole, undivided human being. The authors specifically address the split brain cases and blindsight in their work. Adverting to Kenny's 1971 article: 'Homunculus Fallacy' they concede that it may not strictly speaking be a fallacy to ascribe psychological predicates to the brain or its parts, since it is not in itself a form of invalid reasoning, but it can lead to fallacies, confusions and creates nonsensical descriptions (Bennett and Hacker, 2003, p.73 fn13).

This present work will acknowledge the argument as they make it, but will argue that the case raised by the dissociations against the unity of consciousness should not be blunted on this view by their insistence on an ordinary language description of what is going on in the conditions. If, in the end, we persist with the descriptions of the dissociations given by the neurologists or psychologists, we shall have good reasons for doing so. It may be possible to

re-describe the dissociations of function in such a way that meets their objection, yet presents a case against the claims for a Unity of Consciousness. Bennett and Hacker themselves allow an important counter-argument against their position, suggesting that neuroscientists are not precluded from using verbs such as 'to perceive', 'to believe' etc 'in new ways according to conditions other than the received conditions of their use, as long as they can explain what these new uses mean' (Bennett and Hacker, 2003, p.384). Short of any re-description that might satisfy their objections, we may have to make a case for a conceptual change.

In this prefatory section, however, I wish to outline Hacker and Bennett's objections, to identify what can be rescued from their criticism of Pat Churchland's view that the concepts involved in the area need a thorough review.

Hacker and Bennett make the point that it was a characteristic feature of Cartesian Dualism to ascribe psychological predicates to something other than the whole human being. In Descartes' case, it was to the mind, and only derivatively to the whole person, that psychological predicates were ascribed (Bennett and Hacker, 2003, p.72). They observe that neuroscientists and others are increasingly drawn to talk in this way about the brain, or even its parts.

Against this tendency, they advance the mereological principle – which they describe as being a logical principle, and therefore not amenable to empirical or experimental confirmation or disconfirmation. It is a convention that determines what does or does not make sense. Thus, the idea of attributing knowledge, or belief or agency, visual sensation to each hemisphere is ruled out (as, one might say, a category mistake). 'It is senseless to speak of the right hemisphere observing the actions of the left hand. […] neither hemisphere can observe, [nor] does the left hand perform any actions' (Bennett and Hacker, 2003, p.392).

That there is a temptation to speak in such ways is undeniable. We have already seen Weiskrantz do so. Other neuroscientists and psychologists, even some self-described philosophers, do so. Sperry wrote of split brain patients as having two separate minds, that is, two separate spheres of consciousness. John Eccles, a Nobel laureat for work in physiology, commented, when wrestling with the split brain question, that the right hemisphere can't know because only the left hemisphere can have thoughts or knowledge.

Puccetti speaks of the left hemisphere as being unaware of the other, which considers mind therefore to be unitary, whereas the right hemisphere has 'known the true state of affairs from a very tender age. It has known this because beginning at age two or three it heard

speech emanating from the common body' (Puccetti 1981, p.97). McGilchrist, writes extensively about the functional differences between the two hemispheres, and comments: 'I do think a hemisphere can have a will' (McGilchrist, 2009, p.215).

Bayne, our target in this chapter, and, as we shall see, rather characteristically, keeps the option open: 'Experiences are properties of subjects of experience [..] Talk of 'the left hemisphere doing this' or the 'right hemisphere knowing that' is tempting – indeed I myself have engaged in it on occasions – but it needs to be taken with a large pinch of salt' (Bayne, 2010, p.204). He concedes that his unity thesis is put under a lot of pressure by the split brain cases, in which the best account of their condition might be to accept that a commissurotomy involves the multiplication of subjects of experience. It is the object of this chapter to explore this possibility.

If so many speak in this way, it may be indicative of the claim that this is the best way of making sense of the cases. In fact, to attribute a different psychology to each hemisphere may be the only way we can make sense of something like the 'alien hand syndrome' we shall discuss in this chapter. One of the criteria for attributing psychology to a hemisphere might be whether the hemispheres act with separate discriminable agency, and this seems to be happening in this case. But let us examine the Bennett/Hacker case in more detail.

They give an account of the split brain cases as follows. They concede that what has been discovered by experiments on split brain patients[22] 'is a very strange dissociation of functions that are normally intimately associated' (Bennett and Hacker, p.391). We note their use of the word 'strange' here. Such a word alerts us to expect an account which properly accommodates this strangeness. They go on to say, however, that there is no sense in which one hemisphere of a split brain see things that another hemisphere does not, nor that one hemisphere knows what another does not. Rather 'the transmission of neural signals across the corpus callosum, which is a necessary condition of the person knowing […] has been prevented by the commissurotomy' (Bennett and Hacker, p.393). The patient may still respond to what is visually presented to him [by pointing to objects] even though he does not know why he thus points. 'In short, everything that Sperry and Gazzaniga have discovered can be adequately described without transgressing the mereological principle […]' (Bennett and Hacker, 2003, p.393).

[22] For further details see the section: 'The Split Brain Syndrome' p.134 et seq below.

This is of course, *an* account of the experiment we shall consider in the section to come, but it does not capture the strangeness of the result. As we shall see in the studies that follow, the subject gives different responses to questions in the experiments – (eg what was the object shown?) The left (the verbal) hemisphere has no access to the visual information presented to the right and so because of the commissurotomy does not know what was presented to the right, and says so, or confabulates. The experiment can similarly establish that the right hemisphere has no access to the visual stimulus shown to the left, and cannot identify the object (although it would do so in a non-verbal way). These negative responses are in keeping with the account Bennett and Hacker give. The subject as a whole person fails to have knowledge, because the commissurotomy has interrupted the passage of neural information. But the experiment has positive responses too – each hemisphere has a correct response unavailable to the other, and these responses are best accounted for by allowing that the two hemispheres each have different access to different visual information. What is strange is not what the subject *doesn't* know, but what is to be found in, or concluded from the correct but differentiable responses of each of the hemispheres.

Bennett and Hacker also give an account of blindsight in ways that make it quite clear that the characterisation of blindsight as unconscious perception is ruled out. Thus: 'The conceptual apparatus Weiskrantz invokes to describe the phenomena is defective. [Once cleared up] we must shed the description of blindsight patients as people who have sensory experience of visible objects within their scotoma, but who are unable to apprehend that they do' (Bennett and Hacker, 2003, p.395). We are presented with the question whether subjects like DB or GY can or cannot see in their scotoma, 'but this is precisely the question that cannot be answered'. They go on: '[the conceptual conflict that an insistence on using ordinary language creates means] that one can neither say that the patient sees objects in the scotoma, nor that he does not' (Bennett and Hacker, 2003, p.396). It has been our contention throughout that DB and GY *do* see objects in their scotoma, with a great deal of accuracy, and detail in the sense that they respond accurately as to the detail of the stimuli shown.

Whether they intended to or not, it is evident that this approach lends weight to the neo-cartesian view of consciousness which seems to reject the notion of 'unconscious perception'. This seems to be evident from their comment: '[In the behaviour of a blindsighted person] we resort to paradoxical descriptions such as 'blind-sight' or 'unconscious awareness' (Bennett and Hacker, 2003, p.396) They take the view that there is

nothing paradoxical about blindsight – there is only an indication that a concept is inapplicable.

As we have made clear elsewhere, this thesis is inclined to a view that these dissociations prompt a review of the way they are dealt with by the neo-Cartesians, and possibly conceptual revision, and this is precisely the response we might make to the case presented by Hacker and Bennett. We can cite in support, the position taken by Patricia Churchland in her work: *Neurophilosophy* (2000). We have already quoted her view on the importance of blindsight: 'This is a case where empirical discoveries put pressure on us to make conceptual revisions' (Churchland, 2000, p.228).

She makes a methodological point later in her book: 'To make theoretical progress we precisely expect some degree of meaning change, and certainly, if observation sentences are revised then observational predicates undergo meaning change'. Churchland's own illustration of this point can be improved (Churchland, 2000, p.274), so we advance our own: Newtonian concepts of space and time, were generally accepted by the scientific community down to the end of the nineteenth century. That community regarded them as absolute: time exists and passes at a rate irrespective of any observer at any point in the universe. Space is absolute insofar as it was similar (throughout) and immoveable. Such concepts were deeply embedded in the thinking of the time, and were thought obvious. However, by the beginning of the twentieth century these concepts were challenged by Einstein, with observational evidence from the precession of the orbit of Mercury, and confirmed by Eddington's study of a solar eclipse, which showed light bending in the sun's gravitational field. As scientists gathered evidence to challenge the Newtonian concepts they thereby acquired reason to accept a change of meaning to the concepts in question, and absolute space and time have been replaced by new concepts.

The point of this is to show that science can help to bring about a profound change in conceptual architecture, and Churchland's point is to claim that this can be true of neuroscience and the concepts of mind too. She sides with Feyerabend in thinking that the whole fabric of mental concepts might be systematically improvable and revisable (Churchland, 2000, p.275).

On the other hand, Bennett and Hacker insist that '[…] neuroscience can contribute nothing to the clarification of the concepts of consciousness' (Bennett and Hacker, 2003, p.403), which is a claim which sits oddly with their concession that neuroscientists are not

precluded from using psychological predicates in new ways, so long as they can explain what these uses mean (Bennett and Hacker, 2003, p.384). Of course, we can agree with them that merely adding an asterisk to an existing expression (eg 'perceives*') is inadequate, but it should be possible clearly to specify an extension to the use of an expression. We might agree and establish, for example, that 'perceives' can properly be used when the candidate in question responds to a stimulus in a way other than verbally, or in contradistinction of another response made verbally, or in cases where the candidate in question has no awareness of doing so, or where a response seems to be at odds with his/her verbal account of things. All of these extensions to the use of an expression will need the support of experimental results, but such formulae would allow the 'strangeness' of split brains and blindsight to be fully accommodated. Making such changes is part of our project after all, and this thesis has been able to endorse Hacker's criticism of other concepts, such as qualia and 'what-it-is-like'. The results of the studies we shall look at will provide reason enough for making such conceptual changes.

In summary, Bennett and Hacker were right to criticise the Cartesian claims that psychological attributes could be predicated of mind, but to insist that only the whole person may legitimately be ascribed psychology, seems excessively conservative, especially if a change allows theoretical and empirical progress. Following the spirit of Churchland's own methodology: the preservation of the status quo should not override considerations of empirical adequacy, and it should be possible for discoveries in neuroscience to 'mold, shape and perhaps transmute the language of the mental' (Churchland, 2000, p.275).

Notwithstanding this, in this chapter we shall make efforts to avoid the charge of being in breach of the mereological principle. However, if the empirical evidence makes it necessary to depart from this, or if doing so is the only way to make proper sense of some of the dissociations, then we shall allow the descriptions of the neuroscientists and psychologists some licence. What we are here exploring are phenomena which are, in several ways, extraordinary, and properly to describe what is going on may require a change to ordinary or canonical ways of speaking of them. Let us now turn to Tim Bayne and the Unity Contention.

Tim Bayne and the Unity of Consciousness.

In 2000, Tim Bayne shared authorship of a paper 'What is the Unity of Consciousness' with David Chalmers - a paper which featured ultimately in Chalmers' book: *The Character of Consciousness* (2010). Bayne's book: *The Unity of Consciousness* (2010) is a substantial restatement and defence of the position they jointly took up in the 2000 paper. It is this book, and the position taken by Bayne and Chalmers which will be at issue in this chapter.

In their work, Bayne and Chalmers argue for a certain structure of consciousness which puts a constraint upon wider theories of consciousness, which specifically militates against atomistic, or Higher Order Thought theories of mind. This is in keeping with an overall commitment to the more dualistic picture of mind as they have defined it. The paper makes this wider motivation clear: '[We argue that the unity thesis] is incompatible with these theories: if the unity thesis is true, then these theories are false' (Chalmers, 2010, p.498). The basic claim is this: if consciousness has a unity to it which means that 'the basic notion of consciousness is that of a total phenomenal state: what it is like to be a subject at a time. This yields a holistic rather than an atomistic view of consciousness' (Chalmers, 2010, p.538). Using Bayne's own distinction, an holistic picture will be one which denies that there are any independent conscious states that need to be bound together to form part of a phenomenal field, while an atomistic picture might hold that there are discriminable features of the phenomenal field, which need to be bound together to form the whole. We shall look further into the atomistic/holistic question at the end of this chapter, but for the time being we note that an atomistic picture is one which is consonant with the idea of stray independent conscious states not bound holistically. We shall present synaesthesia as such a case in point.

As noted, Bayne and Chalmers aim their unity thesis against those Higher Order Thought theories which give an account of phenomenal consciousness, specifically those put forward by Rosenthal and others. Their claim is that these Higher Order Thought theories advance the view that: 'A mental state M is phenomenally conscious iff a subject has a higher order thought about M' (Chalmers, 2010, p.531). Chalmers and Bayne suggest that nothing about this position requires that 'phenomenally conscious states will always be unified. Certainly, it will not be necessary that they be unified'. Thus it is entirely possible for a subject to have a phenomenally conscious state A1, phenomenally conscious state A2, phenomenally conscious state A3 and so on, and the corresponding higher order thoughts: 'I

am in A1', 'I am in A2', and 'I am in A3', and yet without also having the higher order thought: 'I am in A1, *and* A2, *and* A3' (Chalmers, 2010, see pp.532-533). Such a subject could have contradictory beliefs that she never puts together, and so never has the thought: 'I believe in x, and in not-x'. Chalmers adds: 'This might be strange or unusual, but there would be nothing contradictory about it.'

A higher-order thought position would allow that there might be phenomenal disunity at times. If it is true, then it would contradict the central intuition behind the Unity Thesis, that, necessarily, there is something it is like to be in A1, A2 and A3 all at once. Bayne and Chalmers conclude that the two positions are at odds with each other. 'If the unity thesis is true, the higher order thought thesis is false', and vice versa.

Anyone defending the higher order thought position can reply in a number of ways. He can argue that the theory agrees that there could be cases of unified experiences, but disagree on what constitutes that unity, and explore this line of enquiry. Or he can deny the Unity thesis and embrace the view that unity holds typically, but not necessarily.

This work will investigate the latter course. We aim to put pressure on the idea that the unity thesis can be sustained, and in so doing, release ourselves from neo-cartesian constraints upon a wider theory of consciousness. We are then free to explore atomistic and higher order thought theories. The chapter will follow this plan: I will firstly identify the unity principle, and how it sits with the general position on mind that Bayne endorses; I will then argue that the dissociations we are interested in present more of a problem for his position than he would allow. On this view, Bayne might be more open to blindsight and other dissociations, such as alien hand syndrome. The chapter will examine his own account of how split brain cases may be understood (the so-called 'switch model'). We shall then turn to other challenges to the unity of consciousness, in synaesthesia. Finally, we shall examine the criticism made of the notion of Unity by S Zeki and his claim that consciousness is disunified.

The claims involved.

The central contentions of Bayne's work are these:

- The kind of unity enjoyed by human subjects and which is the primary subject of his examination, is defined as *phenomenal* unity.

- Phenomenal unity is to be understood as there being something-it-is-like for a subject to have multiple experiences together.
- 'Phenomenal consciousness' is pleonastic – all consciousness is phenomenal (Bayne, 2010, p.6).
- There are no obviously decisive counter-examples to the contention of unity, and so, we never have dis-unified experiences (Bayne, 2010, p.16).
- (This allows the move to the stronger claim that) Consciousness is necessarily unified, with the implication that a human being can only have one stream of consciousness at any one time.
- The claim of *phenomenal* unity (and not representational unity, or any other kind of unity) of consciousness delivers on the substantive claim that Bayne wants – that atomistic theories stressing the role of neurological sub-systems, or higher-order thought theories are ruled out of account by the Unity Contention.

Let us first be clearer about the Unity Principle and what is being argued. Bayne considers that there is something-it-is-like to see the saxophone, and something it-is-like to hear it being played, and there is a third thing: there is something-it-is-like to see it, and hear it being played at the same time – to have both experiences, but also to have them together. Bayne describes the relationship between the experiences as being 'unified when, and only when they are co-subsumed' (Bayne, 2010, p.20).

In other words, there is a phenomenal state which 'subsumes' all the other experiences which are part of this state – 'My total experiential state is a whole that includes within itself various experiential parts, such as my overall perceptual experience, my overall auditory experience, (and experiences in other modalities)' (Bayne, 2010, p.21). In defining subsumption, in the same place, Bayne gives a mereological account – in terms of the relationship between parts and the whole: 'conscious states are phenomenally unified in virtue of the fact that they occur as the parts of a single conscious state.' It is clear that, although this single unified state is a conscious state in its own right, it does not exist as separate from, or independent of its constituent parts.

A further feature of phenomenal unity as defined by Bayne is its relationship to representational content. More will be said on this issue later, but we can now briefly introduce the idea. He comments: 'Experiential states don't merely have a phenomenal character, or 'what-it-is-likeness'; they also possess representational content. Indeed it is

plausible to hold that the phenomenal character of conscious states is intimately related to their representational content. And this suggests it might be possible to analyse phenomenal unity in terms of relations between the representational contents of unified states' (Bayne, 2010, p.48). In this connection, and against his claim, Bayne entertains the possibility that there might be 'phenomenological danglers' – 'states whose phenomenal character outruns whatever representational content they might possess' (Bayne, 2010, p.51-2). If there are phenomenological danglers as defined, there is a problem for the claim that phenomenal unity can be analysed in terms of representational content. The central commitment to unity may be undisturbed (I could have an undirected mood of discontentment unified with the feeling of a woolly jumper on my skin), but danglers could create a difficulty for the view that 'we can fully capture the phenomenal character of consciousness by an appeal to the propositional contents of representational states'. In other words, he can fall back upon the what-it-is-like characterisation, or the mereological account of unity, but he concedes that, in cases where we cannot define phenomenal states in terms of representational content, we can count those a 'diagnostic of failures of phenomenal unity'(Bayne, 2010, p.72).

Within this context, Bayne discusses moods and pains - states which lack representational content altogether, or are such that their phenomenal character is not fully captured by their representational content. One example might be boredom. Bayne however, leaves it open that there could be plausible responses to these phenomenological danglers, and so sets them aside. We shall return to this area in a later section – there may be more serious danglers which present a problem for the representationalist picture Bayne and Chalmers endorse and which may be diagnostic of a failure of unity.

Some preliminary critical comments on the Unity Principle.

Before we embark on the main line of attack on the unity principle, - its handling of dissociation cases - I offer some further commentary identifying the position taken, and how it fits into a wider philosophical discussion of mind.

Returning, for a moment, to the characterisation of unity as one in which unity is a further state, over and above individual experiential states, and using the what-it-is-like formulation that the neo-cartesians are wedded to, they would argue as follows. There is something-it-is-like to hear the samba drummer, (v1) and there is something-it-is-like to see

the drummer hit her drum, (v2); there is something-it-is-like to smell the barbeque (v3); there is something-it-is-like to feel the pain in one's head (v4), and there is something-it-is-like to wonder when the band will stop (v5). In addition, there is something-it-is-like (R1) to hear the drumming, while seeing the drum being hit, while smelling the barbeque, feeling the pain and wondering when they will stop. (v1, v2, v3, v4 *and* v5) This total state R1 captures what it is like to be the subject at the time.

There is a telling criticism of the principle, which is broadly shared by Tye (2003) and Hurley (1998) and which has become known as the 'phenomenal bloat' problem. If phenomenal unity is meaningful, then there must be a sense in which I experience it. Suppose R1 is the unity relation between experiences v1, v2, v3, v4, v5. A condition of me enjoying phenomenal unity is that I have the experiences, and also the relation R1 between them, or there is no compelling reason to think they are unified at all. The subject must be conscious *that* they are conscious of v1-5 and R1. But then, in order to account for the fact that R1 is unified with v1-5, we need to posit a further unity relation R2. The arguments (recalling the third man argument of classical philosophy) suggest that the content of one's experience is multiplied beyond necessity, (just more content) or even fatally, if the argument ends in regress (Bayne, 2010, p.30).

In short form then:

- Suppose phenomenal unity is a relation between experiences v1-v5
- This unity relation (R1) between them must itself be experienced, or there would be nothing it is like for v1-v5 to be unified.
- If R1 is itself experienced, it must have its own phenomenology.
- If R1 has its own phenomenology, its phenomenology must be unified with that of v1-v5
- In order to account for the fact that R1 is unified with v1-v5, we must posit a new unity relation, R2
- R2 must itself be experienced, for if there were no experience of the unifying relation, there would be nothing it is like for v1-v5 to be unified with R1
- We have thus begun a vicious infinite regress.

(Bayne, 2010, p.30, referring to Tye's argument in 2003 pp.21-22).

Bayne deals with the objection by suggesting that we deny premiss 2 above, that R1, the phenomenal unity of v1-v5 must itself be experienced. He argues that there is nevertheless a unity relation, between v1-v5, without there being an experience of the unity relation itself, that phenomenal unity is a relation that makes no additional contribution to what it is like to be the subject in question. (Bayne, 2010, pp.31-2). He then acknowledges a contribution made by Ian Phillips – that understanding unity is not a question of experiencing that unity as an *object*, so much as a *manner* of experiencing'. No further development of this key point is given.

This now seems to be something of a tangle, partly caused by the attachment that Bayne (and Chalmers) have to cashing in the idea of being in a total conscious state in terms of there being 'something-it-is-like' to have a unified consciousness. This thesis takes the view that this phrase is unhelpful and contributes little beyond an appeal to intuition in defining any aspect of consciousness or experience.

We now have a claim that what it is to experience the unity of v1-v5, is for the subject to consider those experiences in a certain way, or fashion, not that they present a unity in themselves. At the very least, this seems to weaken the claim that experiences are necessarily unified, as their unity would seem to turn on the vicissitudes of the subject's manner.

There are other issues. If a commitment to the unity principle is a commitment to analysing 'what-it-is-like' in terms of multiple experiences, there is then a question of how many experiences there may be, synchronously. A holistic approach suggests many, or a lavish view of consciousness. An atomistic approach, such as we might find in the higher order thought account, might make a claim for a 'sparse' view of consciousness. In considering this, we encounter the 'lavish/sparse' debate.

Lavish/Sparse conceptions of consciousness.

The unity principle, in presenting 'what-it-is-like' in terms of multiple experiences unified together, might be thought to be a picture in which consciousness is rich, or lavish – suggesting that, at any one time, a moment of consciousness contains a detailed and multi-modal representation of the world. If this were the case, then Bayne would appear to be siding with other non-reductive monists, who also subscribe to an 'abundant', or 'lavish' conception of experience. John Searle, for example, describes his experience as having an

attended centre, and a broad periphery of consciously experienced but unattended objects and sensations (Searle, 1992, pp.137-8).

The ground for this position is, as Searle and Bayne admit, how consciousness appears from within, ie as a consequence of introspection. Thus: Bayne (2010, p.75) 'The plausibility of the unity thesis derives largely from introspection' and: 'the unity of consciousness that is revealed to introspection is […] a feature it enjoys all the time, even when one doesn't introspect.' Schwitzgebel makes a similar observation: 'Those who see consciousness as abundant generally provide little positive argument. They tend simply to state the position and expect the reader to agree' (Schwitzgebel, 2011, p.94).

We have dealt with the problems of a reliance on introspection elsewhere, and so need not repeat too much of that here. Bayne concedes the weakness of such reliance (Bayne, 2010, p.76 *et seq*)[23] and ultimately comes to declare that: 'Although introspection appears to lead some subjects to embrace a rich account of consciousness, [...] I myself incline to some degree of modesty here. As best I can tell, my typical phenomenal field involves a small band of focal experience surrounded by an experiential penumbra. This focal experience is usually dominated by at most two or three modalities at a time, with only a few objects and features represented in any detail across these two or three modalities. [..] This sketch might not qualify as an endorsement of the 'sparse' view consciousness, but it is certainly some distance from the lavish model' (Bayne, 2010, p.79).

In saying this he suggests he is going some way to repudiating the lavish or abundance model, but another reading of this picture is not a middle course between sparse/abundance – it leans directly towards abundance. A 'few' (objects and features) times 'two or three' (modalities) starts to look fairly lavish. This is only to be expected; the unity thesis must lean in some way to a lavish conception of consciousness, as unity is necessarily the unification of several experiences at the same time.

Bayne's co-author, Chalmers, does not dissemble. He certainly embraces lavishness in his speculative chapter on Consciousness and Information (Chalmers, 1996, ch 8) in which he equates consciousness with information processing. According to this, consciousness is everywhere.

[23] Bayne is nothing if not generous to counter-positions to his arguments. So much so, that he seems more often to side with them rather than take too trenchant a line. This makes it harder to dismantle his arguments.

Why is this distinction and the positions taken in it relevant and indicative? Schwitzgebel hints at the answer. A commitment to abundance is more likely to come from those who hold a Chalmersian concept of phenomenal consciousness – the kind of concept which 'makes dualism tempting and the mind-body problem interesting' (Schwitzgebel, 2011, p.94).Those who lean to the sparse conception are, unsurprisingly, those who are inclined to the possibility of reduction – Dan Dennett, (1991), Mack and Rock (1998).

Why is this the case? If defenders of abundance are right, our consciousness is awash with detail in several different modalities at any moment in time, and on such an account, there is no place for unconscious perception, if it is conceded that it exists at all. We are conscious of all our perceptual states in every modality – the position we ascribe to Chalmers. A narrow, sparse view of consciousness is consonant with the picture held by HOT theorists – it is the view that our experience is limited to a few events or objects at any one time. On this view, most of our perception is unconscious. It is perhaps no accident that the same divide in the philosophical community over the question of reduction and the idea of unconscious perception, is the same divide on the sparse/abundance question.

Schwitzgebel concedes that resolving the sparse/abundance question may not be easy, it is arguably very important, since settling the range of conscious experience (and thus the range of unconscious experience) might be prior to any commitment to a more general theory of consciousness. Resolving it without resort to a war of intuitions (Schwitzgebel, 2011, p.96) needs a thorough-going evaluation of the possible deliverances of introspection. Necessary too, is some agreement on how we might individuate experiences and ways of counting the number of experiences we have at any one time.

This last problem is nicely exposed by the flashlight objection. First mooted by Jaynes (1976), he argued: 'Consciousness is a smaller part of our mental life than we are conscious of, because we cannot be conscious of what we are not conscious of.' […] 'It is like asking a flashlight in a dark room to search for something that does not have any light shining on it. The flashlight, since there is light in whatever direction it turns, would have to conclude that there is light everywhere. And so consciousness can seem to pervade all mentality when actually it does not' (Jaynes, 1976, p.23).

Wherever we look, whatever we attend to, there is experience. There is no easy way of settling the question of how much content is involved. Introspection is necessarily flawed as a line of inquiry into this.

Empirical research in the form of 'numerosity judgements' may have a bearing. These are much replicated studies which show that subjects are only able to identify four items or fewer in displays, with any degree of accuracy of confidence, while error rates and reaction times are poorer with displays using five items or more (Sperling 1960; also Atkinson 1976, Livermore and Laing 1998) Again, Bayne generously allows that these studies have some purchase: 'Taken as a whole, this research certainly provides *some* support for the sparse conception of consciousness' (Bayne, 2010, p.78; emphasis in the original).

A final judgement is not delivered, however; Bayne leaves the question of empirical support for the lavish/sparse debate open. Nevertheless, a good deal of his book deals with the challenges to the unity contention provided by empirical studies of dissociations and other conditions. It is these conditions that we may argue provide an empirical challenge to the unity contention and then help to settle the lavish/sparse debate. If we can make an argument for unconscious perception and establish counter examples to the unity thesis, show that experience is limited to a very small number of attention-occupying objects at a time, we dismantle some of the motivations for non-reductive monism, and the constraints dualists impose on a wider theory of mind.

It is now time to consider whether the unity contention stands up against the evidence of dissociated minds.

The Unity Thesis and the evidence of dissociation.

