The University Of Sheffield.

The Functional Role of Phenomenal Consciousness

By:

Damiano La Manna

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
Faculty of Arts and Humanities
Department of Philosophy

October 2017

For my parents, Agostino and Virgilia, and my siblings, Davide and Miriam

Abstract

This thesis explores what I call *the functional question*: what is the functional role of phenomenal consciousness (p-consciousness)? Over the last thirty years or so, philosophers (and non-philosophers) have been occupied with trying to provide an account of p-consciousness in scientifically acceptable terms. One question that has not received as much attention, however, is the question of what p-consciousness contributes to the functioning of the mind. This question is important, for we tend to think that p-consciousness must contribute something to our mental economy. How might we attempt to address it? The strategy that I adopt in this thesis is to look at some important accounts of the functional role of p-consciousness, and see whether they offer a promising answer to the functional question. The accounts that I consider in this thesis satisfy an important requirement: they both endorse Ned Block's distinction between p-consciousness and access consciousness (a-consciousness). Even though we still lack a satisfactory account of the former, the distinction has allowed us to make significant progress in explaining the latter. Today, several researchers from different fields agree that the Global Workspace Model provides an empirically-grounded, and theoretically-satisfying picture of a-consciousness. So, which are the accounts of the functional role of p-consciousness that I have in mind? Uriah Kriegel's, and Block's very own. According to both accounts, p-consciousness figures, somehow, in the causal explanation of a-consciousness. Unfortunately, as I will argue, both accounts fail to provide a tenable answer to the functional question. While Kriegel's account is incompatible with certain empirical results, Block's ends up providing the functional role of a property other than p-consciousness. In fact, Block's account may even rest on a misunderstanding of the functional question. I conclude that

Kriegel's and Block's are not the accounts that we should look at for a promising answer to the functional question.

Acknowledgments

Where to start? I thought this would be the easiest part. And yet, judging by how long it is taking me to even think about what to write on this page, it is evident that I was mistaken.

If there is an obvious place to start, then that's probably from my supervisors: Dominic Gregory, my primary supervisor, and Luca Barlassina, my secondary supervisor. Without their knowledge, encouragement, dedication, and unlimited patience, this thesis would not have seen the light of day. For all that, I am grateful beyond words. It has been an honour to work under their guidance.

I am also extremely thankful, of course, to Paul Faulkner, who was my secondary supervisor during the first two years of my PhD.

But I am also thankful to all other members of staff in the Philosophy Department, who have made me feel welcome since my arrival in Sheffield a few years back.

Next, I would like to give my sincerest thanks to the AHRC and the Faculty of Arts and Humanities of the University of Sheffield, without whose financial support, again, none of this would have been possible.

One would suppose I have friends to be thankful to. And, indeed, I do. Some of them I have known for a few years now. Others I have known for slightly less. In no particular order, I would like to thank Miklos Kurthy, Ahmad Fattah, Neri Marsili, Josie Mcquillan, Stephen Bolton, Ryo Yokoe, Anton Eriksson, Francesco Antilici, Armin Kameh, Lewis Brooks, Mher Kazandjian, and Alexandre Duval. Why should I be thanking them? Well, they are my friends, right?

Last, but not least (and I would like to stress the 'not least') I would like to thank my family. Their support has been immense. In every respect. Among many other things, I would like to thank my mum for having taught me to love philosophy. I would like to thank my dad for having always believed in me. I would like to thank

my brother, for having been a brother. And I would like to thank my sister, for having been a sister. In addition, I am grateful to my dearest grandparents, Peppu', Eluccia, Nonna Rosa, and Nonno Damiano. And I am also grateful to my beloved aunt, Cinzia.

Table of Contents

Word Count (exclusive of bibliography): 55103

Introduction

Some mental states are phenomenally conscious (p-conscious). To use Thomas Nagel's (1974) famous phrase, *there is something it is like for one* to be in them. As I sit at my desk, there is something it is like for me to see the words that I am typing on the screen, to smell the coffee that's coming out of my new moka pot, and to hear the sound of the carpenter's shop just outside. A question that has kept many philosophers (and non-philosophers) occupied over the last thirty years or so is how to account for p-consciousness in naturalistic terms; that is, roughly, in scientifically adequate terms (Carruthers 2002, xiii). In this context, one topic that has not received the extensive treatment that it deserves is *the functional role of p-consciousness*. To a first approximation, we may understand the phrase 'the functional role of phenomenal consciousness' as referring to what p-consciousness *does*. To a second approximation, we may understand the same phrase as referring to what p-consciousness *contributes to the functioning of the mind*.

The main topic of this thesis is exactly the functional role of p-consciousness. More precisely, on the quite plausible assumption that p-consciousness has a functional role *at all*, the question that I seek to explore is *what* that functional role is; for the purposes of this introduction, call that the *functional question*. The issue is huge and extremely complex. This probably explains – at least in part – why it has not received the extensive treatment that it deserves. I will further develop and, more importantly, assess two important accounts that purport

to answer the functional question. Which accounts are the ones that I have in mind, we will see in a few paragraphs.

There is one main reason why exploring the functional question is important. For we tend to think that p-consciousness must make a contribution to the functioning of the mind, or, equally, that it must play some role in our mental economy. Indeed, p-consciousness is such a central feature of our mental lives that it would be very surprising – let alone counterintuitive – if it did not have a functional role. As a consequence, on the assumption that the ultimate aim of the study of p-consciousness is to provide a comprehensive, naturalistic account of the phenomenon, such a study can hardly ignore the question of what the functional role of p-consciousness is.

(One note. For the purposes of this work, I assume a rough distinction between accounts of p-consciousness – which sometimes I will also refer to as *accounts of what p-consciousness is* – and accounts of the functional role of p-consciousness – which sometimes I will also refer to as *accounts of what p-consciousness does*.[1] According to what I said in the paragraph immediately above, however, the two are best seen as different – albeit tightly-related – parts of an overall, comprehensive account of p-consciousness).

There are at least two potential strategies that one might adopt to address the functional question. To start with, one might first defend – or assume – a specific account of p-consciousness, and then address the functional question by developing an account of the functional role of p-consciousness against the background of the former. Alternatively, one might look at the different accounts of the functional role of p-consciousness on offer, and then attempt to understand

[1] In talking of accounts of what p-consciousness *is* and accounts of what p-consciousness *does*, I am following Uriah Kriegel (2004b).

which account – if any – provides a satisfactory answer to the functional question.[2] The strategy that I adopt in this work is the latter.

For present purposes, one issue that I need to deal with right from the start is the following. For obvious reasons, considering *all* the accounts of the functional role of p-consciousness on offer is not something that can be done in one thesis. As a consequence, I have no choice but to find one or more criteria on the basis of which I can pick out *some* accounts to consider. There is one specific criterion that I have treated as crucial: the accounts that we should take into consideration will have to respect Ned Block's distinction between p-consciousness and access consciousness (a-consciousness). What is a-consciousness? To a first approximation, a mental state is said by Block to be a-conscious if it can be used in reasoning, report, and the (rational) control of action (1995, 382). For example, my perceptual experience as of a glass of water on the table is said by Block to be a-conscious if I can use that experience in my reasoning about whether I should be drinking from it or not.

Of course, now I need to say something about why the distinction between p-consciousness and a-consciousness is so important. The idea is that although there is still no agreement as to what a promising account of p-consciousness should look like, that distinction has allowed researchers working in the broad discipline of *Consciousness Studies* to make some genuine progress. In particular, it has enabled researchers from fields as different as philosophy, psychology, and cognitive neuroscience, to develop a shared, empirically-grounded, and

[2] Perhaps there is also a third strategy. In the *Oxford Companion to Consciousness* entry on "Functions of Consciousness", Tim Bayne seems to suggest that addressing the functional question might help us develop an account of what p-consciousness is (2009, 315).[2] His suggestion appears to imply that one might be able to handle the functional question without providing, or endorsing, an underlying account of the nature of p-consciousness. Personally, I am not so sure how – or even whether – this strategy could be pursued, however. Indeed, I know of no one who has pursued it.

theoretically-satisfying model *of a-consciousness*: the Global Workspace Model (GWM) (e.g., Dehaene and Naccache 2001; Dennett 2001; Carruthers 2015a).

Of all the accounts of the functional role of p-consciousness available, which are the accounts that presuppose Block's distinction? Two main accounts come to mind: Uriah Kriegel's (2007, 2009, 2015), and Block's (e.g., 1995, 2009) very own. According to both accounts, p-consciousness figures, somehow, in the causal explanation of a-consciousness. That is, p-consciousness is somehow involved in making a mental state usable in reasoning, report, etc. Roughly, on Kriegel's account, p-consciousness is that property *in virtue of* which a mental state is a-conscious. On Block's account, p-consciousness is that property that *contributes* – in a way to be duly explained – to a mental state's becoming a-conscious.

How do Kriegel's and Block's account fare? Alas, as I will argue, neither account provides a promising answer to the functional question. In a nutshell, there are two main issues with Kriegel's account. First it is incompatible with GWM. Second, given certain empirical results, the account entails a view of the relation between a-consciousness and attention that is highly implausible. Block's account, too, is affected by two main problems. First, it misidentifies the functional role of p-consciousness. Second, even though he sells it as an account of the functional role of p-consciousness, it may not be an account of the functional role of p-consciousness at all.

Here is a chapter-by-chapter outline of the thesis:

Chapter 1 is divided in three parts. In the first part, I rehearse a number of familiar distinctions between kinds of consciousness (*simpliciter*); e.g., creature consciousness vs state consciousness, and transitive consciousness vs intransitive consciousness. In the second part, I briefly close in on the very notion of p-consciousness. The biggest chunk of the chapter is the third, where my main aim is to provide a rigorous definition of the phrase 'the functional role of p-

consciousness', a definition that I will work with throughout the remainder of the thesis.

In chapter 2, as in chapter 1, I do three things. First, I zoom in on Block's distinction between p-consciousness and a-consciousness, and clarify my commitments to the distinction. Second, I offer a first pass at GWM and its relation with a-consciousness. GWM is the model of a-consciousness that Block endorses and that, because of the impressive amount of evidence in its support, I, too, assume throughout the thesis. Last, I look at some empirical studies on the relationship between a-consciousness and attention, and argue, together with the advocates of GWM, that attention is necessary for a-consciousness.

Chapter 3 considers, and assesses, Kriegel's account of the functional role of p-consciousness. In the first part of the chapter, I outline Kriegel's self-representationalist theory of p-consciousness. This is a necessary step to appreciate his answer to the functional question, which he offers against the background of that very theory. In the second part of the chapter, I introduce Kriegel's account of the functional role of p-consciousness, and argue that it is untenable. As anticipated, Kriegel argues that a mental state's being p-conscious is that property *in virtue of* which the state is also a-conscious. In turn, this allows him to construe a-consciousness as the functional role of p-consciousness. As I also anticipated, I believe that Kriegel's account is untenable for two reasons. First, it is incompatible with GWM. Being GWM a well-established model of a-consciousness, this gives us reason to reject the account. Second, the account entails a view of the relationship between a-consciousness and attention that, given the results of some of the studies discussed in chapter 2, is empirically implausible.

In chapter 4, I come back to Block's distinction. In the first part of the chapter, I discuss his most recent view on how, exactly, p-consciousness relates to a-consciousness (e.g., 2007a; 2008). This allows me to introduce a number of notions that I will then put to work in chapter 5, where I will deal with Block's

account of the functional role of p-consciousness. In the second part of the chapter, I defend Block's most recent view on the relation between p-consciousness and a-consciousness from a criticism advanced by Peter Carruthers (2015b). Carruthers's criticism threatens to show that Block's view is affected by serious issues that would clearly make it untenable independently of its potential as an account of the functional role of p-consciousness. By removing these potentially devastating obstacles to Block's view, my response to Carruthers will allow me to focus precisely upon its relevance for the functional question that is this thesis's main concern.

In chapter 5, which is also the last, I consider, and assess, Block's account of the functional role of p-consciousness. As we will see, Block has addressed the functional question on several occasions, but his position suffers from some important unclarities. A large part of my job in this chapter, then, consists in refining Block's ideas in important ways, and using many of the notions introduced in chapter 4 to develop his proposals to a significant degree. As noted, Block's view is that p-consciousness *contributes* to a mental state's becoming a-conscious. With my development of Block's account in place, however, I then argue for two claims. The first is that Block provides an account of the functional role of a property *other than* p-consciousness. The second is that that account may even rest on a fundamental misunderstanding of the functional question.

1

Setting the Stage

1. Introduction

In this chapter, I do three main things. First, I rehearse a number of familiar distinctions between kinds of consciousness. In particular, I distinguish between creature and state consciousness, transitive and intransitive consciousness, look at some of the relations between them, and characterize p-consciousness as a kind of state consciousness. All that, I do in §2. Second, I briefly close in on the notion of p-consciousness. As usual, I characterize p-consciousness in terms of the notion of *what it is like for one*, and provide some examples of p-conscious states. This, I do in §3. The other major aim of this chapter is to provide a clear definition of the notion of *functional role of p-consciousness*. I argue that, in a way to be duly explained, that notion may be defined as follows: the fund of appropriate dispositional properties that p-consciousness endows a mental state with. The task of defining the notion of functional role of p-consciousness, I undertake in §4, which is the largest section of the chapter.

2. Kinds of Consciousness

The phrase 'is conscious' has various meanings, and is also applied to different sorts of things (Kriegel 2009; Van Gulick 2009). The complex usage of 'is conscious' is apparent in everyday language, but is arguably even more apparent in philosophy

of mind/psychology. When the doctor says of the patient that he is still conscious, she plausibly means that the patient is still awake. Moreover, the doctor attributes the property of being conscious to a whole creature, namely, the patient. When philosophers say that some mental states are p-conscious, they mean that there is something it is like for a creature to be in them. Moreover, they attribute the property of being conscious to the creature itself, not the creature's mental states. This leads us to the distinction between two concepts of consciousness (and the properties they putatively pick out): *creature consciousness* vs *state consciousness*. In this section, I elaborate upon this and a further distinction: that between *transitive consciousness* vs *intransitive consciousness*. Both distinctions are due to David Rosenthal (e.g., 1986, 1997), but have been further developed since. Here's how I will proceed. First, I introduce the concept of creature consciousness, and provide some examples of the different ways in which a creature may be said to be conscious. Second, I introduce the distinction between transitive and intransitive consciousness as it applies to creature consciousness. I then introduce the concept of state consciousness, provide some examples of the different ways in which a mental state may be said to be conscious, and explain how the distinction between transitive and intransitive applies to state consciousness.

Let us begin, then, with the concept of *creature consciousness*. Creature consciousness is consciousness as pertains to creatures; e.g., humans and other animals. In saying that Suzy is conscious, I am saying of a human, Suzy, that she is conscious. Analogously, in saying that my cousin's dog, Heiwin, is conscious, I am saying of a dog, Heiwin, that she is conscious. There is more than one sense in which a creature may be said to be conscious (Van Gulick 2009). Below, I distinguish three.

In everyday life, when one attributes consciousness to a whole creature, what one has sometimes in mind is *wakefulness*. That is, the phrase 'is conscious' is construed as a rough synonym of 'is awake'. Plausibly, as noted, this is how the

doctor uses the phrase 'is conscious' when she says of the patient that he was still conscious after the accident, or that he has just regained consciousness. On this understanding of creature consciousness, a creature can be in different states of alertness, which vary in how responsive the creature is to its environment (Van Gulick 2009, 163). It is only when it is in a high state of alertness – that is, *awake*, as opposed to being asleep or in a coma, for example – that the creature counts as conscious. In the literature, this specific notion of creature consciousness was isolated by Rosenthal (1997).[3]

A second sense in which a creature is sometimes said to be conscious is if it is "*perceiving* such-and-such" (Carruthers 2002, 10). On this construal, the phrase 'is conscious' is treated as a rough synonym of 'is perceiving'. This is how we understand creature consciousness when we say that Suzy is conscious of the book on her desk, or that my cousin's dog, Heiwin, is conscious of her bone on the mat.

A third, possibly less common, sense in which a creature may be said to be conscious is if it is *self-conscious*. Carruthers (2002) distinguishes between a weaker and a stronger variety of self-consciousness. For a creature to be self-conscious in the weak sense is for it to (be able to) be self-aware of itself "as an *object* distinct from others" (Ibid., 12). According to Carruthers, this variety of self-consciousness is not too demanding in terms of the creature's conceptual capacities. That is, roughly, it does not require that the creature possess and be able to deploy a large/complex conceptual repertoire. Accordingly, many animals can probably be said to be conscious in this sense. On the other hand, for a creature to be self-conscious in the strong sense is for it to (be able to) be self-aware of itself "*as a self*, as a being with mental states and a subjective inner life" (Ibid.). Unlike the weaker

[3] It is debatable whether, on the present understanding of creature consciousness, wakefulness really is a *necessary* condition for consciousness. While this is the position espoused by Rosenthal (1997), Carruthers (2002, 9) aptly points out that whether wakefulness really is necessary for consciousness depends on whether we wish to say that creatures are conscious when dreaming –this, however, is far from being uncontroversial.

variety, this variety of self-consciousness is conceptually very demanding, and is probably possessed by humans and (possibly) great apes only (Ibid.).[4]

Creature consciousness can be *transitive* or *intransitive*. Consider the following two reports: (i) 'Suzy is conscious of the noise'; (ii) 'Suzy is conscious'. (i) says of a creature, Suzy, that she is conscious *of* something, namely, the noise. This is the transitive sense. (ii), instead, says of a creature, Suzy, that she is conscious *simpliciter*. This is the intransitive sense. Transitive creature consciousness is always consciousness "*directed at* an object" (Van Gulick 2009, 737). Intransitive creature consciousness, instead, is consciousness *simpliciter*. While (i) attributes to Suzy a relational property (Kriegel 2009, 25) – the property of standing in a consciousness-of relation to something[5] – (ii) attributes to Suzy a non-relational property – the property of being conscious *simpliciter* (Ibid.).