In Bayne's defence of his claim, and in the empirical elimination of counter-examples, he considers cases of neglect, agnosias and the 'split-brain' syndrome, which is the product of a commissurotomy. Blindsight also figures in his discussion, but as part of an argument explaining the extent to which we can use, or rely on the testimony of subjects, and the kind of report we may admit as evidence for perception. We shall consider his rejection of these dissociative conditions, and whether he is right to contend that there are no counter-examples to his Unity Thesis. If any of them do constitute a counter-example, that would be sufficient to undermine the claim of necessity.

Bayne on Blindsight

As we have noted, Bayne is a generous, non-dogmatic advocate for the Unity Thesis, and he is careful to concede ground where the evidence or strength of argument demands it. One such concession comes in his discussion of blindsight (Bayne, 2010, chp 5). For reasons that may be understandable, he does not consider blindsight as an empirical test for the unity claim, but it features in a ground clearing discussion of how to evaluate the unity claim. It is mentioned in context of: what emphasis should we put on the subjects' reports of their experience, (the 'introspective criterion') and whether other forms of reporting may be allowable. Further, to what extent can intentional agency be a criterion of consciousness? (the 'agentive criterion').

In the course of his discussion, he makes a number of claims which are germane to our investigation, and we shall look at these before we acknowledge his concession. He acknowledges the well-rehearsed problems for the introspective criterion: introspective reports can be wrong, they cannot be made by higher animals and pre-linguistic children etc. It is in connection of the agentive criterion that he deals with the dissociations.

Firstly, he recognises that the agentive criterion does not have widespread support, since cognitive sciences have come to the belief that much of our cognitive and behavioural control is the product of 'zombie systems' – systems that operate on the basis of unconscious representations' (Bayne, 2010, p.98). In short, Bayne goes on to repudiate this claim, and to make a defence for the agentive approach to the ascription of consciousness. Blindsight, DF's visual agnosia, and continuous flash suppression are the conditions considered in detail. Bayne assesses the evidence for these as examples of 'zombie systems' and finds it wanting on the following grounds.

He argues that they are none of them entirely 'zombie'. The conditions 'are not homunculi – 'mini-me' who form and execute their own plans. Instead, their operations are very much under the control and guidance of the contents of perceptual and intentional consciousness' (Bayne, 2010, p.100). This is in context of a rejection of the idea that Milner and Goodale's DF unconsciously perceives the letterbox in order correctly to orient her hand to post a letter. 'The vast majority of DF's actions involve a great deal of conscious perception' (ibid).

Blindsight is also unconvincing as an example of unconscious perception: 'subjects are unable to discriminate complex shapes in their blindfield, and even those properties that they can unconsciously identify are not spontaneously integrated into their practical and theoretical reasoning'. (They have to be prompted). Perceptual content is not manifest to the subject or available to them, if they take themselves to be 'merely guessing' (Bayne, 2010, p.100). Helen, Humphrey's monkey, did seem to exhibit unprompted spontaneous behaviour, but this superblindsight is not replicated in humans[24].

Let us pause for a moment to summarise and examine these points. The claims are these:

- Blindsighted subjects are unable to discriminate complex shapes
- Perceptual content is not manifest or available to the subject
- Availability of content might be conceded if ever a case of superblindsight were found, but none has.
- Further, in the case of DF, in her visual agnosia, Bayne claims that a lot of her unconscious representation is parasitic upon conscious perception.

On each of these points there is something to add. Firstly

i) On the attempt to suggest that, whatever blindsight is, it is confined to simple stimuli and paradigms, and is not therefore to be taken seriously – a few recent studies call this dismissal into question. Tamietto *et al* (2009) shows that blindsight subjects are able to detect the emotional valence in a projected face projected to the blindfield, whether it was smiling or frowning. There is evidence here for a 'zombie system' identifying projected shapes, and responding with the twitching of muscles in the subject, uniquely used for smiling or frowning, and doing so without the awareness of the subject. Weiskrantz also reports the increased sensitivity in DB after some years (*Blindsight* 2009 edition p 18), and found that he, DB, could successfully identify low-contrast outline images of objects and animals, and make same/different discriminations of pairs of such images with an 89% accuracy. This is clear evidence of an ability to identify complex shapes, rather than simple light stimuli.

[24] In point of fact, that Helen could move around under visual guidance suggests to Bayne that she was actually visually conscious, this despite the fact that Helen's visual cortex was removed. Her perceptual success is too successful. This is redolent of the difficulty of satisfying conspiracy theorists. Helen was so accomplished, she must be recruited as a defence witness, not as a case for the prosecution.

ii) On the claims above concerning superblindsight – as noted earlier, recently the case of TN has come to prominence (de Gelder *et al*, 2008), - a bilateral case of blindsight, and the case closest to a condition of superblindsight yet encountered. The absence of superblindsight in humans is accounted for by the fact that lesions, the product of brain damage or strokes rarely disable the V1 to the extent that a subject loses sight in both eyes and completely (usually a lesion leaves part of the visual field affected – an area called a scotoma), so that subjects continue to have some visual function in a sighted field, and so never have to develop a spontaneous response, or exercise a reliance on the subcortical pathways to the extent that Helen was.

iii) On DF's case, and the claim that much of her perceptual accomplishments were parasitic upon conscious perception, and thus: 'very little behaviour is under the exclusive control of zombie systems' (Bayne, 2010, p.100)- a closer reading of Milner and Goodale (2005) is prompted. DF's condition is not one where the zombie systems can be separated out so easily. It is understood that there are linkages between the ventral and dorsal processing systems, so that the vision-for-perception (ventral) route does get some feedback information, so yes, there is some sense in which DF's behavioural success can be put down to piggy-backing off conscious perception. However, her condition is such that much of her success in orientation of her hands, in placing her feet in rough terrain while walking, is accomplished without awareness in any proper sense of the term. DF cannot identify objects or separate them from their background. She fails to register edge or shape. As Goodale and Milner comment, (2005, p.10) 'it is not a question of her failing to interpret a clear visual experience; her problem is that she does not have that clear visual experience to begin with'. In the absence of this, DF was nevertheless able to walk up an uneven and steep pathway, and walked confidently without stumbling over roots and without colliding with branches. She was able to hold her hands in the correct position to receive plates, knives, forks despite the fact that could not verbally identify them. She was able to catch a ball thrown at her. She did not have much conscious perception off which to piggy-back, and on this view, Bayne is too quick to dismiss her case, and eliminate the idea of zombie systems at work.

However, it is clear Bayne does not regard blindsight and DF's agnosia as being challenges to his unity contention, or they would have been given their own chapter. As noted, they are considered in connection with the question –how are we to evaluate the unity claim - to what extent should we rely on the reports or introspective judgements of subjects, or on their behaviour as agents.

His eventual conclusion on this question (and this is the concession we remarked comes in this section) that other forms of behaviour may count as evidence for successful registration of stimuli, including examples of intentional agency (that is – actions of subjects mimicking movement, pressing commentary keys, orienting hands, and reactions to emotional valence etc.) This seems right, or we would rule out of account important evidence, and in particular, studies which concern animals or pre-linguistic children or others who cannot give verbal responses based on introspective judgements. Philosophers post-Wittgenstein are unlikely to think it sufficient to base contentions solely on 1st person reports of introspective judgements[25], and Bayne takes the same view: 'Introspective reports may play a core role in the ascription of consciousness [..] but they are *not* the sole basis on which conscious states can be legitimately ascribed' (Bayne, 2010, p.96; emphasis in original).

He does concede that sometimes, introspective reports might dissociate from other forms of behaviour, and it is a matter of some delicacy on the question of which to prefer, in this event. A conservative approach would be to ascribe consciousness to a creature iff each of the possible markers point in the same direction, but he acknowledges that there is no obvious reason why we should be looking to prefer a conservative approach.

These are important questions that have a bearing on other parts of this thesis, but here at least, it is important to note, and to rebut his resistance to 'zombie systems' in these conditions, since it is our goal to establish that such automatic unconscious systems processing individual features of experience exist, and account perhaps for a big proportion of human behaviour.

However, to put maximum pressure on the unity contention we should now turn to the dissociative cases which Bayne deals with, as potential empirical counter-examples to his

[25] Even so, Bayne thinks it is an open question as how to respond when the different markers of consciousness point in different directions (when, for example, an agentive response (eg orienting hands, or indicating a light source's movement) points in the opposite direction to introspective judgements ('I don't see anything')). It seems to me we could be less equivocal.

claim. Recall that Bayne says that consciousness is necessarily unified, and so one counter-example is sufficient to create a difficulty for him.

The Split Brain Syndrome.

The brain has two hemispheres, connected by a network of fibres – the corpus callosum. There are two kinds of divided brain patients – ones who are born without a corpus callosum (acallosal subjects) and those who have had this network to some extent surgically severed, usually to control epilepsy (split brain subjects). The operation– a commissurotomy – is today infrequently performed, but was much studied in the 1960s and 1970s after operations performed by Sperry became the subject of studies conducted by Gazzaniga and others. It is an operation of last resort today, as epilepsy is now largely controllable by drugs.

The commissure is a differentiable structure, and it is specialised in the sense that different information is transferred by different sections of the collosum. Different operations in which different parts of the commissure is severed, identify the functions of the several parts: the anterior mid-body transfers motor information, the posterior mid-body transfers somatosensory information, the isthmus transfers auditory information, and the splenium transfers visual information. (Funnell *et al,* 2000). It is not possible completely to separate the two hemispheres, and some connectivity always remains, even in cases of more radical sectioning – the brain stem maintains some linkage. Some studies involve incomplete transections and are interestingly different from those in which complete transections are done, but all cases may be relevant to the question of the unity of consciousness, depending on the nature of the dissociation.

What is the practical outcome of the operation? Third party observers are surprised at the limited nature of the changes seen in a post-operative patient, and those who are uninitiated often remain unaware that subjects have had the change, but since the way the outcomes are to be understood by the casual witness is one of the things at issue, we will discuss this below.

Under testing and laboratory conditions, however, the Split Brain syndrome emerges more strongly. An easy-to-repeat paradigm shows this. In the majority of normal human brains, speech centres are located in the left hemisphere, and can report verbally on objects presented to the Right Visual Field (RVF), but in the case of the split brain patient, in whom

the inter-hemisphere connective channels have been severed, information concerning objects in the LVF, does not reach the speech centres, and the subject cannot verbally report on the contents of any experience exclusive to the LVF, however, the subject can single out and retrieve the object with her left hand (Sperry 1974). If asked what she had (correctly) retrieved, the speech equipped left hemisphere fails to identify the object.

The studies invite the following comparison and interpretation.

In the studies of blindsight, the subject is arguably not aware of a stimulus, but is conscious of her behavioural response to it. She maintains that she sees nothing, but is aware of her arm pointing in a certain direction, or making a guess as to the colour of a stimulus. Given the surprising degree of success with these guesses, we argue that she is unconsciously perceiving the stimulus shown. Blindsight shares some similarities and some differences with the split brain cases. In split brain scenario, there is a case for saying that one hemisphere is not aware of the stimulus given to the other, and often not aware of the response either. In any event, a stimulus is correctly processed, and a successful discrimination made, but one hemisphere at least is unconscious of it. If it is the left hemisphere that is excluded, as it were, from the stimulus-response story, it often confabulates – the subject creates a spurious account of why the response was made. To this extent, there are some parallels with the blindsight cases. But there are differences - in every split brain paradigm, wherein the hemispheres are separated in their stimulus-response stories, consciousness goes with perception. Each hemisphere perceives a stimulus, and correctly responds to it; our interest arises in the dissociation created by the splitting of the brain, and the fact that the hemispheres can operate independently.

The relevant interpretation of the cases as far as the Unity Contention is concerned is this: It may be that the callosal section creates two separate streams of consciousness, and provides us with a counter example to the Unity Thesis. This is indeed the claim made by Sperry, saying: 'both the right and left hemispheres may be conscious simultaneously in different, even in mutually conflicting, mental experiences that run along in parallel' (Sperry 1974). It is for this reason that Bayne dedicates a sizeable chapter to the split brain syndrome, and our next task is to evaluate his response to it.

Bayne's response to the Split Brain Syndrome.

Before defending his own case for the split brain syndrome, Bayne first considers the case for the prosecution. He considers three arguments – firstly, the empirical evidence for agentive disunity; then two *a priori* arguments, which he considers valid, and which for sake of brevity, we shall set aside.

We shall look at his dismissal of the case for disunity, and then his own account of the everyday appearance of unity – the switch model. According to this position, consciousness in the split-brain moves or switches from one hemisphere to the other. 'Although both hemispheres can process information concurrently, they take turns in supporting consciousness' (Bayne, 2010, p.210).

Bayne's case against the disunity claim

i) Agentive disunity

Bayne is ready to concede that, on the face of it, there is agentive disunity in Split Brain cases, with evidence to be found in manifestations of inter-manual conflict, or anarchic hand syndrome (not unlike the unruly gloved hand in Peter Seller's depiction of Dr Strangelove). Bayne however stresses that the most salient feature of the Split Brain cases is not a dysfunction so much as their 'social ordinariness'.

On this view, however, Bayne is too quick to dismiss the case for agentive disunity – he suggests that the anarchic hand condition typically subsides within a few months (Bayne, 2010, p.195, citing Bogen 1998), but the condition can persist – Baynes (K) *et al* 1997 is a

report of a study of AW – a woman with a chronic case which had persisted over 10 years. In the same paper, they review other studies of long-term inter-manual conflict or anarchic hand syndrome, some lasting years. The latter study reported: 'the patient's chief complaint was that her left hand did not obey her. On several occasions while driving, the left hand reached up and grabbed the steering wheel from the right hand. The problem was persistent and severe enough that she had to give up driving. She reported instances in which the left hand closed doors the right hand had opened, unfolded sheets the right had folded, snatched money the right had offered to a store cashier, and disrupted her reading by turning pages and closing books.'

The persistence of the condition sits at odds with the claim of social ordinariness, about which we will have more to say. There are other symptoms, which include: signs of indecision, or being internally conflicted, problems with short-term memory, and absent-mindedness. Bayne concedes these symptoms are long-term (Bayne, 2010, p.199) arguing however that these symptoms do not challenge the unity principle in the way long-term inter-manual conflict might seem to - absent-mindedness, even indecision, is still consistent with a unified phenomenology. But the case of the anarchic hand strongly suggests a disunity, born of separate competing intentions in the different hemispheres. Recall the mereological account of phenomenal unity given by Bayne from the outset – Unity is described as having an experience e1, and an experience e2, and for there to be a further experience R which is not separate or independent of e1 or e2, which is e1 *and* e2, - summed up as what it is like to be the subject in question. In the anarchic hand situation, we have a subject who has difficulty in tracking the contents of their subjective states, and who has lost a sense of (some of) them being their own. What it is like to be the subject for whom one hand is steering right and the other pulling left is hard for the subject to give an account of, much less anyone else, but it seems we must agree that the situation lacks an essential unity.

Bayne gives this account of the syndrome: it is 'best thought of as triggered by stimulus-driven intentions' (Bayne, 2010, p.195), but the subject herself need not be conscious of the intentions, (just as one can be unconscious of navigating down a street whilst engaged in conversation). Motor independence (of the anarchic hand) seen here involves only subconscious mechanisms, and we do not need to posit conscious intentions in each hemisphere, or think of each hemisphere as harbouring an autonomous agent.' This makes the syndrome no more interesting than an account of the automatism sometimes involved in

long distance driving, and seems to have no compelling account of the conflict that clearly exists in the subject, the feeling they have that their hand is out of control.

It is not easy to see why this characterisation makes the anarchic hand less of a problem for the question of phenomenal unity. An account which sits more naturally with the facts of the case is this: the sundering of the neurological links between the two hemispheres sunders the mechanisms by which the processing of, and response to situations are integrated into a single agentive behaviour. The anarchic hand, or inter-manual conflict is evidence of agentive disunity, and agentive disunity is more compellingly explained by phenomenal disunity. As Bayne concedes, there are subconscious mechanisms at play in the behaviour of people with the condition, and stimulus driven intentions may be unconscious, but a natural explanation of this is because such intentions and behaviour are linked with different hemispheres with poor access to each other's contents.

ii Spatial decoupling.

Bayne introduces a second form of agentive disunity – spatial decoupling, citing the study of JW (Franz *et al*, 1996) in which a split-brain subject is shown to be capable of drawing distinct patterns simultaneously with left and right hands. Bayne: 'Spatial decoupling is [..] consistent with phenomenal unity. [..T]he kind of motor independence seen here involves only subconscious mechanisms; we needn't posit separate conscious goals or intentions in each hemisphere' (Bayne, 2010, p.195).

Bayne concludes 'that there is no evidence that each of the patient's hemispheres thinks of itself […] as an autonomous agent', and that 'the evidence suggests that split brain patients are single agents who are attempting to get by as best they can.' Indeed, he repeats approvingly a quote from JW: 'Are you guys trying to make two people out of me?', as evidence that despite the researchers' probes, there is but one agent present in the study.

In this section, the argument and evidence is somewhat thin. JW's comment can be quickly dealt with as the product of a linguistically competent and assertive left hemisphere speaking on behalf of itself, and unaware of its relationship with the other hemisphere. There is only one *verbal* agent present in the study. Moreover, a further study of JW takes his spatial decoupling further, in a study which makes a stronger case for phenomenal disunity as well. In Eliassen *et al* (2000), JW is shown tachistoscopically, two figures either side of a

focus point and asked to draw the figures with pens in both hands. JW responds quickly and without error – a feat which strongly suggests that each hemisphere is conscious of the object shown to its corresponding visual field, and that JW himself demonstrates agentive disunity, and that this is the product of a divided phenomenal field. Gazzaniga: 'It was as if two people were present, one guiding each hand with absolutely no interference from the other' (Gazzaniga 2015).[26] Bayne asserts, in this section, that, were JW 'deluded in taking himself to be a single agent, […] this would be an unwelcome result' (Bayne, 2010, p.196).

In the anarchic hand cases however, this is where the evidence does point, even if the case is not, and perhaps cannot be completely conclusive either way. The subject does not report herself as having a fragmented self, or recognise competing wills within, but we have third-person evidence of competition, we know that the vocal left hemisphere is dominant, and inclined to speak for both, and confabulate where it is denied information exclusively available to the right. In such circumstances, we might be as sceptical of verbal reports as we are when we ask the blindsight subject if they can see. Bayne concludes the section: 'Split brain patients are single agents who are attempting to get by as best they can' (Bayne, 2010, p.197). But it may be at least as credible to say: split brain patients have separate, hemisphere-based agents, lacking the structures of integration and inhibition, whose left hemispheres are creating a narrative of unity.

Everyday Integration.

Bayne goes on to weigh the evidence for the 'two streams' models' ability to account for everyday integration and behavioural unity in the split brain. He claims: 'it is far from clear how the disunity models might account for 'social ordinariness' (Bayne, 2010, p.200). In this section, there is a presumption of 'social ordinariness' or behavioural unity, and he finds neither of the two models offered adequate, but if we can account for social ordinariness in other ways, or establish that it is perhaps an illusion, then his argument against these two positions is diminished. This is the aim of this section.

In many other contexts, we know that how things seem is not necessarily a reliable guide to how things are. It may well be the case that 'naïve observers are rarely aware that the subjects suffer from cognitive impairments' (Bayne, 2010, p.199), but this is not a reason for

[26] See: https://vimeo.com/96628409

concluding that the subjects themselves have a unified experience in some way. We have already noted that the subjects themselves do have difficulty in making decisions, or are absent-minded. Other subtle effects include: deficits in ability to solve geometrical problems; associating names with faces, poorer memory. Prima facie, this, together with our discussion of the long-term alien hand syndrome, is some evidence for thinking the subjects are not as ordinary as depicted. Not all of these indicators point in the direction of disunity of experience, it is true, but all suggest a lack of usual integration and cooperation between the hemispheres, which might begin to point to a severing of experience. But there is, in the idea of cross-cuing, a means for explaining what ordinariness they do exhibit. Bayne mentions cross-cuing, but discusses it briefly, and without assessing the evidence (Bayne, 2010, p.196).

Cross-cuing is a way of each hemisphere providing the other with information about its experiences, exogenous to the brain. It can be subtle and hard to detect, and is something which investigators make efforts to eliminate from their studies. Springer and Deutsch provide us with one example (1985 p.36). They report a study in which a subject was shown numbers to the right visual field, and asked which number was seen. The vocally equipped left hemisphere had no access to the information shown to the right, and so could not at first say what the number was. After a while, however, the subject was able to come up with a correct response. The investigators concluded that the left hemisphere was counting subvocally, and when the right number was reached, the right hemisphere would signal the left to stop, and report the signal out loud. This was confirmed when larger numbers were used in the experiment. When the larger numbers were presented to the right visual field, it took longer for the subject to respond; this fitted well with the suggestion that the left hemisphere was counting subvocally. 'The larger the number, [..] the longer the list of numbers the left hemisphere would have to go through before reaching the correct one' (Springer and Deutsch, 1985, p.37).

Gazzaniga (1967) gives us another example. A red or green light was exclusively shown to the right hemisphere (RH), and asked to say what was seen. The RH saw red, and heard the vocal LH guess 'green'. Knowing this was wrong, the RH shook the subject's head and frown, and cued the LH to say 'red' and correct the response. If the first responses were taken out of the study, the correct response rate dropped to 50%.

Given the fairly extensive literature on cross cuing (Bittner 2011, Gazzaniga 1969 as examples) and that many of the sources cited by Bayne feature in his bibliography, it is a

little surprising that cross-cuing does not feature more as a way in which the split brain can restore the unity of behaviour of the subject. Of course, the unity of behaviour of a subject does not mean that the unity of experience is restored.

This then, is an answer to the question of 'social ordinariness' of split brain patients raised by Bayne. In such cases, we evidently have two hemispheres with different mental contents, and an inability of one hemisphere to get direct access to the contents of the other. (See Sperry's 1979 study of LB). However, in ordinary, non-laboratory circumstances, a split brain subject finds no difficulty in equalising the contents of both hemispheres and behavioural unity is the product. It is only natural for the patient to use whatever information she has, to make sense of her situation. In practice, the fact that the two hemispheres share a single body constrains the extent to which any individual can behave in a disunified fashion (Schechter p.217). The subject will naturally want to be seen to be normal, and behave in a socially acceptable manner. Finally, it would be natural for the subject to want to create a unified life story, even with events originating exclusively in the right hemisphere, and unshared with the left. For all these reasons, the subject might present as ordinary, and remarkably unaffected, but the underlying reality is subtly different.

We do not need to labour the differences between the two accounts Bayne suggests explain this ordinariness (contextualist and duplicationist) as cross-cuing would appear to apply in both cases.[27] Nor do we need to be committed to Puccetti's (1981) claim that normal brains are divided in the same way. Our purpose is met by finding a reason to question Bayne's position on disunity models. Our next task is to examine Bayne's own account of Split Brain patients, in which he maintains that they enjoy phenomenal unity – the 'switch model'.

The 'Switch Model'.

According to this account, there is still a single stream of consciousness in the split -brain subject, but the physical basis for this moves or switches from one hemisphere to the other, as

[27] See Puccetti (1981). Puccetti would appear to be a duplicationist, but accounts for laboratory mental duality by cross-cuing of information from speaking to mute hemisphere. He also maintains that this is a mechanism which would give the *subject* an illusion of mental unity, since the mute or minor hemisphere has known its subordinate status 'from a very tender age. (…) Post-surgically, little has changed for this mute hemisphere (other than the loss of sensory information about the ipsilateral half of bodily space). It does find that it can now sometimes undertake an independent action (…) but on the whole it has no way of expressing itself, except through a nervous system much under the control of another. Being inured to this status as a cerebral helot, it goes along.'

a function of which hemisphere has won the competition for attentional resources. Bayne concedes that both hemispheres can process information concurrently, but each takes turns in supporting consciousness – over time, a hemisphere has one and one only centre of consciousness, that never diachronically overlaps with one sustained by the other. If they do not overlap in time, they do not constitute a threat to the unity thesis. The picture created is one in which there is a 'fluctuating perceptual extinction in each hemisphere' (Bayne, 2010, p.210), as stimuli are favoured in each visual field, and the switch takes place. At any one time, there is one subject, but 'consciousness switches rapidly and effortlessly between hemispheres in response to the demands of the patient's context.'[28]

This section will review the arguments against this picture, regarded as implausible and ad hoc by Van Gulick (in Bennett and Hill 2014), and with scepticism by Prinz (2013). The best attack comes from Schechter (2012). We begin with one which Bayne himself concedes is the most powerful.

Bayne considers the model thus: 'The switch model holds that the contents of consciousness are sequentially informed by processing in each hemisphere. If this were so, then one would expect patients to report sudden changes in the contents of their consciousness. [..] Why are split brain patients apparently unaware of the alteration between rival percepts?' (Bayne, 2010, p.218).

In non-laboratory situations, and where the contents of each hemisphere are unlikely to be too different, the subject might not be expected to report a shifting experience, but in those experimental conditions, where different words or symbols projected to different hemispheres, and a different experience is contrived for each, one might expect a subject to report discontinuities. Schechter puts the question: 'why don't they say things like: "moments ago I saw a boat, but now that's gone, and instead I see a fork; no – wait, there's the boat again"' (Schechter, 2012, p.219). In fact, this is an unlikely response, given that each hemisphere would not be aware of the other's perceptions. (In the story given, there is no observer in a position to see the boat change to the fork and back again.) On the other hand, one might expect them to report a jerky, discontinuous experience, as the subject switches from one to the other, depending on whether she is asked to focus on the function, or the

[28] This is not unlike the imagined situation in Dan Dennett's well-known story 'Where am I?', in which on-line access to, and control by 'Fortinbras' is switched between 'Yorick' and 'Hubert', with neither the wiser.

emotional response to the object shown. If a clock is built-in as part of the paradigm, she might report a certain 'gappiness' in her experience.

Bayne concedes he is not sure he has a lock-down answer to this point, but defends his thesis by saying that there are several other occasions where other neurologically compromised subjects can be quite unaware of radical changes to the contents of their consciousness – he offers achromatopsia, in which subjects fail to realise they have lost the experience of colour; another example might be unilateral neglect, so it is not unusual to find patients unaware of deficits or changes in their perceptual experience. This of course does not meet the objection head-on, and Schechter makes the point that this would put the split brain phenomenon on a par with, and essentially no different from, other disorders of attention, and fails properly to deal with the charge that, if a switch model does indeed imply that there would be gaps in experience, then we might expect the subject to comment on it.

The switch model and psychology in the split brain.

Schechter makes another point which we must amplify. If the two hemispheres preserve a unity as Bayne contends, we might expect little or no psychological difference between the two hemispheres, as far as we can detect it. But there is a psychological difference, and the idea of disunited hemispheric consciousness is rather better able to account for psychological difference than unity. Schechter puts it thus: 'There is evidence that they have different emotional propensities, access to different memories, perhaps even long term dreams and goals in life' (Schechter, 2012 p.219). Even Bayne himself concedes that 'split-brain patients may have distinct RH and LH self-conceptions' (Bayne, 2010, p.196). Some evidence for this is given in Schiffer et al (1998), a study in which a Split Brain patient AA was asked emotionally sensitive questions, and required to respond with his hands, hidden by a screen, by touching pegs in a row, indicating how he felt about the issue. AA gave significantly different answers to emotionally sensitive questions, to those given verbally. Other studies have been able to detect differences between the hemispheres on the question of belief in God, or favourite colour. McGilchrist devotes the whole of the first half of his book (*The Master and his Emissary* 2009) to asymmetries in function, processing and cognitive differences, and so on, between the two hemispheres. Arguably, it is easier to account for this psychological distinctness, by assuming a disunified inner experience.

Perhaps more tellingly, in Schechter's view, what the switch model implies is that there is no important *psychological* distinction between consciousness in the split brain subject and in anyone else. '[It suggests] there is a single stream of consciousness per human being, any human being, realized dynamically in the brain, and what you are conscious of depends upon the outcome of a competition for attentional resources. So, [the switch model seems to promise] there is no genuine psychological puzzle of split brain consciousness' (Schechter, 2012, pp.220-1).