How does the transitive/intransitive distinction apply to the varieties of creature consciousness considered above? The first is of the intransitive kind, while the second and the third are of the transitive kind. To say of Suzy that she is conscious in the sense of being awake is to attribute to Suzy a non-relational property. To say of Suzy that she is conscious in the sense that she is perceiving such-and-such, instead, is to attribute to Suzy a relational property, namely, the property of being conscious of such-and-such. The same considerations that apply to creature consciousness as perception of such-and-such apply to creature

[4] The reader should bear in mind that not everyone understands self-consciousness exactly in those terms.

[5] The term 'relational' here might suggest that the item of which Suzy is conscious has to exist. This is not what I – and Kriegel – have in mind when using that term, however. For the transitive sense of 'conscious' is meant to apply also in cases in which, for example, Suzy hallucinates something. When I say that the report 'Suzy is conscious of the noise' attributes to Suzy a relational property, then, I mean the property of standing in a consciousness-of relation to an *intensional* object. Intensional objects may or may not exist. In this sense, the object picked out by 'O' in reports like 'Suzy is conscious of O' is like the object picked out by 'O' in reports like 'Suzy believes that O is beautiful': Both objects need not exist. Analogous considerations apply below, when I discuss transitive *state* consciousness.

consciousness as self-consciousness. To say of Suzy that she is conscious in the sense of being self-conscious is to attribute to Suzy a relational property, namely, the property of being aware of herself "as a being with mental states and a subjective inner life" (Carruthers 2002, 12).

Now for *state consciousness*. While creature consciousness is consciousness as pertains to whole creatures, state consciousness is consciousness as pertains to a creature's mental states. In saying that Suzy's experience is conscious, I am attributing the property of being conscious not to a whole creature, but to one of its mental states. The most discussed – as well as the most mysterious – variety of state consciousness is *p-consciousness*. As anticipated, a mental state is said to be *p-conscious* if there is something it is like for one to be in it. A second variety of state consciousness is *access consciousness* (a-consciousness). As noted, a mental state is said to be *a-conscious* if its content is available for report, deliberation, reasoning, etc.

Like creature consciousness, state consciousness can be *transitive* or *intransitive*. When, as above, I say that Suzy's experience is conscious, I say of a mental state of Suzy's that it is conscious *simpliciter*. In so doing, I take consciousness to be a non-relational property of Suzy's mental state. On the other hand, when I say that 'Suzy's experience is conscious of the cup', I say of a mental state of Suzy's that it is conscious *of* something. In so doing, I take consciousness to be a relational property of Suzy's mental state. As should be apparent, ordinary language struggles to express transitive state consciousness. Indeed, reports of the form 'S's mental state is conscious of O' are nonsensical (Kriegel 2009, 27). But that does not mean that transitive state consciousness is conceptually problematic. One way to express the concept of transitive state consciousness in ordinary language may be by characterizing transitive state consciousness as "the property mental states exhibit when, and only when, their subjects are transitively conscious of something in virtue of being in them" (Ibid.). The idea is that a mental state's being

transitively conscious is the property that accounts for a creature's being transitively conscious. It is in virtue of the fact that Suzy's experience of the cup is transitively (state) conscious that Suzy is transitively (creature) conscious of the cup.

How does the transitive/intransitive distinction apply to p-consciousness and a-consciousness? As we will see in more detail in the next chapter, a-conscious states are essentially representational (Block 1995). This entails, among other things, that a-conscious states are essentially transitive states. To say of a mental state that it is conscious in the sense of being a-conscious is thus to attribute to the state a relational property. But what about p-consciousness? Things are a little bit more complex here. Everyone seems to agree that p-consciousness is *essentially* intransitive. As I will explain in the next chapter, however, Ned Block (Ibid.) thinks that p-consciousness is often *also* transitive. Differently put, although it is not of the essence of p-consciousness to be directed at an object, he thinks that many p-conscious states are also so directed.

Before moving on to the next section, there is one worry that I would like to address: Why does p-consciousness only figure on my list of the varieties of *state* consciousness? I think few philosophers, if any, would deny that a creature can be said to be p-conscious. More strongly, few philosophers, if any, would deny that at least some creatures *are* p-conscious. In contemporary philosophy of mind, however, phenomenal creature consciousness (PCC) is seldom discussed. Better, insofar as philosophical accounts of p-consciousness are concerned, philosophers have consistently been focusing on phenomenal *state* consciousness (PSC). But why?

It is noteworthy that most of the time it is simply *assumed* that to enquire about p-consciousness is to enquire about PSC, or that at least it is more convenient to do so by enquiring about PSC. But even from those who do offer an explanation for focusing on PSC, it is hard to find answer to our question. According to Jesse

Prinz (2013), for example, a creature's mental state's being p-conscious at a time T is *sufficient* for that creature's being p-conscious at T. This supposedly justifies the focus on the latter. According to Kriegel (2004a, 2009), on the other hand, a creature's mental state's being p-conscious at a time T is *necessary* for that creature's being p-conscious at T. Thus, in his view, PCC can be said to *depend on* PSC. This, Kriegel takes to be the reason why philosophers have normally focused on PSC.[6]

As far as I am concerned, I do not need to take a stance on this issue. Thus, given the typical focus on PSC, I will put talk of PCC aside, if only to avoid confusing the reader. Nothing substantial hinges upon this choice.

3. Closing in on the Notion of Phenomenal Consciousness

In this section, I briefly close in on notion of p-consciousness. The *locus classicus* for characterizing (or attempting to characterize) p-consciousness is Thomas Nagel's (1974) *What Is It Like to Be a Bat?* There, Nagel characterizes p-consciousness (or conscious experience, as he calls it) in terms of *what it is like for an organism* (or *for one*). There is something it is like for one to hear the sound of a trumpet, to see the Colosseum, and, more generally, to be in certain mental states. Following standard practice, in the remainder of this section, I provide some examples of p-conscious states.

One uncontroversial place to start (if not *the obvious* place to start) is p-conscious perceptual states. As I sit at my desk, there is something it is like for me to see my laptop screen slowly being filled with words, to see a glass full of water to my left, and my iPad to my right. But there is also something it is like for me to chew a square of my 90% cocoa chocolate bar, to hear the noise coming from the

[6] In fact, some philosophers have recently started to put the focus on PSC into serious question (e.g., Piccinini 2007).

carpenter's shop just outside, and to smell the coffee that my dad is brewing in the kitchen.

Another paradigmatic example of p-conscious experience is pain. There is, as we all know too well, something it is like for one to be in pain. Pain experiences might range from light pricks to severe burns, or from stabbing pains in the bottom of one's feet, to intense pains in one's chest (D. J. Chalmers 1996, 9).

Other uncontroversial examples of p-conscious states are quasi-perceptual states, like visual images – as when you visually imagine a white rabbit hopping in the grass – and auditorily imaginings (not sure that's the right term?) – as when you auditorily imagine Bach's *Sonata no 5*. The list could go on, of course. But what I have said should suffice to give the reader an idea of the variety of states p-consciousness attaches to.

Before turning to the next section, I should note that, given the philosophical (as well as scientific) literature's almost exclusive focus on p-consciousness as pertains to visual perceptual states, that is the kind of states that I will focus on in this work.

4. Getting Clearer on the Notion of Functional Role of Phenomenal Consciousness

What exactly is the meaning of the phrase 'the functional role of p-consciousness'? This is the question that I aim to address in the remainder of this chapter. When philosophers ask about the functional role of p-consciousness, they are normally interested in one of two things: What is the causal contribution that p-consciousness makes to the functioning of the mind? Or, what was p-consciousness "engineered" by evolution to do – that is, what was it selected for? When one asks the former question, one understands the notion of functional role in terms of the notion of *causal role* (Polger 2004, chap. 5; Piccinini 2010). When

one asks the latter question, one understands the notion of functional role in terms of the notion of *teleological* function (Ibid.).

On the former understanding of the notion of functional role, for a thing X to have a functional role is for X to have a causal role. Tentatively, for X to have a causal role, in turn, is for X to be apt to enter into certain causal relations that explain (or contribute to explain) the functioning of the system X belongs to. For a certain gear G to have a causal role is for it to be apt to enter into certain causal relations that explain, for example, how the watch G belongs to works (Polger 2004, 150). On the teleological function understanding of functional role, instead, for a thing X to have a functional role is for X to have a teleological function. Assuming that X is a biological trait of an organism, for X to have a teleological function, in turn, is for X to have one or more effects "that [explain] its presence […] due to evolution by natural selection" (Ibid., 166).[7]

In this work, I am interested in the functional role of p-consciousness in the *causal role* sense of 'functional role'. That is, I am interested in the question of *what p-consciousness contributes to the functioning of the mind.* For present purposes, I assume that 'functioning' stands for *normal functioning*. To ask what p-consciousness contributes to the functioning of the mind is thus to ask *what it contributes to its normal functioning*. A watch may be said to function normally when it tells the (right) time. A carburettor may be said to function normally when it mixes air and fuel (in the right way). What does it mean for the mind to function normally? Things are a bit trickier here. For there seems to be no non-vague and non-circular way to address the latter question. The best one can do, it seems, is say things like: The mind functions normally when it operates as it usually does; or the mind functions normally when it does not malfunction. Although answers such

[7] For an account of the relevant notion of function, which sometimes is also known as 'etiological' or 'proper' function, see, for example, Neander (1991). For an account of the notion of teleological function as applies to both human artefacts and biological organisms, instead, see Wright (1973).

as these are, as noted, are both vague and circular, there is also a sense in which they do *intuitive* justice to the notion of normal functioning (or so I hope). For my purposes, that is all that I should need.

Being p-consciousness a property of mental states, I suggest we begin by closing in on the notion of functional role as it applies to the latter, and then move to the same notion as it applies to p-consciousness.

4.1. The Functional Role of Mental States

In the contemporary philosophy of mind, when philosophers talk about the functional role of mental states, they normally understand 'functional role' precisely in terms of the notion of causal role (Polger 2004, 150). Before introducing the notion of functional role as it applies to mental states, it will thus be convenient sketching a picture of the notion of causal role. [8] I should note that, unless otherwise specified, I will be concerned with *mental state types* (or with *types of things*). The considerations that I will make in relation to types, however, will equally apply to *tokens* of those types.[9]

There are two main things to note about the notion of causal role. First, the causal role of a thing X is the *fund of dispositional properties* of X (Cummins 1975). More specifically, the causal role of X is X's *fund of dispositions to enter into certain causal relations* (*fund of dispositions/dispositional properties* for short). Second, the causal role of X is normally a *proper subset of X's total fund of dispositional*

[8] The best-known account of the notion of causal role is Cummins's (1975). According to Cummins's account, the causal role of an X is always relative to an analytic account A of a system's capacity to C (Ibid., 763). For present purposes, however, I am going to adopt a simplified version of Cummins's account.

[9] What is a *token*? Tokens are normally understood precisely in opposition with *types*. Consider the following string of words: 'Bad and bad and bad and bad'. In one sense of the term 'word', the above string contains two words – 'bad' and 'and'. In another sense of term, it contains seven words: Four 'bad' and three 'and'. Differently put, the above string contains two word types (the word type 'bad' and the word type 'and') and seven word tokens (four tokens of the word type 'bad' and three tokens of the word type 'and') (Bickle 2016).

properties.[10] This is because X's causal role is always *relative to a theory T of the functioning of a system S* (Godfrey-Smith 1996, chap. 1; Polger 2004, chap. 5). Let me expand on these two points in order.

In general, a dispositional property can be defined as the tendency of a thing X to behave in characteristic ways, under certain circumstances. As an example, consider the solubility of salt. Roughly, this is the tendency of salt to dissolve when poured into water (or some other suitable solvent). An important aspect of dispositions is that for X to have a certain disposition, X need not (ever) *actually manifest* that disposition, but only be apt to behave in certain ways under certain circumstances. What I mean is that for a pinch of salt to be disposed to dissolve in water, for example, it need not (ever) actually be poured into water and dissolve. The only thing that matters, instead, is that if that pinch of salt were poured into water, it *would* dissolve. As concerns causal role, for a thing X to have a certain causal role is for X *to possess a fund of dispositional properties*. In other words, for X to have a certain causal role, is for X to be capable of entering into a number N of causal relations, under certain circumstances.

As anticipated, the second thing to note about causal role is that the causal role of a thing X is normally a subset of all the causal relations that X is disposed to enter into. The reason is that the causal role of X is always *relative to a theory T of the functioning of a system S to which X belongs*. We can think of T as a theory that tells us/gives us two things. First, it gives us an inventory of the "components" or "parts" X's of the system S. Depending on the nature of T (e.g., on its being a folk or broadly-scientific theory),[11] 'component' or 'part' will refer to different kinds of things; e.g., subsystems, properties/states, and processes of S. Second, a theory T tells us which, of all the dispositions that X has, are those dispositions that explain,

[10] I am assuming here that dispositions are always causal dispositions. That is, that to be disposed to X is always to be disposed to do something, where by 'do', I mean enter into certain causal relations.

[11] More on this later.

or contribute to explain, the functioning of S (or how we should go about finding out what those dispositions are).[12] That is, aside from telling us what X's belong to S, T also tells us what the causal roles of those X's are (or how we should go about finding out what they are).

Consider a watch. According to a theory of the watch TW, a watch has many different components; e.g., a number of gears G_n, a pallet, a hairspring, and two hands. Now, consider a gear G_1. G_1 has several dispositions. That is to say that G_1 is apt to enter into a number of different causal relations. Suppose that G_1 is: (1) disposed to cause G_2 to rotate clockwise; (2) disposed to cause G_3 to rotate anticlockwise; (3) disposed to cause G_4 to rotate at a certain speed S_1; and (4) it is disposed to move the air particles around it, making a soothing noise. Of all the dispositions that G_1 has, only some will be included in G_1's causal role. The reason is that the causal role of G_1 is relative to our theory TW. According to the latter, of the all the dispositions that G_1 has, it is only (1) and (3) that figure in an explanation of how the watch functions. What this means is that, among (1), (2), (3), and (4), only (1) and (3) are included in the causal role of G_1.

In sum, first, the causal role of a thing X is a fund of dispositional properties of X. That is, for X to have a causal role is for it to have a fund of dispositions. Second, not all the dispositions that X has are included in its causal role. The reason is that the causal role of X is always *relative to a theory T of the functioning of a system S to which X belongs.* With this general picture of the notion of causal role in hand, let us now turn to the functional role of mental states.

A standard definition of the notion of functional role of a mental state M – where 'functional role' is understood in terms of the notion of causal role – is the following:

[12] I shall say more about the phrase between parentheses later.

Functional Role of M: M's fund of appropriate dispositional properties.

Let me unpack this definition. First, in line with I said earlier about the notion of causal role, M's functional role is taken to be a *fund of dispositional properties* of it (Kriegel 2004b; Block 2007b; Kim 2010). As explained, for a thing X to have a certain disposition, X need not (ever) actually manifest that disposition, but only be apt to behave in certain ways, under certain circumstances. The same applies to the functional role of a mental state M. M can (and does) have a certain functional role regardless of whether it actually (ever) enters into any of the causal relations it is disposed to enter into. Thus, for M to have a certain functional role, it is only required that M *be disposed to* enter into those causal relations (Kriegel 2009, 213).

The second point which we should consider concerns the meaning of 'appropriate'. In this context, a dispositional property of M's counts as *appropriate relative to a theory T of the functioning of a system S to which M belongs*. As concerns mental states, T will be a *psychological theory* PT. Analogously to what I said above, here we can think of PT as a theory that tells us what the "components" or "parts" *X's* of the mind are (I will call these X's *mentalia*), and which, of all the dispositions that those mentalia have, are those dispositions that explain, or contribute to explain, the functioning of the mind (or how we should go about finding out what those dispositions are). Below, I introduce two PT's. I should note, however, that I will only provide a very broad (and simplistic) outline of such PT's. That, after all, is all that I need for my purposes.

One PT is what is usually known as *Folk Psychology* or *Common-Sense Psychology* (e.g., Lewis 1972). Folk Psychology is an *analytic* PT. It is a PT constructed by analysing the concepts we ordinarily use when thinking about the mind (Polger 2004, 154). The theory can be seen as a set of generalizations of the form "When someone is in so-and-so combination of mental states and receives sensory stimuli of so-and-so kind, he tends with so-and-so probability to be caused thereby to go into so-and-so mental states and produce so-and-so motor

responses" (Ibid.). An instance of such generalizations might be "someone who is pricked with a pin will likely feel pain and flinch" (Ibid.).

Folk-Psychology tells us, first, what mentalia there are. Such a list would certainly include mental states/properties such as pains, beliefs, and desires, as well as capacities such as perception and memory. What Folk-Psychology also tells us is the causal roles of those states, properties, etc. For example, as concerns pain, Folk-Psychology tells us that the causal role of pain is its fund of dispositions to i) be typically caused by tissue damage; 2) cause the belief that something is not right with one's body; 3) cause winces and groans; 4) cause the desire to abandon the state, etc. (Levin 2016, n/a).

Now, in telling what the causal roles of mentalia are, Folk-Psychology, in effect, tells us which, of all the dispositional properties that – for example – a mental state M has, are the dispositional properties that are relevant to the functioning of the mind. Plausibly, the total set of pain's dispositional properties is much larger than the set of dispositions that are included in its causal role. For example, pain might always be disposed to cause the desire to *remain* in the state, rather than be out of it. According to Folk-Psychology, however, the disposition to cause the desire to remain in pain does *not* enter into the causal role of pain.

The second psychological theory PT that I wish to consider is what we may call *Cognitive Science*. Analogously to Folk Psychology, Cognitive Science can be seen as a set of laws and generalizations about the mind. Unlike the former, however, the generalizations of Cognitive Science are not derived (entirely) a-priori – derived from the analysis of the concepts we normally use when thinking about the mind – but (partly) derived from an empirically informed study of the mind. Cognitive Science, too, tells us which mentalia there are. In general, it seems that the inventory of mentalia provided by Cognitive Science partly overlaps with the one provided by Folk Psychology (Prinz 2013, 281). Thus, for example, a list of mental states as provided by Cognitive Science will certainly include states such as

pains, perceptions, beliefs, and desires. The inventory of mentalia provided by Cognitive Science, however, is larger than the provided by Folk-Psychology. This is because, being an empirical PT, Cognitive Science's inventory of mentalia will include, for example, several "empirically discovered distinctions that are not recognized in Folk Psychology (episodic vs semantic memory [...] iconic memory [...] and so on)" (Ibid.). Being an empirically informed PT, Cognitive Science is far from being a complete theory of the mind. On the contrary, it is to be seen as an *approach* to the study of the mind.

This last point becomes quite evident the moment we turn to what Cognitive Science tells us about the causal roles of mentalia. Unlike Folk Psychology, Cognitive Science holds that those causal roles are to be discovered – at least in part – empirically (Block 2007b; Levin 2016). While Folk Psychology derives the causal roles of those states, processes, and so on, by analysing the concepts of the folk, Cognitive Science holds that those causal roles should be discovered through an empirically informed study of the mind.

Cognitive Science so characterized is the psychological theory PT that I assume for the purposes of this work. To enquire about what the functional role of p-consciousness is, is thus to enquire about it against the background of such a PT (more on this below). Having introduced the notion of functional role as it applies to mental states, I now turn to the notion as it applies to p-consciousness.

4.2. The Functional Role of Phenomenal Consciousness

The things that a mental state can do depend on the properties that the state has. To put it with Kriegel, "the state has various properties, F_i, ..., F_n, and each property F_i contributes *something* to [...] the state's fund of causal powers" (2004b, 172). In other words, the state has various properties, and each property *endows the state with certain dispositions*. Consider, again, a watch's gear G. The things that G can do depend on its properties. They depend, for example, on G's shape, dimensions,

where exactly in the watch G is positioned, and what G's rotating speed is. Each of G's properties can thus be said to *dispose* G to enter into certain causal relations.

Some mental states have the property of being p-conscious: There is something it is like for one to be in them. If metaphysical epiphenomenalism about p-consciousness is false, then p-consciousness will contribute something to the fund of dispositions of the mental states that have it (Ibid.). As applies to p-consciousness, metaphysical epiphenomenalism claims that the latter is *causally inert*. That is, it denies that it contributes *anything* to the fund of dispositional properties of the mental states that have it. The view is usually associated with property dualism, the doctrine that p-consciousness is a non-physical property of physical states (Polger and Flanagan 2002, 35).

On the basis of the considerations made thus far, we may define the functional role of p-consciousness, relative to a mental state M, as follows:

> *Functional Role of P-Consciousness Relative to a Mental State M: The fund of appropriate dispositional properties that p-consciousness endows M with.*[13]

Differently put, the functional role of p-consciousness is the fund of appropriate dispositional properties that M has *in virtue of being p-conscious*. To my understanding this definition should capture what many philosophers normally have in mind when discussing the functional role of p-consciousness. On top of that, the definition should be in line with what we have in mind when we ask what the functional role of some non-mental property in a non-mental system is. Consider the following analogy. The things that a watch's gear G can do, as we have seen, depend on G's properties. Suppose that G has a certain shape. Call this property *X-shape*. Suppose also that, on the assumption that X-shape has a functional role at all, we wish to address the question of what that functional role

[13] My definition is akin to the one offered by Kriegel (2009, 252).

is. I take it that to address that question is just to identify the appropriate dispositional properties that G has *in virtue of being X-shaped*.

One more thing to note about the definition that I have offered is that it construes p-consciousness as what, in effects, *plays* (or *realizes*, or *occupies*) (at least part) of the functional role of M. For it is precisely *in virtue of being p-conscious*, that M has certain appropriate dispositional properties. To return to our gear G again. Suppose that, in virtue of being X-shaped, G has the appropriate disposition to enter into certain causal relations with G_n. We can then say that the property of being C-shaped *plays* or (or *realizes*, or *occupies*) part of the functional role of G.[14]

Having provided a definition of the notion of the functional role of p-consciousness, let me now briefly put forward a working hypothesis.

4.3. A Working Hypothesis

For the purposes of this work, I am going to assume, as a working hypothesis, that p-consciousness does have a functional role *at all*. Ideally, that will make it easier for us to focus on the question of *what* that functional role is, which is the question that I am interested in exploring. Importantly, such a hypothesis should not be seen as a blind assumption. Instead, as noted in the introduction, we tend to think

[14] It may also be interesting to note that the proposed definition is *neutral* as to whether the functional role of p-consciousness is, or is not, a *proper part* of the functional role of M. The idea is this. The functional role of M, as I explained in §4.1., includes M's fund of appropriate dispositional properties. Now, on top of being p-conscious, M may, or may not, have other properties. *If* M has other properties, then, possibly, some of the appropriate dispositions that M has, M has in virtue of having those other properties. From this, it follows that the fund of appropriate dispositions that M has in virtue of being p-conscious is smaller than M's total fund of appropriate dispositions. In other words, the functional role of p-consciousness is only a proper part of the overall functional role of M. On the other hand, if M has no properties other than p-consciousness, then the functional role of M is identical to the functional role of p-consciousness. Nothing in the definition that I have proposed, however, implies that p-consciousness is or is not a proper part of the functional role of M. Moreover, the issue of whether p-consciousness is or is not a proper part of the functional role of M will play no role in this work.

that p-consciousness must have a functional role. Indeed, p-consciousness is such a central aspect of our mental lives that it would be very surprising – let alone counterintuitive – if it did not have a functional role.

It is worth pointing out that to work under the assumption the p-consciousness has a functional role at all is to assume the falsity of two forms of epiphenomenalism about p-consciousness. One form, we have already encountered: *metaphysical* epiphenomenalism. According to metaphysical epiphenomenalism, as we have seen, p-consciousness is causally inert. To work under the assumption that p-consciousness has a functional role at all, then, is, in part, to assume that p-consciousness *in effect* endows the mental states that have it with certain dispositions.

But to assume that metaphysical epiphenomenalism is false does not by itself guarantee that p-consciousness has a functional role. For that to be the case, it is also necessary that we assume that another form of epiphenomenalism about p-consciousness is false, too. Why? For granted the falsity of metaphysical epiphenomenalism, p-consciousness might only endow the mental states that have it with dispositions that are *not contemplated* by (a *complete* version of) the psychological theory that we have endorsed. So, even though p-consciousness would still contribute something to what those states can do, it would not endow them with any *appropriate* dispositions. If that were the case, although metaphysical epiphenomenalism about p-consciousness would be false, *causal role* epiphenomenalism about p-consciousness would be true.

What is causal role epiphenomenalism about p-consciousness? As characterized by Polger and Flanagan (2002, 34–35), causal role epiphenomenalism about p-consciousness is exactly the thesis that p-consciousness contributes nothing to the functioning of the mind. Consider an analogy. Suppose you have a car whose engine is particularly noisy. Being particularly noisy makes your car engine easily audible from 800 meters away. Still, being extremely noisy probably

contributes nothing to the functioning of the car. In a similar way, causal role epiphenomenalism allows that p-consciousness disposes its mental states to enter into certain causal relations, but denies that it endows them with the appropriate dispositions.

To sum up, in order to focus on the question of what the functional role of p-consciousness is, I am going to work under the (plausible) assumption that p-consciousness has a functional role at all. To assume that p-consciousness has a functional role at all, in turn, is to assume that both metaphysical and causal role epiphenomenalism are false. Before moving onto the next chapter, there is one last thing that I would like to do. That is to clarify that the claim p-consciousness has a functional role is to be kept clearly distinct from the claim that *functionalism* about p-consciousness is true (Kriegel 2004b, 2009). Let me explain.

4.3. Functionalism about Phenomenal Consciousness

What is functionalism about p-consciousness? Addressing this question will be easier if we explain what functionalism about the mind is first. By 'functionalism', in this context, I mean a metaphysical view about the mind, which goes back to at least Hilary Putnam (1975) and Jerry Fodor (1974). Qua metaphysical view, functionalism concerns what kinds of things mental states *are*, what their *ultimate nature* is (Block 2007b, chaps. 1, 2; Polger 2012).

Functionalism arises as an alternative to the mind-brain identity theory (e.g., Smart 1959). While the latter identifies mental state types with brain state types (e.g., pain with C-fibre activation), functionalism holds that a system could have mental states without having brains states like ours, or any brain states at all (Polger 2004). A system may be in pain regardless of whether it is in a brain state identical to C-fiber activation, whether it is in a hydraulic state of a hydraulic computer, etc. According to functionalism, then, mental state types are *not* identical to brain state types. Instead, mental states are said to be *realized* or

implemented by other (typically, although not necessarily, physical) states, be these physico-chemical states of brains, hydraulic states of hydraulic computers, etc. Thus, on the functionalist view, mental states are *multiply realizable* states. Different physico-chemical states, as well as hydraulic and electronic states, for example, can all realize or implement the same mental state type.

But what does functionalism take mental states to be, exactly? Functionalism dictates that mental states are (nothing but) *functional states* of systems (Putnam 1975). To say that a mental state is a functional state is to say that there is nothing more to the state than its having a certain functional role (Block 2007; Polger 2004). The notion of functional role, as we have seen, can be understood in different ways. Commonly, it is understood in terms of the notion of causal role (Kim 1996; Polger 2004). On this understanding, to say that there is nothing more to a mental state than its having a certain functional role is to say that there is nothing more to that state than its having a certain causal role.

Importantly, by construing mental states as functional states, functionalism takes the former to be *second-order* states of systems. In this context, a second order state is the state of "*having some first-order state or other which plays the specified causal role"* (Pryor 2002, n.a.). Consider pain, again. Pain is the state of having some first-order state or other that plays or carries out the pain role. The first-order state is precisely what realizes, implements, or plays the pain role. As noted, given multiple realization, this may be a certain brain state, electronic state, hydraulic state, etc. As for the pain role, instead, this is the functional role of pain as specified by psychological theory PT (Block 2007a).[15]

[15] Although the variety of functionalism that purports to reduce mental states to their functional role is the most prominent, there is also another, less popular, variety of functionalism, due to David Lewis (1980). While the former variety identifies mental states with second-order states of systems, Lewis functionalism holds that mental states are to be identified with *first-order* states. That is, according to Lewis functionalism, a mental state is a state of a system that *realizes, fills,* or *occupies* a certain functional role. Considering the example of pain again, Lewis functionalism identifies pain with whatever

I have sketched a picture of functionalism as applies to mental states. But the reason I have talked about functionalism at all is to make one specific point. The point is that to hold that p-consciousness has a functional role is *not* to commit to functionalism about p-consciousness. So, what does functionalism about p-consciousness say? According to the thesis, what makes a mental state p-conscious is (plausibly *part of*) its functional role. In other words, there is *nothing more* to a mental state's being p-conscious than its having a fund of appropriate dispositions.

With a picture of functionalism about p-consciousness in hand, it should be apparent why the claim that that p-consciousness has a functional role is not to subscribe to functionalism about p-consciousness. For it is one thing for p-consciousness to endow the mental states that have it with certain dispositions, and it is another for it to be *identical to* (plausibly part of) the functional role of those states.

5. Concluding Remarks

I begun this chapter by rehearsing a number of familiar distinctions about kinds of consciousness, and clarified that, for my purposes, I will treat p-consciousness as a kind of *state* consciousness. After that, I provided some examples of p-conscious states, and explained that, as is customary, I will understand p-consciousness in terms of Thomas Nagel's *what-it-is-like-for-one* notion. In the third – and largest – part of the chapter, my main task was to offer a rigorous definition of the notion of

---

state – be this a brain state or a machine state – realizes or fills the pain role. One consequence of this is that what pain *itself is* will be population-specific: It will be a certain brain state in humans, a different brain state in bats, a hydraulic state in a hydraulic automaton, and a non-physical state in a ghost. But in what sense can all those states be said to qualify as pain? Lewis addresses this question by suggesting that the *concept* of pain is a non-rigid designator. In other words, the concept of pain is identical with the concept 'the state that occupies such and such a functional role', where the functional role occupied by the state is specified by a theory of the mind T (Block 2007b). In this sense, the concept of pain has multiple referents – it applies to different things: A certain brain state in us, a certain hydraulic state in a hydraulic automaton, etc.

functional role of p-consciousness, a definition that I would be able to work with in the rest of the thesis. In the specific, I defined the functional role of p-consciousness as *the fund of appropriate dispositional properties that p-consciousness endows a mental state with*. Having offered a definition of the functional role of p-consciousness, I then did two more things. First, I suggested that, for the purposes of exploring the question of what the functional role of p-consciousness is, we work under the assumption that p-consciousness does has a functional role *at all*. Second, I clarified that the claim that p-consciousness has a functional role is to be kept distinct from the claim that functionalism about p-consciousness is true.

2

Block's Distinction and the Global Workspace Model of Access Consciousness

1. Introduction

This chapter has three main parts. In §2, I introduce Block's (1995) distinction between p-consciousness and a-consciousness. As anticipated, that distinction is a fundamental assumption of this work. A mental state is said by Block to be p-conscious if there is something it is like for its subject to be in it (Ibid., 228). A mental state is said by Block to be a-conscious, instead, if its content is available to a number higher-level cognitive processes/capacities such as report, reasoning, and the rational control of action (Ibid., 231).

In the second part of the chapter (§3), I draw a first, general picture of the Global Workspace Model of a-consciousness (GWM) (e.g., Dehaene et al., 1998; Dehaene and Naccache 2001). As I have also anticipated, this is the model of a-consciousness that I assume in this work. According to GWM, a representation's being a-conscious is to be explained in terms of its content's being broadcast in a global workspace. The latter, as I will duly explain, is a resource that allows modular systems of the mind/brain to exchange information.

In the third and last part of the chapter (§4), I argue, together with the advocates of GWM, that attention is necessary for both global broadcasting and a-consciousness. As we will see, different experiments suggest that when subjects fail to attend to a certain stimulus S, they also fail to report having seen S. From this, it is inferred that attention is necessary for a-consciousness and – given certain assumptions – for global broadcasting as well.

The reader should note that, in this chapter, I will only offer a *first pass* at defining GWM and its relationship with a-consciousness, and that I will be offering a more detailed account of both over the next few chapters.

2. Block's Distinction Between Phenomenal Consciousness and Access Consciousness

Following an explosion of books and papers on consciousness between the late 1980s and the early 1990s, the publication of Ned Block's paper *On a Confusion about a Function of Consciousness*, in 1995, compelled philosophers and cognitive scientists working in the field to take stock. In his paper, Block charged researchers with having (unwittingly) conflated two concepts: the concept of p-consciousness and the concept of a-consciousness.

In Block's (1995) view, the confusion between p-consciousness and a-consciousness can be vividly brought to light by considering a number of hypotheses about the functional significance of consciousness *tout court*. Arguably, the most influential of the hypotheses that Block has in mind has to do with the phenomenon of blindsight. On the basis of blindsight, philosophers and psychologists alike (Marcel 1986; Flanagan 1992, chap. 7; Van Gulick 1994) have argued that part of the functional role of consciousness must be to enable information to be used in reasoning, the rational control of action, and other

higher-level cognitive processes/capacities.[16] To give a reader a sense of what Block has in mind, below I outline what is sometimes known as the *Argument from Blindsight*.

2.1. The Argument from Blindsight

What is blindsight? In humans, the optic nerve that leaves the eye flows directly into an area of the thalamus known as the *lateral geniculate nucleus*. From there, new fibres originate that terminate into a specific region of the occipital lobe – one of the four lobes of the cerebral cortex, situated in the "back of the head" – known as the *primary visual cortex*, *striate cortex*, or *V1* (Weiskrantz 1990, 3-6). V1 is the visual area that receives most of the information coming from the eyes. It is responsible for processing some of that information, and for channelling the rest to other areas of the visual system where further processing will occur.

Lesions to V1 and to the occipital lobe more generally typically result in the appearance of a blind region in the patients' visual field. A lesion can be caused by the occurrence of an accident, a stroke, or by the surgical removal of portions of the cortex for medical purposes, for example (Ibid.). If the lesion is such that the V1 is completely destroyed, the patient will lose her sight altogether. If the lesion affects only *part* of the V1, instead, the patient will suffer from partial blindness; that is, she will be blind in circumscribed regions of her visual field.[17] A local region of blindness is typically known as a *scotoma* – plural, *scotomata* (Ibid., 9).

[16] What does the term 'information' mean here? What philosophers as well as psychologists (and arguably also many neuroscientists) normally have in mind when using the term is *semantic information*; that is, the *content* of a representation (Wu 2014, 14–15). I will come back to this point below.

[17] A lesion to the right portion of the primary visual cortex will result in a blind area corresponding to part, or the whole of the left field of vision of both eyes, depending on the extent of the damage. On the other hand, a lesion affecting the left part of the primary visual cortex will cause the formation of a blind area covering the right field of vision of both eyes (Weiskrantz 1990, chap. 1).