But, as she goes on – there *is* a psychological difference, and there is a problem of accounting for that difference, and the switch model seems to recognise neither.

The switch model and JW

Lastly, in a section defending the duality model, Schechter notes that the split-brain experiments identified different responses in the left and right hemispheres, but without determining that the hemispheres were simultaneously, separately conscious, and if the switch model was to be defeated, 'it would be nice to find some cases in which right and left hemisphere conscious events were simultaneously manifested'. On this view, the case mentioned above, of Gazzaniga's patient, JW, seems to provide just the kind of case Schechter needs. JW seems to be briefly aware of two different shapes, and then each hemisphere, controlling the different hands, can draw the shapes without difficulty. 'It was as if two people were present, one guiding each hand with absolutely no interference from the other', notes Gazzaniga. Van Gulick makes a similar point, without using the example of Gazzaniga's JW, arguing that the split-brain patients seem capable of carrying out independent and contrary actions with their right and left hands at the same time, each of which is complex and non-habitual to a degree that would indicate conscious control rather than control by a "zombie system" according to Bayne's own criteria. It is hard to avoid the conclusion that collectively, these empirical cases concerning split brains put a question mark against the Unity Thesis, and the switch model.

The switch model and attention 1

In a separate paper. Prinz makes a short but telling point (Prinz 2013) '[If consciousness depends on attention, that means that, in the switch model interpretation of the split brain experiments] attention is shifting from one hemisphere to another. This is empirically

implausible'. Here is why. In the split brain studies, patients are presented with two stimuli simultaneously – one to each hemisphere - and they successfully respond to both. 'The stimuli in [the] studies are usually presented for 150 milliseconds, and while attention can switch in as little as 60 ms, usually attention doesn't shift as quickly as that. When staring at an unchanging stimulus, attention often remains fixed by about three seconds […] Even a saccade to a suddenly appearing stimulus takes about 200 ms to initiate' (Prinz, 2013, p.221). His point is that there is not enough time for consciousness to shift between hemispheres in the experiments, and so the switch model is unlikely as an explanation of unified behaviour in the cases. There are signs that Bayne has anticipated this line of attack and he may reply that this argument is predicated on the idea that there are two attentional systems that the split brain patient switches between. The evidence for this is uncertain. Indeed, he boxes himself in when he argues: 'some attentional systems remain unified in the split-brain' (Bayne 2010, p216). He goes on, rather less convincingly: 'it is possible that these systems play an important role in 'shuttling' consciousness between hemispheres as and when required.' Sadly, this remains undeveloped.

The switch model and attention 2

There is a further collision between Bayne and his critics on the question of attention. Again, he opens the flank himself – 'Given the intimate connection between attention and consciousness, surely attentional disunity in the split brain constitutes strong evidence for phenomenal disunity' (Bayne, 2010, p.214-5). Bayne concedes that there is some evidence for split brain patients having independent attentional systems. He and Schechter cite the same papers in support. But after consideration of papers pointing in a different direction, Bayne declares: 'I am not sure there is any tidy story to tell. […] At the very least, split-brain patients do *not* have two attentional systems' (Bayne, 2010, p.216; emphasis in original).

It is easy to reach a different conclusion, which indeed Schechter does. There is evidence that the corpus callosum itself plays a role in creating a unified spatial attentional system, and once severed, it can no longer provide the neurological basis for a unified attentional system. How the switch model is grounded in neuroanatomy remains a problem.

We have examined the case for unity and the hard case that Bayne suggests is the most likely counterexample to unity of consciousness. We will turn now to a line of criticism

which presents a different hard case, one not considered at all by Bayne. This is the suggestion that in cases of synaesthesia, there is a phenomenal superunity.

Synaesthesia as a challenge to the Unity contention?

In brief, synaesthesia is a condition some have in which the subject engages in some act of perception, and is successful in doing so, but at the same time she/he also experiences something extra. Sounds can be accompanied by tastes, or by colours, tastes can be accompanied by feelings, and in the most common version, letters and digits can induce unexpected experiences of colour. The many different variants make generalising difficult, but for the purposes of this section, we can mostly concentrate on this 'grapheme-colour' synaesthesia, which constitutes about 66% of all cases (Cytowic and Eagleman p.24). This is a form of synaesthesia known as 'projective' – in which colours are projected upon objects, or items in the visual field, or sometimes to nowhere geographically specific[29]. Another type of synaesthesia is 'associative', where colours, or other unexpected perceptual experiences are associated with nothing that one could say is part of the perceptual field – thus some synaesthetes might describe a day of the week as being associated with a colour.

Recent developments in brain scanning (fMRI) and testing have made the study of synaesthesia more respectable, and case studies are starting to feature in the thinking of philosophers of perception, and neurologists. There is still debate about whether the additional experiences are illusory or perceptually veridical, but there is no question that the subjects themselves are often highly consistent in their claims to have experiences; these experiences are ones they cannot 'turn off' at will, the experiences are not the product of some early associative learning (Gray in Robertson and Sagiv, 2005), and there are studies in which the colour experience improves the discriminatory performance of subjects – for example the 'pop-out' effect (see Blake *et al* in Robertson and Sagiv, 2005). Illustrating this last point is a study now almost regarded as a clinical test for synaesthesia – a subject (with the grapheme-colour variant) is shown a matrix of 5s with a triangle of 2s embedded in it.

[29] Some report that they experience the concurrent (colour experience) as transparent, and as being located just on top of the inducer grapheme's true colour. (..) Other participants report that they experience the concurrent as an after-image floating some distance above the grapheme. Often the language used by subjects is confused: Thus, FS, quoted by Brogaard (2012), 'The colours are not out there [..] I think it's related to imagery. It feels like imagining something has a colour. But I am not just imagining it. I think it's perceptual. The phenomenology is sensory.'

Control subjects find it hard quickly to find the triangle, but because synaesthetes see the 5s as one colour, and the 2s as another, the triangle of 2s 'pops out' and can be quickly identified (see illustration below). Other subjects find that synaesthesia facilitates memory, (see Cytowic, 2011, p.80) to the point where subjects have extraordinary powers of recollection. This paper will proceed on the basis that synaesthetes really do have the experiences they claim to have, and are not suffering an illusion; enough brain-imaging studies point in this direction. 'Projectors' show increased activity in the brain's colour regions (V4/V8) when looking at graphemes (Simner 2012).

Bayne is silent on the phenomenon of synaesthesia, and one explanation might be that he does not consider it a challenge to the unity thesis as he has framed it. Recall that it is phenomenal unity that he is focussed on: 'a subject has a unified consciousness iff at a certain time t, every one of their conscious states is unified with every other. [..] The total state will capture what it is like to be the subject in question' (Bayne, 2010, p.12). He might therefore say: in the case of the consistent synaesthete, the visual experience of the number 5 is unified with a (veridical/illusory, but consistent) visual experience of the colour green. The fact that it is hard for the non-synaesthetic normal to imagine what that is like is neither here nor there – there clearly is something it is like for the synaesthete to experience the two together. In this way, he may consider that synaesthesia is not a challenge. However, there are some observations which make it worthwhile discussing the condition. The first point I propose to make is upon a position requiring some careful description.

Synaesthesia and representationalism.

At the outset of his book, Bayne (2010, p.12) stipulates: 'We want a [unity] thesis that might capture the thought that the simultaneous experiences of a single subject must bear some kind of unity relation to each other.' The unity he is committed to is phenomenal unity, wherein the unity, or the overall phenomenal field, is understood as what-it-is-like to be the person in question at that point. Recognising, perhaps, that this might not be entirely satisfactory, he explores the question that this might be captured by appealing to the representational content of phenomenal states. 'Experiential states don't merely have a phenomenal character, or 'what-it-is-likeness', they also possess representational content' (Bayne, 2010, p.47).

A little further into the chapter he says: 'I will adopt the claim that the phenomenal character of consciousness can be fully captured by appeal to the propositional contents of representational states' and: '[It ought to be possible] to analyse phenomenal unity in terms of relations between the representational contents of unified states.'

Bayne declares he is following a number of other philosophers in making these claims, and the wider commitment to representationalism. However, he does face a difficulty in being attached to his specific version. It may be claimed that synaesthesia presents a difficulty that other positions in the Philosophy of Mind may have to deal with, including ones opposed to Bayne's wider position. We may find that the dissociative condition in question again requires us to think about the theory and concepts to be adopted.

Representationalism takes many forms and as we note, is a position in its broadest form which can be endorsed by philosophers of mind from many different commitments. In that broadest form it is the view that mental states can be understood as representing outer objects, or be about them in some sense. This much is shared by Chalmers (2010), Bayne (2010), Rosenthal (2005), Gennaro (2012), but it is just the point at which their versions of representationalism diverge. Unsurprisingly, Chalmers and Bayne subscribe to a version of representationalism which is non-reductive – that is, they hold that phenomenal properties (the qualitative character of some mental states) 'are identical to representational properties, where these cannot be understood without appeal to phenomenal notions' (Chalmers, 2010, p.350). Other philosophers see representationalism as the first stage in a reductivist or naturalistic account of consciousness. If consciousness can be explained in representational terms, and if representation can then be explained by neurology, then we can hope for a complete account of consciousness in naturalistic terms. Here, I hold that synaesthesia

presents difficulties, but perhaps more for the non-reductive version of representationalism than the latter.

The specific form of representationalism endorsed by Bayne is a 'weak' version rather than strong, where weak representationalism is the claim that 'the phenomenal character of consciousness can be fully captured by appeal to the propositional contents of representational states.' This is distinct from strong representationalism according to which phenomenal states just *are* those representational states. This commitment to weak representationalism is a view shared by Chalmers (2004).

As we have already noted, any conscious states which might lack representational content are a challenge to this position; these would be 'phenomenological danglers' to use his own term, (Bayne, 2010, p.51), and any failures in the representational account should be taken as 'diagnostic of failures of phenomenal unity' (Bayne, 2010, p.72).

If the following argument works, I maintain that it concerns representational unity only. Insofar as it does work, in Bayne's terms, the danglers are diagnostic of the failure of phenomenal unity, rather than being outright refutations. To pose a threat, however, we have to establish whether concurrent synaesthetic experiences represent anything or nothing. That being said, we turn now to synaesthesia as a possible example of a phenomenal dangler.

The problem posed by synaesthetic experiences is this: in the case of the green concurrent 'hovering above' the inducer number 5s – the experience of the 5 can be said to have representative content, but it is much harder to say what the experience of green in this situation *represents*. One way of spelling out the argument is captured by Brogaard (2015). She is committed to a version of representationalism which works on the premise that the character of an experience is constituted by the properties of the *objects* experienced, and so she gives a definition thus:

> If E consciously and visually represents o as *F*, then visually attending to o regularly and non-deviantly causes a visual appearance of o being *F*.

She argues that: the experiences of the projector synaesthetes 'involve a property that is not represented as being instantiated by objects.' Take, for example a subject who has a visual experience as of a brownish orange volume in front of a grapheme R printed in black on a page. It never occurs to her that the grapheme *is* brownish orange. 'It follows from [the definition], that there is no property F instantiated by the grapheme that regularly and non-

deviantly causes her concurrent experience. So her experience does not represent the grapheme as instantiating brownish orange [..] so the phenomenology of projector synaesthesia is not exhausted by its representational content' (Brogaard, 2015, p.8).

On this reading of representationalism, synaesthesia presents a challenge; the concurrent experiences are phenomenological danglers. This in turn threatens Bayne's claim that 'phenomenal unity can be analysed in terms of relations between representational contents of unified states' (Bayne, 2010, p.48). Bayne already concedes that danglers such as boredom, nausea, or certain moods, - states whose phenomenal character outruns representational character – might be less easy to set aside than he imagines, and if synaesthesia joins them, we have an additional reason to doubt whether phenomenal character can be captured by appeal to the propositional content of representational states, as Bayne claims. On boredom, he admits that, 'if the experience of boredom has intentional objects, then its objects are not 'open to view' in the way that those of perception, imagery or bodily sensations are.' Elsewhere, he seems committed to a kind of 'object-representationalism' (Bayne, 2010, p.10), so the line of attack that Brogaard makes has purchase here.

There are possible lines of defence, however, which may not involve the kind of representationalism Bayne is committed to. We might pursue the line advanced by Tye (2000) – recognising that, in the case of indiscriminable hallucinations, there may not be objects which our experiences represent. According to his version, the properties which constitute the character of an experience are indeed the properties of objects, but may not necessarily be properties that are currently being instantiated by objects. Instead, they are properties that the experience represents as being instantiated[30]. This version of representationalism would allow that the synaesthetic colour could be represented as being present, but not as a colour belonging to the object. This however, is not the situation as described by synaesthetes, who all say that their occurrent colour 'belongs' to the object experienced. It is not clear, in any event, that this version of representationalism is one which Bayne can endorse.

[30] I owe this distinction to William Fish (2010)

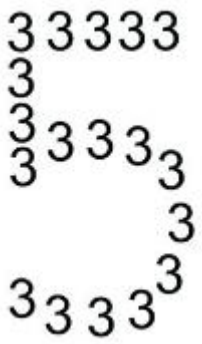
Synaesthesia, representationalism and supervenience

There is a second claim which Bayne appears to endorse, and which is an implication of representationalism – ie that a subject's phenomenal states supervene on her representational states. In other words, there would be no phenomenal change/difference without a representational change/difference (Bayne, 2010, p.69).

In synaesthesia studies, it has been shown that, although the relevant inducer does not change, there is not always consistency in the synaesthetic experience generated. Evidence comes in a study of Cassidy C (Cytowic, 2011, pp.68-70) – 'when Cassidy C encountered the word 'phthalocyanine' his synaesthetic colour experience was determined by the word's individual letters. However, upon learning that phthalocyanine is the name of a vivid blue-green pigment used in paint, his synaesthetic experience became quite different. Those letters which once had a red or pink experience attached, were suddenly projecting blue. The colours of the letters of the word 'synaesthesia' appeared differently to him when the word as a whole was encountered.

Ramachandran and Hubbard presented a stimulus involving an ambiguous letter as per the picture below. The figure can be seen as either an H or an A. When flanked by T, E it is read as an H, and when flanked by C, T, it is read as an A. When the synaesthete subject read the first she saw the ambiguous letter as an H and had the concurrent experience of the colour blue; when she read the second, she saw the letter as an A, and had the concurrent experience of the colour pink (in: Robertson and Sagiv, 2005, p.160).

A final example. In the figure below, a Five composed of threes. When two synaesthetes were shown this, they reported that their concurrent experiences flickered backwards and forwards from red to green as they attended to the whole (5) as compared to the parts (3s).



The point here is this: according to the principle of supervenience in this connection, there should be no phenomenal change without a change in representation. But there would appear to be no change in the object represented, - the ambiguous letter remains as it is, equally the figure above, but there is a change in the phenomenology of the subject. This would again appear to create difficulties for Bayne's position – the phenomenology is detached from the object represented.

We can find the same problem applicable to Chalmers' position here. In a densely argued chapter of Chalmers 2010 (*The Representational Character of Experience*), Chalmers argues for an identity between phenomenal properties and representational properties of a certain kind, admitting that the only obstacle to this identification is the possibility that two distinct phenomenal properties might correspond to the same representational property. He writes: 'This will happen if two perceptual experiences in different modalities [a visual experience and an auditory experience] could have the same content. [It would also happen] if two phenomenally distinct experiences in the same modality [two visual experiences, for example] could have the same content' (Chalmers, 2010, p.347). In this context, the content of a representation is to be understood as given by what it is the representation concerns.

Chalmers confesses: 'It is not clear whether any of these cases can arise'. In the case of the first, he writes: 'One might argue that a visual experience and an auditory experience might have the same content, perhaps representing the fact that there is an object on one's left. In this case, it is far from clear that such states can have the same specific representational content, which is what is needed for a counterexample to equivalence' (Chalmers, 2010, p.348).

It is the view of this thesis that just the kind of example needed here is provided by synaesthetic cases. A substantial minority of synaesthetic cases are sound-to-colour cases, constituting 33% of all types presenting, with Liszt, Rimsky-Korsakov, Sibelius and Messiaen all as exemplars. In all such cases, with the playing of a certain note or chord, a concurrent experience of colour, a visual experience, is occasioned.

On the second issue, ie: 'It is not easy to come up with specific cases of phenomenally distinct visual experiences with the same content', again, cases from synaesthesia studies provide telling and obvious examples. Cytowic and Eagleman (2011) cite a particularly good example – a subject, presented with a 5 comprised of smaller (black) 2s (not unlike the figure immediately above), had a concurrent perception of a red composite 5 if the subject focussed on the 2s composing the 5, and a concurrent perception of a green composite 5 if the subject attended to the figure that the 2s composed. In this case, one might argue that the content of the experience was the black figure as described, but we have a subject whose phenomenal experiences were phenomenally distinct, and who could shift between red and green as attention was shifted.

Both Bayne and Chalmers are evidently committed to positions which track phenomenal properties with representational properties. The purpose of Bayne's version is to work towards an analysis of phenomenal unity in terms of representational unity. The purpose of Chalmers' version is to ground intentionality in consciousness, without allowing for a version of representationalism that paves the way for reduction. It is clear from a reading of their work that they neither of them yet consider synaesthesia worthy of incorporation into their thinking, as the condition does not feature at all in their arguments. Because of this absence, it is not clear that they cannot accommodate it with appropriate changes.

It is also unclear whether a reductive form of representationalism is not also troubled by synaesthesia. It is not easy to weave through the bewildering varieties of representationalism on offer, but the issue is addressed generally by Wager (1999). He makes the position clear: the reductive representationalist must give an account of phenomenal properties which explains them in terms of properties of the external objects giving rise to them. In the case of a coloured-hearing synaesthete who has a synaesthetic visual concurrent of redness on hearing Middle C, this is a problem, since an objective account of Middle C requires Middle C to 'possess integrated reflectance properties', which of course, it does not

(Wager, 1999, p.274). There are some responses to this, but it is not clear that they rescue a reductive representationalism. Brogaard suggests that if synaesthetes 'perceive the inducer as instantiating the concurrent, then the perceptual experience probably ought to be taken to represent the inducer as instantiating the concurrent' (Brogaard 2014). But of course the inducer/object doesn't actually *have* the (concurrent) property in question, so the experience has to be counted as illusory.

Torin Alter also contends that synaesthesia does not undermine representationalism (Alter 2006), but does so in a way flatly at variance with Wager. The core of the argument he presents is built on the case of DS, who is a sound/vision synaesthete. DS is someone who, when stimulated by sounds, also has concurrent visual experiences of shapes. Alter objects: '[..] why couldn't DS's experiences also represent shapes? I see no basis for inferring from (a) DS's experiences represent sounds to (b) these experiences do not represent shapes. Further, it seems intuitively plausible that DS's visual experiences represent shapes' (Alter, 2006, p.5). Alter repeats the testimony of DS in support: 'The shapes are not distinct from hearing; they are part of what hearing is […] that's what the sound is; it couldn't possibly be anything else.' Alter would again be defending a version of representationalism in which a subject experiences a stimulus which can in no way be said to be spatially extended, but which has a concurrent experience of something which is. In spite of DS's testimony, it is hard to reconcile this with any kind of representationalism that would fit neatly into a reductive account of perception. It would seem such an account would need to defend diaphanousness – the subject would, ideally need to 'see through the experience' to the property of an object, or some other distal property: in DS's case, to the sound of Middle C for his concurrent experience. This does not seem to capture what is happening in synaesthesia cases.

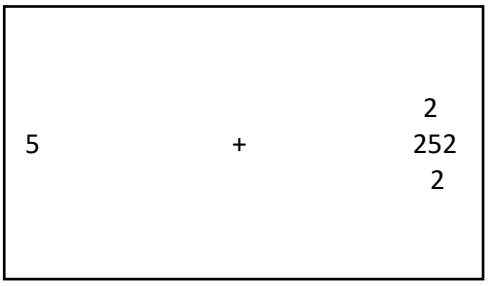
Short of a Brogaard-inspired account in which it is claimed that the concurrent is an illusion, it might therefore be contended that synaesthesia, like the other conditions in this thesis, prompts a reconsideration of concepts in the field, but a fuller discussion of the debate concerning it and representationalism is beyond the scope of this work.

Synaesthesia and the neurological evidence.

We must return to the difficulties posed by synaesthesia for Bayne, and we turn now to an argument from the neurological evidence from synaesthesia studies. Bayne defends his approach to consciousness as being holistic– an approach which seems central to his unity thesis. Holists, he claims, 'hold that the components of the phenomenal field are conscious only as the components of that field. [They] deny that there are any independent conscious states that need to be bound together to form a phenomenal field' (Bayne, 2010, pp.225 et seq).

Bayne dismisses the atomistic approach on grounds that their position requires a mechanism which binds together the 'atoms of consciousness' - a mechanism which the holistic approach does not need to posit. Such a mechanism would be responsible for binding together the elements of consciousness, but Bayne claims there is little evidence of its existence. 'If there were such a mechanism, we would expect it occasionally to malfunction, with the result that the subject would be left with phenomenal fragments, (un-integrated into phenomenal wholes.) To the best of my knowledge neuropsychology furnishes us with no examples of phenomenal fragmentation. [..] there are no syndromes in which the phenomenal unity of consciousness breaks down' (Bayne, 2010, p.236-237).

Here, my claim is that synaesthesia is a candidate for phenomenal fragmentation. To begin with, it challenges Bayne's analysis of phenomenal unity (see above), - we can argue that the phenomenal extra (the greenness) is a stray, not connected to the object (the 5) represented. Studies of the syndrome have more to say in the subject of atomism. The neurological/fMRI evidence coming in from studies of grapheme-colour synaesthesia suggest that it is caused by cross-activation between adjacent brain regions – the LH fusiform gyrus, which is linked to word processing, and the V4 colour area. (Ramachandran and Hubbard in Robertson and Sagiv, 2005). It is still entirely possible that other, even more exotic forms of synaesthesia might be the result of such cross-activation, since there are connections between areas of the brain more geographically remote from each other. But one particular study is of interest. Ramachandran and Hubbard asked some synaesthesia subjects to report their experience of the study below:



Fixating on the cross at the centre, it is easy to see and identify the number 5 on the left hand side, even if not looking at it. The number 5 on the right hand side, however, is 'crowded' with number 2s, and much less easy to see and identify. Control volunteers are at chance identifying this number, when the display is shown. Synaesthetic subjects however, gave a different account. The first question was – did they see the number as coloured? (and did the colour rescue the number from the crowding effect?) Synaesthetes did indeed perform better than controls. But, As Ramachandran and Hubbard comment, 'the really exciting part came when we asked subjects about their experience. Instead of saying that they saw the number, and then the colour, [..] they said "I did not see the middle number – it was fuzzy. But it looked red so I guessed it must be a 5." This suggests they were not consciously registering the number as a number, but it was nonetheless being processed somewhere in the brain at an unconscious level.'

This can be presented as another example of 'zombie systems' perception, - systems that operate on the basis of unconscious representations - and even more tellingly, it seems to show that the processing of the number, in the fusiform gyrus takes place, and evokes the appropriate colour early in the sequence of sensory processing, and before the number itself is consciously perceived.

This suggestion of there being different stages in sensory processing sits badly at odds with the holistic picture that Bayne advocates. Further evidence that the attributes of the visual world are processed at different speeds and with different pathways is provided by the work of Zeki, which will feature in the next section. In Zeki 2003, he claims that it has been

shown, for example, that colour is perceived before motion by about 80ms. Colour itself is processed after location, and location precedes orientation[31].

The robust identification of atoms, or composite parts, of experience is a difficulty for Bayne. He throws down a last gauntlet (Bayne, 2010, p.248). 'The real pressure on atomism derives from the fact that conscious states occur as the components of a single phenomenal field. How might atomism account for the phenomenal unity of consciousness? They can deny that consciousness is unified […] but this flies in the face of introspection.' But even if we discount the evidence from synaesthesia, Prinz points out that this cannot be challenged: disunity can never be introspected, since attending to two stimulus features to see whether they are unified would unify them. We are faced with a refrigerator light illusion – unity is always present simply because it is present when we look for it (Prinz 2013).

In synaesthesia then, we find amongst other things, evidence working against the holism of the unity thesis.

Synaesthesia and Access Disunity

The Unity Contention faces a second, more direct challenge from Synaesthesia. The line of attack is provided by Bayne himself (2010, ch 5), where he identifies the criteria by which we can evaluate the unity thesis:

'A [more potent] line of argument for phenomenal disunity concerns the uses to which a subject's conscious states can be put. In general, the contents of a creature's conscious states are available to the same range of cognitive and behavioural systems. But suppose we came across a subject who appeared to be in two conscious states at the same time (e1 and e2) where the contents of these two states were not available to the same systems of cognitive and behavioural consumption. It would be tempting to take this selective accessibility – this breakdown in access unity – as evidence that e1 and e2 were not phenomenally unified' (Bayne, 2010, p.107).

Clearly, what is understood as consuming systems in this context is central to whether synaesthesia provides us with a case of phenomenal disunity. Bayne is not altogether clear.

[31] This hierarchy of processing helps to explain the sequence of events as I remember them in the incident of the snake. I reacted to the processing of the shape (?) of the snake, before the whole thing was bound up together and brought to consciousness *as a snake*.

He gives two examples, suggesting that different experiences might be available for verbal report, or available for memory consolidation. He also discusses the problem of individuating consuming systems (Bayne, 2010, p.110) in a section in which he gives himself a line of retreat. In it, he makes it clear that he is thinking along the lines of systems identified as neuropsychological processes, which produce a specific behavioural response. He discussed the suggestion that such systems might be individuated by types of motor responses, but considers that there might not be a one-to-one relation between consuming systems and motor responses, as 'a single consuming system might be implicated in multiple behavioural responses.' He finally concludes that individuating systems is an empirical matter, but that arguments from access disunity are 'hostage to fortune, for our current assumptions about the borders between consuming systems might be quite wide of the mark.' As we advance our knowledge of cognitive and behavioural systems, we may find that what we once thought were two discrete consuming systems, might turn out to be linked in some way. It is the difficulty in individuation of consuming systems, and the possibility of future advances that gives him cover.

With such prefatory considerations in place, we may imagine that Bayne would agree that consuming systems might be individuated by neurological processes involved, by different cognitive accomplishments, especially if these were accompanied by different behavioural responses. If so, then the following empirical evidence taken from synaesthesia studies may provide us with evidence of access disunity and therefore phenomenal disunity.

Firstly, there is now considerable neuro-imaging evidence that different brain areas are in involved in synaesthetic perception in the same modality. Thus, for those synaesthetes who have a concurrent experience of a colour, when triggered by seeing a number (colour-graphemic synaesthesia), '[In synaesthetes] synaesthetic colours were associated with activation in area V4/V8 in the left hemisphere, whereas visually detected colours activated the corresponding areas in the right hemisphere' (Triesman in: Robertson and Sagiv 2005, quoting Nunn *et al* 2002).[32]

[32] Interestingly, this study established that while the V4/V8 colour areas of the brain were activated during the synaesthetic experience, the primary visual cortex (V1) was not active, which on this one point, brings synaesthesia into the same neurological territory as blindsight.

Evidence is still coming in, and the picture is confused by the wide variety of synaesthetic experiences, but we can conclude that it is likely that different and segregated consuming systems at the neurological level are responsible for synaesthetic and physical colour perception. This is even more likely to apply in synaesthetic cases where different sensory modalities are involved – for example in phoneme trigger – colour concurrent synaesthesia cases. In the same 2002 study by Nunn and colleagues, it was found that, for a spoken word trigger – concurrent colour case, spoken words, which would presumably have been processed by an auditory/neurological consuming system, was associated with brain activation in an area specialised for colour vision – an entirely separate consuming system.