Among those patients that have suffered from lesions that exclusively affect portions of V1, some have been discovered to have a very special ability. When presented with relatively simple stimuli in their scotoma, all patients claim to see nothing. Yet, if asked to guess whether a presented stimulus is either a vertical or horizontal line, for example, some patients will perform at a rate significantly above chance. In a similar fashion, if forced to attempt to grasp an object in their blind field, the same patients will preadjust their wrist, fingers, and arm suitably to the shape, orientation, and location of that object (Marcel 1988, 136, 146). These patients are thus "able to respond to visual stimuli without consciously perceiving them" – hence the name of 'blindsight' (Weiskrantz 2007, 175).[18]

Crucially, unless probed, blindsight subjects will never spontaneously report or act on the relevant stimulus. Thus, as Anthony Marcel (1986, 41) points out, if a glass of water is placed in the blind field of a thirsty blindsight patient, the latter will fail to reach for the glass. But why? Why is it that, despite her remarkable abilities, the patient will not *spontaneously* grasp the glass of water? For Marcel, the answer lies in the patient's lack of consciousness in her blind field. Consciousness, he contends, enables the initiation of a voluntary action: "[p]eople will not themselves initiate voluntary actions which involve some segment of the environment *unless* they are phenomenally aware of that segment of the environment" (Ibid., 146, emphasis mine). Thus, consciousness is somehow causally responsible for the deliberate initiation of actions with respect to a portion

[18] Why is it only *some* patients who are able to make visual discriminations within their scotomata? According to Weiskrantz, there may be different reasons for that: one such reason, for example, "is that the location and extent of lesions are not uniform across patients" (1998, 154). Another reason may be related to the age at which a patient has suffered damage to her primary visual cortex: younger patients may be more likely to have the ability to make discriminations within their scotomata than older ones (Ibid.). (See also Weiskrantz 1995).

of the environment. This makes the facilitation of the guidance of voluntary action part of its functional role.[19]

As Block (1995, 241-242) notes, in speculations about the functional role of consciousness from blindsight, the reasoning is typically the following: (1) blindsight patients lack consciousness of the stimuli in their blind field; (2) information about the stimuli in their blind field is not used in report, reasoning, or the deliberate control of action; (3) therefore, (part of) the functional role of consciousness is to facilitate report, reasoning, etc. In Block's view, however, the above reasoning is fallacious. This is because it misses the distinction between p-consciousness and a-consciousness. 'Consciousness' is used in one way in (1), and in another way in (3). More specifically, it is used to mean what Block calls *p-consciousness* in (1), and *a-consciousness* in (3). For this reason, the argument slides "from an obvious function of A-consciousness to a non-obvious function of P-consciousness" (Ibid., 232).

2.2. Zooming in on Phenomenal Consciousness and Access Consciousness

2.2.1. Phenomenal Consciousness

In this section, I close in on Block's take on p-consciousness. One thing that I should note from the start, however, is that although I endorse Block's distinction, I wish to remain neutral as to whether, as Block contends, p-consciousness is neither functional nor identical to a mental state's having a certain representational content (I will explain what those qualifications mean in a couple of pages).

Block acknowledges that it is hard, if not impossible, to define p-consciousness in non-circular terms. The best one can do, he concedes, is "point to the phenomenon" (1995, 230). In his own words, "P-consciousness is experience. P-

[19] Van Gulick, too, suggests that "[i]nformation needs to be presented to us phenomenally for it to play a role in the choice, initiation, or direction of the intentional action" (1994, 33).

consciousness properties are experiential ones. P-conscious states are experiential, that is, a state is P-conscious if it has experiential properties. The totality of the experiential properties of a state are 'what it is like' to have it'" (Ibid.). Paradigmatic examples of p-conscious mental states are *sensations*. Consider, for example, what it is like for you to have a throbbing pain in your right hand, or what it is like for you to be cold, or to have an orgasm.

Another way of fixing the reference of the phrase 'p-consciousness' is by appealing to the famous *explanatory gap*.[20] P-consciousness, notes Block, is "the entity to which the explanatory gap applies" (1995, 232). The idea of an explanatory gap in connection to p-consciousness seems to have been around for quite a while – at least since Leibniz's *Monadology* (1898).[21] A very vivid expression of the gap is due to John Tyndall:[22]

> Were our minds and senses so expanded, strengthened, and illuminated as to enable us to see and feel the very molecules of the brain; were we capable of following all their motions, all their groupings, all their electric discharges, if such there be; and were we intimately acquainted with the corresponding states of thought and feeling, we should be as far as ever from the solution of the problem, "How are these physical processes connected with the facts of consciousness?" The chasm between the two classes of phenomena would still remain intellectually impassable (1872, 120).

What Tyndall calls *chasm* is what philosophers today seem to have in mind when they use the phrase 'explanatory gap'. In contemporary philosophy of mind, the

[20] Perhaps *fixing the reference* is too strong. For as I will explain below, Block thinks that the explanatory gap might apply to phenomenal consciousness only *contingently*.

[21] See *Monadology* §17.

[22] See also Huxley (1900).

notion of an explanatory gap in relation to p-consciousness is typically associated with the work of Joseph Levine (1983).

Consider the two following identity statements: (1) "[h]eat is the motion of molecules"; (2) "Pain is the firing of C-fibers" (Ibid., 354). As Levine notes, (1) seems to explain everything there is to explain about heat.[23] Thus, for example, (1) explains why metal expands at high temperatures, why heating up water makes it boil, etc. It seems, however, that the same cannot be said about (2). Surely, notes Levine, (2) seems to explain something about pain: it explains the mechanisms underlying the fact that, for example, pain "warns us of damage, it causes us to attempt to avoid situations we believe will result in it, etc." (Ibid., 357). Yet, (2) also seems to *leave something out*. "There is a 'gap'", says Levine, "in the explanatory import of [statements like (2)]" (Ibid.). This gap stands out rather vividly when we consider questions such as "Why should pain feel like *this* rather than *that?"*. Or "Why should pain feel like *anything at all?"*. There seems to be nothing about C-fibres that would enable us to answer these questions. Differently put, the relation between c-fibres and pain seems *contingent*. We can easily imagine a world where there are C-fibres firing in your brain, but you do not feel any pain, and a world where you feel pain, but there are no C-fibres' firing in your brain. Hence "the chasm between the two classes of phenomena", between physical and p-conscious phenomena.[24]

In going back to Block's characterization of p-consciousness, Block thinks that the *p-conscious properties* of mental states are distinct from the states' *intentional properties* (1995, 232). Let me immediately say something on p-conscious properties and intentional properties, respectively. By 'p-conscious

[23] Importantly, what we mean by 'heat' here is *not* the p-conscious experience that we enjoy when we are sitting by the fireplace, but the phenomenon as it exists independently of our experience (Levine 1983, 355).

[24] So why does Block think that the explanatory gap might apply to p-consciousness only *contingently*? Roughly, the idea is that one day we might be able to produce the concepts that will allow us to close the gap (1995, 231).

properties', Block means "experiential properties" (Ibid., 230). Examples of p-conscious properties are the reddish character of a perceptual experience as of a red sunset, the painfulness of pain, and the pleasantness of orgasms. The totality of the p-conscious properties of a mental state, Block takes to be identical with what it is like to be in that state (Ibid., 232).

Now for intentional properties. An intentional property is the property of being *directed at*, or *about*, certain objects, properties, or states of affairs. We will encounter talk of intentional (or *representational*) properties in the following chapters as well. For this reason, it is worth briefly expanding on the subject right away. Mental states that have intentional properties are normally referred to as *intentional mental states* (intentional states, for short). Classic examples of intentional states are propositional attitudes; e.g., beliefs and desires. Consider, for example, my belief *that Mars is a planet*. My belief is directed at something, namely, Mars and its being a planet. As a second example, consider my desire *that it be sunny outside*. My desire is about something, namely, today's weather being a certain way.

The term 'intentional' is normally considered to be on a par with 'representational'. The intentional properties of mental states are the states' representational properties. To say that intentional states are directed at, or about, objects, properties, or states of affairs is to say that they *represent* them. Because 'intentional' is normally considered on a par with 'representational', philosophers normally use the terms 'intentional states' and 'representational states' interchangeably. Representational states are said to have content. Typically, the latter is taken to be a proposition. To come back to the examples above, my belief *that Mars is a planet* has, as its content, the proposition *that Mars is a planet*. My desire *that it be sunny outside*, instead, has, as its content, the proposition *that it be sunny outside*.

(One note: for the purposes of this and the following chapters, I will treat the notion of *information* as equivalent to the notion of *representational content*, and the notion of *carrying information* as equivalent to the notion of *having representational content*. To my understanding, this is how the authors that figure in this work understand 'information' and 'carrying information').

In light of what I have just said about p-conscious conscious properties and intentional properties, Block's claim that the p-conscious properties of mental states are distinct from their intentional properties is to be understood as the claim that the experiential properties of mental states are distinct from the states' having a certain representational content. But if, according to Block, the p-conscious properties of mental states are distinct from their intentional properties, what kind of properties does Block think p-conscious properties are? I will address this question at length in chapter 4. For now, however, we can make some progress towards addressing it by putting things as follows.

Consider the painting *La Mort de Marat*, by the famous Neoclassical French painter Jacques-Louis David. The painting represents Marat, a leading French revolutionary, lying dead in a bathtub. The painting has a certain content; e.g., *that Marat lies dead in a bathtub*. The property of having that content is a representational property of the painting. But the painting also has other properties. What I have in mind are the paint, the canvas, etc. But what kinds of properties are these? They are *intrinsic*, non-representational properties of the painting. Now, according to Block, the p-conscious properties of a mental state are akin to a painting's intrinsic properties (1996): they, too, are intrinsic properties of the state.

Thinking of p-conscious properties as intrinsic properties of mental states will also help us appreciate a further claim of Block's. The claim is that p-consciousness can be – and often is – both *intransitive* and *transitive* (1995, 232). That is, although p-consciousness is *essentially* intransitive, it often is *also*

consciousness *of*. According to Block, the reason is this. Sometimes, p-consciousness is like a blob of paint on the very edge of a painting: it represents nothing (or so we may assume). Quite often, however, p-consciousness is like the paint that represents Marat: it represents/has content. It will also be helpful here recalling what I said in the previous chapter about transitive state consciousness. Following Kriegel, we can think of transitive state consciousness as "the property mental states exhibit when, and only when, their subjects are transitively conscious of something in virtue of being in them" (2009, 27). As concerns p-consciousness, Block's idea seems to be that, because p-consciousness often represents, one can be *p-conscious of* something in virtue of being in a p-conscious state.[25]

Two brief points before moving onto a-consciousness. First, according to Block, transitive p-consciousness is *non-conceptual*, in the sense that one need not possess the relevant concepts in order to be p-conscious of something (Ibid.). Thus, for example, I can be p-conscious of a square without seeing the square *as* a square, and I can be p-conscious of the sound a trumpet without hearing it *as* the sound of a trumpet.

The second point is that not only does Block deny that p-conscious properties are distinct from intentional properties; he also denies that they are distinct from *functional properties* (1995, 230). We can understand this claim as the claim that functionalism about p-consciousness is false. As explained in the previous chapter, this means denying that what makes a mental state p-conscious is its functional role.

[25] One issue which I mention only to set aside is this. Given Block's characterization of p-consciousness as *what it's like*, it is not clear how p-consciousness might also be said to be transitive. For *prima facie*, at least, consciousness as what it's like is only an intransitive kind of consciousness. When I have a p-conscious visual experience as of a yellow bird, what it is like to have that experience is not a matter of being conscious *of*. Rather, it seems to be exclusively a matter of my experience's being conscious *simpliciter*.

2.2.2. Access Consciousness

To a-consciousness now. A mental state is said by Block to be a-conscious if its representational content is "made available to the brain's 'consumer' systems: systems of memory, perceptual categorization, reasoning, planning, evaluation of alternatives, decision-making, voluntary direction of attention, and more generally, rational control of action" (2005, 47).[26] The term 'rational' here is meant to exclude the guessing-guided behaviour of blindsight patients (1995, 228). For there is a sense in which information about the stimuli presented in the blind field of a blindsight patient *is* available to the relevant consuming systems. If the patient is forced to guess whether she has just been presented with an 'X' or an 'O', for example, she will somehow be able to access her perceptual state and make an accurate report. Still, there is another sense in which information about stimuli presented in her blind field is not available for her to "deliberately" or "rationally" use. The kind of access that Block has in mind when talking about a-consciousness, however, is full-blooded access in the latter sense. Hence, a blindsight patient's reporting a perceptual state on the basis of a "hunch" does not count as *rational* guidance of her behaviour.

I will say much more about of a-consciousness when discussing GWM (the Global Workspace Model), which Block and I endorse. For now, let me briefly mention a couple of things. First, unlike p-consciousness, a-consciousness is always consciousness *of*. A-conscious mental states are always states in virtue of we are transitively conscious of something. Relatedly, while Block thinks that it *might* be possible for a mental state to be p-conscious without having representational

[26] Block (1995, 231) originally defines access consciousness as follows: "[a] state is access conscious (A-conscious) if, in virtue of one's having the state, a representation of its content (1) inferentially promiscuous, that is, poised for use as a premise in reasoning, (2) poised for rational control of action, (3) poised for rational control of speech". The choice of opting for Block's (2005, 47) characterisation of access consciousness here is due to two main reasons. First of all, it is more developed/informative than the one offered in his (1995). Secondly, it employs some terminology that is more consonant with the terminology I will adopt in the remainder of this chapter, particularly in §3.

content, a mental state cannot be a-conscious if it does not have content. After all, a state counts as a-conscious only if its content can be used in reasoning, the rational control of action, etc. Thus, paradigmatic examples of a-conscious states are propositional attitudes like beliefs and desires (Ibid., 232). Why? For these states *always* have representational content. Third, unlike p-consciousness, a-consciousness is *essentially functional* (Ibid.). A-consciousness, as noted, is a matter of the content of a state being made available to a number of consumer systems. As such, what makes a mental state a-conscious is the role that the state has in the subject's cognitive economy.

2.2.3. Two Numerically Different Properties

From what I have said so far it should be clear that p-consciousness and a-consciousness are different *concepts*. But Block's (1995) aim is also to argue that the two concepts pick out numerically different *properties*. How does he argue for such a claim? The thought seems to be the following: if we can find conceptually possible cases where p-consciousness and a-consciousness come apart, then we will have found – to put it with Kriegel (2009) – "metaphysically possible particulars that instantiate phenomenal consciousness but not access consciousness, and vice-versa" (Ibid., 34). From there, we can then infer that p-consciousness and a-consciousness are numerically different properties. [27] Importantly, in saying that I

[27] Block (1995) says that by introducing conceptually possible cases where p-consciousness and a-consciousness come apart he only wishes to show that the two are different *concepts*. In a later, slightly revised, version of the same paper, however, he adds that it is also his aim to show that p-consciousness and a-consciousness are quite likely numerically different *properties* (2002, 211). Different authors have understood Block's aim in different ways, however. Carruthers (2015b), for example, takes Block to be arguing only for the claim that p-consciousness and a-consciousness are two different concepts. Kriegel (2007, 2009), on the other hand, takes Block to be arguing that the two are also different properties. As is apparent, here, I side with Kriegel. Why? Because if Block's intention were only to show that that p-consciousness and a-consciousness are different concepts, he would not need to present cases where one occurs without the other. Instead, it would probably be sufficient to characterize them in different ways.

assume Block's distinction, I mean that I assume it *both* as a conceptual distinction *and* as a distinction between properties.

Below, I am going to offer a more precise reconstruction of Block's argument. To that end, however, I first need to briefly introduce three notions the understanding of which is fundamental to the understanding of the argument.[28] The notions that I have in mind are the ones of *conceptual possibility*, *metaphysical possibility*, and *metaphysically possible world*. Let's proceed in order. Following Levine (2001), we may say that, given a situation S, "S is conceptually possible relative to a representation R just in case S, when thought of under R, is judged possible" (Ibid., 40).[29] By way of example, consider a situation in which someone drinks water but does not drink H2O. This situation is conceptually possible relative to the representation *that someone drinks water but does not drink H2O* just in case, when I think of someone who drinks water but does not drink H2O, under the representation *that someone drinks water but does not drink H2O,* I judge that it I possible that someone drinks water but does not drink H2O.

As for metaphysical possibility, to say that a situation is metaphysically possible is to say that "it could happen (or could have happened)" (Ibid., 39-40). Importantly, this does not entail that that situation could happen (or could have happened) according to the laws of nature that hold in this world. Thus, a situation's being metaphysically possible is compatible with its being nomologically impossible (Ibid., 40). Last, for metaphysically possible worlds. A metaphysically possible world is just a world that could have been, where 'world'

[28] Plus, we will encounter two of these notions in the next chapter again.

[29] I assume that the notion of *situation* is relatively intuitive. Still, it won't hurt us to be precise about it. Here's Levine on the notion: "[b]y a situation I mean an object's instantiating one or more properties (perhaps at a time), or an ordered n-tuple of objects instantiating one or more relations. It is the sort of entity that serves as the truth condition for a statement. Situations are the subjects of possibility and necessity; this could happen, that couldn't have happened, and this must happen. In all these cases, we are referring to situations. What can happen, I presume, is that I quit my job tomorrow; but it can't happen that I both quit and do not quit my job tomorrow" (2001, 40).

stands for a "maximally inclusive situation encompassing all others" (Menzel 2016, n/a).

With the above notions in hand, we can now take a closer look at what I take to be Block's argument for the claim that p-consciousness and a-consciousness are different properties. Here is one way of reconstructing the argument: 1) if it is conceptually possible for p-consciousness to occur without a-consciousness, and vice-versa, then p-consciousness and a-consciousness are not coextensive in every metaphysically possible world; 2) if p-consciousness and a-consciousness are not coextensive in every metaphysically possible world, then they are different properties; 3) it is conceptually possible for p-consciousness to occur without a-consciousness, and vice-versa; therefore, 4) p-consciousness and a-consciousness are not coextensive in every metaphysically possible world; therefore, 5) p-consciousness and a-consciousness are numerically different properties.[30]

A step-by-step discussion of Block's argument would take us too far afield. As noted on several occasions, it is not my aim to *argue* for Block's distinction. Rather, I am only *assuming* it. Still, it will be worth looking at some of the conceptually possible cases which Block comes up with, for that will give us a better idea of the distinction itself.

Start with p-consciousness without a-consciousness. Block (1995, 233) asks us to imagine an animal that is p-conscious, but whose brain areas underpinning

[30] I should point out that the step from conceptual to metaphysical possibility is not always a legitimate one to make. I strongly sympathise with Block's idea that there being conceptually possible cases of p-consciousness without a-consciousness, and vice-versa, implies that p-consciousness and a-consciousness are not coextensive in every metaphysically possible world. But note that there are uncontroversial cases where conceptual possibility does *not* imply metaphysical possibility. For consider the example provided above. While it is conceptually possible that someone drinks water but does not drink H2O, relative to the representation *that someone drinks water but does not drink H2O*, it is not metaphysically possible for someone to drink water but not H2O. Why? In a nutshell, the reason is that, as Saul Kripke (1980) has convincingly argued, the term 'water' is a rigid designator: it picks out H2O in *all* metaphysically possible worlds.

the systems for reasoning and the rational control of action have been destroyed. Surely, says Block, it is at least conceptually possible that the brain machinery underlying p-consciousness and the machinery underlying a-consciousness are distinct. In this case, the animal's experiences will be p-conscious without being a-conscious.

A second case of p-consciousness without a-consciousness may be the following:

> Suppose you are engaged in intense conversation when suddenly at noon you realize that right outside your window there is – and there has been for some time – a deafening pneumatic drill digging up the street. You were aware of the noise all along, but only at noon you are *consciously* aware of it. That is, you were P-conscious of the noise all along, but at noon you are both P-conscious *and* A-conscious of it (Block 1995, 234).

Thus, although p-conscious all along, your experience of the noise is not a-conscious until noon. That is to say that, although, before noon, there is something it is like for you to hear the noise of the drill, the content of your experience is not made available to the relevant cognitive systems until after noon. [31, 32]

Finding conceptually possible cases of a-consciousness without p-consciousness is more difficult. There is only one such case that Block is able to come up with. This is the case of *superblindsight*, an imaginary extension of blindsight (Ibid., 233). We already know what blindsight is: it is the capacity of subjects that have suffered damage to the primary visual cortex to respond to

[31] It is apparent that on top of being metaphysically possible, the pneumatic drill scenario may be an *actual* case of p-consciousness without a-consciousness.

[32] Here is a similar case from Tyler Burge: "[i]f one 'sees stars' in a drunken stupor but cannot reason with or about that state, the images might be phenomenally conscious, but thoughts with or about such a state might be unconscious in the sense that they are not [access conscious]" (2007, 385).

stimuli in their blind field, albeit only when probed. Superblindsighters are imaginary blindsight subjects that, through intensive training, have acquired the ability to make guesses about stimuli in their blind field *without* being instructed to guess. To put it with Daniel Dennett, superblindsighters have acquired the ability to "guess when to guess" (1991, 331). When a stimulus is presented in a superblindsighter's blind field, "visual information from his blind field simply pops into his thoughts in the way [...] some people just know the time or which way is North without having any perceptual experience of it" (Block 1995, 233). In this respect, superblindsighters are similar to Elijah Chudnoff's blindfolded clairvoyant (2012). When walking into a room, the blindfolded clairvoyant knows how the room is arranged, even though there is nothing it is like for her to (visually) perceive the room. On the assumption that p-consciousness is actually missing in blindsight, superblindsight might then count as a case of a-consciousness without p-consciousness. For even though a superblindsighter's visual perceptual state is not p-conscious, the contents of the state are available to be used in thought, report, etc.[33]

With superblindsight, I conclude my exposition of Block's argument. Before moving onto the next section, however, I should make it clear that to claim that p-consciousness and a-consciousness are different properties is *not* to say that, *in this world*, one can occur without the other, and vice-versa. The two may well be different properties but, as a matter of empirical fact, *always* co-occur. Block's (1995) claim is thus compatible with the possibility that p-consciousness and a-consciousness are "empirically inseparable" (Chalmers 1997, 148).

[33] Unlike Block, Daniel Dennett (1991, 322–44) thinks that a superblindsighter's visual perceptual state would actually be phenomenally conscious. Indeed, Dennett uses the same thought experiment in order to show that there is nothing more to phenomenal consciousness than the functional role of the mental states that have it.

3. Access Consciousness as Broadcasting in a Global Workspace

Since Block's (1995), our understanding of a-consciousness has increased dramatically. Many researchers from different fields (e.g. Carruthers 2015; Dennett 2001; Kanwisher 2001; Kouider and Sackur 2014; Shelton et al. 2008) now agree that GWM (the global workspace model) offers an empirically-grounded and theoretically-satisfying account of our capacity to access our mental states. As Block himself notes, "[there is] impressive evidence that our ability to report our mental states hinges on such a global workspace" (2007, 491).[34, 35] Indeed, the model makes a number of empirically testable predictions, most of which have been borne out by studies of different nature. These include anatomical, physiological, brain imaging, and pharmacological studies (Dehaene et al. 1998; Dehaene and Naccache 2001; Dehaene and Changeux 2011). It should thus be safe to say that the model in question is *the* model of a-consciousness.

GWM was originally put forward by Bernard Baars (1988), and later developed in a neural/neuronal direction by Dehaene and colleagues. (In what follows, I am going to use the terms 'neural' and 'neuronal' interchangeably). Although my focus in this work will be on Dehaene et al.'s version of the model, below I begin setting the stage by outlining some of the model's key tenets as introduced by Baars.

A terminological note. Recently (e.g., 2007; 2008; 2015), Block has replaced the phrase 'access consciousness' by the phrase 'cognitive accessibility': "[a]ccess consciousness was my term for approximately what I am calling 'cognitive accessibility' here" (2007, 486). The main difference between the two phrases is that while 'access consciousness' implies that there are two different concepts of

[34] Block only mentions report, but he also has in mind reasoning, the rational control of action, etc.

[35] In line with Block, Carruthers writes that "there is now extensive evidence supporting the global broadcasting account" (2015a, 52).

consciousness, 'cognitive accessibility' does not (Coates and Coleman 2015, 329). For my purposes, I will overlook the difference between the two phrases and use them interchangeably. This means that I will sometimes talk of GWM as a model of a-consciousness, and sometimes as a model of cognitive accessibility, or accessibility *tout court*.

3.1. Baars's Global Workspace Model

GWM construes the cognitive system as comprising two major functional spaces: a "*massive parallel set of specialized processors*" (Baars 2005, 46, emphasis mine), and a *global workspace*.[36] A processor is defined "as a relatively unitary, organized set of processes that work together in the service of a particular function" (Baars 1988, 50). In this sense, processors are *functionally unified* or *modular* sets of processes. The global workspace, on the other hand, might tentatively be characterized as a system that allows processors to exchange information with one another (Ibid., 87). Let us close in on processors and the global workspace in order.

Processors have one job: to compute or transform mental representations.[37] Among others, there are processors tasked with computing sensory or perceptual representations; processors tasked with computing representations of objects or events stored in different kinds of memory (long-term memory, recognition memory, etc.); and processors dedicated to the computation of mental representations relevant to language recognition and production. (Ibid., 43–64).[38]

Processors have a number of important features. For one thing, as mentioned, they (typically) work in *parallel*. In this sense, processors are quite *autonomous* or *independent* from one another: one processor need not rely on

[36] See also (Baars 1988, chap. 2).

[37] What is it to compute a mental representation? Following Jerry Fodor, we may understand the computation of a mental representation as "a transformation of representations which respects these sorts of semantic relations [i.e., implication, confirmation, and logical consequence]" (1981, 5).

[38] This is by no means supposed to be an exhaustive list.

another processor's work. Second, processors are *highly specialized*: each processor or small set of processors is able to process a very specific kind of input only. For example, some processors may only be able to process specific types of visual information, while others may only be able to process specific types of auditory information. Third, because of their high level of specialization, processors are relatively *rigid*. What this means is that they are quite bad at processing unfamiliar or novel kinds of information. At the same time, in virtue of the fact that they always process the same kind of information, processors are very *fast* and *efficient*. Last, processors are characterised by a *variable composition*. As Baars explains, "processors are like Chinese puzzle boxes: they are *structured recursively*, so that a processor may consist of a coalition of processors, which in turn may also be a member of an even larger set of processors that can act as a single chunk" (Ibid., 62).

Alongside specialized processors, the cognitive architecture of the mind/brain comprises a further component, namely, the *global workspace*. In Baars's own words, the global workspace "is a fleeting memory capacity that enables access between brain functions that are otherwise separate" (2005, 46). The global workspace is a system that enables otherwise isolated processors to disseminate information across the whole cognitive system, thereby allowing them find (or at least look for) common solutions to problems that individual processors are unable to handle by themselves (1988, 87). Baars aptly compares the global workspace to a "television broadcasting station in a human community" and to a "blackboard in a classroom" (Ibid., 74). At any one time, a number of processors compete or cooperate to "broadcast" a message – some piece of information – across the entire system for further processing (Ibid.).

An analogy from Baars himself will help make the picture clearer (Ibid., 87–88). We can compare specialized processors to a very large community of human experts gathered in an auditorium. Suppose now that the whole community is

presented with a problem, and that none of the experts who understand (or partly understand) that problem, are, however, able to solve it all by themselves. Call the set of the experts that understand (or partly understand) the problem *E*. How might a member of E, E1, find and communicate with E2, E3, and En, and get to work on that problem with them? A blackboard in the front of the auditorium might do the trick. E_1 might write a message on the blackboard, thereby making it potentially available for everyone else in the auditorium to read. In this way, E2, E3, and En will be able to read what's written on the blackboard, and react. Thanks to the blackboard, E's members will then also be able to keep exchanging information, and work towards a common solution to the relevant problem.

In an analogous way, the global workspace allows specialized processors to disseminate information across the entire cognitive system, and to solve problems that they would not be able to solve by themselves. Specialization has several advantages: one knows exactly what to do in particular routine situations. Yet, specialization implies a lack of flexibility in dealing with new situations. A specialized processor is very good at dealing with what is known, but very bad at dealing with novelty (Ibid., 77). By providing specialized processors with a space to distribute information to, and react to information from, other processors, the global workspace allows the cognitive system to effectively deal with problems that arise in *novel* domains. Examples of novel domains are learning a new language or learning how to ride a bicycle. In cases such as these, the rigid, highly specialized nature of the single processors may be an issue. Novelty requires new strategies to solve new problems. It requires a kind of flexibility that the single processors lack. Thanks to the global workspace, information from otherwise separate sources of knowledge may then be put together to reach solutions to new problems (Ibid., 89).

More generally, the global workspace allows specialized processors to solve "problems whose solutions are *under*determined" (Baars Ibid., 92). Aside from

arising in novel domains, these problems commonly arise in *ambiguous* domains – that is, domains that allow for multiple interpretations. Ambiguities, suggests Baars, are often found both in language and visual processing, for example. To deal with ambiguities, a unified and coherent interpretation has to be found. This, in Baars's view (Ibid., 77), requires the cooperation and coordination of multiple specialized processors at the same time.

To summarize, according to Baars, the functional architecture of the mind/brain comprises a set of specialized processors and a global workspace. Processors are highly specialized, functionally unified sets of processes that (typically) operate in parallel. The global workspace, on the other hand, is a fleeting memory capacity that allows specialized processors to disseminate information across the whole system, and to cooperate towards the solution of problems that each individual processor would not be able to solve by itself. In this way, the global workspace allows the mind/brain to, among other things, effectively deal with novel situations, and readily resolve ambiguities.

But in what sense is GWM supposed to be a model of cognitive accessibility? I have said nothing about this issue yet, at least not *explicitly*. This is the aim of the section immediately below. There, I introduce Dehaene et al.'s further development of GWM and close in on the relationship between the latter and cognitive accessibility. I should point out again, however, that in what follows I only offer a *first pass* at both GWM and its relationship with a-consciousness, and that I will be offering a more detailed account of both over the next few chapters.

3.2. The Global Neuronal Workspace Model

When speaking of the GWM, what philosophers and cognitive scientists normally have in mind today is the model as has more recently been developed by the cognitive neuroscientist Stanislas Dehaene and colleagues (e.g., Dehaene et al. 1998; Dehaene and Naccache 2001; Dehaene et al. 2006; Dehaene and Changeux

2011). What Dehaene and colleagues have done is develop Baars's original version of GWM in a neural direction, and provide a massive amount of evidence in its support. (It is GWM as further developed by Dehaene and colleagues that I will assume in the remainder of this work). Like Baars, Dehaene and Naccache (2001, 12) hold that the mind/brain is to be construed as a massive set of specialized processors on which "automatic [...] cognitive processing rests". At the level of the brain, these processors can be found both in cortical and subcortical regions (Dehaene and Changeux 2004, 1146). Processors dedicated to the processing of motion, for example, can be found in area V5 (or MT) of the visual cortex, while processors dedicated to the processing of faces in the so-called 'fusiform-face area'.

On top of this large collection of specialized processors,

> The human brain also comprises a distributed neural system or "workspace" with long-distance connectivity that can potentially interconnect multiple specialized brain areas in a coordinated though variable manner. Through the workspace, modular systems that do not directly exchange information in an automatic mode can nevertheless gain access to each other's content. The global workspace thus provides a common "communication protocol" through which a particularly large potential for the combination of multiple input, output, and internal systems becomes available (Dehaene and Naccache 2001, 13).

According to the model, the global workspace consists of a set of cortical neurons with long-distance connections – that is, with particularly long axons. Through their long-distance connections, these neurons allow a variety of modular systems (processors) to exchange information. In this way, the global workspace "break[s] the modularity" of the mind/brain (Ibid., 1147). Indeed, as Dehaene et al. (1998, 14530) point out (but see also Dehaene and Naccache 2001, 14), at least five categories of modular systems are interconnected through the workspace: 1) perceptual systems; 2) motor systems – which include speech-production and

language systems; 3) long-term memory systems; 4) evaluation systems; 5) attentional systems.

But what is the connection with a-consciousness? Focusing on the accessibility of *perceptual* representations, the idea is that, by broadcasting the content of their output representations in the global workspace, perceptual processors can make such content "*available to a variety of processes* including perceptual categorization, long term memorization, evaluation, and intentional action" (Dehaene and Naccache 2001, 1, emphasis mine). Cognitive accessibility is thus to be explained in terms of the content of a mental state/representation being globally broadcast.

A very clear picture of the relation between the global broadcasting of perceptual information and cognitive accessibility is offered by Block, who has endorsed the neural GWM in many of his papers (e.g. 2002; 2007; 2008). Block (2007a, 491) usefully distinguishes between *suppliers* of and *consumers* of information. Perceptual systems can be thought of as supplying information to consuming mechanisms. These include "mechanisms of reporting, reasoning, evaluating, deciding, and remembering" (Ibid.). Perceptual systems supply information to the consuming mechanisms by sending information to "an active storage system" (Ibid.) – the global workspace – to which the consuming mechanisms are connected. Information that is sent to this storage system is *directly* available to all cognitive mechanisms. That is, it is "available to all cognitive mechanisms without further processing" (Ibid.) (see also Block 2009, 1111).

Even though above I focused on the global broadcasting of perceptual information, it is important to note that it is not just perceptual systems that can broadcast information in the workspace. In other words, it is not just perceptual systems that act as *suppliers* of information. On the contrary, other systems, too, can (and do) function as suppliers of information. An uncontroversial example is systems of long-term memory. These systems work *both* as suppliers *and*

consumers of information. When the content of a perceptual representation is globally broadcast, it can, among other things, be stored in long-term memory. In this case, long-term memory systems will act as consumers of information. But information stored in long-term memory can also be globally broadcast, and thus made available for report, for example, as when one is asked to report a past experience. In this case, the long-term memory systems will work as suppliers of information. In the words of Dehaene and Naccache (2001, 14), these systems can "reinstate past workspace states". That is, they can broadcast information that was previously broadcast by the perceptual systems, for example.

One thing that is worth noting is that even though GWM as developed by Dehaene and colleagues is described in neural/anatomical terms, it is still a *functional* model of cognition. There are two main reasons for this claim. First, as Dehaene and Naccache acknowledge (Ibid., 14), the global workspace itself does not coincide with a single brain structure. Although certain brain areas, like the prefrontal cortices, the anterior cingulate, and parietal regions seem to be more "important" or "dominant" than others – workspace neurons are particularly dense in those areas, which is where the consuming systems are located (Dehaene et al. 1998, 14533) – there is no "sharp anatomical delineation of the workspace system. In time, the contours of the workspace fluctuate as different brain circuits are temporarily mobilized, then demobilized. It would therefore be incorrect to identify the workspace [...] with a fixed set of brain areas" (Dehaene and Naccache 2001, 14). Secondly, and perhaps more importantly, the fundamental tenets of the model "abstract away from the neuronal details" (Block 2009, 1111). Nothing in the model requires the specific biological makeup of the brain. Instead, like any other functional model, the GWM can be implement in systems of a wide variety of

makeups. That is, it "can just as easily be realized in silicon-based computers as in protoplasm" (Ibid.).[39]

Before moving onto the next section, I should point out one last thing. According to its advocates, GWM is not just a model of a-consciousness. On the contrary, it is also meant to be an account of p-consciousness. Essentially, the idea is that p-consciousness just is "global information availability" (Dehaene et al. 2011, 56).[40] A representation is thus said to be p-conscious iff it is cognitively accessible; that is, iff its content is directly available to the consuming systems (see also Block 2007a, 492).[41] Of direct relevance here is the fact that although GWM is also meant to be an account of what Block calls 'phenomenal consciousness', nothing prevents us from construing it as a model of cognitive accessibility *alone*. One can endorse the model as a model of accessibility, while holding that the model fails to capture the nature of phenomenality. I will return to GWM's identification of p-consciousness with cognitive accessibility in chapter 4.

[39] Thus, as Dennett (2001, 233) writes, "The proposed consensual thesis is [...] that this global availability [...] *is*, all by itself, a conscious state". See also Block (2001, 203–4).

[40] That what Dehaene and Naccache are trying to explain is phenomenal consciousness, they make it quite clear when they write: "[w]e postulate that this global availability of information through the workspace is what we subjectively experience as a conscious state" (2001, 1).

[41] As we will see in chapter 4, on the other end of the spectrum, we find what Block calls 'non-cognitive' theories of consciousness (2015, 165–68). These theories share a commitment to the claim that a representation can be phenomenally conscious *independently* of its being broadcast in the workspace. In this sense, these theories do retain the distinction between phenomenal consciousness and cognitive accessibility. Importantly, this should not be taken to imply that, once globally broadcast, a perceptual representation is no longer phenomenally conscious. Instead, the claim is just that global broadcasting is not "part of what it is to be conscious" (Block 2007a, 481). Among others, non-cognitive theories of consciousness have been advocated by Block (e.g., 1995, 2007); Lamme (2003), and Bronfman et al. (2014). I will discuss these theories at length in chapters 4 and 5.