Thirdly, and lest it be argued that these cases do not involve separate consuming systems since they are commonly perceptual, there are synaesthetic cases where what distinguishes the consuming system in question is a role in interpretation or involves meaning. Cytowic and Eagleman (2009) discuss a case in which a phoneme trigger-colour concurrent synaesthete Cassidy C had a different colour experience of the word Phthalocyanine on learning that the chemical was a vivid blue-green pigment. The synaesthetic experience before the meaning of the word was learned, was of letters which were predominantly red or pink in colour. After the meaning was disclosed, the red and pink were replaced by blues, and the green became more pronounced in the synaesthetic experience. The case would suggest that the phoneme was processed by the usual visual consuming system, in both before and after cases, but at some point a different consuming system, one involving meaning, was involved in producing the concurrent. Recall again, the case of the ambiguous letters in the presentation below:

Cytowic and Eagleman note - the middle letter is graphically identical in the two words, yet most synaesthetic (and non-synaesthetic) subjects will automatically interpret the first as an H, and the second as an A. in the case of the synaesthetes, however, a different colour concurrent is triggered according to the context. This would indicate that it is not the shape of

the letter on the retina that generates colour, but the consciously attended concept of the letter' (Cytowic and Eagleman, 2011, p.75).

In these cases, we may argue that there is one experience, e1, of the phoneme itself, alongside a second experience of colour, e2, which has been routed, not only through a discrete neurological area, but also through interpretive mechanisms.

Finally, a case using the very example of a different consuming system provided by Bayne himself – memory consolidation. A number of studies have shown that memories are more easily formed and consolidated by the synaesthetic experience. Merikle *et al* (2002) found that, on showing a random matrix of numbers to a synaesthete, he could more readily memorise the numbers than non-synaesthetes because of the extra clue provided by the colour. Blake *et al* tell the story of WO who 'relies on his colours for remembering technical terms as well as for recalling telephone numbers' (in Robertson and Sagiv, 2005, p.48). But all of these cases are trumped by that of Daniel Tammet, who combines his synaesthesia with other conditions, but who has a prodigious power of memory as a by-product of him having concurrent sensations of colour, shape and feel, with numbers. He can, for example, recite the expansion of pi to 22,514 digits. (see Baron Cohen 2007, Bor *et al* 2008, and Matey in Bennett and Hill 2014). In Tammet's case, V4 is not activated by the visual presentation of numbers and Bor *et al* conclude that he has a rarer, more conceptual form of synaesthesia.

Hupe (2012) had similar results to the Tammet case with a further study, concluding that, in some cases, synaesthetic colour experience lies not in the brain's colour system, but instead the area of activation is more distributed, over the right retrosplenial cortex, as opposed simply to V4/V8. This cortex make it an 'ideal candidate region to explain various synaesthetic associations [..since] it has functional and anatomical links with the memory system [..] and is involved in emotion processing' (Hupe 2012, p.1631). This difference between the neurologists in locating the area of activation provides no comfort for Bayne, since it is clear synaesthesia involves a consuming system or systems beyond that simply of perception.

In these cases, it may be that the consuming systems are not separate or distinct, but singular, complex and hard to individuate, in the way that Bayne anticipated. But more likely, the cases are evidence of different experiences associated with distinct consuming systems, neurologically, and psychologically, involving not just perception and concept interpretation, but also memory formation and recall, producing interesting behavioural responses. On this

account, synaesthesia provides an example of access disunity and phenomenal disunity. Tied together with other observations suggesting that synaesthetic experiences are phenomenological danglers, we may conclude that, at the very least, Bayne must acknowledge and find some way to meet the challenge of synaesthesia. We now turn to a fuller treatment of the challenge to the unity thesis presented by those who lean towards atomism.

Bayne's holism, and the case for atomism.

Bayne observes that the distinction between holism and atomism in their approach to consciousness is one which cuts across established positions on consciousness, but that 'atomistic accounts of consciousness are at odds with the unity thesis, and (recommends) that we should adopt a holistic conception' (Bayne, 2010, p.228). An atomist would assume that the phenomenal field is comprised of discriminable contributory features, which are independently conscious, while a holist would hold that components of that field are conscious only as part of that field (Bayne, 2010, p.225). Bayne, unsurprisingly, nails his colours to the holist mast. His commitment to holism has a methodological implication – 'a holist will recommend a top-down methodology' (Bayne, 2010, p.226). He does not say so, but this approach makes an assumption of unity as he has defined it, and is clearly predisposed against the idea of unconscious perception, or dissociable modules of consciousness which are the focus of this work. Such a presumption would start from the premise that there are no phenomenal 'danglers', or fragments. At the very least a holist would be disposed to resist the empirical evidence for them.

In debating the case for holism, Bayne asserts that, (if an atomistic picture is right), perceptual features – colour, motion, shape and so on, 'ought to be capable of independent existence within the stream of consciousness, and that doesn't seem to be the case' (Bayne, 2010, p.233).

Going further, he claims that, for atomism to work, there needs to be a binding mechanism which can bring together the constituent units of phenomenal experience for there to be a unified consciousness, but that many atomists seem reluctant to posit such a mechanism. Moreover, there is little evidence of its existence. A few pages later: 'what we know about the role played by local neural activity in the generation of consciousness does

not provide any support for atomism' (Bayne, 2010, p.241). Accordingly, he makes the claim that: 'we should regard total states as the basic units of consciousness' (ibid p.244).

It is the view of this thesis that, the more we have come to understand the dissociations studied, and the neurology involved, the more we have come to identify the functionally specific regions of the brain, and their interactions – inhibition, excitation, feedback, feedforward etc, the more we might need to revise our philosophical concepts. Blindsight would appear to provide us with excellent evidence of the dissociability of all these aspects. To add to the list already given, Carlson (2013, p.4) gives an example of how a patient, protesting his blindness after a stroke, was nevertheless able to orient his hand to a cane presented in his blind field. GY was able to point in the direction a moving light stimulus was projected. The emotional valence of faces was correctly registered etc. Bayne does claim that 'there is in general no perceptual experience prior to feature binding' (Bayne, 2010, p.233) but this can only resist the evidence above if we understand experience always to be conscious.

To make this clear, the *Unity of Consciousness* makes a claim about the basic unit of consciousness as already composite – as a quilt, to use Bayne's metaphor (ibid p.244), but that this is the basic unit of consciousness, and nothing is prior to it. This thesis takes the view that this picture of the structure of consciousness has more evidence against it than Bayne is ready to concede, and that the concept needs revising, indeed jettisoning.

In fact we side with the view made by Zeki: [progress in finding the] "neural correlates of consciousness will remain elusive until we acknowledge that consciousness is not a unity" (Zeki, 2003, p.214). Instead, Zeki may be right to claim that there are multiple micro-consciousnesses that are distributed geographically in the brain, with different processing times.

Neurology aside for a moment, it should be noted that Bayne himself provides us with no account of what it is that makes it the case that we are having a unified experience of seeing-and-tasting the wine. There appears to be nothing advanced to explain why we have a Kantian smooth and unified experience, instead of a Humean tumult. As Platchias asks: 'Is it another intrinsic property of the (allegedly) essentially conscious mental states? In Bayne (and Chalmers') view, it just seems to happen that one is having a unified experience: it is a happy coincidence' (Platchias, 2011, p.89). This is an assumption at the heart of their position, and nothing has been advanced to argue it. If it is made, it would seem to be made

on the strength of introspection, and we have identified the weakness of that line of thinking. The lack of empirical support confirms the view throughout this thesis that the kind of dualism Bayne and Chalmers promote is at one remove from empirical confirmation.

Zeki's Microconsciousnesses

Let us briefly make the case for distributed micro-consciousnesses. Zeki argues that the foundation for such a theory is to be found in the geographical and functional separation of areas of the brain specialising in different roles. The identification of such areas and their roles is not a matter of controversy, but of empirical fact.

Working from a study of the visual system in the brain, Zeki reminds us of studies showing that there are recognisably different geographical areas responsible for motion (V5) and for colour (V4) in the brain. 'A lesion to one area does not invade or disable the perceptual territory of the other' (Zeki, 2003, p.214). And so, a lesion to V5 produces akinetopsia, while still being able to perceive colour, and a lesion to V4 produces achromatopsia, while still being able to see and be conscious of motion. Zeki argues that each system is a distinct micro-consciousness which needs to be bound or integrated into a whole, and it is this binding that leads to the conscious experience – or, 'what-it-is-like'. (This study focusses on the visual system; a wider study (Koch 2016) summarises the different ways lesions to the brain can affect cognitive performance, and produce disabilities of a specific character, each identifying candidates for micro-consciousnesses. Many of these have featured in this thesis[33].)

Zeki's claim is that these two mechanisms are features in a hierarchy of consciousness, which culminates, at its highest level, in unified consciousness, but that it is essential these micro-consciousnesses are recognised as prior to unified consciousness, or we cannot make any progress in neuroscience.

[33] There is some comfort for Bayne in this recent (2016) article. A candidate for the location in the brain in which constituent units of consciousness could be brought together and fashioned into a unity, as per the atomistic picture, was the fronto-parieto cortices. Koch *et al* suggest that other more recent findings put this in doubt. While they now lean towards the 'posterior cortical hot zone', the truth of the matter is that there is still no confirmed location, and we may have to get used to the idea that, because of the complexity and interconnectedness of the brain, there may be more than one.

Bayne however, does have a reply, and while conceding that the neural mechanisms underpinning consciousness 'are distributed across a multiplicity of cortical loci' (Bayne, 2010, p.239), he does not follow Zeki in concluding that this demonstrates the atomic structure of consciousness. In fact, he argues that the various systems are unconscious and are only manifest in consciousness once unified. The systems register different features (colour, shape, motion) and are bound together to produce the unity of experience, but that 'there is in general no perceptual experience prior to feature binding' (Bayne, 2010, p.233).

Bayne allows the following replies: firstly, the atomist can simply deny that the atoms of consciousness are unified, and that the so-called unity of consciousness is no more than an illusion. (He sets this aside quickly, but we note that if introspection was the main reason for concluding that consciousness is unified, and if there are powerful reasons for thinking that introspection is unreliable, we may not want to dismiss this too quickly). The second reply available to the atomist is to posit some neural mechanism responsible for unifying the atomic elements of consciousness – to bind them together to bring them into the total phenomenal state of 'what-it-is-like'. Bayne argues that the holists do not need to posit such a mechanism – it is not needed in their account. (Though note Platchias' point above.) Bayne does remark, however, that 'there is little evidence of the existence of a mechanism for binding, and that, if it did exist, we might expect it occasionally to malfunction, and produce phenomenal fragments' – surviving atoms of consciousness no longer integrated into phenomenal wholes. And, to the best of his knowledge, studies across the realm of neuroscience have nowhere shown phenomenal fragmentation. While other forms of unity have broken down (we have explored problems for representational unity, for example), 'there are no syndromes in which *phenomenal* unity breaks down' (Bayne, 2010, p.237).

What kind of replies are open to those who lean towards Zeki's position? Zeki himself makes two replies: firstly, raising the case of GY, one of our primary blindsight subjects – in GY, there is blindness, because of damage to V1, but there is also a 'crude but conscious vision for fast moving high-contrast stimuli, perception of which is mediated by (direct access of information to) V5' (Zeki, 2003, p.214). Zeki seems to argue that this is an atomic processing-perceptual site.

(We might ourselves add: there need not be a single mechanism solely responsible for binding – there may be many, operating within a sensory modality, and between them. Bayne's malfunction point seems to be reliant on there being a single mechanism only; if

there were several, then the malfunction of one may to some extent be compensated for by others.)

Zeki's second point is this: he claims that the processing of motion (in V5) and the processing of colour (in V4) is 'significantly uncorrelated' (because geographically distant, and distributed in time). He found, in conjunction with others (Moutoussis and Zeki 1997, Zeki and Bartels 1999) that, because of these distinctions, subjects consistently mis-bind colour perceived at time t to motion perceived at time t-1. 'It follows that, over very brief time windows, the brain does not wait for each area to complete its processings, rather, it simply binds what has been processed and reached a perceptual level.' Zeki goes on: 'This in turn suggests strongly that binding is a post-conscious phenomenon' (Zeki in Velmans and Schneider p.584). A further indication that binding might be post-conscious comes from experiments which demonstrate that the binding of colour to motion comes after the binding of colour to colour, or motion to motion (Bartels and Zeki 2006). This would again point to micro-consciousnesses at the level of elementary visual features.

We might add the further replies: reflecting upon the claim that there are no examples of a breakdown in *phenomenal* unity, one is tempted to ask – how would we recognise such a breakdown if one were to happen – what would it be like to have such a disunity? There is an obvious point to make: if phenomenal unity is understood in terms of 'what-it-is-like' – if it is inseparably linked, it might make no sense to ask – ''what would it be like to have phenomenal disunity?' But the reply might be – phenomenal disunity is a condition in which experiences are had sequentially, but not in parallel (I have the experience of drinking coffee and I have the experience of listening to music, but I do not have the experience of drinking-coffee-and-listening-to-music-at-the-same-time.) The critic might reply – this may not be so very far from the truth. Resist the siren call of introspection as a testament of unity; the basic unit of consciousness is more atomic; we bind micro-consciousnesses into a macro-consciousness of combined attributes; attention singles out macro-consciousness of attributes from a rich potential array of possible experiences.

The exploration of various agnosias and pathological conditions we have made tends to endorse the position taken by Zeki insofar as each singles out registrations and processing short of full phenomenal consciousness, and for further progress in identifying neural correlates, it may indeed be wise to abandon the concept of the unity of consciousness, and the holistic approach favoured by Bayne and Chalmers.

Conclusion.

It is this view that Bayne's central contentions, which we began with: that all consciousness is phenomenal; that it is necessarily unified, and that we never have disunified experiences – struggles to hold up under challenge. The split-brain syndrome seems to provide evidence that consciousness can sometimes be disunified, and there is a case for saying that synaesthesia poses a different challenge, which he does not consider. Some recent developments in neuroscience, and studies in that field have posed further questions.

In the view of this work, Bayne is right to consider the contribution of neuroscience, and empirical work to the study of consciousness – this is what makes his work so appealing. (For this reason, Prinz (2013) says it is easily the best thing that has been written on the subject.) But there remains a tension in his work: his concessions to the empirical seem generous, and the evidence for a contrary position mounts up, but he resists it on grounds which are not entirely convincing. This is especially so in his chapter on split brains. Further, introspection is still the deciding factor when considering what credibility to give to reports, and affords the primary intuition about unity. Although he admits that providing introspective justification for the unity thesis is 'challenging', (Bayne, 2010, p.92) ultimately, introspection is the main way in which we gain an insight into the structure of consciousness. His main motivation for the thesis is phenomenological: how things seem. Bayne admits that the inference from seeming unified to being unified is risky, but he makes it anyway.

If we recall, one of Bayne's original motivations for the unity contention, was to put a constraint upon certain theories of consciousness, and if we think his contention fails, then some of those theories come back into consideration. These would include: Higher Order Thought theories, Prinz' attention-engendered AIR theory, or scientific theories looking for a neurological basis for consciousness.

Higher order thought theories do at least give us an account of why, in normal cases, our experience appears structured and unified – As Platchias puts it: 'It is plausible to suggest that having a higher order thought about these qualitative states smooths them out so they are not experienced as particulate or sporadic, but rather as ultimately homogenous and unified' (Platchias, 2011, p.89). But before we consider other candidate theories, there is one further task to complete.

Chapter 5

Ian Phillips and the exchange with Ned Block

This thesis now turns to a third target – the campaign fought by Ian Phillips against the idea of unconscious perception in article and in conversation with Ned Block, during the development of this thesis (Phillips 2016, Block 2016, Block and Phillips in Nanay 2016, and Peters, Kentridge, Phillips and Block 2017). Phillips has emerged as a critic of the idea of unconscious perception in the tradition of Chalmers and Bayne, but in conversation resists being pigeonholed as a partisan in the wider debate over what lessons can be drawn from blindsight and the dissociations. Among several lines of criticism he explores, he picks up a line of thinking from Campion *et al* (1983) and forcefully argues that the experiments establishing a dissociation between function and experience are flawed by a common problem – the so-called problem of the criterion. In doing so, he is asking: is blindsight properly unconscious? In discussion of other candidate dissociations, he brings other criticisms to bear, raising the question whether they constitute perception according to the definition with which we plan to work.

As we recognise, Phillips resists being typed as a neo-cartesian, but his arguments are ones that might assist that cause, and, indeed he admits as much. In defending the view that the dissociations have a lesson to be accommodated by philosophy, we will look closely at Phillips's articles and attempt to referee his debate with Block, with whom this thesis will have some sympathy. As is evident from the exchange, it is a skirmish in Block's attempt to preserve his Phenomenal/Access Consciousness distinction.

Before we begin, a few prefatory comments. In this debate we will be dealing not with the definition of phenomenal character, or qualia, — the more fundamental questions we dealt with in previous chapters. In the chapter on Chalmers, we were at pains to make it clear that the target we were tilting at was the one Chalmers presents – the idea that perception necessarily involves phenomenal character, and the claim that the dissociations could not by that token have a bearing on this. In the questions Phillips raises, the issue is not closed off by this kind of definition, and centres on the notion of unconscious perception, as opposed to unconscious phenomenal character. As we explore it in this chapter, the issue will still be one of whether blindsight and other dissociations pose a problem for neo-cartesian philosophers.

At the outset, it would seem that the debate would be tractable empirically, but as we shall see, ultimately, the choice of definition is crucial.

Phillips adopts a definition of perception advanced by Burge (2010), who in turn borrows upon Kant (Burge, 2010, p.368, referring to the *Critique of Pure Reason* A320/B376). Phillips regards this as a suitable battleground on which to contest Block, since he maintains that Block himself adopts it, and it offers a number of lines of enquiry. We can set aside any consideration of the merits of the definition at this point (some have defended subjective representation as a definition of perception, but I do not think we need to pursue this.)

A careful look at the definition is a convenient place to start, since Burge himself considers it possible that blindsight may be regarded as one of a number of cases of unconscious perception within the definition as he presents it. To be clear, and for the purposes of his argument, Phillips largely accepts the definition, (at least at first) but then contests the view that unconscious perception is permissible within it. Let us in consequence identify the position taken by Burge in some detail – it is the battleground on which Phillips and Block both manoeuvre.

Burge's Definition of Perception

Burge defines perception as constitutively a matter of 'objective sensory representation by the individual' (Burge, 2010, p.368). A case of perception must in consequence have objective content, and be attributable to an *individual*. There is a further question as to whether attribution to an individual means that the content is available to a central co-ordinating agency. Burge allows that subsystems, (for example, the visuo-motor system,) might perceive, but these would themselves be cases of the individual perceiving. An essential part of what it is for an individual to perceive is for that perception to be available such that individuals can 'represent goals of, obstacles to, or threats to their activities and act accordingly' (Burge, 2010, p.370).

For Burge, *objectivity* in perception is achieved by the exercise of perceptual constancies which are: 'capacities systematically to represent a particular, or an attribute as *the same* despite significant variations in registration of proximal stimulation' (Burge, 2010, p.408). Burge offers some examples: Size constancy in a visual system is the capacity to

represent the object's size as the same, even as the stimulus from the object […] moves closer to or further away from the viewer. Shape constancy is a capacity to represent a given shape under various stimulus and perspectival conditions. A circular pattern can be seen as circular whether it is viewed head on or at an angle.' I present this definition in some detail as it will be useful later.

Burge contends that, within this definition of perception, there might be unconscious cases of it. 'There is considerable evidence that individuals, not merely subsystems, have unconscious perceptual states' (Burge, 2010, p.374). 'Blindsight patients perceive environmental conditions. The perception involves perceptual constancies […] The perception guides action. There is strong reason to believe that some of these patients lack phenomenal consciousness in the relevant perceptions.' (There is clearly something of Block's P-Consciousness and A-Consciousness distinction in this.)

As a reminder, Phillips accepts this Burgean starting point, but does not agree that it allows for unconscious perception – specifically that blindsight does not meet the conditions for perception – one of his complaints is that constancies are not involved. Other dissociations are addressed with different objections. One key question at issue will be the question of whether the definition of perception as something attributable *to the individual* necessarily involves a further feature – that the content of the perception is available to central co-ordinating agency.

In summary, Phillips will raise some fundamental questions: First: if blindsight and the other dissociations are cases of unconscious perception, then they must be demonstrably cases of genuine perception, and second, they must be unequivocally unconscious. Block points out that establishing that either one of these conditions is met works against the other: 'the best evidence for a lack of consciousness would be if there was absolutely no effect on the visual system – but that would not be perception. (On the other hand) any effect on vision could be used by opponents to argue that visual registration was not really conscious' (p. 10 Block and Phillips 2016 on-line paper and now in (ed) Nanay 2016).

Phillips does not want to advance the idea of 'unconscious perception' as oxymoronic, (which he claims Campion *et al* do, when they argue that any kind of perceptual sensitivity is evidence that stimuli are conscious, and that therefore 'all subjects in all studies, are by conventional criteria, aware of stimuli to some extent' (see Campion *et al*, p.480)). To work with such a definition would be trivially to block any interpretation of dissociations as

unconscious perceptions. However, if 'unconscious perception' is not to be an oxymoron, then there must be some practical way in which each feature can be established *as* a feature of some case or condition.

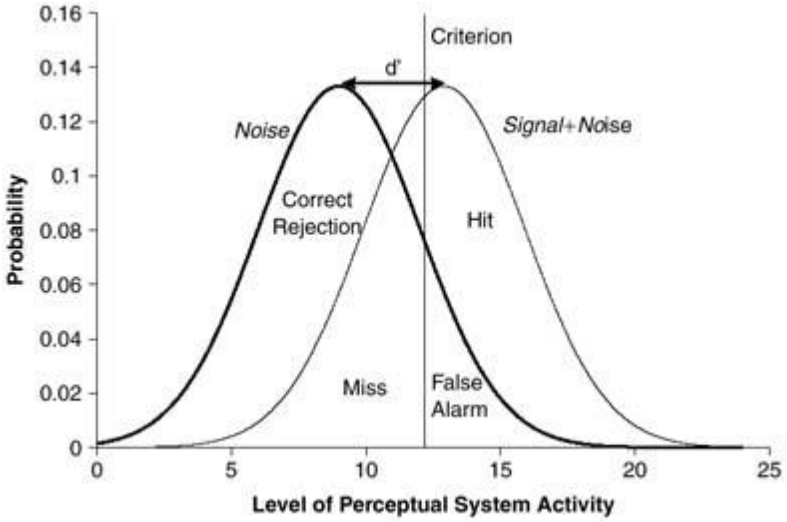
As we shall find, Phillips addresses both questions, asking: are blindsight and other dissociations properly unconscious? And, do blindsight and other dissociations satisfy some definition of perception, and therefore – are they perception? To use his own terms: do they constitute perception of *the same fundamental kind* as that involved in ordinary perception.

In dealing with such matters, Phillips is of course raising important questions about what we mean by terms such as 'consciousness' and 'perception', but throughout the discussions he concedes that the questions can in principle be resolved empirically, and so empirical detail about the paradigms and the various dissociations – especially recent developments – is germane. Let us turn to the first of the two questions Phillips raises.

Phillips and the problem of the Criterion: Are the examples of unconscious perception truly unconscious?

The first issue that Phillips takes with the claim that blindsight etc are cases of unconscious perception is with the idea that they may not properly be cases of *unconscious* perception. As acknowledged above, this is a point first made in the Campion *et al* paper, although in that paper, it is not developed fully. The claim is briefly put thus: the fact that a blindsight subject reports that he did not see a stimulus, or has low confidence of seeing a stimulus, does not mean he had zero subjective experience of it. Some experience may therefore be the root cause of the response of the subject, and their protestations may not be relied upon.

Here is the detail of the argument: Borrowing from Green and Swets (1966) signal detection theory, (as Campion *et al* also do) a subject's responses in any task of the kind explored by blindsight studies are determined by two things: first, the subject's perceptual sensitivity to the stimulus, and secondly, their response criterion. The latter is the subject's own threshold at which the subject is sufficiently aware, or sufficiently confident of the stimulus for him to give a positive response.



(In the diagram above the two curves are response distributions – one for 'noise' (i.e. no significant stimulus present) and the other for signal+noise (i.e. stimulus present). The line marked 'Criterion' is the threshold at which the individual subjectively identifies that a signal is present. d´ is a measure of the subject's perceptual discriminability. The greater the distance depicted by d´, the more it is the case that the subject can identify the presence of the signal within the noise. The distance d is measured in multiples and fractions of standard deviation).

The problem advanced by Phillips is that this response criterion is subjectively decided, not fixed, and may be subject to manipulation. It is not and cannot be objectively located. The threshold at which a subject might begin to report a stimulus may be affected by factors such as fatigue, or preconceptions about the importance of the trial, or how the subject sees the importance placed by experimenters on the trial, and so on. For a variety of reasons, a subject might adopt a conservative criterion, which would mean she might respond 'no' (I haven't seen anything) when her sensory response in the trials is actually due to a stimulus detected, by some low or degraded experience. In other words, she may actually have some awareness of the stimulus, even though she reports otherwise. In such cases, the line representing the criterion in the diagram is further to the right.

The problem of the criterion re-opens the debate around type 2 blindsight cases, in which subjects do report some experience, though we have in an earlier chapter discounted this as unconvincing proof of their experience. (Recall that what GY and others said was that they had 'a feeling' something had moved, like a hand passing in front of closed eyes).

The problem, however, reaches even into type 1 cases. A subject may say, with utter conviction, that they do not see, believe themselves to be totally blind, and assert that they are simply guessing, when asked to make a discrimination, but according to the objection, their success in making a successful discrimination could merely be down to a conservative criterion, and an unreported low-level awareness. As we have seen elsewhere, there are factors which might incline them to adopt such a criterion, as Phillips records. To take an extreme example, if a subject is offered £1 for every time they correctly say that a stimulus was present, but fined £100 for every time they incorrectly say a stimulus was present, the subject will automatically adopt a conservative criterion, and regularly deny that a stimulus was present even in cases when they may have clear perceptual sensitivity. The fines would incline subjects to shift their threshold to a point where it would be represented by a line to the right in the diagram, and under-report their experience.

Factors other than financial considerations might have the same effect, with the consequence that in all cases of blindsight (and other dissociations) the subject could be in possession of under-reported experience, even if she is emphatic that she sees or experiences nothing. So we should: 'reject any approach for distinguishing conscious from unconscious perceptual processes that is based solely on the subjective reports' (Phillips, 2017, p.428 quoting Reingold and Merikle, 1990, p.17-18).

In such circumstances, Weiskrantz' insistence that he trusts his subjects to report their condition accurately, looks beside the point. A subject could be utterly insistent on the absence of any experience in tests, but yet nevertheless have some low-level experience at work. The phenomenon of blindsight may not be truly blind. To put it slightly differently, the claim is that some consciousness is at play, and the perceptive capacity of the subject is not truly unconscious.