4. Global Broadcasting, Access Consciousness, and Selective Attention

Let us assume, for the moment, that global broadcasting is both necessary and sufficient for a-consciousness.[42] Of central importance for the issues that we will discuss in later chapters is the relationship between global broadcasting and selective attention ("attention", for short). Dehaene et al. (e.g., 1998; 2001; 2006) suggest that global broadcasting, and thus cognitive accessibility, are strongly dependent on attention. But what is attention? Is the above claim justified? And in what sense does global broadcasting depend on attention? These are the main questions that I aim to address in this section.

The question of what attention is, has received a great many answers (Mole 2011). It is not my aim here to argue for one view of attention or another, or even review the different answers that have been provided. For my purposes, I will instead assume a very thin notion of attention which, in principle, should be accepted at least by the majority of researchers. The general idea seems to be that attention is a mechanism (or set of mechanisms) whose main purpose is to control how our cognitive system's resources are allocated.[43] The background assumption is that the cognitive system's processing capacity is limited: there is only so much information that can be handled at any one time (Wu 2014, chap. 1). It is thus generally accepted that the role of attention is to *filter* or *select* information relevant to the organism's current goals, so as to allow only the selected information to reach further stages of processing (Lamme 2004; Wu 2014; Carruthers 2015a).[44] Attention, in other words, determines the *depth of processing* of information.[45]

---

[42] I will come back to this assumption at the end of this section.

[43] Even this very general idea, however, has been challenged by some (see Mole 2011).

[44] The so-called 'filter' model of attention was originally proposed by (Broadbent 1958).

[45] As we will see, there are other factors involved in determining depth of processing. Some stimuli, for example, are processed at a deeper level than others simply because they are stronger than others in terms of brightness, colour, or dimension, for example (Lamme, n.d., 4).

A significant body of evidence has been offered in support of the claim that our capacity to report and, more generally, to access our mental states depends on attention. In the language of GWM, this claim can also be understood as the claim that attentional selection is necessary for global broadcasting. Following Wu (2014, chaps. 5–6), we can call this the *gatekeeping view of attention*: attention is the gatekeeper of the global workspace. The basic idea is quite simple. A representation of a stimulus/group of stimuli will be broadcast in the global workspace on condition that the subject attend to the stimulus. Attention determines or selects which representations get into the workspace and, accordingly, which representations become available to the relevant consuming systems. By contrast, representations of unattended stimuli remain inaccessible, at least until (if ever) they are themselves targeted by attention.[46]

A much-discussed phenomenon that is thought to lend support to the gatekeeping view is *inattentional blindness* (Mack and Rock 1998). This can be broadly defined as a subject's failure to have an (*a*-)conscious perception of an object or features of an object in plain sight in the absence of attention.[47] Several experiments are thought to provide evidence for the phenomenon. One of the most well-known sets of experiments is Simons and Chabris's (1999).

In one experiment, subjects were invited to watch a videotape, and were later asked whether they had noticed anything unusual. Prior to viewing the video, subjects were told that they would be watching two teams of three players, one in black, and one in white, passing basketballs. Subjects were also instructed to pay attention to either the team in black or the team in white, and to keep count of the number of passes made by the attended team (the Easy condition), or to keep count

[46] I will say more on this very last point in chapter 5.

[47] Some, like Dehaene and colleagues, would define inattentional blindness a subject's failure to have a conscious (without the 'a') perception of an object or features of an object in plain sight in the absence of attention. As I will explain below, however, one can believe that attentional blindness is a true phenomenon, and understand 'blindness' as having to do with a-consciousness, but not p-consciousness.

"of the number of bounce passes and aerial passes made by the attended team (the Hard condition)" (Ibid., 1066) .

After 44 - 48 sec. from the beginning of the video, an unexpected event took place: an actor wearing a gorilla costume walked through the scene, from one side of the action to the other. The event lasted for 5 sec. Both during, and after the event, the players continued passing basketballs as before. After the video ended, subjects were immediately asked to record their counts on paper. They were then asked whether they had noticed a gorilla walk across the screen. Surprisingly, about half of the subjects provided a negative answer: they had failed to notice the gorilla walking through the scene. Even more surprisingly, the experimenters obtained roughly the same results when, instead of simply walking through the scene, the gorilla "stopped in the middle of the display, turned the face to the camera, thumped its chest, and then continued walking across the field of view" (Ibid., 1069).

There is wide agreement that experiments like the ones above lend support to the inattentional blindness hypothesis (Mack and Rock 1998; Simons and Chabris 1999; Dehaene and Naccache 2001; Carruthers 2015a).[48] Following Wu (2014, 159), we can reconstruct the inference as follows. We begin from the observation that subjects do not report seeing the gorilla. From this, it is inferred that, while the gorilla was walking across the screen, no representation of the gorilla was (a-)conscious. Given that all (or almost all) of the subjects' attentional resources were focused on the ball, it is further inferred that subjects were not attending to the gorilla. Hence, there is evidence that no representation of the gorilla was (a-)conscious, and that subjects were not attending to the gorilla. Proceeding by inference to the best explanation, it is then concluded that a representation of the gorilla was not (a-)conscious because subjects were not

[48] Although see Watzl (2017) for a dissident position.

attending to the gorilla. Hence, attention is necessary for (a-)consciousness. According to the advocates of GWM, the above conclusion further implies that attention is necessary for global broadcasting.

I should note that the inattentional blindness hypothesis is not uncontroversial, though. For failure to report a previously seen stimulus is compatible with the claim that subjects were (a-)conscious of the stimulus. Indeed, in connection to the Simons and Chabris's experiments, for example, one might suggest that whilst watching the video subjects did become (a-)conscious of the gorilla, but failed to report seeing it because no representation of the stimulus was stored in memory. According to this reading of the evidence, inattention results in memory lapses, rather than in a representation's failure to become (a-)conscious (Wu 2014; Carruthers 2015a). Such a hypothesis is normally known as 'inattentional amnesia'. It is not my aim here to engage in a thorough defence of the inattentional blindness against the inattentional amnesia hypothesis. Yet, as Carruthers (2015a) notes – correctly, in my view – the inattentional amnesia hypothesis is a very implausible alternative. For in experiments like Simons and Chabris's, the event or stimulus that subjects fail to report is, to put it with Carruthers, "both unusual and striking" (Ibid., 54). This makes it really hard to understand how, if seen, such an event would fail to be remembered, as the defender of the inattentional amnesia hypothesis contends (Ibid.).

Importantly, it should be borne in mind that, for those who equate a-consciousness with p-consciousness, the experiments above show that attention is necessary for consciousness *simpliciter*. Someone, like Block, who distinguishes between p-consciousness and a-consciousness can endorse the inattentional blindness thesis and understand 'blindness' as having to do with a-consciousness, but not p-consciousness. That is, he can accept that our capacity to access our mental states depends on attention, while holding that we can be p-conscious of stimuli we do not attend to. In the present context, given our endorsement of

Block's distinction, and given our understanding of GWM as a model of a-consciousness alone, the 'blindness' part of 'inattentional blindness' should only be understood as blindness as concerning a-consciousness (hence my choice to place an 'a-' in parentheses next to 'consciousness' above and below).

Let me cover one more piece of evidence that lends support to the gatekeeping view of attention. Lesions to the right parietal region of the brain often lead to a condition known as *hemineglect* (e.g., Driver and Mattingley 1998; Dehaene and Naccache 2001; Driver and Vuilleumier 2001). In most, everyday circumstances, hemineglect patients fail to report and/or act upon stimuli on the left side of space. To put it more vividly, they act as if the left side of space did not exist (Driver and Vuilleumier 2001, 40).[49] Focusing our attention on visual perception, hemineglect patients may fail to read words on the left side of a page, ignore people on the left, etc. Consistently with the above, in experimental settings, when presented with two stimuli side by side, one on the right of fixation, and the other on the left, hemineglect patients only report seeing the stimulus on the right. But there is something very surprising: when the stimulus on the left is presented alone, patients often report seeing it. This is consistent with the fact that, unlike blindsight, in hemineglect patients, most of the neural systems and pathways underpinning visual perception (including the primary visual cortex) are left untouched. But if most of the visual perceptual machinery is there, and if, when the left stimulus is presented alone, patients report seeing it, why is it that they fail to do so, when right and left stimuli are presented simultaneously?

The most credited hypothesis is that hemineglect is an attentional impairment. When subjects are presented with two stimuli, one on the right of fixation, and the other on the left, the stimuli compete for attention. This is thought to happen in hemineglect and normal subjects alike. In the former, however,

[49] In the remainder of this paragraph, and in the next, I will draw on the work of Driver and Vuilleumier (2001).

attention is thought to be *strongly biased* towards the right side of space. For this reason, in simultaneous presentation conditions, stimuli on the right are always going to win the competition for attention, and are thus said to *extinguish* the stimuli on the left.

Different lines of evidence support the attentional impairment hypothesis of hemineglect. For one thing, as mentioned, in hemineglect patients, most of the neural systems and pathways underpinning visual perception are left intact. Consistently with this, as we have seen, when subjects are presented with one lone stimulus in the left side, they report seeing it. Secondly, in experimental settings, hemineglect can be temporarily reversed (Watzl 2017). The reversal of the condition is normally achieved by presenting patients with very salient stimuli on the neglected side, thereby "forcing attention to it" (Ibid., 256). Last, but not least, the right parietal region of the brain has traditionally been thought to underpin the capacity to guide attention into one's visual field (Watzl 2017).

What is the upshot for the gatekeeping view? On the plausible assumption that hemineglect is indeed an attentional impairment, the condition is strong evidence for the claim that there can be no (a-)consciousness – and thus no global broadcasting – in the absence of attention (Dehaene and Naccache 2001). Accordingly, hemineglect is strong evidence for the gatekeeping view.

Having established that attention is necessary for cognitive accessibility (and global broadcasting), *how* does it work? The idea is that attention triggers the global broadcasting of information about the attended stimulus/group of stimuli. We do not know the exact mechanisms behind the triggering yet. The general idea, however, is that when one attends to a certain stimulus, attention amplifies/boosts the activity of selected groups of processors beyond a certain threshold, while suppressing the activity of other groups (Dehaene and Changeux 2004, 1147).[50]

[50] See also Dehaene and Naccache (2001, 19)

Once the activity of the selected groups of processors reaches that threshold, global broadcasting occurs. Following Carruthers (2015b, 4, 2015a, 84) we may thus called such a threshold the "threshold for global broadcasting".

By 'activity' here, I mean *neural* activity (or *activation*). For according to GWM, processors are highly specialized cortical or subcortical groups of neurons, each group being dedicated to the processing of a very specific kind of information. In the case of visual perception, this may be information about a certain kind of horizontal motion, or information about the edges of objects. When one attends to a stimulus/group of stimuli, attentional signals target the activity of selected groups of neurons, thereby boosting it beyond the threshold for global broadcasting, and suppressing the activity of other groups.[51]

One last point. I have assumed that global broadcasting is both necessary and (all other things being equal) sufficient for a-consciousness. As a consequence, I have understood arguments for the conclusion that attention is necessary for a-consciousness, also as arguments for the conclusion that attention is necessary for global broadcasting. As I will argue in the next chapter, however, I think that there is one further condition that must be met for a-consciousness to be present. This is what I will call the 'functionality condition'. Roughly, the idea is that for a state to be a-conscious, it is necessary, alongside global broadcasting, that the consuming systems be functional. But this raises an issue: how can we be sure that attention is a necessary for global broadcasting, rather than for the consuming systems to be functional? What justifies the inference from the claim that attention is necessary for a-consciousness to the claim that attention is necessary for global broadcasting?

[51] It is worth noting that, according to view of attention under discussion, the neural activity targeted by attention is activity in the so-called *mid-level* sensory areas of the brain (Carruthers 2014, 147). In the case of visual perception, these are those areas of the visual cortex that are responsible for processing "color, form, motion, spatial layout, and faces" (Ibid.).

The answer is that the evidence that I have discussed in this chapter is corroborated by evidence of a different nature. Thanks to different brain imaging techniques, it has been possible to establish that it is *only* when one attends to a certain stimulus/group of stimuli, that neural activation extends from the occipito-temporal areas in the "back of the head", to the more "frontal" areas that are thought to underpin/realize much of the global workspace. On the other hand, when the relevant stimulus is not attended, neural activation remains confined in the "back of the head" (e.g., Dehaene et al. 2006).[52] From this, it is inferred that attention is necessary for the content of a perceptual representation to be broadcast in the workspace.[53]

The reader will probably wonder why I have chosen to discuss inattentional blindness and hemineglect, rather the evidence that I just mentioned in the paragraph above. The reason has to do with the fact that, for the purposes of chapter 2, I needed a way to establish that attention is necessary for a-consciousness, and that such a claim can be endorsed even by someone who's not committed to GWM.

5. Concluding Remarks

I started this chapter by introducing Block's distinction between p-consciousness and a-consciousness. As noted on several occasions, Block's distinction is a fundamental assumption of this work. More specifically, as I pointed out in §2.2.3., the assumption is that p-consciousness and a-consciousness are different concepts, but also numerically different properties. The claim that the two are different properties, I should stress, is to be kept distinct from the stronger claim that they are not coextensive in the actual world. Something else that I should stress is that, in spite of my commitment to Block's distinction, I wish to remain neutral as to

[52] I will come back to this point in chapters 4 and 5.

[53] In chapter 4, I will also argue that, all other things being equal, attention is also *sufficient* for global broadcasting.

whether p-consciousness is neither functional nor identical to a mental state's having a certain representational content.

In the second part of the chapter I gave a first pass at defining GWM, the model of a-consciousness that Block and I assume. As explained, according to GWM, access consciousness is a matter of the content of a representation being broadcast in the global workspace. Globally broadcast contents are *directly* available – "without further processing" (Block 2007a, 491) – to systems of report, deliberation, memory, and so on. More will be said on GWM and its relationship with a-consciousness over the next chapters.

In the third (and last) part of the chapter, I argued, together with the advocates of GWM, that attention is necessary for both a-consciousness and global broadcasting. As we will see, the former claim will be particularly important for the purposes of chapter 3, and the latter for the purposes of chapter 4 and 5. We are now ready to consider, and evaluate, Kriegel's account of the functional role of p-consciousness.

3

Kriegel on the Functional Role of Phenomenal Consciousness

1. Introduction

This chapter challenges Uriah Kriegel's account of the functional role of p-consciousness. Against the backdrop of his self-representational theory of p-consciousness (2009), Kriegel (2007, 2009, 2015) argues that the relationship between p-consciousness and a-consciousness is analogous to the one between the molecular structure of salt and salt's solubility: p-consciousness is the categorical basis of a-consciousness. On this basis, Kriegel then construes a-consciousness as (part of) the functional role of p-consciousness.

I offer two arguments against Kriegel's account. First, I argue that the account is incompatible with the Global Workspace Model of a-consciousness (GWM). On the assumption that an account of the functional role of p-consciousness must be compatible with GWM, I then conclude that Kriegel's account is untenable. Second, I argue that the account gets the relationship between a-consciousness and attention wrong. As I argued in chapter 1, the empirical evidence strongly suggests that a-consciousness depends on attention. On Kriegel's account of the functional role of p-consciousness, however, a

representation can be a-conscious even in the (almost complete) absence of attention.

Here is the plan. In §2, I provide an outline of representationalism about p-consciousness in general. In §3, I introduce Kriegel's self-representational account of p-consciousness, and in §4 his account of the functional role of p-consciousness. My arguments against Kriegel's account of the functional role of p-consciousness will be developed in §5, in the order outlined in the paragraph above.

2. Representationalism about Phenomenal Consciousness: A Brief Outline

Representationalism about p-consciousness holds that the *phenomenal character* of a mental state – *its p-conscious aspect* – is explainable in non-phenomenal terms, or, more specifically, in terms of intentionality.[54] As explained in the previous chapter, the intentionality of a mental state is that feature whereby the state is *directed at*, or *about*, a certain object, property, or state of affairs. As is typically understood, the term 'intentional' is to be considered on a par with 'representational'. Intentional mental states are representational states. Representational mental states have content, and this content is normally taken to be a proposition. Thus, my belief *that Mars is a planet* has, as its content, the proposition *that Mars is a planet*.

There are two main brands of representationalism: first-order and higher-order. First-order representationalism (FOR) purports to explain the phenomenal character of an experience, its p-conscious aspect, in terms of its *first-order* content; i.e., in terms of its representing the world as being a certain way. Thus, for example,

[54] Qualifications are already in order. The literature distinguishes between *weak* and *strong* representationalism about p-consciousness (e.g., Levine 2003, 57–76; Chalmers 2010, 339–79). Weak representationalism holds that p-conscious states always exhibit intentionality. Strong representationalism, instead, holds that the p-conscious aspect of a mental state is explainable in terms of the state's intentional properties. As should be apparent, the kind of representationalism that I am concerned with here is *strong* representationalism.

FOR will explain the phenomenal character of a visual experience as of a red tomato on the counter in terms of its representing *that there is a red tomato on the counter*, say. But virtually all advocates of FOR agree that mental representation alone does not suffice for p-consciousness. Thus, typically, it is also required that for an experience to have phenomenal character, the experience or its content satisfy further constraints.[55]

For example, according to one influential version of FOR, due to Michael Tye (e.g., 1995; 2000), an experience has phenomenal character iff, among other things, it is poised to make an impact on the belief-desire system, and its content is non-conceptual. As concerns being poised, the thought is that p-conscious experiences are states that are disposed to inform our beliefs and desires. As for non-conceptuality, the idea is that p-conscious experiences are states one can be in, regardless of whether one possesses the relevant concepts. As an example, I can have a visual experience as of a specific shade of blue – majorelle blue, say – even if I lack the concept of majorelle blue.

Higher-order representationalism (HOR) is a family of theories that purport to explain the phenomenal character of an experience in terms of its being suitably represented by a higher-order state. Suppose I am enjoying a visual experience as of a red patch of colour. According to HOR, the experience's representing that *there is a red patch of colour* will not suffice for it to have phenomenal character. On the contrary, the experience itself must also be suitably represented by some higher-

[55] Thus, the literature distinguishes between *pure* and *impure* (strong) representationalism (Bourget and Mendelovici 2014; Chalmers 2010). Pure representationalism is the thesis that the phenomenal properties of a mental state – its p-conscious properties – are identical with pure intentional properties. A pure intentional property is the property of having a certain content. Impure representationalism, instead, holds that phenomenal properties are identical to impure intentional properties. An impure intentional property is the property of representing something *in a certain manner;* e.g., visually, auditorily, doxastically, etc. Furthermore, as Chalmers notes, "manners of representation may also involve functional characterizations of the representing state" (2010, 342). More on this immediately below.

order state. The main reason for this constraint is that, according to the advocate of HOR, it does justice to the idea "that all (or at least most) mental state-types admit of both conscious and unconscious varieties" (Carruthers 2016, n/a).

One of the main points of disagreement between different higher-order theories is this. Following a tradition which goes back to (at least) John Locke (1975), some philosophers, like Lycan (1996), take the relevant higher-order state to be perception-like. On this view, the relevant higher-order representations are non-conceptual. Others, like Rosenthal (2005), however, take the relevant higher-order state to be thought-like. According to theories like Rosenthal's, the relevant higher-order representations are conceptual.

By requiring that experience be the object of some higher-order representation, HOR purports to do justice to the intuition that p-conscious experiences are experiences their subjects are *aware of*. According to the advocates of HOR, FOR holds the merit of having made progress with respect to the question of why a certain experience has *this*, rather than *that* phenomenal character (Kriegel 2009). Yet, they go on to suggest, FOR is unable to explain why our experiences are p-conscious *at all*. To do this, FOR would have to explain how we come to be aware of our experiences, which it does not.

Traditionally, higher-order theories of p-consciousness have construed the higher-order representation and the experience as numerically distinct states. More recently, however, authors like Uriah Kriegel (2009) have argued for the alternative view that the higher-order representational state is actually *constitutive* of the experience itself.[56] According to this view, known as Self-Representationalism (SR), an experience has phenomenal character just in case it *represents itself* in a suitable manner.

[56] More on this in §3.

While it is true that, unlike traditional HOR, SR does not distinguish a higher-order state from a lower-level one, the latter can generally still be seen as a higher-order theory (Carruthers 2016). Indeed, like standard higher-order theories, SR accounts for the phenomenal character of experience in terms of its having two different contents: a first-order content, whereby the experience represents the world, and a second-order content, whereby the experience represents itself.

3. Kriegel's Self-Representational Theory of P-Consciousness

This section outlines Kriegel's theory of p-consciousness. Following Kriegel's (2012) explanatory strategy, I begin my exposition of his theory by outlining his account of the *explanandum* – p-consciousness. At a later stage, I will then introduce his account of the *explanans* – peripheral inner awareness and, ultimately, self-representation.

Levine (2001) distinguishes between two aspects of phenomenal character: what he calls *subjectivity*, on the one hand, and *qualitative character*, on the other. Consider my experience as of a red sunset. There is something it is like for me to have the experience, a *reddish way it is like for me*. This is the overall phenomenal character of the experience. According to Levine, "subjectivity is the phenomenon of there being something it is like *for me* to see" the red sunset (Ibid., 7). Qualitative character, instead, "concerns the 'what' it's like for me: reddish or greenish, painful or pleasurable, and the like" (Ibid.).

Echoing Levine, Kriegel (2009), too, distinguishes between qualitative character and subjectivity, although he prefers to call the latter *subjective character*. Again, with reference to my experience as of a red sunset, qualitative character is the *reddish* aspect of phenomenal character, while subjective character is its *for-me* aspect or *for-me-ness*. In Kriegel's view, one way of understanding the relation between subjective and qualitative character is as a determinable /determinate relation. In his own words, "My view is that there are many

determinate phenomenal characters – bluish-for-me-ness, greenish-for-me-ness, bitterish-for-me-ness, trumpet-for-me-ness, etc. – and the determinable of all of them is for-me-ness as such" (2012, 443).[57] One consequence of this, notes Kriegel, is that subjective character provides the *existence* conditions of phenomenality, while qualitative character provides its *identity* conditions (Ibid.). While subjective character is what makes a mental state "phenomenally conscious at all (rather than a non-phenomenal state)" (2009, 10), qualitative character is "what makes it the phenomenally conscious state it is (rather than another)" (Ibid.).

Because subjective character is what makes a mental state p-conscious at all, Kriegel identifies it with p-consciousness *as such*; that is, with "that which is common (and peculiar) to all phenomenally conscious states" (2012, 444). For this reason, although a complete theory of p-consciousness will also provide an account of qualitative character,[58] its main concern is to account for subjective character (2009, 12–13).

---

[57] Kriegel (2009, 1, 11) suggests that phenomenal character can also be understood as the *compresence* of subjective and qualitative character. However, I believe this way of understanding phenomenal character is confusing. The reason is that the term 'compresence' suggests that qualitative and subjective character are two *different* properties. But this is not Kriegel's view. In fact, qualitative and subjective character are just two *aspects* of the *same* property (Ibid., 1, 8).

[58] Kriegel (2009, 8) suggests that the qualitative character of a conscious experience at a time T is the sum of the qualitative properties the experience instantiates at T. In his view, qualitative character is explainable, at least in part, in terms of an experience's having a certain representational content. More specifically, an experience's having qualitative character is partly a matter of its representing certain *response-dependent* properties (Ibid., 87). Roughly, a response-dependent property is a dispositional property of objects. It is the disposition to cause certain intra-cranial effects, in specific kinds of subjects, under certain circumstances (Ibid., 84–93). One important thing to note is that, unlike most representational accounts of the qualitative properties of experience, Kriegel takes the content involved in qualitative character to be *narrow* (Ibid., 93). Roughly, what this means is that this content is determined, at least in part, by factors that are *internal* to the subject's head. Another way of explaining the same idea is to say that, according to this view, qualitative character is *locally supervenient* (Kriegel Ibid., 93). That is, roughly, there cannot be a difference in qualitative character without an intra-cranial difference. Crucially, it should be noted that, according to Kriegel (Ibid., 110), an experience's having a certain representational content is but a *necessary* condition for it to have qualitative

Kriegel's ultimate aim is to provide a *reductive* account of subjective character; that is, an account of p-consciousness in non-phenomenal terms (Ibid., 12). [59] To that end, Kriegel begins by explaining subjective character in terms of a certain kind of awareness. In his own words, "to say that my experience has subjective character is to point to a certain *awareness* I have of my experience" (Ibid., 8). Consider my current visual experience. I am visually aware of my laptop, the alarm clock to its left, and the red lamp to its right. Because the objects of my awareness are objects and features in my environment, Kriegel calls this awareness *outer* awareness (Ibid., 16). However, one's awareness can also be directed at one's inner goings on. More precisely, one can also be aware "of internal events and states in one's own mental life" (Ibid.). The latter form of awareness, Kriegel calls *inner* awareness. Inner awareness is fundamental to p-consciousness. Indeed, in Kriegel's view, the *right kind* of inner awareness is normally both necessary and sufficient for a mental state to have subjective character (Ibid., 16-17). But what is the right kind of inner awareness? Kriegel thinks it is *peripheral* inner awareness. Let me explain.

Normally, notes Kriegel, inner awareness is *peripheral*. To see what this means, consider my current visual experience again. I am visually aware of my laptop and a host of other items. But the way I am aware of my laptop differs from the way I am aware of the red lamp to its right, and the alarm clock to its left. In particular, I am currently *focally aware* of the laptop, but only *peripherally aware* of the other items. In a more picturesque manner, we can say that the laptop occupies the centre stage of my consciousness, while the lamp and the clock occupy one of its outer sides.

---

character. Indeed, for an experience to have qualitative character, it is also necessary that the experience be *for* the subject. As we will see immediately below, an experience is *for* the subject just in case the latter is aware of the former in the right way.

[59] I should point out that the claim that subjective character is the main explanandum of a theory of p-consciousness was already emphasized by Levine (2001). According to both Levine and Kriegel, it is subjective character – or 'subjectivity' as Levine calls it – that is responsible for the explanatory gap, and that, as a consequence, is where the deepest problem of consciousness lies.

The focal/peripheral distinction extends to inner awareness (Ibid., 47). On occasion, we can be focally aware of our experience. This happens when we introspect. In that case, we bring our experience to the "forefront of consciousness". In all other circumstances, however, we are only peripherally aware of our experience. That is to say that, typically, the awareness we have of our own experiences is analogous to the awareness I have of the alarm clock beside my laptop, the traffic noise in the background, and so on. While being introspectively (focally) aware of my experience is sufficient for it to have subjective character, it normally is not necessary. That is precisely because, aside from those rare cases in which we introspect, we are normally only peripherally aware of our experiences. In this sense, normally, peripheral inner awareness is both necessary and sufficient for subjective character. More strongly, in Kriegel's own words, "subjective character *just is* normally peripheral inner awareness" (Ibid., 50, emphasis mine).

(One note. Because focal inner awareness is of secondary importance – for both Kriegel's theory and my critique below – we can now put it aside, and focus exclusively on peripheral inner awareness. Moreover, unless otherwise specified, in the remainder of this chapter, I use 'inner awareness' and 'peripheral inner awareness' interchangeably).

But how is inner awareness to be explained? Kriegel argues that it should be explained in self-representational terms. As anticipated, like traditional HOR, Kriegel holds that being aware of one's experience is a matter of being in a mental state that represents the experience. Unlike traditional HOR, however, he rejects the idea the state representing the experience – on the one hand – and the experience itself – on the other – are numerically distinct states. On the contrary, in his view, to be aware of one's experience is a matter of having an experience that *represents itself* (in the right way).[60]

[60] In Kriegel's view, a mental state is p-conscious iff it represents itself non-derivatively, specifically, and essentially (2009, 157–164). As Kriegel (2012, 445) explains, "The first

One of the most important aspects (if not *the* most important aspect) of Kriegel's theory is his account of the *ontology* of self-representation. That is, his account of just what *kind* of property self-representation (and hence p-consciousness) is (Ibid., 200). What is it for a mental state to represent itself? Kriegel opts for an account that appeals to the notions of *broadly causal relation* and *indirect representation*. Let me explain.

Roughly, Kriegel's idea is that p-conscious states are complex, two-part states, where 'complex' means that the overall state is not a mere sum of the two parts, but a single entity of which the two parts are *proper* parts. P-conscious states comprise a lower-order part (P1) and a higher-order part (P2). Suppose you are having a p-conscious experience as of a shining star. P1 represents the star as shining – it is, so to speak, world-directed. P2, instead, represents the whole mental state – that is, P1 plus itself. The question now is just how P2 does what it does. Kriegel's idea can be explained in two steps. The first step involves P2's representing P1. Kriegel's thought is that P2 represents P1 by standing in a broadly causal relation with it. The kind of broadly causal relation that Kriegel has in mind is that of *covariation*. Thus, for example, the number of rings on a tree's trunk represents the tree's age in virtue of a covariation between the former and the latter.

The second step involves P2's representing the whole mental state. This, Kriegel explains in terms of the notion of *indirect representation*. What is it to represent indirectly? Consider a painting of a house. The painting depicts a house, even though a small part of the house is occluded by a bush. Kriegel thinks it is natural to suggest that the painting represents the entire house (indirectly) in virtue of representing a big part of it (directly). In this sense, notes Kriegel, "it seems

---

qualification is designed to rule out merely derivative, or conventional, self-representation (as in, e.g., the sentence 'this very sentence is written in black'); the second to rule out generic self-representation (as in, e.g., having the thought that all thoughts are neurally realized); the third to rule out purely accidental self-representation (as when I think that my mother's nieceless brother's only nephew is probably short before realizing that I am my mother's nieceless brother's only nephew)".

that *x* may sometime represent *y* in virtue of representing *z* because *z* is a part of *y*" (Ibid., 225).61 In roughly the same fashion, P2 represents the entire mental state (and hence itself) indirectly, in virtue of representing P1, directly.

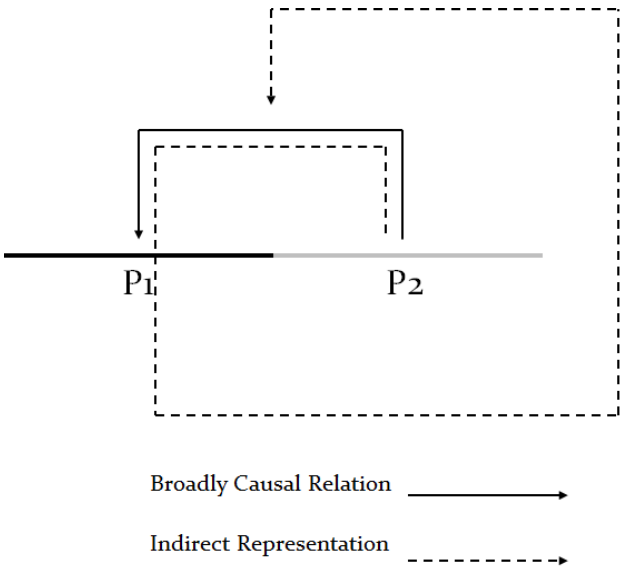

Figure 1. An illustration of Kriegel's account of self-representation. I should stress that P1 and P2 form a *complex* of which they are *proper* parts.

With Kriegel's account of p-consciousness in hand, we can now turn to his account of the latter's functional role.

4. Kriegel's Account of the Functional Role of Phenomenal Consciousness

This section introduces Kriegel's (2007, 2009, 2015) account of the functional role of p-consciousness. In a way to be duly explained, Kriegel thinks that the functional role of p-consciousness is a-consciousness. In his view, this claim is meant to follow directly from another claim, namely, that p-consciousness is the *categorical basis*

[61] Importantly, as Kriegel points out, "one thing's being part of another does not guarantee that by representing the former we represent the latter" (2009, 225). The house in the example above is part of our galaxy, but this does not mean that by representing the house we represent the galaxy. As a consequence, for something to represent something else by representing one of its parts, further conditions must be met (Ibid.).

of a-consciousness. At a first pass, to say that a property P is the categorical basis of another property, P1, is to say that P is related to P1 in the same way in which, for example, the molecular structure of salt is related to its solubility (or the physico-chemical structure of glass is related to glass's fragility).

4.1. Phenomenal Consciousness as the Categorical Basis of Access Consciousness

Kriegel introduces the claim that p-consciousness is the categorical basis of a-consciousness as part of his attempt to vindicate current scientific practice. As explained in the previous chapter, Block (1995) believes that scientists have systematically failed to target p-consciousness, and studied a-consciousness instead. In Kriegel's view (2009, 34-35), Block's argument for this claim can be roughly reconstructed as follows: (1) a-consciousness and p-consciousness are distinct properties; (2) scientists have only studied a-consciousness; (3) therefore, scientists have not studied p-consciousness. Is Block's argument sound? Kriegel thinks it is not. The idea is that if we manage to establish that p-consciousness is the categorical basis of a-consciousness, then we will be able to claim that scientists *might* have been studying the former *via* studying the latter.[62] As a result, from premises (1) and (2) above, Block's conclusion, (3), does not follow.

But why should we believe that p-consciousness is the categorical basis of a-consciousness? Here's a reconstruction of Kriegel's argument:

1. A-consciousness is a dispositional property.
2. Dispositional properties have categorical bases that ground them.
3. P-consciousness grounds a-consciousness. Therefore,
4. P-consciousness is the categorical basis of a-consciousness.

[62] As Kriegel points out, his claim is not that "research into consciousness *has* targeted phenomenal consciousness, but, more modestly, that its doing so is not excluded by the facts that it has targeted access consciousness and that phenomenal and access consciousness are distinct" (2009, 35).

We begin with the premise that a-consciousness is a dispositional property. As you will remember from chapter 1, a dispositional property can be defined as the tendency of a thing T to behave in characteristic ways, under certain circumstances. As an example, Kriegel considers fragility (Ibid., 37). Fragility is a dispositional property of certain materials. It is the tendency of a wine glass (Ibid.), for example, to shatter when subject to a force of a certain degree.

But on what grounds should one think of a-consciousness as a dispositional property? Earlier we saw that, in Block's view, a state is a-conscious if its content is available to be used in processes such as report, reasoning, the rational control of action, etc. More generally, a state is a-conscious if its content is available to a number of higher-level cognitive processes/capacities. In this sense, a state's being a-conscious seems to be a matter of the state's being *disposed* (or having the potential) to be accessed by its subject. As Kriegel points out:

> Nothing has to actually happen with a mental state or event for it to qualify as access-conscious: the state or event need not actually be access*ed*; it needs only to be access*ible*. To become access-conscious, a mental state need not be *actually* used in the control of reasoning and action, and so on; it need only be *poised for* such use (Kriegel 2009, 37).

In the same fashion, nothing has to actually happen for a wine glass to be fragile. The glass need not actually break; it only needs to have the *tendency* to break under suitable conditions.

The second premise of the argument says that dispositional properties have *categorical bases* that ground them. As Kriegel explains, "dispositional properties are often taken to require *categorical bases*. A categorical basis is a non-dispositional, occurrent property that accounts for and grounds certain

dispositions" (Ibid.).[63] To appreciate the above passage, let us first briefly consider the question of what a non-dispositional, occurrent property is. As an example, Tim Maudlin compares two properties of a specific window: massiveness and fragility. "Massiveness", he says, "is just a matter of how the window *is*" (2007, 72). Fragility, on the other hand, "is a matter of *how* [the window] *would behave under certain circumstances*" (2007, 72). The latter, as we have seen, is a dispositional property of the window. The former, instead, is a non-dispositional, occurrent property of it; that is, it is a property that concerns how the window is, as opposed to how it would behave were one to throw a stone at it.

Occurrent, non-dispositional properties that account for or ground certain dispositions are known as *categorical bases*. Consider a wine glass again. One of the dispositional properties of wine glasses is fragility. Fragility is the tendency to break under certain circumstances. But *in virtue of what* does a wine glass have such a tendency? What are the non-dispositional, occurrent properties that *ground* such a tendency? The answer lies in the "physico-chemical properties the glass it is made of" (Kriegel 2009, 37). It is "*because,* or *in virtue of its* physico-chemical constitution" that the wine glass is fragile (Ibid.). The physico-chemical constitution of the wine glass is thus the *categorical basis* of its fragility (Ibid.).