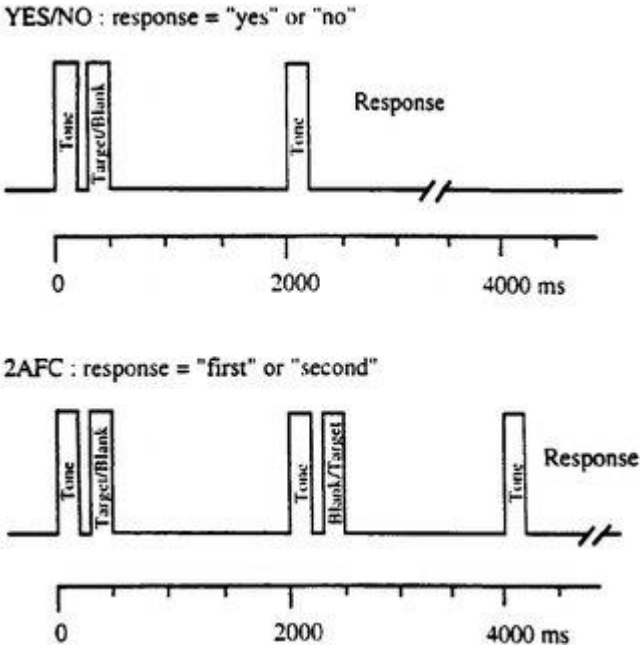
Phillips applies the same criticism to the neglect case of PS as studied by Marshall and Halligan (1998). Their study concluded that PS failed to make a distinction between cards depicting line drawings of a house where some had flames emerging from the house on the left of the cards shown and some did not. When asked further, which house would she prefer to live in, she replied that that was a 'silly question'. But when forced to choose, she reliably picked out 9/11 houses where there were no flames involved. Marshall and Halligan concluded that she unconsciously perceived the flames emerging from the relevant cards, and chose accordingly. Phillips however comments that the paradigm has not ruled out the

possibility that PS was operating with a conservative criterion and had some degraded conscious perception below the response threshold, or that she may have perceived some low-level pre-attentive feature, such as asymmetry. PS was asked a y/n question whether she detected anything wrong with any of the cards. She replied: 'no', but Phillips dismisses these questions as 'biased' and potentially subject to the criterion problem (Phillips, 2016, pp.432-433).

Other dissociations may also be subject to the same criticism where y/n or same/different (s/d) questions are put. This is indeed strongly suggested by studies looking at tasks which use s/d discriminations (cited in Phillips, 2015, p.428). He records the conclusions of Macmillan and Creelman 2005: 'participants seem naturally to adopt a preference for "same" in hard to discriminate stimuli'. Neglect patients might be particularly subject to this, given that, (as Phillips suggests) 'stimuli presented to their sides showing deficits are stimuli plausibly near threshold', and 'we should therefore anticipate strong biases towards "same" responding.' PS might, therefore, be inclined to say the pictures of the houses presented are the same as each other, and her response to the question of where she would rather live is the product of some degraded experience under-reported because of her conservative response criterion.

The question of s/d, or y/n tasks being more prone to the influence of a conservative response bias is explored by Phillips in a longer section referencing two papers by Azzopardi and Cowey (1997, 1998) concerning a study of GY. These studies were alive to the criterion problem, and sought to circumvent it by a change in task. Alongside a series of y/n tests (did you see the stimulus?), GY was asked to reply to a two-alternative forced choice (2afc) test. He was given 2 intervals, indicated by an auditory tone across 4000ms, (each interval = 2000ms). The stimulus was presented for 200ms, but randomly varying with a blank within each period. GY was then asked to say which temporal period the stimulus was most likely to have been presented. (See the diagram below). This task was chosen for its minimal bias (d' = 0.01 compared to d' = 1.867 in the y/n task)

.



Azzopardi and Cowey's findings were instructive: GY's sensitivity (i.e. percentage success rate) was significantly higher in the 2afc test than in the y/n task, scoring percentages consistent with blindsight findings in other paradigms (between 81% and 96% correct depending on the contrast of the stimulus used. In the y/n test (the test subject to the criterion problem) GY's sensitivity ranged over 66%-75% depending on contrast.

The study showed that GY had a moveable response criterion, depending on task given, but that success in discrimination was unimpaired – indeed improved – by using a bias-free task. A similar study using moving stimuli did not show this effect; instead success in discrimination 'did not differ significantly between y/n and 2afc detection.'

Azzopardi and Cowey conclude that while possible response bias presents a difficult problem for any investigation of visual awareness, 2afc responding is bias-free. GY showed significant bias in the y/n task, but in the bias-free task, he showed stimulus-sensitivity which could not be dismissed as degraded normal vision unreported due to response bias.

However, this does not settle the matter for Phillips. Consistent with his strategy in dealing with other dissociations (as we shall see later) and having been unable to make one charge stick, he explores another confound. His main objection is the argument that Azzopardi and Cowey have only shown that GY's sensitivity is not the product of degraded *normal* vision. His point is that they have not shown that GY's sensitivity is not the product of degraded *conscious* vision (Phillips, 2015, pp.438-439). For the purposes of clarity: Phillips is here suggesting that there may still be some unreported conscious experience to

account for GY's sensitivity. He writes: '[some] deny that there is any simple association between detection theoretic thresholds and consciousness, typically granting only that $d' = 0$ is a *sufficient* condition to the absence of consciousness. [..] The crucial point is this: where a subject performs well on an unbiased task (eg 2afc) but poorly (as judged by percent correct) on a biased task, this pattern of responding does not entail unconscious perception unless we make the very strong assumption that consciousness is found only above the subjective threshold. If we reject that assumption, we need to consider seriously the possibility that the response pattern reflects conscious perception unreported due to a conservative response criterion' (Phillips, 2017, p.428).

This returns us to the status of the reported testimony of the subject ('I did not see anything'), and doubt on this question potentially affects every paradigm. Phillips explores the case of GY in considering this question (2017, pp.440-1), and says of him: 'His experiences are the visual equivalent of the faintest of murmurs, he sees through a glass darkly' and concludes: 'it would be rash to conclude that GY lacks any form of conscious experience in other trials from this data.'

Response to Phillips' contention

How should we respond to this line of attack? I offer four replies.

The first addresses the question of individual results being possibly compromised by an unstable response criterion (by, for example, fatigue) or advance expectation (criterion 'jitter'). It applies especially to perimetry – i.e. scotoma-defining tasks at the outset of any study of blindsight, which necessarily are y/n in character. The argument takes its lead from commentary upon Descartes' First Meditation. In it, and in relation to the veridicality of sense impressions, Descartes writes:

> Reason convinces us that we should withhold assent just as carefully from whatever is not completely certain and indubitable as from what is clearly false; if I find some reason for doubt in each of my beliefs, that will be enough to reject all of them. (*1st Meditation*, 1998, p.18)

In other words, Descartes invites us to conclude that, if we are wrong about our sense impressions in some situations, that is sufficient to doubt all. We might argue that, in the

same way, those who cleave to the criterion problem are encouraging the same view of the testimony of dissociation subjects. But – is the move from *some* to *all* legitimate or supportable? It might be that some paradigms have results which are false positives, the product of unreported experience coupled with a conservative criterion, but what is the likelihood that all the explorations in all the paradigms have this flaw at their centre? The move from *some* to *all* in Descartes' application of doubt is often regarded as being too quick, or even illegitimate, and our first response to the criterion problem might be to take the same line. We have elsewhere argued further that the neo-cartesians play a double game with subjects' experience and their testimonies concerning it – when it suits their argument, they are quick enough to build a case upon them. Recall that Chalmers says that it is via this introspective route that we know ourselves better than anything else.

Concerning the question of whether the data set in the studies is more systematically compromised by consistently held response bias, we might argue with the following observations. This is also a very practical approach to the paradigms and their results. Let us return to Weiskrantz' late study of DB in 1999, after a break of 15 years. As part of this study, (Trevethan *et al* 2007) DB was asked to identify animals portrayed in low contrast (2%) outline drawings, presented in his blindfield (at a 9.4% distance from his fixation point). DB's subjectively reported awareness with low contrast stimuli is in the Type 1 blindsight category – he reports no awareness and says of himself he is 'just guessing'. Nevertheless, DB was able correctly to identify 25/28 items, failing only in identifying a deer as a horse. Several germane points can be made. Of course, according to the criterion problem objection, DB could have had some experience, but under-reported it, owing to a conservative criterion. Typically, in higher-contrast paradigms, DB characteristically reports that he is 'aware' that something had been presented, or metaphorically as 'moving waves', but that he did not 'see' it. To repeat, he did not report any awareness in the paradigm with 2% contrast animal drawings, but it is possible he could have had some, allowing him to achieve the high success rate, but look at the paradigm again. He was able to make quite fine discriminations requiring some degree of detail, in order to (say) distinguish between a horse and a cow, which he is clearly able to do. Even if he has some under-reported vague experience, it seems most unlikely that it would allow him to make such fine discriminations of form. Identifying a cow as a cow (and not as a horse) requires not just a limited experience of a something vaguely represented at the threshold of ordinary vision, but details – the cow's horns, an udder, a

shortened neck line. Furthermore, Weiskrantz displayed the line drawings at a very low contrast – at 2%, the drawings were subjectively invisible to the experimenters.

Additionally, Weiskrantz records that DB was better at identifying the animals when presented to his blindfield compared to when he was shown them in his intact field (at the same degree of eccentricity in his peripheral vision), and performed significantly better than five control subjects. Weiskrantz concedes – 'a cynical view, [would be] that DB has normal vision and is simply simulating' but no-one could predict a better performance than in the intact field, or in comparison with normal controls.

To underscore the main point, remember the claim in Phillips's objections. That DB could be consistently under-reporting experience due to response bias. This might be a persuasive complaint if the tests dealt with discriminations between gross features – an X or an O, for example —, but in this paradigm the task concerned the identification of animals, and discriminations on quite fine detail[34].

A third response is this – made in this case by Block. To some extent this addresses Phillips's final point, the claim that all trials rely on the 'very strong assumption that consciousness is found only above the subjective threshold'. We do indeed lack a scientific criterion for consciousness, and there is no proof to establish that any particular type of state is unconscious. 'But what holds for unconscious states also applies to conscious states' – there is no scientific proof that the reader of these words is conscious and 'this ought to put the first [claim] into perspective' (Phillips and Block, 2015, p.5). To amplify: no conclusive scientific test exists to establish that a neurotypical subject is experiencing stimuli, even if they show an appropriate or typical response to them, so it is unreasonable to demand a test for the absence of experience in neurologically less typical subjects.

We can make a fourth point, which is not an argument so much as a strategy. As we saw with Azzopardi and Cowey's (1997, 1998) study of GY a final response could lie in finding paradigms wherein the subject finds the right balance between being conservative and being liberal in their choice of response criteria. As those papers show, it is possible to arrange paradigms in this way. Or we can find conditions and paradigms in which the subject does not, or cannot distinguish between a signal and noise, or is not asked to do so, and yet is

[34] There is a possible explanation for DB's improved performance in the 15 years he was away from trials, which is that, over that time, his brain reconfigured itself, developing new pathways by which the images were processed. It is not clear however, that this provides any defence for Phillips' position. DB still reported 'no experience' when shown line drawings with low contrast.

still capable of showing awareness of a stimulus. In other words, cases where the detection distribution curves, in the Green and Swets diagram above, coincide, and $d' = 0$. Establishing cases of unconscious perception and confirmed in these paradigms by-passes at least part of the criterion problem, even though residual questions may remain. We shall explore this in connection with the Block/Phillips debate over perception. As we shall see, many of the paradigms at issue in that debate now contrive a situation in which $d' = 0$.

Phillips and the debate with Block: Is Unconscious Perception still Perception?

The Burgean definition of perception through two case studies – the role of subsystems.

Firstly, as a starting point, some further detail about the definition of perception offered by Burge, and accepted by both Block and Phillips as their criterion for what counts as perception. Here again is Burge: 'Perception is a type of objective sensory representation by the individual' (Burge, 2010, p.368). In this first section we will be examining the suggestion that perception is necessarily by the individual (as opposed to a product of subsystems of the individual), and whether that involves central co-ordinating agency.

It has been the contention of this thesis so far that unconscious perception is best understood by regarding subsystems of the individual as being at the root of the phenomenon. In other words, we understand what is happening in cases of blindsight etc if we hold out the possibility that the perceptual performance of the subject is the product of the survival of one subsystem, in the context of a failure of another. (Blindsight = surviving v4/v8, or v5 in context of loss of v1; visual agnosia = surviving dorsal stream in context of damaged ventral stream; Prosopagnosia = surviving v1 in context of compromised right fusiform gyrus. The differential responses in split brain patients is best explained by the operation of the two hemispheres as separate subsystems.) In this way, we can account for the perception on the one hand and the lack of consciousness on the other. If we insist that all perception is necessarily by the individual, and if that blocks explanation using neurological subsystems, then this line of explanation is denied to us.

Phillips is clear that he endorses perception as applying to individuals, saying; '[..N]ot all useable transduced information constitutes perception. Perception is objective sensory representation by the individual'. 'It must be attributed to the individual, not merely, say, to

their visuo-motor system'. (Block and Phillips' 2016 on-line paper, p.1). This last comment would appear to be a clear reference to Milner and Goodale's study of Dee Fletcher, who suffers from visual agnosia, and relies on her (intact) dorsal visual processing stream for her navigation around her environment and which allows her to orient her hand correctly when posting items through angled letter boxes (even though she claims not to be conscious of those targets). Phillips would rule this out as a case of unconscious perception as only a sub system of DF would appear to be involved in her action in manoeuvring her hand, or navigating a pathway. With this understanding of perception, Phillips will easily dispense with a variety of dissociations which might otherwise count as unconscious perception.

Insofar as Burge acknowledges blindsight, prosopagnosia, extinction-neglect cases and states in early vision (Burge, 2010, pp.374-5) as possible cases of unconscious perception, he does so, not on the grounds that they are the product of the survival and failure of subsystems – he entertains them on grounds that they might nevertheless be examples of unconscious perception *by individuals*. (They are cases of unconscious perception insofar as they exhibit unconsciousness in certain respects, but certain perceptual constancies are preserved, and the perception often guides action.)

While Burge is ultimately insistent that it is the individual that perceives and this is fundamental[35], he does nevertheless provide some helpful qualifications. 'Perception is the product of subindividual processing' (Burge, 2010, p.368) and 'I do not claim that all perceptions are perceptions by an individual.' Going on: 'I claim that all perceptions, including *any that are not strictly attributable to the individual*, serve perception by the individual' (ibid, my emphasis). These qualifications are more than Phillips would concede, as we shall see and his comment suggests, and they would seem to allow that DF's condition is another case of unconscious perception. On Burge's account, DF makes fine motor adjustments on the basis of her intact visuo-motor system, and can act accordingly, although it is precisely the absence of a complete visual system that means she does this unconsciously. We might say she acts as an individual in manoeuvring her hand, but her perception via her dorsal stream is one not strictly attributable to DF as individual.

This point is worth labouring; we are considering the question of what counts as perception, and the definition we arrive at matters as far as the notion of unconscious

[35] 'The fact that individuals perceive, and have perceptual states is basic to the explanatory roles of the notions of perception and individual' Burge, 2010, pp.369-70.

perception is concerned. How we understand perception to be construed is pivotal in deciding what if anything counts as *unconscious* perception. Here, as we have commented at other points in this thesis, our definitions act as an important constraint on our wider theories of mind, and empirically investigable anomalies concerning mind. [36]

Let us now consider another example: my own story of the snake-flinching in the garden. Recall the sequence of events: I approached the grassy mound; I felt uneasy; I leapt sideways away from the mound, and then became consciously aware of the snake sunning itself on the mound.

It is likely some processing took place in me by some primitive subsystem, allowing an early representation of the snake to be fashioned, and recognised as such and as potentially dangerous. This representation was made available to response mechanisms allowing evasive action to be taken, by me, the individual; all within the 250-300 msecs it takes for one to become conscious of seeing the snake.

A number of assumptions are locked into this story. Of course, against its anecdotal character – was the sequence of events (becoming conscious of seeing the snake only after I had flinched) just as I remember it? (The situation cannot be beyond repetition in laboratory circumstances, but may be subject to an objection based on a different type of response criterion). Was the representational content of the perception that of a snake? (one presumes so, as ordinarily, coiled pieces of wire, for example, do not have the same impact.) Was the evasive action taken by me, taken as an individual? I would claim: yes. I bodily leapt

[36] Before we leave DF's story, in Peters *et al* 2017, Phillips makes further claims against the claim that unconscious perception is involved in her case, saying: 'it is far from obvious that vision-for-action constitutes genuine perception.' Then, quoting Milner and Goodale 2008: 'although [they] do think perception can occur unconsciously, they insist that "the visual information used by the dorsal stream for programming and on-line control … is not perceptual in nature"'

It is not clear that this comment works to endorse Phillips' dismissal of visual agnosia as a case of unconscious perception. Milner and Goodale describe perception as 'the conscious experience of seeing' and unconscious perception as 'mental representations that potentially could reach conscious awareness'. On this score, visual agnosia does not count as unconscious perception on the simple ground that no part of the sensory information processed by the dorsal system could reach awareness. If we work with perception on the Burgean model, according to which consciousness or potential access is not considered a necessary feature of the definition, then Dee Fletcher's condition still counts as a case of unconscious perception. Dee's dorsal system processes low-level information concerning orientation and angle, and this instructs the action of manoeuvring her hand. It again turns on the definition of perception we elect to work with.

sideways. Was the early processing of the snake completed by a subsystem? In all likelihood, yes – as we have noted elsewhere, there are various speeds at which different aspects of objects are processed – shape, colour, etc, some markedly faster than others, all accomplished by subsystems of the visual processing stream. The amygdala is usually identified as responsible in cases such as these.

We should not be surprised at the sequence of events in this story. As Jeffery Gray points out (Gray, 2004, pp.6-9) consciousness comes late and extra. Fast processing by subsystems and fast reaction times are evolutionary necessities, allowing us to avoid being prey to some predator; conscious awareness of the predator comes almost superfluously, allowing admiration for its pelt and teeth, and other features, best appreciated at a distance.

Does the case count as one of unconscious perception? The definition provided by Burge allows this. He comments: 'certain states in early vision [states in the first micro-seconds of visual processing] may count as perception by the individual but fail to be conscious […] Perception occurs and figures directly in guiding action [but] at the level of conscious access, individuals are oblivious to what they perceive' (Burge, 2010, p.375).

I offer some comments as a summary of this section: Burge presents a definition of perception which has no bar to unconscious perception insofar as certain other features are preserved. He also would seem to suggest that, while perception is individual, there are subsystems which help constitute perception at the individual level. This is close to the spirit of this thesis, which sees the role of processing subsystems as an explanatory tool in accounting for how the dissociations arise and manifest themselves. As we shall see, Phillips is selective in borrowing upon the Burgean definition, in his exchange with Block; in doing so, he works with a definition of perception which closes off some dissociations as forms of unconscious perception. The concept as adopted frames and constrains the reading of empirical situations; this is just as we have found elsewhere – the concepts as chosen and defined are at the root of the interpretation of the dissociations studied. The adoption of certain concepts such as 'qualia' 'phenomenal character' and 'what-it-is-like' and here, 'perception', works to constrain our conclusions about unconscious perception. The adoption of a more accommodating concept of 'perception' might allow that unconscious perception is a possibility. Of course, preferring the one approach over the other has to be argued for: given that it is unlikely we can say whether one approach or the other is nearer to the truth, we might have criteria such as: does the theory of which this concept is a feature, have greater or

less explanatory width; does it accommodate or deny the most likely interpretation, or the best explanation of the dissociations we have looked at; how does it fit with growing neurological knowledge about the way the brain works.

Phillips, Block and their reading of Jiang et al – Perception as by an individual.

Let us return to the question we began with – whether unconscious perception is still perception. The debate between Block and Phillips ranges over a number of issues and paradigms, but one suitable route into their debate, is their respective treatment of a paradigm conducted by Jiang *et al* (2006). This particular paradigm side-steps the problem of the criterion since it presents us with a study 'in which subjects display no preserved discriminative capacity in respect of some feature (ie $d' = 0$) and yet that feature continues to exert some perceptual influence.' (Block/Phillips on-line paper 2016, p.3). Other examples might include masked priming cases, in which a masked stimulus is shown very briefly and yet the stimulus shown subliminally has a priming effect on supraliminal targets. Something like this effect is seen in the paradigm worked on by Jiang *et al*.

In this study, the subject is shown a fixation point, and then each eye gets a pair of stimuli for 800 msecs. One eye receives a pair of Mondrians, while the other eye gets a nude figure (male or female) and a patch of flesh-coloured fragments. The Mondrians presented to one eye suppress the conscious perception of both the nude and the nude texture presented to the other eye. (If the subjects do report any difference between the one and the other, on such presentations, the data on this is excluded.)

What the paradigm then goes on to measure is the subject's attraction to or repulsion by the nude, depending on their sexual orientation. This is prompted by the presentation of a final target briefly presented (100 msec) – a Gabor patch rotating clockwise or anti-clockwise. The patch is located in the same field in which the nude was shown. Subjects are required to say which way it is rotating. It is known that accuracy in this judgement is increased by attention, and that the subjects directed their attention can be assessed by whether they are right in their clockwise/anticlockwise judgement. The finding in Jiang was this: attention in men and women was affected by their gender preference for the gender of the nude shown, whether the subject was homosexual or heterosexual.

In sum, the study aims to create the unconscious perception of nude figures, and tests for the reaction of the subject to them according to their gender preferences. This is accomplished by judging the accuracy of their judgements in their final test, having been prompted by the rotating Gabor patch.

The study is a variation on a continuous flash suppression paradigm which forms a large part of the debate between Block and Phillips. One powerful advantage of this paradigm over blindsight cases is that the effect can be replicated in neuro-typical subjects, and is not reliant upon brain damage for its causal origin. This is an advantage if brain damage creates uncertainty about how to understand the response.

The question of course is whether the effect on attention is a form of unconscious perception. In this part of the exchange between Block and Phillips, we shall be concerned with whether the perception in the study conforms to features of the Burgean definition of perception - ie whether it is individual-level, and whether it involves central-co-ordinating agency.

The Jiang result is at first dismissed by Phillips on grounds that the differential response as a result of unseen nudes does not obviously 'implicate central agency'. Availability to, and response by central agency is another feature of Phillips' criterion for perception drawn from Burge. He makes a reference to Burge (2010, p.333), which Phillips reads as follows – a key requirement of representations being attributable to an individual turns on whether the representational content is available to central co-ordinating agency.

Let us look carefully at what this involves. Behaviour associated with perception, imputable to individuals needs to be available to central agency, ruling out reflexive responses such as muscle twitch, spasms, the firing of neurons, eye saccades. This conceptual stipulation would seem to discount as cases of unconscious perception paradigms such as those testing for affective blindsight as in for example Tamietto *et al* 2006, and skin conductance tests in prosopagnosia cases. These are cases in which it can be argued a central co-ordinating agency was not involved, and subsystems were responsible for the response. It is then claimed that the behavioural response could not be imputable to the individual. Phillips asks the same question of the Jiang paradigm – is it obvious that the response involves central agency: 'Not obviously, if the effect is due to the automatic attraction of eye saccades'(Block/Phillips on-line paper, 2016, p.11).

There are two questions to raise – first, the empirical question of whether eye saccades are the reason why perception takes place, and whether they are automatic. The second point to raise concerns the conceptual understanding of what counts as perception.

The points are linked in the sense that, if eye saccades are automatic, then they can hardly be evidence of central co-ordinating agency being involved. On the simpler point - whether eye saccades might be responsible for the effect in Jiang: it may be true that eye saccades are automatic and might therefore not be a product of central agency, but it seems improbable that they would produce the response. The claim in the paradigm is that attraction or repulsion associated with gender preference is impacting upon attention, and it is hard to see how this could be automatic 'in any relevant sense of the term', as Block puts it. Phillips defends his position saying: 'personal level gender preferences correlate closely with many reflexive autonomic responses. For example, [there are studies which] examine the differential pupillary responses elicited by gendered erotic stimuli.' (Block/Phillips on line paper, 2016, p.18). But the evidence would seem now to suggest that pupillary responses are not autonomic, and are integrated with higher level mental phenomena such as spatial attention and contextual processing (Binda and Murray 2015).

We are down to a narrow point in the exchange between Phillips and Block, on the question of whether the Jiang paradigm provides us with an example of a stimulus response which is more than an automatic reflex, and satisfies the Burgean definition of perception. Block argues yes, and that the Jiang results cannot be dismissed as an automatic response, or as sub-cortically mediated, and, on the contrary, is the product of the involvement of central agency. It is thus a case of perception. Phillips argues that we should discount the Jiang result as the product of autonomic reflexes, and so therefore not involving central agency, and is thus *not* a case of perception.

The dispute illuminates the point of difference - what is it that counts as perception in this case. This is a conceptual, stipulative question. Burge identifies the link between active behaviour by an individual and the role of central co-ordinating agency, but is quick to qualify it, and it is clear from subsequent pages (Burge, 2010, p.334-5), that he does not think that involvement in central agency is a necessary condition of unconscious perception. Block explicitly agrees with this reading. Phillips is less attached to it, saying: 'I did not propose that involvement in central agency is a necessary condition of perception, but when a

representation is unavailable to central agency we lack a positive ground for attribution.' (Block/Phillips 2016 on-line paper, p.18).

This thesis has maintained that definitions or conceptions are material to the recognition of dissociation as problem cases. If our definition of perception includes a stipulation that it necessarily involves consciousness then the question whether there can be unconscious perception is closed, and there seems little to be gained in philosophers studying the dissociations further. If we stipulate that perception is something done by an individual, and that that means it must prompt or guide action by the individual, but at the same time stipulate that these actions must be under *conscious* voluntary control, then we have another definitional reason to down tools – we seem to have a reappearance of the requirement that 'perception must be conscious'. Phillips does not take this final step explicitly, but he is demanding that a high bar is met.

Perception as objective *representation. The constancies question.*

Phillips notes that, for Burge, perceptual objectivity in perception is achieved by the exhibition of perceptual constancies, that is, "capacities to represent environmental attributes or environmental particulars, as the same, despite radically different proximal stimulations" (Burge, 2010, p.387 but see especially pp.408-413). Burge himself argues for blindsight as a case of unconscious perception on grounds that, while it seems to be such a case, 'blindsight patients perceive environmental conditions. The perception involves perceptual constancies' (Burge, 2010, p.374). He holds out the possibility that other conditions might also qualify: 'in several other cases, for example in prosopagnosia and extinction-neglect syndromes [..] the perceptual state meets conditions for perception; in particular, perceptual constancies are exhibited' (Burge, 2010, p.375).

Phillips addresses this issue, asking if it really is the case that blindsight involves constancies. Much early work did not address the question – Weiskrantz (2002) concedes that none of the 'visual constancies has ever been addressed in any blindsight studies of which I am aware.' Phillips references several more recent studies, but concludes that none of them establish that subjects with blindsight exhibit perceptual constancies. This from Peters *et al* 2017: 'DB perceives neither surface colour, nor chromatic contrast, matching coloured stimuli purely on the basis of wavelength' (Kentridge *et al* 2007). Motion detection in GY is

limited to ‘objectless’ first order motion energy, as opposed to changes in position and shape (Azzopardi and Hock 2011).

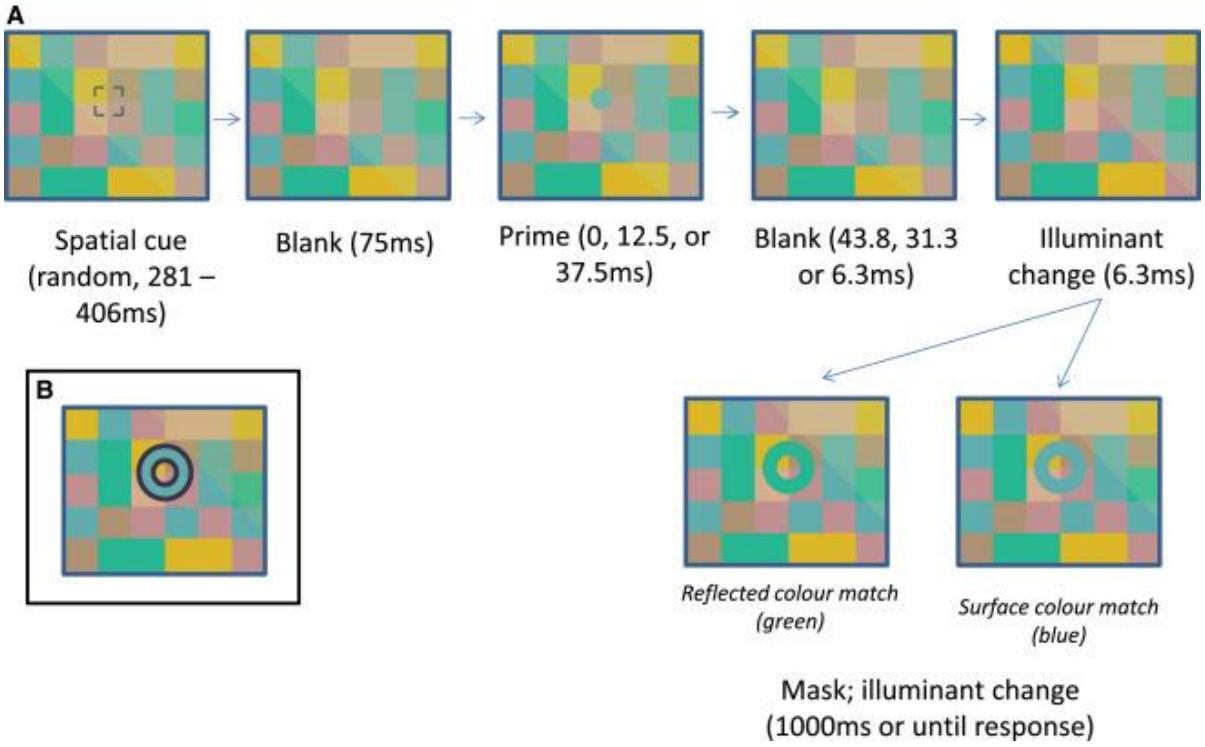
However, there is a still more recent study which Phillips must address more fully and at the point of writing concedes that it works against his conclusion. The study in question is: Norman *et al* 2014. Before we examine the study in detail let us remind ourselves what is meant by constancies, and colour constancy in particular.

For Burge, perception is: objective sensory representation by the individual. The criterion of objectivity in perception is met by the exhibition of constancies – ‘perceptual constancies are capacities for objectification’ (Burge, 2010, p.399). Various constancies are advanced – motion, size, distance, shape, but the one we are here interested in is colour constancy. This is the ‘capacity to represent a colour as the same under various conditions, including different illuminations’ (Burge, 2010, p.410). It is this last specification that the Norman study looks at.

The study is another using continuous flash suppression, and so benefits from being able to use neuro-typical subjects, thus sidestepping any potential misinterpretation as a consequence of damage to the brain. The study also makes an effort to avoid the problem of the criterion, as we shall see (*d´* is not significantly different from zero).

Here are the details. Subjects were shown a coloured Mondrian pattern as in the illustration below (downloaded from the internet)[37]

[37] https://www.sciencedirect.com/science/article/pii/S0960982214012858



They were given a spatial cue for a randomly chosen time duration between 281 and 406 ms, then shown a blank for 75 ms, then a prime. The prime was a colour circle shown in the region of the spatial cue, and presented for one of three durations – 0 ms, 12.5 ms, or 37.5 ms. A blank was again shown, during which time there was a change in illuminance. Finally, a mask was shown, - a colour target which varied in its colour – either matching the prime, or not. In this way, the prime and the mask were presented under different illuminance conditions – one resembling direct sunlight, the other, average daylight. Subjects were asked simply to identify the colour of the mask at the conclusion of the sequence. Under normal circumstances, subjects are quicker to respond when asked to identify the colour of the mask, when the prime and the mask match in colour.

Subjects were asked, over three sessions, to say whether they saw the prime as being present or not, on a confidence scale of 1 to 8. Recall that sometimes (in half the trials), the prime was not present at all (= 0 ms presentation time). Results established that the subjects were not able to detect the presence of the prime for all those presentations in which the prime was shown for 12.5 ms. They were little better at 37.5 ms, but two subjects could detect the presence of the prime in the longer interval, but taken over the study as a whole, mean d' was 0.006for the short prime duration, and 0.186 for long prime duration. These figures, not significantly different from $d' = 0$, indicates strongly that the subjects had no experience of the prime, and so the study is not prey to the response criterion problem

discussed above. The exposure time was very short, and so we may conclude that the colour presented as the prime was unseen – the subjects were not conscious of it.

Was there nevertheless success in perception? Subjects were asked to say what the colour of the mask was at the end of the sequence shown. As noted, subjects are quicker to identify the colour of the mask if they have been primed with the same colour as the mask. This would indicate that the mask has been unconsciously perceived. The key feature of this experiment however, is the change in luminance between the prime and the mask (This is not easily rendered in the diagram above, but the change in illuminance changes the reflected colour of the mask). The results of the study showed that subjects were still able to respond as quickly to identify the colour of the mask despite the change in luminance – with the conclusion that colour constancy was preserved. How the surface colour was perceived remained unchanged, and the luminance was discounted. This is just as we would expect if colour constancy is exhibited.

The study and its additional controls also allow the authors to speculate upon the neurological mechanisms responsible as the basis for colour constancy. Although nothing is entirely conclusive, it is clear that accepting the possibility of unconscious perception allows here as elsewhere, permits scientific investigation into the neural correlates of consciousness.

As we remarked above, at the time of writing, Phillips has yet to respond fully to the Norman *et al* paper. He concedes that it is an exemplary study in finding a methodology in which $d' = 0$, but very briefly suggests that the response is not a result of a representation being available to central co-ordinating agency, and so not evidence of individual-level perception. (Block/Phillips on-line paper, 2016, p.4). The Norman *et al* paper is unlikely to be the last word on the issue; it would appear to be simply a question of devising the right kind of paradigm to explore the preservation of other constancies in blindsight, or blindsight inducing techniques.

Block and the Anna Karenina Principle

The final case to make concerning Phillips' campaign against unconscious perception is to pick up and add to Block's charge in Block 2015, an argument that also figures in the final pages of Block/Phillips on-line paper in 2016. This is the claim that Phillips has offered

several different objections against the various candidates of unconscious perception, but without a common theme or denominator.

He deals with blindsight by raising the possibility that the effect has an explanation in unacknowledged and degraded experience, linked with a conservative response criterion. This is the charge that it is not unconscious. It is also, as we have seen, dismissed for failing to be a case of perception. Further, in taking Burge's definition of perception as the pivot of his argument, (appearing on p.2 Block/Phillips on-line paper 2016 and in fn 20 in Phillips 2016) he maintains that there is no clear evidence that constancies are preserved in blindsight cases, and he repeats the charge in cases in which blindsight is induced by continuous flash suppression (Peters *et al,* pp.5-6).

Against the Jiang paradigm (continuous flash suppression) and the idea of unconscious gender-based responses to primes, Phillips offers the view that the responses are the product of automatic reflexes, and so therefore works against that feature of the Burgean definition of perception which suggests perception is individual-level.

Against visual agnosia (DF's condition) he claims that the responsible mechanism is a subsystem and the effect is therefore not attributable to the individual. He writes: '[there is a] concern whether pertinent modulations of behaviour witness genuine control and guidance "by the individual" and so meet [the] requirement for perception' (in Peters *et al,* p.6). He could make the same case against subjects with prosopagnosia, who are found to have a differential skin conductance response to known faces. Presumably in these cases, the response could be described as reflexive, not the product of central agency. As we saw in the footnote above, Phillips also objects to visual agnosia that it may not be a genuine case of perception on other grounds.

Neglect, as in the Marshall and Halligan paradigm, is partially dismissed as a case in which unacknowledged conscious perception is involved. He cites studies which concede that flames might not be experienced, but explore the idea low-level features such as asymmetry were perceived to account for the preference data (Phillips, 2016, p.432). Phillips also launches the criterion problem at the paradigm, complaining that no neglect study comes close to the gold standard of $d' = 0$. 'I know of no study of unconscious priming in neglect which rules out the possibility that the prime is consciously perceived.'

In brief, these are various candidates for unconscious perception for which Phillips has found a reason to dismiss. Block, who has defended several of these candidates, points

out that there are many other ways of producing unconscious perception – citing Breitmeyer, (2015) who identifies 24 methods of producing the effect in neuro-typical subjects, using methods such as: crowding, the attentional blink (and other attentional blinding methods), backward pattern masking, metacontrast masking, continuous flash suppression, sandwich masking, and single-flash interocular suppression. We can add to this list with more recent studies into blindsight-inducing trans-cranial magnetic stimulation (Christensen *et al* 2008).

Phillips asked whether, in all these cases (including the cases arising from brain malfunction), perception *of the same fundamental kind* as we find in conscious perception is to be found. It is clear that, in each of the cases explored, Phillips has found an ad hoc criticism, which collectively share few commonalities. Against the notion of unconscious perception, he has catalogued some cases as not unconscious, and against others, he as identified them as not being cases of perception. Under the last category, he has claimed that some do not qualify as *objective* representation in which constancies are preserved, and others are found not to be *individual-level*.

Against this collection of objections, Block's point is that there is no single common flaw to each of the conditions or cases, and raises questions about the totality of the case Phillips makes against them. We can amplify this. There is much in common with our understanding of the dissociations, and only a series of ad hoc objections that Phillips makes in reply. Block observes: 'Successful conscious perception is a dance of oscillating feed-forward-and-back loops.' (Block/Phillips on-line paper, 2016 p.23, Block 2016). Most of the paradigms discussed in this chapter have involved a compromise of, or interruption of these feed-back and forward loops. In the case of blindsight, the feed-back loops are stalled by the absence of the primary visual cortex. In the case of continuous flash suppression, the loop has not had to time to establish itself before a stimulus is masked, or disappears (< 100 ms). In others, attentional distraction has had the same effect. Importantly, in these mechanisms we have a neurological underpinning of the same fundamental kind; there is substantial evidence of this underlying commonality (Lamme 2001) to the point where investigators have begun to talk about and develop a unified model of conscious and unconscious perception, with the difference between the two to be found simply in whether the processing of stimuli enters working memory, or is experienced for long enough to consolidate (Kiefer *et al* 2011). With such a unified model of perception becoming available, we might expect any theoretical breakdown of, or principled objection to unconscious perception to have a similar unity.

Conclusion

At the beginning of this chapter, we noted that, in the view of Phillips, the question of unconscious perception was empirically tractable. He does not trivially block the possibility of the idea by defining perception in phenomenal terms, involving the idea that there is something it is like to perceive a red patch. The strategy adopted by Phillips, however, is to adopt a tight reading of Burge's definition of perception, as objective representation at the individual level with the involvement of central agency in the response to stimuli.

His first tactic is to suggest there are no clear cases of *unconscious* perception – that some degraded experience is possibly at work in explaining discriminative success. This is launched in the form of the criterion problem – that subjects have, notwithstanding their protestations, weak conscious awareness unreported due to a conservative response criterion. This is met by those investigations and laboratory simulations which aim to contrive a $d' = 0$ result ie in which it is established that the subjects' subjective criterion is neutralised.

Phillips then shifts to the question of whether the dissociations qualify as *perception* in the relevant sense. Here, the objections are:

a) That the cases in question are not ones in which representation is at the individual level, or ones in which central agency is involved in responses to stimuli
b) That the cases in question do not involve *objective* representation by the individual ie involve constancies.

On these grounds, DF's visual agnosia was questioned insofar as there are worries that her 'modulations of behaviour witness genuine control and guidance by the individual' (Peters *et al* 2017, p 6). In the Jiang study, Phillips claims that the relevant responses were the result of automatic reflexes and so therefore not individual-level. Both of these objections run counter to Burge's definition, for Burge makes it clear that blindsight can count as unconscious perception. We may suppose that this applies to other dissociations. As we have noted, at the time of writing, Phillips acknowledged that the Norman *et al* paper works against him, and so his objection that constancies are not preserved in forms of unconscious perception fails.

Phillips' objections, in questioning the methodology of the paradigms and other issues has led to improved experimental approaches. There is much to be gained from caution and the careful elimination of confounds, and for this his contribution has been most valuable.

But on each count his objections have been met, and one suspects the Burgean definition no longer serves. He seems ready to distance himself entirely from the Burgean conception when he remarks: ‘whilst it has been convenient here to adopt a broadly Burgean conception of perception, that conception is hardly beyond dispute’ (Block/Phillips on-line paper 2016, p.22). Robert Kentridge believes that his CFS study (in Peters *et al* 2017) satisfies all Burge’s criteria, and that ‘if it walks like a duck and quacks like a duck, then it is a duck – that unseen primes can be perceived, and so colour perception can be can be unconscious […]’ (Peters *et al* 2017 p 5*)* In the same paper, Megan Peters takes the view that, with the sum of his objections, Phillips is now raising ‘philosophical objections to the very concept of unconscious perception (Peters *et al* 2017 p 3).

Chapter 6

Higher Order Thought theories

This thesis has trailed at points throughout that higher order thought (HOT) theories stand at odds with the neo-cartesian position, and it is time to develop these more fully. This section will give a sketch of what is involved in the theories and how they hold out a promise of accommodating the dissociations. It will explore some of the more obvious objections to HOT theory, and show how those objections can be blunted. It is important to stress that this can only be an outline of HOT theories, and what they can offer.

It is part of the appeal of HOT models that they offer the possibility of a reductive account of consciousness. The claim is that they explain conscious mental states in terms of other mental states, and leave it open for neuroscience to complete the programme of reduction at some later date - a reductive account of non-conscious mental states would intuitively seem to be easier than a reductive account of conscious mental states. We have also suggested that the dissociations can be understood as presenting non-conscious mental states, and so our dissociations should have no difficulty finding a home in HOT theories. I shall also argue that HOT models provide us with an account of the evolutionary development of consciousness, and the process of learning in children and adults.

Conceptually, the starting point for higher order thought theories lies in jettisoning the idea that consciousness is an intrinsic feature of all mental states. The cartesian view is to claim that consciousness is definitive of the mental. With such a presumption, it is obvious that the very idea of a mental state we are not conscious of, is problematic. But a counter proposal exists in exploring the idea that consciousness is not essential to mental states.

Let us build this case carefully. The neo-cartesians have argued that consciousness is the possession of a something (mental qualitative properties or qualia) - some essential element which is intrinsic to the relevant mental states. Experience has been built into the concepts of perceptual or sensory states such that there is something it is like to have them. The proposal here is to uncouple the phenomenal from what we might call ‘first order’ sensory states, such that we leave it a possibility that such first order sensory states can be unconscious.

Our excursions into dissociations have made the case for the possibility that there could be mental states (sensory registrations, successful perceptive discriminations) which

can occur unconsciously. But this has, in any case been anticipated by Freud and others for some time. Since the nineteenth century it has been a commonplace to think of propositional attitudes such as desires or beliefs being held by someone, but suppressed, banned from consciousness. The conditions we have explored have made a strong case for thinking that mental states of the perceptual or sensory kind need not be conscious either. The best interpretation we can put on the dissociations is that in those cases, some people have an ability to respond appropriately to visual stimuli in the absence of any awareness of having seen them. In the terms offered by Chalmers, there is nothing it is like for the blindseers to have had those stimuli. Put slightly differently, and in terms other neo-cartesians would use, in the cases studied, Dee Fletcher's visual form agnosia, the cases of hemispheric neglect and so on, a very natural way of explaining these results is to say the subjects lack awareness of their mental states, but are nevertheless capable of responding correctly. It is a very natural thing to say – that they have a kind of sensory state of which they have no awareness, and which there is nothing it is like to have, to use the term our target philosophers are inclined to use.

Of course, the neo-cartesian model would disallow such a description, but if we begin with a different pre-theoretic commitment, such that we uncouple experience from mental states, we create a class of mental states which are unconscious. We have a category into which we can naturally position the kind of states which are involved in blindsight. Such sensory states would rub shoulders with non-occurrent propositional attitudes, such as my belief (when I am not thinking about it) that the battle of Waterloo took place on June 18th 1815, or my murderous contempt for certain politicians, or other suppressed desires. Many sensory states can join them, such as the sensory state that I am in when a car has slowed down in front of me while I am driving on the motorway in my automatic mode, or the headache I have had all afternoon, but which I am currently unaware of, since I am concentrating so hard. Include too, Ken Park's tragic somnambulism.

All of these mental states remain unconscious until (according to HOT theory) I have a higher order mental state concerning them. In doing so, I become aware of them; I have a thought concerning them. On this model, a conscious mental state is a mental state I am conscious ***of.*** And it is in virtue of the higher order thought that I am conscious of that mental state. We now have two classes of mental states – unconscious mental states, (to which we will give the term: ά) and conscious mental states (to which we give the term: T[ά]). A

conscious mental state is a state about which I am having a thought. In such an event, the thought concerned is intentional in character.

Self Consciousness

It is argued that it is quite possible to be in a conscious mental state without being conscious of being *in* that state[38]. One imagines this is true of the vast majority of our mental states at any one time. According to the HOT model therefore, in order to be aware of being aware, a further and still higher order thought is necessary.

To put this more carefully: in the model as described by Rosenthal (2005), our thought concerning a mental state, T[ά] need not itself be an introspectively conscious state. It is possible for me to be introspectively conscious of a conscious state by having a further, still-higher order experience of the conscious state T[ά] by having a further thought about that state (ie: T[T[ά]]) In such a condition, I am aware of being aware of being in ά.

We now have three kinds of mental states: unconscious states, non-introspectively conscious states, (aka first order experiences) and introspectively conscious states (aka higher order experiences). These distinctions are important to avoid the charge that the model generates a vicious regress of nested conscious states. I look out of the window and I see a green tree. T[ά], where ά is the perceptual state, and T[ά] is the experience of seeing a green tree. This does not necessarily involve my being aware of myself seeing a green tree, and for this to be so, I would need to have the further thought that I am having the experience of seeing a green tree.

Importantly, on this account, there is a place in the menagerie of mental states for unconscious perceptual states. They are not dismissed out of account or regarded as a contradiction in terms. Instead**,** they are a foundational state for experiences – or to use the neo-cartesian term: together with the higher order thought they are constitutive of the feeling of what-it-is-like[39].

[38] Dretske also makes the distinction between being aware of a stimulus and being aware *that* one is seeing it (Dretske 1998).

[39] Rosenthal clings to the idea that the 'what-it-is-like' notion can be analysed or explained in this way, and so the term is one he is comfortable using. This thesis remains sceptical of the term, whether it is useful, or ampliative, or even properly a synonym for phenomenal consciousness, or qualia. The question of whether what it is like can be expressed in HOT terms forms the basis of the Block/Rosenthal debate in Analysis in 2011.

We can clarify still further: What is it to have a higher order thought? HOT theorists maintain that to direct a higher order thought towards a mental state requires the application of concepts. The thought that I am aware of seeing a green tree T[T[ά]] requires the possession not only of the concepts of: 'tree', 'green', but also: 'seeing', and 'self'. The first order experience of T[ά] has fewer concepts involved, ('tree', 'green'), some other, primitive ones will be deployed. This position has obvious similarities with Kantian thinking, in the same vicinity. We can say it is the synthesis of the higher order thought and associated concepts together with the mental state ά that constitutes consciousness, at least in the case of perceptual states.

This chapter will outline a case for HOT theories, but before we do so, let us consider obvious problems with the model, especially those levelled by the neo-cartesians.

The HOT model – drawbacks.

What is Chalmers' position on HOT theories? He suggests that the HOT account leads to a picture of the mind with unnecessarily duplicative features. 'To suppose that there are two separate cognitive states for every detail of experience, a first order and a second order judgement, leads to a cluttered picture of the mind' (Chalmers, 1996, p.230-31). To illustrate this, he reproduces a cartoon in which a character, at the end of a long queue of observers, looking over each other's shoulders, says (of himself) 'I'm observing myself observing myself, as I observe you observing yourself observe yourself as you draw yourself observing yourself.'

I submit that this is mistaken. It seems to hint at a vicious regress in the HOT picture, but if so, it is easily deflected. The reader is invited to make the assumption that each observer in the cartoon is a conscious, or self-observing homunculus, and this of course would be a regress. But recall that, in the version of model promoted, the higher order thought need not be a conscious state. Only the combination of the higher order thought with the mental state (a perception, or a propositional attitude) results in a conscious mental state.

The regress is only a problem if we assume that higher order states, by virtue of which one is aware that one is in ά, must be conscious. If Chalmers makes this assumption, he possibly does so from the point of view that every mental state must be conscious, - the very thing we have suggested is a feature of his own account. But that he seems to make this

assumption is borne out by further comments in the same section. He writes: ‘A system with a second order judgement for every detail in the visual field would seem quite redundant’ (Chalmers, 1996, p.230). In our reading of Rosenthal’s system, there would not be a second order judgement about every detail of the visual field. There would be plenty of items in the visual field about which we would not have a second order judgement; there would be many more items in the wider sensory field about which we would not have a second order judgement. Rather at any one time, there would be perhaps very few ά of which we were forming a T[ά], perhaps in the way a torch picks out items in a dark cellar. To the observer in question, then, there would not be a cluttered picture of the mind.

He goes on: ‘it is hard to see why evolution would bother to build these second order judgements across the board’ (Chalmers, 1996, p.231). In the position we are taking, the Chalmers understanding of consciousness has the bigger problem with evolution. According to his position, all mental states are essentially conscious, and presumably always were. In this we lack the mechanisms to explain the passage from lower life forms which may have had the capacity to taste fruit, or see a predator (capable of having ά) to our early hominid ancestors who then became able to form beliefs about the taste (capable of having T[ά]) and who then in turn evolved to be able to form a T[T[ά]] – acts of introspection about myself and my biography. Chalmers’ account – his commitment to all mental states being conscious seems a curiously static, and assumes undeveloping mental sophistication. By contrast, HOT theory provides an attractive cognitive architecture which may be able to capture our likely evolutionary development. We shall say more about this later.

Chalmers does not give HOT theory much more attention in *The Conscious Mind*, but of course was committed to the Unity Thesis in later works, which were specifically levelled against HOT theories. We can now turn to another problem which HOT theory must address.

The Problem of the Rock

This problem was advanced by Alvin Goldman (1993), and it seems to anticipate Chalmers’ argument concerning the Hard Problem. His question is simply put – How is it possible for a higher order thought to ‘confer subjectivity of feeling on a lower-order state that did not otherwise possess it? […] A rock does not become conscious when someone has a belief

about it. Why should a first-order psychological state become conscious simply by someone's having a belief about it?

On the face of it, this would seem to return us to a version of what some have called the generation problem which lies at the centre of all theories of mind. As Seager puts it: 'The Generation Problem is to explain precisely how the possession of property P generates or produces (or underlies, subvenes, constitutes, realises, whatever) consciousness in those systems that possess it' (Seager, 1999, p.18). How can it be that a thought of a certain character concerning the unconscious registration of sensory information produce the feeling of what-it-is-like to taste coffee, see a sunset. It is easy to see this as a relative of the Hard Problem.

How can HOT theory respond? The first reply is to point out that what we are having a thought concerning is not a rock, but a mental state. There is something especially different about the first order mental state – it is *mental*, not a physical state, and the gap to be leaped is not one where we begin with the physics of the brain (objective facts of the matter) and ask how can it be that physical facts of the brain, obeying the laws of physics, can generate the qualitative, subjective character of consciousness. We will encounter this point again when we discuss reduction and HOT theory. Goldman posits a false analogy in assimilating rocks to first order mental states, in his question. There are features of the mental states in question which are quite different to physical states – one obvious difference between first order mental states and rocks is that the first is already a subjective mental state of *mine,* whereas a rock is a distal physical state of matter. Part of the gap has already been crossed.

But it may be that we have not quite answered Goldman's question. If he was asking *how* does having a higher order thought about a mental state, make that mental state conscious? then the problem remains. What can it be about a higher order thought such that having one can confer consciousness upon lower order mental states?

Gennaro (2005) makes a beginning with this issue by taking our invocation of Kant a little further. He invites us to consider: our conscious awareness of a brown tree begins with the passive reception of sensory information by the senses (through the Kantian faculty of sensibility). Some of this information rises to the level of unconscious mental states (ά), states which can, (just as it did in our study of dissociations) have causal influence upon our behaviour. However, these mental states do not become conscious until the faculty of understanding operates on them by application of concepts. Thus I consciously experience

the brown tree as a brown tree because I deploy the concepts of 'brown' and 'tree' in my higher order thought to the sensory information in ά. The conscious state: T[ά] becomes possible by the contribution of both faculties (sensibility and understanding) – the application of concepts gives rise to conscious experience. Gennaro concedes that the neo-cartesian might reply: this account might go some way to explaining *how*, but *why* does the higher order application of concepts give rise to conscious experience? At which point, Gennaro believes that the neo-Cartesian is asking an illegitimate question – 'the chain of explanation has already come to an end [..] it does not make sense to ask *why* this is so' (Gennaro, 2005, p.11).

It is possible to have some sympathy with Gennaro's conclusion. There may indeed be a limit to explanation: 'why is there something rather than nothing?'; 'why are the laws of physics such as they are?'; 'why is seeing red like *this*, rather than like *that*?'; 'why does a proton have the mass that it does?' all seem to be in search of answers that cannot be given.

In pressing for the deeper explanation, it may be argued that neo-cartesianism has no answers to the same questions which are any more convincing. While the HOT account might be criticised as being incomplete, the account offered by Chalmers as to how it is that consciousness arises from the brain's physics is reliant on unconvincing relationships such as emergence, and the idea that consciousness is ontologically novel, (1996, p.130) or an expansion of the fundamental features of physical particles to include experience alongside mass, or charge (2010, p.125).

What are the main arguments for HOT theories?

There are several. The claim often offered to recommend HOT theories is their role in a plausible naturalistic account of mind. We shall consider this first.

HOT and Reduction

For philosophers and others who find dualism intuitively hard to accept, the prospect of a completed reductive account of mind is appealing. HOT theory does not itself attempt to explain a consciousness in materialist terms, but it can be part of a wider strategy that does. Carruthers (2005 p.6) argues that attempts to reduce consciousness to physical or neural terms is implausible insofar as they attempt to leap over too many explanatory levels at once. Its advocates claim that HOT theory can be helpful in providing one link in a chain of

explanation. Explanation is the ultimate goal in any reductionist strategy – a non-reductive position leaves gaps and mystery, and is unsatisfactory on that count alone.

Let us see what contribution HOT makes. The goal of any explanatory account is to explain a phenomenon in terms of something else. As we have seen, according to HOT theory, what it is like to be introspectively conscious (in our terminology: T[T[ά]]) can be cashed in in terms of the deployment of a HOT or higher order conceptual resources towards a conscious mental state. That mental state (T[ά]) in its turn, can be given an account in terms of a world-directed non-conscious sensory state (ά), towards which we might ordinarily deploy lower-order conceptual resources. These states (ά) are unconscious, and involve the kind of perception we see in the dissociations, and other states. This is already reductive in the sense that higher order mental states are being given an account in terms of lower order mental states. However, such (ά) states themselves look amenable to reduction – certainly more than the conscious, introspective higher order kind. Theories such as those suggested by Gennaro (2012) suggest there may be a route through some version of strong representationalism[40], but one advantage of HOT theory is that it is not committed to any, or any particular reductivist strategy, and can leave the final step in explanation to others. So, while many of the HOT theorists might ultimately be materialists, their contribution is simply to provide an explanation of conscious mental states in terms of unconscious mental states (mentalistic reduction).

The elimination of an underlying pre-systematic conceptual architecture.

It is maintained by HOT theorists that the pre-systematic commitment to the idea that mental states are intrinsically conscious has all the hallmarks of a pre-Keplerian commitment to the idea of *anima motrix*. This might share similarity with Dennett's story of the rainforest tribe committed to the idea of *fatigues*. Sometimes concepts get off on the wrong foot, and the persistent commitment to them, can become a barrier to further progress. The idea of *anima motrix* as principle of motion intrinsic to the planetary body itself was not just wrong-headed; it forestalled scientific advance; we have made more progress in understanding cosmology by rejecting it in favour of other conceptual structures. In the same way, the HOT theorist will argue that the neo-cartesian commitment to mental states as intrinsically conscious is also a

[40] Where strong representationalism contends that: having representations of a certain kind suffices for having qualitative states.

blind alley. Philosophers who subscribe to it are constrained to speak of the neural *correlates* of consciousness, and the principle of *coherence* between consciousness and cognition. Framing consciousness in phenomenal terms puts consciousness beyond third-person, or scientific reach. To think of consciousness as having a necessary unity turns a blind eye to the richly interesting dissociations, and lessons learned from split brain and synaesthesia cases, which might prompt a more modular understanding of the workings of the human brain.

If a complete case can be made for them, HOT theories could offer a pre-systematic framework in which the advances of neuroscience are welcomed as part of the overall theory of mind, rather than partitioned off, as Chalmers does, into a supporting work on psychology, which for him does no more than stand in a relationship of structural coherence with phenomenology.

Learning

In a short but provocative paragraph in Rosenthal 2005 he writes: 'We typically come to make more fine-grained discriminations as we master more subtle concepts pertaining to various distinct sensory qualities. Experiences from wine tasting to hearing music illustrate this process vividly. An account in terms of higher-order thoughts explains the bearing these concepts have on our very awareness of sensory differences. [However] if consciousness is intrinsic to sensory states, the relevance of concepts remains mysterious. The cartesian might just deny that sensory differences exist when we are unaware of them. But it will be even more difficult to explain how learning new concepts can actually cause sensory qualities to arise that previously did not exist' (Rosenthal, 2005, p.40).

In the interests of exploring HOT theories, let us amplify this argument. The claim is that learning new musical theory for example, in mastering concepts such as augmented/diminished chords, major/minor keys, hemiolas, syncopation, aeolian/lydian/mixolydian (etc) modes, consonance/dissonance, the listener is equipped to hear music differently, and with greater richness. This can be accounted for in two ways, according to the position taken by our different theorists – either: our learning of new concepts results in our perceptual or sensory states being revealed as having qualitative properties they had, but were not unlocked, or brought to the surface, or: all the qualitative properties were already present and transparent to the mind in perceptual states before the

moment of our learning a new concept, and the learning of a new concept involved the creation of an entirely new qualitative property. The two routes are very different.

The neo-cartesian, in maintaining that consciousness is intrinsic to sensory states, would have to subscribe to the latter; qualitative properties are present in our sensory states and transparent at all times, and the role of new concepts learned is to create entirely new properties that were not there before, and the process by which this happens is mysterious.

If we subscribe to a higher order thought theory, which has, in the foundation of its account, unconscious mental states, (including sensory states), our explanation of the new experience, upon learning new concepts, is the process of acquiring new concepts amplifies the content of our perceptual experience – bringing to the fore properties already there, but so far unnoticed. The qualitative character of our music appreciation, the taste of the wine on our tongue can vary in proportion to our conceptual sophistication.

We can now begin to make a preliminary judgement concerning this point. If a neo-cartesian commitment to consciousness being intrinsic to mental states *de facto* commits them to the view that learning new concepts can create new sensory qualities beyond what is already there, then those philosophers owe us an account of how this can happen. They must account for the idea that concepts and concepts alone can somehow create new sensory qualities that were not there before the concept was learned. No-one has offered such an account. On the other hand, the HOT account would appear to be in a better position to describe how, although our basic experiences must remain broadly the same, through conceptual refinement and sophistication, we have developed a richer appreciation of experience. A developmental account of consciousness may be possible, contrasting with the neo-cartesian position which seems far more static. Let us pursue this further.

Animals, pre-linguistic children and evolution.

Seager points out that those who subscribe to HOT theory, must hold that beings without thoughts are not conscious (Seager, 1999, p.83). He points out that this is problematic for animals, and we can extend this difficulty to children in the early stages of development – we might want to hold that higher animals and children are conscious; but how can this be? They would lack the conceptual sophistication to have a thought about their sensory states, and so

fail to have a higher order thought about them. Rosenthal offers an answer, suggesting that it depends on what is involved in a higher order thought at this level.

Clearly, to have a higher order thought about a propositional mental state (I believe I have drunk too much coffee) requires a set of sophisticated concepts which it would be unreasonable to expect of non-verbal children and animals. However, a higher order thought concerning a sensory state does not require such sophistication; 'Relatively weak conceptual resources will suffice for a higher-order thought to refer to one's own sensory states' (Rosenthal, 2005, p.186). Presumably, one would need to be able to individuate, to compare, to locate sensations in the appropriate sensory field, and so on, and these would be resources available to higher animals.

If we accept this answer as sufficient, it is possible to see how it can be the beginnings of an evolutionary and developmental account of consciousness. As our ancestors (and ourselves in early stages of development) developed language and conceptual sophistication, they/we became equipped with the ability to have higher order thoughts, and more complex higher order thoughts about our mental states. They/we have been able to extend our consciousness to include propositional states, and increase the range of sensory states we have become conscious of. We have an account which has within it, a sense of degree and progression.

Here then, is the beginnings of a HOT account of the evolution of consciousness. Our early ancestors would have been capable of sensory registration and perceptual states that allowed them to evade predators and identify food. Such states would have been, at the relevant stage in our evolution, much like the states of unconscious perception or automatism we have been dealing with. We have identified them in our menagerie of mental states as: ά. Gradually, as our conceptual repertoire and language developed, we would have become capable of projecting thoughts upon our perceptual states and the beginnings of consciousness would have flickered into being. Such thoughts, together with the sensory state of which they are the focus, can be designated as non-introspectively conscious states. We have identified these as T[ά] – occurrent thoughts about our sensory states. We can speculate that such higher order thoughts were made possible as the result of new neurological processes and structures– feedback and reflexive mechanisms between responsible centres in the brain. There is evidence that new neural pathways open up as ways of thinking are developed, and attention is focussed (eg Lau and Rosenthal 2011). Our range of concepts

would also have expanded to include *belief* and other intentional terms, allowing us to frame propositional attitudes, to create further first order mental qualitative states (If there is a distinctive what-it-is-like character to the belief that the Eiffel Tower is tall, then this would feature here.)

Eventually, still more conceptual resources are developed, - terms such as 'self' and 'I', and the ability to reflect upon one's condition becomes possible – the thought that I am aware of seeing a green tree becomes a feature of human mental life: I am now introspectively aware of there being mental states I am in. We designate them as T[T[ά]]. This is the culmination of the evolution of consciousness; a process made possible throughout by the development and acquisition of greater, more sophisticated conceptual resources. Those concepts allow us to pay closer attention to (possibly 'to activate') qualitative states which are already inherent in our sensory states.

By contrast, the neo-cartesian view would appear to be far more static. It takes the view that consciousness is an all-or-nothing condition – it is an intrinsic property of mental states. This feature of neo-cartesianism is committed to the idea that new concepts can somehow create entirely new phenomenal qualities in our experience that were not there before. This has yet to be fleshed out with an explanation of how it can be accomplished. Short of this, the neo-cartesian position lacks a sense of degree, and thus of development. If consciousness is an intrinsic property of mental states, it was always an intrinsic property of mental states. But how, and at what point did it leap into existence?

The Higher Order Thought thesis and the Unity Principle.

Recall that Bayne and Chalmers are committed to the proposition that our experiences seem to be tied together in a deep way – by being unified as part of a single encompassing state of consciousness (Chalmers, 2010, p.497). They take the view that when we have three experiences concurrently, (A, B and C – savouring the taste of red wine, looking at its colour and hearing Shostakovich quartet) there is something it is like to taste wine, see the colour and hear the quartet, but there is also something it is like to have all three together. They advance the idea that these experiences are necessarily unified (Chalmers, 2010, p.498). Their unity is defined phenomenally – it can be characterised in what-it-is-like terms. Bayne and Chalmers take the view that the unity claim puts a constraint on theories of consciousness.

Specifically: 'If the higher order thesis is true, the unity thesis is false. And if the unity thesis is true then the higher-order thought theory is false' (Chalmers, 2010, p.533). Our chapter on Bayne's position was directed mainly at the claim of necessity, and instead argued that a subject's consciousness is often or usually disunified: the claim of unity is mistaken. If that succeeded, the constraint is lifted, and the way is clear for the HOT model to give its own account of the 'the robust intuition we have that our conscious mental states constitute in some important way a unity' (Rosenthal, 2005, p.340). This section will present Rosenthal's case.

Rosenthal argues that higher order thoughts sometimes operate on clusters of mental states, as evidenced by the 'cocktail party effect'. He suggests that one suddenly becomes aware of hearing one's name in a conversation that one had until then consciously experienced only as part of a background din. 'For one's name to pop out from that seeming background noise, one must all along have been hearing the separate articulated words of the conversation. But, since one was conscious of one's hearing of the words only as an undifferentiated auditory experience, the higher order thought in virtue of which one was conscious of one's hearing all those words must have represented the hearing of them *as* a single undifferentiated bunch, that is, as a background din.' (Rosenthal, 2005, p.341). This therefore makes a claim that we can have higher order thoughts of a collection of experiences, which might help to explain the sense of unity they seem to share.

This would only impose a sense of unity on an otherwise fragmented collection of mental states. We might also consider a philharmonic concert. It does not seem plausible to say: there is a unity in experiencing the concert, and that that unity is comprised of what-it-is-like to hear the oboes, what-it-is-like to hear the violas, what-it-is-like to hear the trumpets, what-it-is-like to see the conductor set the tempo, what-it-is-like to smell the perfume of a neighbouring listener, all taken together with what it is like to experience the concert. Not only does this seem unnecessarily to overload me with an overwhelming amount of what-it-is-likes, it is not really how it feels. On considering the concert, it does not feel as if all the contents of these perceptions are integrated into a conscious whole – the contents are experienced on a piecemeal basis and not all at once (Plachias, 2011, p.88). Instead of thinking that our sensory state is a unified whole, conscious all the time, it seems much more credible to say (and it is certainly how it feels) that we apply a higher order thought to different features of the sensory tumult. Only some of our sensory states at any one time are conscious – those to which a higher order thought is directed – and there is likely a limit to

this number. We flit around the sensory manifold, and form higher order thoughts of the violins, now the conductor's shirt tails, now back to the double bass, miss the delicious suspended chord in the woodwind because a competing higher order thought brought to consciousness a perception of the falling music score.

If this is right, the HOT model is more compelling; it is, in the view of Platchias, more explanatory (2011, p.89). In the account offered by Bayne and Chalmers, Platchias points out that there is no principle of unification: it just seems a happy coincidence that we seem to have a smooth and unified experience of the philharmonic concert. Nothing about the various mental qualitative properties of each component part of the experience seems to be able to account for the apparent unity. However, by appealing to the HOT model we have a mechanism for structuring our experience. Kant argued that it was through the application of concepts that our chaotically confused and unstructured sensations became organised and comprehensible. The concepts brought to bear by the specific higher order thought launched have that kind of impact on the sensory states we direct our higher order thought towards.

Blunting the challenge of the phenomenal zombie.

One of Chalmers more famous creations is the phenomenal zombie (Chalmers, 1996, pp.96 *et seq*). This is a being which is functionally and behaviourally indistinguishable from other human beings, but which lacks internal phenomenology – they generate no pain qualia on sitting on the drawing pin, or colour qualia on encountering the tomato; there is nothing it is like to be a zombie. Chalmers first postulates their conceivability, saying that it seems logically possible that they could exist. 'I can discern no contradiction in the description' (Chalmers, 1996 p 96).

From this premise, he then argues that their conceivability entails their possibility. (Critics of this argument have focussed their attention on this move.) If this is conceded, then the argument proceeds to the conclusion that physicalism is false. This can be done by invoking Kripke (1972), who contended that all identity statements are necessarily true – ie true in all possible worlds. The same should apply to mind-brain identity claims. The possibility of a world in which zombies exist demonstrates that the mind-brain identity claim is not necessarily true, since as, postulated, it has no internal phenomenology. Hence physicalism is false. Physicalism fails because phenomenal zombies show that something

extra needs to happen for consciousness to be a feature, even of suitably complex physical systems. Consciousness (as defined in phenomenal terms) is thus a non-physical, non-reducible feature unexplained by the laws of physics and properties of matter.

Much ink has been spilt on the zombie question (Kirk 2008 and many others), but it not my project here to engage with these arguments. I want to argue that the HOT model might offer a way of blunting or even embracing the zombie hypothesis.

It seems clear that, if we consider lower forms of life, there is really not much it is like to be them – this would be the case with molluscs or insects, perhaps. At some point in our own phylo-genetic biography, we were presumably creatures for whom it could not be said that there was much that it was like to be that creature. At some point in our evolutionary history, we perhaps were in the same class of animal as zombies. Our behavioural responses to snakes, predators and other stimuli was all of it at the level of early vision described in chapter two. As we developed, our range and sophistication of conceptual resources contributed understanding to our sensory states, and brought conscious awareness into being - late and extra, and for many functions, unnecessary. The zombie state is part of the natural history of humans.

Today, although we now do enjoy conscious awareness of many mental states, there are still many lower sensory states that do not reach awareness, precisely because a thought is not directed to them, as we have argued – the somnambulist, the long-distance driver, the blindsighted subjects.

Although he was not defending HOT theory when he made the comment, Dennett has claimed something similar – 'Are Zombies possible? They're not just possible; they are actual. We're all zombies' (Dennett, 1991, p.406)[41]. This is the claim I wish to make – HOT theory can embrace zombies.

Furthermore, for the kind of zombies we were and sometimes are, HOT theory can provide a compelling rationale for their behaviour, and Chalmers cannot. In the picture presented by Chalmers, the zombie has no qualia (necessarily conscious qualitative properties) to account for all the sophisticated behaviour they show. But then, what story can

[41] Dennett footnoted this: 'It would be an act of desperate intellectual dishonesty to quote this assertion out of context'. Referencing him here is slightly out of context (he was dealing with epiphenomenalism at the time) but my supposition is that this is sufficiently true to the spirit of what he is claiming to escape the charge of desperation. It might in any case be said that Dennett subscribes to a kind of HOT model (Seager, 1999, p.83).

Chalmers present to explain that behaviour? The conceivability of such creatures depends at least in part on a convincing explanation of their behaviour. Without such an account, their acts of snake-flinching will remain a mystery.

HOT theory, as we have argued elsewhere, does have an account of behaviour at this level. Our claim is that we have established that perceptual states can occur unconsciously. Discrimination tests suggest that red sensory states can be distinguished from blue sensory states, even when there is no awareness, ie if there is nothing it is like for one to be in that state. To put it slightly differently, our lower order mental states (ά) are characterised by qualitative properties sufficient to make distinctions, even if they are not brought to consciousness by a HO thought. As Platchias puts it: '[..] I do not see why we should not take such ascriptions of unconscious mental states as literally and realistically as we take ascriptions of conscious mental states in psychological explanations of ordinary behaviour' (Platchias 2011 p. 70).

There is a further, and rather odd consequence to HOT. If the kind of zombie that Chalmers is imagining were somehow spontaneously created today, but equipped with sufficiently powerful conceptual resources[42] to apply to its perceptual states, our argument is that it would not stay a zombie. We may contend: the true possession of a concept C is found in the ability to discriminate between instances of Cs from non-Cs, and the ability to recognise or identify instances of Cs from sensory input. If the zombie has this facility, we can expect it would be capable of directing a higher order thought towards its perceptual states, and, in keeping with the HOT thesis, its lower order mental states would become conscious. Put like this, we expose a problem with the zombie argument as developed by Chalmers – if the HOT model is right, a zombie with the capacity for perceptual states and sufficient conceptual resources, capable of self-reflection and launching these resources in the direction of the perceptual states, is a contradiction in terms.

In short, in subscribing to the HOT model, we naturalise zombies, make them more credible by providing them with a psychology, and ultimately, extinguish them. In doing so, we deprive the zombie hypothesis all of its appeal for dualists.

[42] ie resources which we can assume it has, if it is capable of talking about its inner life, and wields such concepts such as 'qualia' etc

HOT theories – a conclusion.

This thesis concedes that this section is only a rough outline of higher order thought theories and their place in the broader picture of the philosophy of mind. We have identified some of the main claims of the position, and some of the more obvious problems, - and indeed some of the advantages - but the section here cannot be, and does not claim to be a comprehensive treatment of HOT theory. The case made here is to suggest that the dissociations under study might help remove some of the obstacles which neo-cartesianism level against the HOT model. If our case against the necessary unity of consciousness is successful, then, as Bayne and Chalmers concede, then the ground is opened for a more sympathetic reception of the HOT model. Certainly, further work with dissociations offers the prospect of further development of the HOT model.

Conclusion

This thesis has been an exploration of the difficulty presented by our dissociations for certain neo-cartesian philosophers. If it is possible to generalise, their approach to the philosophy of mind has a difficulty with the dissociations insofar as they can be considered as unconscious perception. There is, further, the challenge those dissociations present to the idea of the unity of consciousness.

Working back through our targets, we saw that Ian Phillips argued the empirical ground for classifying our dissociations in such a way that takes them off the table as cases of unconscious perception. At first, his efforts extended simply to raising questions about the paradigms working to explore them – problems of the criterion and other confounds. It was claimed that there was legitimate doubt about that the paradigms established that cases of blindsight constituted real cases of unconscious perception. His arguments then extended to the definitions of perception and consciousness, which, once in the frame, work to raise the bar for any condition to be considered as a case of unconscious perception. If our discussion dealt with this successfully, we kept alive the possibility of unconscious perception against his argument, but it should be noted that Phillips's case against unconscious perception does not rule out exemplars in principle. He puts a great deal of pressure on the proponents of and arguments for unconscious perception, such that cases of it may not be so clear cut, but he is open to the possibility that paradigms might one day be conclusive. Accordingly, while it might be important to blunt his challenges, our argument against him need not be conclusive either.

Against the claim made by Tim Bayne and David Chalmers, that there is a necessary unity of experience, and the concomitant claim that the dissociations do not constitute a case for thinking differently, we found that the dissociations were not so easily dismissed. Cases such as the alien hand syndrome, and other conditions work against the unity thesis. Matters are acute when we enlist the split brain cases as variants in the dissociation stable. Maintaining the unity thesis against the evidence derived from such cases is tortuous and difficult to maintain. The unity thesis also faces problems from recent interest in synaesthesia. Bayne does not consider the syndrome, but if synaesthesia cases are cases of phenomenological 'danglers', then they also constitute a problem for his version of unity. If

our discussion here was successful, we removed the constraint that his unity contention puts on higher order thought theories.

In the discussion on Chalmers and *The Conscious Mind*, we saw that Chalmers sets out with a pre-systematic commitment in identifying consciousness with the possession of mental qualitative properties – qualia. According to this position, mental states are necessarily conscious – an approach, which neo-cartesians hope, is fleshed out by the 'what-it-is-like' locution: there is something it is like to smell a rose, or coffee. Perceptual states to him are by definition the possession of qualia. Chalmers would extend this characteristic to propositional attitudes as well – there is something it is like to think of a lion: when I do so, 'there is a whiff of leonine quality to my phenomenology' (Chalmers, 1996, p.10). However, it is clear that, if perceptual states are by definition conscious states, then that definition rules out dissociations as cases of unconscious perception. There can be no such thing. Our approach in this section was to consider the problems associated with neo-cartesian phenomenal concepts, and on the strength of this, a case for extending the conception of perception to allow it to embrace unconscious perception.

Where Chalmers does allow that blindsight and other dissociations might be problematic, is for his principle of structural coherence between psychological properties and phenomenal properties. With this principle, Chalmers hopes to establish a parallel between the scientific understanding of consciousness ('cognition') and what he understands as the phenomenal aspect of consciousness. As he remarks, the dissociations are sometimes put forward as cases in which consciousness and the associated functional role come apart. 'If there were truly a case in which the functional role and experience were dissociated, it would clearly create problems for the coherence principle' (Chalmers, 1996, p.227).

It has been the contention of this thesis that blindsight and other dissociations do create problems for the coherence principle; his dismissal of blindsight, can, on the present view, be met with new paradigms and fresh evidence. In this section, the argument turns to more empirical matters. Chalmers hoped that the coherence principle might be helpful in providing an indirect way to research consciousness – it 'acts as a kind of epistemic lever in allowing researchers to infer conclusions about experience from third-person data' (Chalmers, 1996, p.234). In arguing for this, Chalmers wants to preserve an account of mind which is at once grounded in materialism – a functional account of the workings of mind, relying on the neural correlates of consciousness, together with a commitment to the non-

reductive phenomenal account of consciousness – as he puts it: 'a *non-reductive functionalism* […] a way of combining functionalism with property dualism' (Chalmers, 1996, p.249). However, to repeat, Chalmers himself concedes that blindsight could create difficulties for this coherence. We can now see that it (and the other dissociations) do indeed challenge his notion of coherence.

Chalmers also calls his position: 'naturalistic dualism' (Chalmers, 1996, ch 4). The non-reductive, phenomenal account of consciousness (the dualism) is hard to square with the naturalism: the commitment to some version of physicalist monism. Such a '*cakeist*' position then also requires costly, exotic accounts of the causal relationship between phenomenal properties and the physical world. Chalmers wavers between epiphenomenalism (Chalmers, 1996, p.160) and panprotopsychism (Chalmers, 1996, p.293). Our arguments here should be understood as contributing in a small way, to the general case against Chalmers' philosophy of mind – in removing some of the neo-cartesian impediments to other positions, we open the door to theories which do not have such a tension in them.

The dissociations.

We return now to the main thrust of this thesis. The argument has been that the dissociations here – blindsight, agnosia, neglect, optic ataxia and so on make a powerful case for a stipulative treatment of perception that allows for unconscious perception. They also provide empirical evidence for unconscious perception where philosophers are open to the idea. It is clear to see why neo-cartesianism is uncomfortable with these phenomena, why they might closely examine the paradigms exploring them, and be inclined to dismiss them. Philosophers who incline to a version of neo-cartesianism and who are open to testing their position empirically, (here, Phillips and Bayne) have put the dissociations very much under the microscope. Despite their efforts, the dissociations make a real case for thinking that some mental states do not have some intrinsic property of consciousness.

The pre-theoretical, stipulative position taken by Chalmers, - the view that perception is necessarily conscious – is of course immune to the empirical evidence presented by dissociation studies. But a different conceptual position would seem, on the strength of what we have dealt with, seems all the more reasonable. Chalmers' phenomenal concepts ultimately commit him to problematic or costly consequences – his stipulation involves the

non-reducibility of qualitative mental properties and this has led him and others to problematic accounts of their relationship to the physical world - emergence, the possibility of epiphenomenalism, or a commitment to the view that experience is a fundamental feature of matter, alongside mass, spin, and charge.

On the other hand, a pre-theoretical stipulation that does not regard unconscious perception as an oxymoron is more plausible in that it would seem to permit an account which grounds consciousness more convincingly in the physical world. That account, it is contended, makes better sense of the empirical work in the field. While Chalmers and others build the notion of consciousness into verbs of perception, and experience, our argument has been – the dissociations constitute cases of unconscious perception, and prompt the idea that we should uncouple consciousness from its role as a necessary feature of perceptual states. The same case has already been made for intentional states. Once this uncoupling is done, a new architecture of explanation becomes possible, with greater explanatory power. That architecture offers a more naturalistic account of mind, and it also provides for a developmental account of consciousness.

We can accordingly argue thus: to be conscious is to have a mental state which we are conscious *of*. Not all mental states are ones which we are conscious of. Our experiential field is sparse, not lavish. Mental states we are conscious of are ones (according to the HOT hypothesis) towards which we direct a higher order thought. In the early stages of our development, we might lack the conceptual resources that brings mental states to consciousness. Gradually, as we acquire the conceptual resources necessary for a higher order thought, we make it possible for more of our perceptual and intentional states to become conscious. In certain situations, a subject has suffered some trauma – blindsight cases, for example, some mental states are perceptual states, which it is difficult to become conscious of in virtue of damage to the feedback mechanisms in the brain's neurology. The role of V1 in the visual system is part of the neurological mechanism allowing us to fashion a thought about our visual states. Once compromised, in blindsight, we remain partially functionally capable of perceiving stimuli, but without the capacity to generate consciousness of that perception.

Of course, this kind of perception (if that is what it is) is not like that in standard cases, and describing blindsight successes, and successful discriminations in neglect and prosopagnosia as cases of *perception* is yet regarded as a stretch by some. In each case,

something is generally absent or unusual – subjects have to be prompted, or they lack subjective awareness of the stimuli, or they have an unaccountable superfluity of phenomena, as in synaesthesia. On the other hand, as we have seen, this is often accompanied by an astonishing success – TN's navigation down an obstacle strewn corridor, DF's ability to catch a ball she is insistent she does not see, and so on. It is this feature of the cases that merits a re-think of our definitions. Whatever definition of perception we arrive at must be broad enough to accommodate these exceptional cases.

We might also advert to our prefatory remarks at the beginning of the chapter on Bayne's Unity Contention. There, we wrestled with the temptation to ascribe psychological predicates to hemispheres of the brain, arguing that in the split brain cases, a case can be made for thinking that one hemisphere perceived stimuli that were denied to the other. If in the end we were not entirely successful in avoiding the temptation, that perhaps makes a further point about the concept or definition of perception as it is used by philosophers and in unexceptional cases. Dramatic changes to the brain's workings of the kind we see in all these cases put considerable pressure on us to consider the extension of concepts such as perception.

In *The Conscious Mind*, Chalmers urges us to take consciousness seriously, and that we should avoid the temptation to redefine it as something it is not (Chalmers, 1996, p xii). In a similar way, this thesis urges that we should take the dissociations seriously, and not define them, and the concepts they concern, in such a way as to take them out of consideration. They present a more interesting challenge than that. Philosophy is richer if it accommodates them, and synaesthesia, properly.

Bibliography

Akins K (1993[a]) 'What it is like to be boring and myopic? In: Dahlbom B (ed): '*Dennett and his Critics'*.

Akins K: (1993[b]) 'A bat without qualities*?* In: M Davies and G Humphreys (eds) '*Consciousness*'.

Andino, Menendez, Khateb, Landis and Pegna (2008) 'Electro-physiological correlates of affective blindsight' in: *Neuroimage* 44 (2).

Atkinson, Campbell and Francis (1976) 'The magic number 4 ± 0; a new look at visual numerosity judgements' *Perception* 5, 327-34.

Azzopardi and Cowey (1997) 'Is blindsight like normal, near threshold vision?' *Proceedings of the National Academy of Sciences* Vol 94 no 25, 14190-4

Azzopardi and Cowey (1998) 'Blindsight and Visual Awareness' in*: Consciousness and Cognition* 7 (3), 292-311

Azzopardi and Hock (2011) 'Illusory Motion Perception in Blindsight' in: *Proceeding of the National Academy of Sciences* 108 (2), 876-881

Balog, K (2009) 'Phenomenal Concepts' in: (eds) McLaughlin, Beckermann, Walter '*The Oxford Handbook of Philosophy of Mind*' Oxford: Oxford University Press.

Barbur, Watson, Frackowiak and Zeki (1993) 'Conscious Visual Perception without V1' in: *Brain*, 116 (6), 1293-1302

Baron-Cohen, Bor, Billington, Asher, Wheelwright, Ashwin (2007) "Savant Memory in a Man with Colour Form-Number Synaesthesia and Asperger". *Journal of Consciousness Studies 14 (9–10),* 237-252

Bartels and Zeki (2006) 'The temporal order of binding visual attributes' in: *Vision Research* 46 (14), 2280-2286

Bauer RM (1984) 'Autonomic recognition of names and faces in in prosopagnosia: a neuropsychological application of the Guilty Knowledge Test.' In *Neuropsychologia* 22 (4), 457-469

Bayne, T (2010) '*The Unity of Consciousness'*. Oxford: Oxford University Press.

Bayne T and Chalmers D (2010) '*What is the Unity of Consciousness*?' in: Chalmers 2010

(eds) Bayne, Cleeremans and Wilken (2009) '*The Oxford Companion to Consciousness'*. Oxford: Oxford University Press.

Baynes, Tramo, Reeves, and Gazzaniga (1997) 'Isolation of a right hemisphere cognitive system in a patient with anarchic (alien) hand sign' *Neuropsychologia*; 35 (8), 1159-1173

Beckers and Zeki (1995) 'The consequences of inactivating areas V1 and V5 on visual motion perception' in: *Brain* 118 (1), 49-60

(eds) Bennett and Hill (2014) '*Sensory Integration and the Unity of Consciousness'*. Cambridge MA: MIT Press.

Bennett and Hacker (2003) '*Philosophical Foundations of Neuroscience'*. Blackwell.

Berti A. and Rizzolatti G. (1992) 'Visual processing without awareness: Evidence from unilateral neglect' in *Journal of Cognitive Neuroscience, 4* (4), 345-351

Binda P and Murray S (2015) 'Keeping a large pupilled eye on high-level visual processing' in: *Trends in Cognitive Science* 19 (1), 1-3

Bisiach E and Luzzatti C (1978) 'Unilateral neglect of representational space'. *Cortex* 14 (1), 129-133

Bittner T (2011) 'Parity cuts both ways: Split Brains and Extended Cognition', *teorema* vol 30/2, 19-33

Block N (1978) 'Troubles with Functionalism' in: (ed) CW Savage: *'Perception and Cognition'*.

(eds) Block, Flanagan and Gulzeldere (1999) '*The Nature of Consciousness'*. MIT: Bradford.

Block N (2002) 'Concepts of Consciousness' in: D Chalmers: '*Philosophy of Mind; classical and contemporary readings*'.

Block N (2011) 'The Higher Order approach to consciousness is defunct' in: *Analysis* 71 no 3, 419-431

Block N (2014) 'The Anna Karenina theory of the unconscious' *Neuropsychoanalysis* 13 (1), 34-37

Block N and Phillips I (2016) 'Debate on Unconscious perception' in: B Nanay (ed*) Current Controversies in Philosophy of Perception.*

Block N (2015) 'The Anna Karenina Principle and Skepticism about Unconscious Perception' – *Philosophy and Phenomenological Research* 93 (2), 452-459

Bogen JE (1997) 'Physiological consequences of complete or partial commissural section' in: *Surgery of the third ventricle* Apuzzo MLJ (ed) (1997)

Bor, Billington, Baron-Cohen (2008) "Savant Memory for Digits in a Case of Synaesthesia and Asperger Syndrome is Related to Hyperactivity in the Lateral Prefrontal Cortex". *Neurocase: the Neural Basis of Cognition. Taylor & Francis Online.* ***13*** *(5–6).*

Breitmeyer BG (2015) 'Psychophysical "blinding" methods reveal a functional hierarchy of unconscious visual processing' *Consciousness and Cognition* 35, 234-250

Brogaard B (2012) 'Non-visual consciousness and visual images in blindsight' *Consciousness and Cognition* 21 (1), 595-6

Brogaard B (2012) 'Seeing as a Non-Experiential State: The case from Synaesthesia and visual imagery' in: '*Consciousness inside and out: Phenomenology, Neuroscience and the Nature of Experience'* (ed) Brown R.

Brogaard B (2014) 'Color synaesthesia and its Philosophical Implications' in: *The Routledge Handbook of Philosophy of Colour* (eds) Brown D and Macpherson F.

Brogaard B (2015) 'Type 2 Blindsight and the nature of visual experience', *Consciousness and Cognition*, vol 32 March, 92-103

Brogaard B: (2015) 'Synaesthesia as a challenge for Representationalism' in: *Blackwell Companion to Experimental Philosophy* (eds) Buckwalter W and Sytsma J.

(ed) Brown R (2012) '*Consciousness inside and out: Phenomenology, Neuroscience and the Nature of Experience'* Dordrecht: Springer.

(eds) Buckwalter W and Sytsma J (2016) *'Blackwell Companion to Experimental Philosophy*' Oxford: Blackwell.

Burge T (2010) '*Origins of Objectivity'*. Oxford: Oxford University Press.

Byrne A (1997) 'Some like it hot: Consciousness and Higher Order Thoughts' in: *Philosophical studies* 86, 103-129

Callwood J (1990) '*The Sleepwalker*'. Toronto: Lester and Orpen Dennys.

Campion Latto and Smith (1983) 'Is blindsight an effect of scattered light, spared cortex, and near-threshold vision?' *Behavioural and Brain Sciences*, 6 (3), 423-486

Carey S (2009) '*The Origin of Concepts'*. New York: Oxford University Press.

Carlson N (2013) '*Physiology of Behavior'*. Needham Heights MA: Allyn and Bacon.

Chalmers D (1996) '*The Conscious Mind'*. Oxford: Oxford University Press.

Chalmers D (2002) '*Philosophy of mind; contemporary and classical readings*'. Oxford: Oxford University Press.

Chalmers D (2004) 'The representational character of experience' in: *The future for Philosophy* (ed) B Leitner.

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

Chalmers D 'First Person methods in the science of consciousness' (accessed from consc.net/papers/firstperson.htm on 3/10/2016

Christensen, Kristiansen, Rowe, and Nielsen (2008) 'Action-blindsight in healthy subjects after transcranial magnetic stimulation' in: *Proceedings of National Academy of Sciences* 105 (4), 1353-1357

Cleeremans A (2003) '*The Unity of Consciousness: Binding, Integration, and Dissociation'* Oxford: Oxford University Press.

Cook N (1986) '*The Brain Code: Mechanisms of information transfer and the role of the corpus callosum*'. London: Methuen.

Corso JF (1967) '*The Experimental Psychology of Sensory Behaviour'*. Holt: Rinehart and Winston.

(eds) Cottingham, Stoothoff and Murdoch (1985) '*The Philosophical Writings of Descartes*' Cambridge: Cambridge University Press.

Churchland P (2000) '*Neurophilosophy*'. MIT: Bradford.

Cytowic RE and Eagleman DM (2011) '*Wednesday is indigo blue; Discovering the Brain of Synaesthesia'*. MIT: Bradford.

(ed) Dahlbom B (1996) '*Dennett and his critics'*. Blackwell.

de Gelder, Pourtois and Weiskrantz (2002) 'Fear recognition in the voice is modulated by unconsciously recognised affective pictures' in*: Proceedings of the National Academy of Sciences* 99 (6) 4121-4126

(eds) Davies M and Humphries G (1993) '*Consciousness*'. Oxford: Blackwell.

de Gelder, Tamietto, Boxtel, Goebel, Sahraie, van den Stock, Stienen, Weiskrantz and Pegna (2008) 'Intact Navigation Skills after bilateral loss of Striate Cortex' in: *Current Biology* 18 (24), 1128-1129

Dennett D (1978) '*Brainstorms*'. Brighton: Harvester.

Dennett D (1992) '*Consciousness Explained'*. St Ives: Penguin.

Dennett D (1994) 'Quining Qualia' in: (eds) Marcel and Bisiach: '*Consciousness in Modern Science'*.

Dennett D (1998) '*Brainchildren: Essays on Designing Minds'*. London: Penguin.

Dennett D (2017) '*From Bacteria to Bach and Back – the Evolution of Minds*'. St Ives: Penguin Random.

Descartes R (1998) '*Meditations and other metaphysical writings'* London: Penguin Classics.

Dretske, F (1995) '*Naturalising the Mind'*. MIT: Bradford.

Dretske, F (2006) 'Perception without Awareness' in: *Perceptual Experience* (eds) Gendler and Hawthorne.

Eliassen JC, Baynes K, Gazzaniga MS (2000) 'Anterior and posterior callosal contributions to simultaneous bimanual movements of the hands and fingers' *Brain* 123 (12), 2501-11

Farrell J (2016) "'What it is like' talk is not Technical Talk" *Journal of Consciousness Studies* 23 (9-10), 50-65

Feinberg and Mallatt (2017) '*The Ancient Origins of Consciousness'*. Cambridge MA: MIT

Fish W (2010) '*Philosophy of Perception'*. Abingdon: Routledge.

Flanagan O (1992) '*Consciousness Reconsidered'*. MIT: Bradford.

Franz EA, Eliassen JC, Ivry RB, Gazzaniga MS (1996) 'Dissociation of spatial and temporal coupling in the bimanual movements of Callosotomy patients' *Psychological Science* 7 (5), 306-310

Funnell MG, Corballis PM, and Gazzaniga MS (2000) 'Insights into the functional specificity of the human corpus callosum' *Brain* 123 (5), 920-6

Gazzaniga MS (1967) 'The Split Brain in Man' *Scientific American* vol 217, 24-29

Gazzaniga MS (1969) 'Cross-Cuing Mechanisms and Ipsilateral Eye-Hand Control in Split-Brain Monkeys' *Experimental Neurology* 23 (1), 11-17

Gazzaniga MS (2015) '*Tales from both sides of the Brain'*. New York: HarperCollins.

Gendler S and Hawthorne J (eds) (2006) '*Perceptual Experience'* Oxford: Oxford University Press

Gennaro R (2005) 'The HOT Theory of Consciousness: Between a Rock and a Hard Place?' in: *Journal of Consciousness Studies* 12 (2), 3-21

Goldman A (1993) 'Consciousness, folk psychology, and cognitive science' in: *Consciousness and Cognition* 2 (4), 364-382

Gonzalez Andino, Menendez, Khateb, Landis and Pegna (2009) 'Electro-physiological correlates of affective blindsight' in: *Neuroimage* 44 (2), 581-9

Goodale and Milner (2005) '*Sight unseen'*. Oxford: Oxford University Press.

(eds) Graham and Stephens (1994) '*Philosophical Psychopathology'*. MIT: Bradford.

Gray J, Chopping, Nunn, Parslow, Gregory, Williams, Brammer, Baron-Cohen (2002) 'Implications of synaesthesia for Functionalism' *Journal of Consciousness Studies*, 9 (12), 5-31

Gray J (2004) '*Consciousness; Creeping up on the Hard Problem'*. Oxford: Oxford University Press.

Gray J (2005) 'Synaesthesia: a window on the Hard Problem of Consciousness' in: Robertson and Sagiv.

Gray J, Parslow, Brammer, Chopping, Vythelingum, ffytche (2006)'Evidence against Functionalism from Neuroimaging of the Alien Colour Effect in Synaesthesia' *Cortex* 42 (2), 309-318

Gray, R (2004) 'What Synaesthesia really tells us about Functionalism' *Journal of Consciousness Studies*, 11 (9), 64-69

Green D and Swets J (1966) '*Signal Detection Theory and Psychophysics'*. Oxford: John Wiley.

Guariglia, Palermo, Piccardi, Iaria and Incoccia (2013) 'Neglecting the Left Side of a City Square, but not the Left Side of its Clock: Prevalence and Characteristics of Representational Neglect' in: *Plos One* 8 (7), e67390

Guttenplan S (ed) (1998) '*A Companion to the Philosophy of Mind'*. Oxford: Blackwell.

Hacker PMS (1986) '*Insight and Illusion'*. Oxford: Oxford University Press.

Harman G (1990) 'The Intrinsic Quality of Experience' in *Philosophical Perspectives* 4, 31-52

Hassin, Uleman and Bargh (2005) 'The New Unconscious'. Oxford: Oxford University Press.

Heil J (1998) 'Privileged Access' in: *Mind* 97, 238-251

Heil J (2013) '*Philosophy of Mind: A Contemporary Introduction'*. Abingdon: Routledge.

Holt J (2003) '*Blindsight and the nature of consciousness'*. Ormskirk: Broadview.

Humphrey N (1974) 'Vision in a monkey without striate cortex.' *Perception* 3, 241-255

Humphrey N (1998) 'Cave art, autism and the evolution of the human mind' *Cambridge Archaeological Journal* 8 (2), 165-191

Humphrey N (1992) '*A History of the Mind'*. New York: Copernicus.

Humphrey N (2006) '*Seeing Red'*. Cambridge MA: Belknap, Harvard University.

Hupé, J-M, Bordier C, and Dojat M (2012) "Neural basis of grapheme-colour synaesthesia"*. in: Cerebral Cortex 22 (7), 1922-1233*

Hurley S (1996) 'Unity and Objectivity' in: (ed) Peacocke C: 'Objectivity, Simulation and the Unity of Consciousness'.

Hurley, S (2003) 'Action, the Unity of Consciousness, and Vehicle Externalism' in: '*The Unity of Consciousness: Binding, Integration, and Dissociation'*, (ed) Axel Cleeremans.

(ed) Hyman J (2017) '*Investigating Psychology'*. London: Routledge.

(eds) Ingle, Goodale, and Mansfield (1982) '*Analysis of Visual Behavior'* Cambridge MA, MIT

Jackson F (1982) 'Epiphenomenal Qualia' in: *Philosophical Quarterly* 32.

Jaynes J (1976) '*The Origin of Consciousness in the Breakdown of the Bicameral Mind'*. Boston: Houghton Mifflin.

James, W (1890) '*The Principles of Psychology'*. New York: Dover.

Jelicic, Bonke, Wolters and Phaf (1992) 'Implicit memory for words presented during anaesthesia' in: *European Journal of Cognitive Psychology*, 4, 71-80

Jelicic, de Roode, Bovill and Bonke (1992) 'Unconscious learning during anaesthesia.' in: *Anaesthesia*, 47 (10), 835-7

Jiang et al (2006) 'A gender and sexual orientation-dependent spatial attention effect of invisible images. *Proc Nat Acad Sci* 103 (45), 17048-17052

Kentridge RW (2015) 'What is it like to have type-2 blindsight? Drawing inferences from residual function in type-1 blindsight' in: *Consciousness and Cognition* 32, 104-128

Kiefer, Ansorge, Haynes, Hamker, Mattler, Verleger, and Niedeggen (2011) 'Neuro-cognitive mechanisms of conscious and unconscious visual perception: from a plethora of phenomena to general principles' in: *Advances in Cognitive Psychology* 7 (2), 55-67

Koch C (2012) '*Consciousness*'. Cambridge MA: MIT Press.

Koch C, Massimini M, Boly M, Tononi G (2016) 'Neural Correlates of Consciousness; progress and problems' in: *Nature Reviews Neuroscience* 17 (5), 307-321

Kornblith H (1998) 'What is it like to be me?' in: *Australasian Journal of Philosophy* 76 (1), 48-60.

Kirk R (1994) '*Raw Feeling'*. Oxford: Oxford University Press.

Kirk R (2005) '*Zombies and Consciousness'*. Oxford: Oxford University Press.

Kuhn T (1996) '*The Structure of Scientific Revolutions'* 3rd edition. University of Chicago Press.

Ladavas *et al* (1993) 'Implicit associative priming in a patient with left visual neglect' *Neuropsychologica* 31, 1307-1320

Lamme, VAF (2001) 'Blindsight: the role of feedforward and feedback corticocortical connections' *Acta Psychologica* 107, 209-228

Lau and Rosenthal (2011) 'Empirical support for higher-order theories of conscious awareness' in: *Trends in Cognitive Science* 15 (8), 365-73

Leitner, B (ed) (2004) '*The Future for Philosophy'* New York: Oxford University Press.

Levine J (1983) "Materialism and qualia: the explanatory gap". *Pacific Philosophical Quarterly*, 64 (4), 354-61

Levine J (2001) '*Purple Haze; the puzzle of consciousness'*. Oxford: Oxford University Press.

Lewis CI (1929) '*Mind and the World Order'*. New York: Charles Scribner's Sons.

Livermore and Laing (1998) 'The influence of odor type on the discrimination and identification of odorants in multicomponent odor mixtures.' *Physiology and Behavior*, 65 (2), 311-320

(eds) Ludlow, Nagasawa and Stoljar (2004) '*There's Something about Mary; Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument'*. MIT: Bradford.

Lycan W (1996) '*Consciousness and Experience'*. MIT: Bradford.

Lycan W (2001) 'The Case for Phenomenal Externalism' in *Nous* 35, (1).

Mack and Rock (1998) '*Inattentional Blindness*'. MIT: Bradford.

Macmillan, Creelman (2005) '*Detection Theory; a user's guide'*. Mahwah NJ: Lawrence Eribaum.

(eds) Marcel A and Bisiach E (1994) '*Consciousness in Contemporary Science'*. Oxford: Oxford University Press.

Marcel AJ (1998) ‘Blindsight and shape perception : Deficit of visual consciousness or of visual function?’ *Brain* 121, 1565-1588

Marks C (1981) ‘*Commissurotomy, Consciousness and Unity of Mind’*. MIT: Bradford

Marshall and Halligan (1988) ‘Blindsight and Insight in visuo-spatial neglect’ *Nature* 336 p 766-767

Martino and Marks (2001) ‘Synaesthesia, strong and weak’ in: *Current Directions in Psychological Science*, 10 (2), 61-65

Maslin KT (2001) ‘*An Introduction to the Philosophy of Mind’*. Oxford: Polity.

Matey J (2014) ‘Can Blue mean 4?’ in: (eds) Bennett and Hill.

McGinn C (1991) ‘*The Problem of Consciousness’*. Oxford: Blackwell.

McPhail, E (1998) ‘*The Evolution of Consciousness’*. Oxford: Oxford University Press.

McGilchrist I (2009) ‘*The Master and his Emissary’*. Yale University Press.

Merikle P, Dixon MJ, Smilek D (2002) ‘The role of synaesthetic photisms on perception, conception and memory’ – Lecture presented at the 12th Annual meeting of the Cognitive Neuroscience Society at San Franscisco. See: *Journal of Cognitive Neuroscience*, 2001 13 (7), 930-936

(eds) Milner and Rugg (1992) ‘*The Neuropsychology of Consciousness*’. London: Academic Press.

Mole C (2008) ‘Attention and Consciousness’ in: *Journal of Consciousness Studies* 15 (4), 86-104

Moutoussis and Zeki (1997) ‘Functional segregation and temporal hierarchy of the visual perceptive systems’ in: *Proceedings of the Royal Society of London: Biological Sciences* 264, 1407-1414

Nagel T (1974) ‘What it is like to be a bat?’ *Philosophical Review* 4, 1974. References are taken from Chalmers 2002, in which the paper features.

Nagel T (1986) ‘*The View from Nowhere’*. New York: Oxford University Press.

(ed) Nanay B (2014) ‘*Perceiving the World’* Oxford: Oxford University Press.

(ed) Nanay B (2016) ‘*Current Controversies in Philosophy of Perception’* New York: Routledge

(ed) Noe (2002) ‘*Is the visual World a Grand Illusion*?’ Thorverton: Imprint Academic.

Norman, Akins, Heywood and Kentridge (2014) ‘Color Constancy for an Unseen Surface’ in: *Current Biology* 24 (23), 2822-2826

Nunn JA, Gregory LJ, Brammer M, Williams SCR, Parslow DM, Morgan MJ, Morris RG, Bullmore ET, Baron-Cohen S, and Gray JA (2002) ‘Functional magnetic resonance imaging of synaesthesia: activation of V4/V8 by spoken words’ in: *Nature Neuroscience* 5 (4), 371-375

O'Shaughnessy, B (2000) '*Consciousness and the world'*. Oxford: Oxford University Press.

Overgaard, Fehl, Mouridsen, Bergholt, and Cleeremans (2008) 'Seeing without seeing? Degraded vision in a Blindsight Patient.' *PloS ONE* 3 (8), e3028

Overgaard and Grunbaum (2011) 'Consciousness and modality: on the possible preserved visual consciousness in blindsight patients.' *Consciousness and Cognition* 20 (4), 1855-1859

Paillard, Michel and Stelmach (1983) 'Localisation without Content: a Tactile analogue of Blindsight' in: *Archives of Neurology*, 40, 548-551

Papineau, D (2011) 'Phenomenal concepts and the private language argument' in: *American Philosophical Quarterly* 48 (2), 175-184

(ed) Peacocke C (1996) '*Objectivity, Simulation and the Unity of Consciousness'*. British Academy: Oxford University Press.

Persaud and Lau (2008) 'Direct assessment of qualia in a blindsight participant' in: *Consciousness and Cognition* 17 (3), 1046-1049

Peters, Ro, and Lau (2016) 'Who's afraid of response bias?' *Neuroscience of Consciousness* doi: 10.1093/nc/niw001

Peters M, Kentridge R, Phillips I and Block N (2017) 'Does unconscious perception really exist? Continuing the ASSC20 debate' *Neuroscience of consciousness* 2017 (1) nix015

Phillips I: 'Unconscious Perception Reconsidered' Draft paper accessed 2016 from Ian Phillips' website: https://www.ianbphillips.com/papers.html. Permission to quote granted.

Phillips I (2016) 'Consciousness and Criterion: On Block's case for Unconscious Seeing' *Philosophy and Phenomenological Research* 93 (2), 419-451

Phillips I and Block (2016) 'Debate on Unconscious Pereception' in: Nanay B (ed): Current *Controversies in the Philosophy of Perception'*.

Platchias D (2012) '*Phenomenal Consciousness'*. Durham: Acumen.

Prinz, J (2010) 'When is Perception Conscious?' in Nanay, B. (ed.) *Perceiving the World* Oxford.

Prinz J (2013) 'Attention, Atomism and the Disunity of Consciousness' *Philosophy and Phenomenological Research*, 86 (1), 215-222

Prinz J (2015) '*The Conscious Brain'*. Oxford: Oxford University Press.

Prinz M and F-I Pris: 'Phenomenal concepts and Wittgenstein's Private Language Argument' in: Papers of the 36th Wittgenstein Symposium 2013. Accessed online: philpapers/rec/PRIPCA-6

Puccetti R (1981) 'The Case for Mental Duality: Evidence from Split Brain Data and other considerations.' *Behavioural and Brain Sciences* 4. 93-123.

Ramachandran VS and Hubbard EM (2005) 'The Emergence of the Human Mind: Some clues from Synaesthesia' in: Robertson and Sagiv.

Reingold and Merikle (1990) 'Recognition and lexical decision without detection: Unconscious perception?' in: *Journal of Experimental Psychology*: *Human Perception and Performance*, 16 (3), 574-583

Ro, Lee, Shelton and Chang (2004) 'Extrageniculate mediation of unconscious vision in transcranial magnetic stimulation-induced blindsight' in: *Proceedings of the National Academy of Sciences* 101(26), 9933-9935

(ed) Robinson H (1993) *'Objections to Physicalism'*. Oxford: Clarendon Press.

(eds) Robertson LC and Sagiv N (2005) '*Synaesthesia; Perspectives from cognitive neuroscience'*. Oxford: Oxford University Press.

Rosenthal D (2005) *'Consciousness and Mind'*. Oxford: Oxford University Press.

Rosenthal D (2011) 'Exaggerated Reports: reply to Block' in: *Analysis* vol 71 no 3.

Russell B (1912) '*The Problems of Philosophy'* Oxford: Oxford University Press (1959 ed)

Sacks, O (1985) '*The Man who Mistook his Wife for a Hat'*. London: Picador.

Schechter E (2012) 'The switch model of split-brain consciousness' *Philosophical Psychology*, 25 (2), 203-226

Schiffer, Zaidel, Bogen and Chasan-Taber (1998) 'Different psychological status in the two hemispheres of two split-brain patients.' In: *Neuropsychiatry, Neuropsychology, and Behavioural Neurology*, 11 (3), 151-156

Schwitzgebel E (2011) '*Perplexities of Consciousness'*. MIT: Bradford.

Searle J (1992) '*The Rediscovery of the Mind'*.

Searle J (1997) '*The Mystery of Consciousness'*. London: Granta

Searle J (2010) '*Consciousness and Language'*.

Seager W (1999) '*Theories of Consciousness'*. London: Routledge.

Shallice T (1988) 'Information processing models of consciousness' in: Marcel and Bisiach '*Consciousness in Contemporary Science'* Oxford: Oxford University Press.

Shallice T (1998) '*From Neuropsychology to Mental Structure'*. Cambridge: Cambridge University Press.

Simner J (2012) 'Defining Synaesthesia', in: *British Journal of Psychology* 103 (1), 1-15

Smith and Jokic (eds) (2003) '*Consciousness. New Philosophical Perspectives.'* New York: Oxford University Press.

(eds) Smithies and Stoljar (2012) '*Introspection and Consciousness'*. Oxford: Oxford University Press.

Sollberger M (2013) 'Rethinking synaesthesia' in: *Philosophical Psychology*, 26 (2) 171-187

Sperling G (1960) 'The information available in brief visual presentations.' *Psychological monographs*, 74 (11), 1-29

Sperry RW (1968) 'Hemisphere Deconnection and Unity in Conscious Awareness' in: *American Psychologist* 23 (10), 723-733

Sperry RW (1974) 'Lateral Specialisation in the surgically separated hemispheres' in: Schmitt and Worden (eds) *Neuroscience, 3rd Study Programme*. 1 (3), 5-19

Sperry RW, Zaidel E and Zaidel D (1979) 'Self-recognition and social awareness in the disconnected minor hemisphere.' *Neuropsychologia* 17 (2) 153-166

Springer and Deutsch (1985) '*Left Brain, Right Brain'*. WH Freeman

Stalnaker R (2010) '*Our Knowledge of the Internal World'* New York: Oxford University Press.

Stoerig and Barth (2001) 'Low Level Phenomenal Vision despite unilateral destruction of primary visual cortex' in: *Consciousness and Cognition* 10 (4), 588-593

Strawson G (1994) '*Mental Reality'*. Cambridge MA: MIT Press.

Strawson G *et al* (ed Freeman A) (2006) '*Consciousness and its place in Nature'*. Exeter: Imprint Academic.

Strawson G (2008) '*Real Materialism and other essays'*. Oxford: Oxford University Press.

Strawson PF (1959) *'Individuals'* Routledge.

Tamietto, Castelli, Vighetto, Perozzo, Geminiani, Weiskrantz and Gelder (2009) 'Unseen facial and bodily expressions trigger fast emotional reactions.' *Proceedings of the National Academy of Science* 106 (42), 17661-6

Trevethan, Sahraie and Weiskrantz (2007) 'Form discrimination in a case of blindsight' *Neuropsychologia* 45 (9), 2092-2103

Tye M (1999) '*Ten Problems of Consciousness'*. MIT: Bradford.

Tye M (2003) '*Consciousness and Persons'*. MIT: Bradford.

Tye M (2010) 'Attention, Seeing and Change Blindness' *Philosophical Issues* 10.

Ungerleider LG and Mishkin M (1982) 'Two cortical visual systems' in: (eds) Ingle, Goodale, and Mansfield: '*Analysis of Visual Behavior*'

Van der Klaauw et al: 'Divergent effects of central melanocortin signalling on fat and sucrose preference in humans' *Nature Communications* doi: 10.1038/ncomms13055

Velmans M (2000) '*Understanding Consciousness'*. Hove: Routledge.

Velmans M (1991) 'Is human information processing conscious?' *Behavioural and Brain sciences* 14 (4), 651-69

(eds) Velmans and Schneider (2007) *'The Blackwell Companion to Consciousness'*. Oxford: Blackwell.

Wager A (1999) ‘The extra qualia problem: Synaesthesia and representationism’ in: *Philosophical Psychology’* 12 (3), 263-281

Wegner D (2002) ‘*The Illusion of Conscious Will’*. MIT: Bradford.

Weiskrantz L (1999) ‘*Consciousness, Lost and Found’*. Oxford: Oxford University Press.

(eds) Weiskrantz and Davies (2008) ‘*Frontiers of Consciousness’*. Oxford: Oxford University Press.

Weiskrantz, L (1986) ‘*Blindsight; a case study and implications’*. Oxford: Oxford University Press.

Weiskrantz, L (2009) ‘*Blindsight’*. Oxford: Oxford University Press.

Wittgenstein, L (1953) ‘*Philosophical Investigations’*. Oxford: Blackwell.

Wittgenstein, L (1958) ‘The Blue and Brown Books; Preliminary studies for the Philosophical Investigations’. Oxford: Blackwell.

Wittgenstein, L (1980) ‘*Remarks on the Philosophy of Psychology; Vol 1’*. Oxford: Blackwell.

Zeki and Bartels (1999) ‘Towards a theory of visual consciousness’ in: *Consciousness and Cognition* 8 (2), 225-259

Zeki and fftyche (1998) ‘The Riddoch Syndrome: insights into the neurobiology of conscious vision’ in: *Brain* 121 Pt1, 25-45

Zeki S (2003) ‘The Disunity of Consciousness’ in: *Trends in Cognitive Science* 7 (5), 214-218