But if one accepts the idea that for every dispositional property there is a non-dispositional property that grounds it, this suggests that there should be a property of this kind for a-consciousness as well (Ibid.). This gets us to the next step of Kriegel's argument: p-consciousness *grounds* a-consciousness. There are two main reasons for this claim. First, p-consciousness is an occurrent property. Second, the phenomenological structure of p-consciousness makes it the ideal candidate to ground a-consciousness. What makes it the case that p-consciousness is an occurrent property? The idea seems to be that, unlike a-consciousness, p-

[63] In taking dispositional properties to require categorical bases, Kriegel follows Armstrong (1968, 86).

consciousness is a property that concerns the way mental states *are*, rather than concerning some *tendency* that they may have. In this respect, to say that a mental state is p-conscious is analogous to saying that the window is massive.[64]

As for the second reason, you will recall that, according to Kriegel, the phenomenal character of an experience – i.e., its p-conscious aspect – is analysable into subjective and qualitative character. The former is what makes a mental state p-conscious at all, whereas the latter is what makes a mental state the p-conscious state that it is. You will also recall that Kriegel identifies subjective character with a special kind of awareness: peripheral inner awareness. Now, Kriegel explains:

> [S]ubjective character seems to play the right explanatory role vis-à-vis access consciousness. The reason why a mental state is poised for the subject's free use in personal-level reasoning and action control, it is reasonable to suppose, is that the subject is already [peripherally] aware of it. Once the subject is [peripherally] aware of the state, she can freely make use of it in her deliberate reasoning and action control. Thus the state's free usage to those ends can be *explained* in terms of its subjective character or for-me-ness [...]. It is *because* (or *in virtue of* the fact that) I am [peripherally] aware of my bluish experience of the sky that the experience is poised to be freely used in my reasoning about the consistently nice weather and in guiding my vacation plans (Kriegel 2009, 37–38).[65]

[64] In fact, as Kriegel himself concedes, categorical bases need not always be occurrent properties (2009, 37). In other words, although, normally, categorical bases are non-dispositional, occurrent properties, the categorical bases of some dispositions are dispositions themselves. In light of this, the claim that p-consciousness is an occurrent property is not the strongest of reasons to believe that it is the categorical basis of a-consciousness.

[65] Importantly, although, in some rare cases, subjective character can take the form of inner focal awareness, when Kriegel speaks of subjective character above, he means *peripheral* inner awareness. This he makes explicit when he says that "the subject's *peripheral inner awareness* of her conscious state is the categorical basis of the state's direct availability to verbal report, introspective appraisal, intentional action, and rational deliberation" (Kriegel 2015 n/a, emphasis mine). Understanding Kriegel otherwise would commit him to a very strange position. As the passage above specifies, among the cognitive capacities that a state with subjective character is accessible to is introspective appraisal. But in 4.1 I explained that introspecting involves being focally aware of one's experience. As a consequence, if Kriegel's thesis were understood as suggesting that both focal and peripheral awareness ground a-consciousness, it would follow that focal awareness

A subject's peripheral awareness of her mental states is thus what grounds the states' accessibility. It is because a subject is peripherally aware of her experience that she can access it, and use its content in a number of higher-level cognitive processes/capacities. Together with (1) and (2), (3) entails (4), namely, that p-consciousness is the *categorical basis* of a-consciousness.[66] To use a more precise (and recent) formulation of Kriegel's conclusion, (1), (2), and (3) entail that p-consciousness "is the categorical basis of [an experience's] informational content being able to be processed by such modules/capacities as verbal report, introspective appraisal, intentional action, and rational deliberation without first being processed by other modules/capacities" (Kriegel 2015 n/a).

According to Kriegel, the proposed account has a number of advantages (2009, 40–42). One is that it preserves the metaphysical distinction between p-consciousness and a-consciousness, while also preserving the existence of an intimate relation between the two. This is because the account "casts the two as metaphysically but not nomologically independent" (Ibid., 41). The point is that dispositions supervene on their categorical bases only with nomological necessity. In a world where gravity was much weaker, say, a wine glass would not be as disposed to break as it is in the actual world. In other words, the wine glass would

grounds, among other things, the possibility of focal awareness itself. But this would commit Kriegel to a very strange position. It would be like saying that the reason why a wine glass is fragile is that it is shattered – that the categorical basis of fragility is its *manifestation*. At any rate, this is *not* Kriegel's view.

[66] But how is this supposed to justify the claim that scientists might have been studying p-consciousness via studying a-consciousness? Kriegel notes that, in the history of science, it is quite common "that scientists labor around a dispositional property by way of trying to learn about its categorical basis" (2009, 38). As an example, he cites the study of hereditary properties. For centuries, notes Kriegel, "geneticists have been studying hereditary properties, which are dispositional, by way of trying to reach an understanding of their categorical basis, which we have only recently identified as DNA" (Ibid.). Analogously, it is possible that researchers have been trying to study a-consciousness, which is dispositional, by way of studying its categorical basis, which is p-consciousness.

not be fragile. Such a world is metaphysically possible, but nomologically impossible. In the same fashion, in a world where the psychological laws are different from ours, p-conscious states may well not be available to be used in reasoning, in the rational control of action, etc. Such a world too is metaphysically possible, but is clearly not nomologically so (Ibid.).

4.2. The Functional Role of Phenomenal Consciousness

What about the functional role of p-consciousness? Kriegel thinks that another advantage of (4) is precisely that it accounts for p-consciousness's having a functional role. More strongly, it accounts for p-consciousness's having a functional role *at all*. He observes that, as drawn by Block (1995), the distinction between p-consciousness and a-consciousness has a rather unhappy consequence. This is that "any function we may wish to attribute to phenomenal consciousness would be more appropriately attributed to access consciousness, leaving phenomenal consciousness devoid of any functional significance it can properly call its own" (2009, 40).[67]

Kriegel continues by pointing out that, "[t]he source of this unhappy consequence is the picture of phenomenal and access consciousness as two separate properties sitting side-by-side at the same theoretical level" (Ibid.). I am not entirely sure as to what Kriegel has in mind here. Presumably, by "sitting side-by-side at the same theoretical level", he means that, according the picture offered by Block, neither property is *more fundamental* than the other. By that I mean that neither property can be construed as grounding the other, in the sense explained in §4.1. But even if that is the right way of understanding Kriegel's phrase above, I find it hard to see why a picture of p-consciousness and a-consciousness as the one he attributes to Block would leave the former without any functional significance.

[67] See also Kriegel (2007, 203).

At any rate, our main concern here lies with Kriegel's own proposal about the functional role of p-consciousness. Having established that (4) p-consciousness is the categorical basis of a-consciousness, Kriegel takes this very claim to entail that a-consciousness is part of the functional role of the former. To appreciate what he has in mind, recall what I said in chapter 1. There, I argued for the following definition of the functional role of p-consciousness:

> *Functional Role of P-Consciousness Relative to a Mental State M: the fund of appropriate dispositional properties that p-consciousness endows M with.*

In light of the above definition, if p-consciousness is the categorical basis of a-consciousness, it is only natural to think that the latter is part of the functional role of the former. More slowly: let D be a-consciousness. That is, let D be an appropriate disposition of a mental state M,

i. D of M's is part of the functional role of p-consciousness iff p-consciousness endows M with D;
ii. If p-consciousness is the categorical basis of D, p-consciousness endows M with D;
iii. P-consciousness is the categorical basis of D (as established by Kriegel's argument above); therefore,
iv. P-consciousness endows M with D; therefore,
v. D is part of the functional role of p-consciousness.

What should we think of Kriegel's account? In the remainder of this chapter, I argue that it is untenable.[68]

[68] Note that Kriegel is not the first one to suggest that a-consciousness can be construed as the functional role of p-consciousness. In the commentary to Block's paper (1995), Atkinson and Davies (1995) suggest that p-consciousness "may figure in the causal explanation of [access] consciousness" (1995, 248). Like Kriegel, they begin by noting that a-consciousness is a dispositional property. States that have dispositional properties, they continue, also have "more intrinsic" properties in virtue of which they have the dispositions that they do (Ibid.). Thus, in response to the question "In virtue of what property of my

5. Against Kriegel's Account of the Functional Role of Phenomenal Consciousness

In what follows, I offer two arguments against Kriegel's account of the functional role of p-consciousness (henceforth, 'AF'). Both arguments will challenge AF by challenging Kriegel's account of the relation between p-consciousness and a-consciousness (henceforth, 'AR'). The first argument concludes that AR should be rejected on the ground that it is incompatible with GWM. The second argument argues for the same conclusion, but this time on the ground that AR is committed to a view of the relation between a-consciousness and attention that is empirically implausible.

5.1. The Objection from the Incompatibility with the Global Workspace Model of Access Consciousness

This section argues that AR should be rejected, on the grounds that it is incompatible with the global workspace model (GWM), a well-established model of a-consciousness. In order to show that AR and GWM are incompatible, I suggest we show that the two provide different sets of jointly sufficient conditions for accessibility. Let us start from GWM. What are the necessary and sufficient conditions for cognitive accessibility, according the model? As noted in the previous chapter, GWM explains a mental state's being accessible in terms of its content being globally broadcast in the workspace. But although necessary, global broadcasting does not, by itself, seem sufficient for cognitive accessibility. To see why, imagine a cognitive system whose consuming mechanisms responsible for

---

pain state am I in a position to report that I have pain?" (Ibid.), the answer, Atkinson and Davies suggest, is that my being disposed to report that I have pain "is grounded in my pain's being a phenomenally conscious state" (Ibid.). It is because I am in pain that I can report that I am in pain. Importantly, as the authors suggest, seeing things this way does justice to the "asymmetric dependence relation" (Ibid., 249) that, in Block's view, exists between p-and a-consciousness. If p-consciousness is the categorical basis of a-consciousness, then we will expect that, in those cases where the machinery linking the two is disrupted, there will be p-consciousness without a-consciousness. But because a-consciousness is grounded in p-consciousness, we won't expect cases of a-consciousness without p-consciousness.

reporting, evaluating, deciding, remembering, etc. were damaged or absent. Now suppose that the content of a mental state was globally broadcast in the system's workspace. Would that state count as being accessible? No: the way I understand it, one of GWM's implicit claims is that cognitive accessibility is to be understood as accessibility *to* the consuming systems connected to the workspace. It is those very systems that, once the content of a state is globally broadcast, are in a position to process it.

The occurrence of global broadcasting will thus *not* guarantee accessibility. What is required, in addition, is that the relevant consuming systems be present.[69] More strongly, it is necessary that the consuming systems not be impaired, at least as regards their capacity to take the broadcast content of a representation as input. Call this the 'functionality condition'. As an example, consider the system of reporting. This system can be said to meet the functionality condition if, were the subject asked to report an experience E, the system would be able to take the content of E as input. But why require that the consuming systems be functional, rather than just present? For a present, and yet non-functional, system is as good as an absent system. To say that accessibility is accessibility *to* the consuming systems is to say that, were the content of a mental state to be globally broadcast, the consuming systems would *at least* take its content as input.

The occurrence of global broadcasting and the satisfaction of the functionality condition will thus guarantee the presence of cognitive accessibility.

[69] That the consuming be present is a condition that seems to be required by Block (1995), too. Recall, for example, one of the conceptually possible cases adduced by Block in support of the claim that p-consciousness and a-consciousness are numerically different properties. I am talking about the case of an animal "for which brain damage has destroyed centers of reasoning and rational control of action" (Ibid., 233). In Block's view, on the *assumption* that the neural underpinnings of p-consciousness are distinct from the neural underpinnings of a-consciousness, this would count as a case of the former *without* the latter. It seems clear then, that for Block (1995), too, accessibility is accessibility *to* the consuming systems. This, in turn, suggests that the presence of the consuming systems is a necessary condition for cognitive accessibility.

We can express this in biconditional form as follows: a mental state is access conscious iff (1) its content is globally broadcast in the workspace and (2) the consuming systems are functional.

A natural question now is what AR takes the necessary and sufficient conditions for accessibility to be. Kriegel (2015, n/a) suggests that "if the phenomenality is present, and the relevant machinery is in place, the disposition [accessibility] is guaranteed to be present". How are we to understand this claim? Phenomenality, as we have seen, Kriegel construes as the categorical basis of accessibility. As for the condition that the relevant machinery be in place, there are two things to note. First, by 'relevant machinery', Kriegel means what I have been calling consuming systems/mechanisms. In this sense, the condition that the relevant machinery be in place is akin to what I have called the 'functionality condition'. Second, such a condition is what Kriegel calls an 'enabling condition'. In general, enabling conditions are conditions that, together with a dispositional property's categorical basis (or bas*es*), guarantee/suffice for the presence of the disposition.[70]

For present purposes, I make the (unharmful) assumption that Kriegel would agree with me in requiring not just that the consuming systems be present, but also that they be functional. In light of this and of what I said in the previous paragraph, Kriegel's (2015, n/a) claim that "if the phenomenality is present, and the relevant machinery is in place, the disposition [accessibility] is guaranteed to be present" can thus be rephrased as follows: a mental state is a-conscious iff (1*) it is p-conscious and (2) the consuming systems are functional. We now have the following two biconditionals:

[70] Thus, notes Kriegel (2009, 41), the fact that wine glasses are characterized by a specific physico-chemical constitution does not, by itself, guarantee their *actual* fragility. For fragility to be present, it is also necessary that the *gravitational constant* be what it is: "[i]f gravity was a thousand times weaker [...] the glass would not *be* fragile – in that its disposition to break would be very limited" (Ibid., 41).

> GWM: a mental state is a-conscious iff (1) its content is globally broadcast in the workspace and (2) the consuming systems are functional.
>
> AR: a mental state is a-conscious iff (1*) it is p-conscious and (2) the consuming systems are functional.

While both GWM and AR identify (2) as one of the two necessary conditions for accessibility, they differ in respect to the other condition – (1) vs (1*). According to GWM, it is also necessary that the content of the relevant mental state be globally broadcast; according to AR, instead, it is necessary that the relevant mental state be p-conscious.

Importantly, the difference between (1) and (1*) is substantive – they are not merely two different ways of picking out the same property. Let me explain. According to GWM, for the content of a mental state to be broadcast in the global workspace is, roughly, for it to be "distributed" across the cognitive system, in such a way this content is made available to systems of report, deliberation, etc. According to Kriegel's account of p-consciousness, on the other hand, the latter is (roughly) a matter complex state's two proper parts, P1 and P2, standing into certain suitable relations with each other and with the overall state (see §3 of this chapter).

I should note that the neural details do not matter here. For one might think that in order to establish that the difference between (1) and (1*) is substantive, we should go "further down", and look at what GWM and Kriegel have to say about the neuronal underpinnings of global broadcasting and p-consciousness respectively. As concerns GWM, I have already explained that although Dehaene and colleagues have developed the model in a neural direction, there is a sense in which the neural details are irrelevant to it (and, of course, global broadcasting). To put it with Block, "the substantive claims of the model abstract away from the neuronal details" (2009, 1111).

The way I understand it, something analogous can be said with respect to p-consciousness. Although I have said nothing about it, Kriegel, too, does tell a story about the neural underpinnings of p-consciousness (and hence self-representation).[71] As with GWM, however, it seems that the neural details are irrelevant to the essence of p-consciousness. What *is* of the latter's essence, instead, is that it has an abstract self-representational structure, as specified in §3 above.[72]

I conclude that AR is incompatible with GWM. Together with the assumption that any model of the relation between p-consciousness and a-consciousness must be compatible with GWM, this entails that AR – and hence AF – ought to be rejected.

5.1.1. Objections and Replies

In this section, I present and respond to two objections. (In fact, I will also present a third objection, but will respond to it at the end of §5.2.). The first objection challenges the claim that we should reject AR on the basis of its incompatibility with GWM. The second objection challenges what I called the 'functionality condition'. Let us proceed in order. One might object that the claim that we should reject AR, and thus AF, on the ground that it is incompatible with GWM, is a trivial claim.[73] As I mentioned, in the hands of its theorists (e.g., Baars 1988; Dehaene et al. 1998; Dehaene and Naccache 2001), GWM is not just a model of *a*-consciousness. Instead, it is also a model of *p*-consciousness. Indeed, as I also noted, according to GWM, there is no real difference between the two. Although, conceptually, the

[71] See Kriegel (2009, chap. 7), for his account of the neural underpinnings of p-consciousness. For a concise overview of the same account, see Kriegel (2011, 59).

[72] The reader should note, however, that this is *not* to say that Kriegel endorses a *functionalist* view of p-consciousness, at least insofar as a functionalist picture of p-consciousness is understood as entailing that the latter is a dispositional property of mental states (2009, 209). In this respect, Kriegel's metaphysical commitments about p-consciousness are quite complex, and they need not concern us here.

[73] I would like to thank an anonymous referee for *Philosophical Psychology* for suggesting this objection.

distinction between p-consciousness and a-consciousness might even make sense, in practice, there is only one empirically tractable phenomenon, and this is what we typically refer to by the term 'a-consciousness'.

If one assumes the truth of GWM, the objection goes, then one will have to endorse the model both as a model of a-consciousness and as a model of p-consciousness. As a consequence, my claim that we should reject AR (and thus AF) on the basis that it is incompatible with GWM will follow quite trivially. For if I assume that GWM is true, then I will also have to be committed to the claim that there is no real distinction between a-consciousness and p-consciousness. And if I am committed to that claim, then it is obvious that I should reject AR, for this assumes that there is a distinction between the two.

But the objection is easily rebutted. Dehaene and Naccache have provided a model that explains our capacity to access our mental states. According to this model, such a capacity hinges on global broadcasting. To claim that such a model is also a model of p-consciousness is a further step. That is, to claim that p-consciousness, too, hinges on global broadcasting is a further step, one that is not entailed by the claim that a-consciousness hinges on global broadcasting. One can thus accept GWM as a model of a-consciousness without having to accept it as a model of p-consciousness. Indeed, as I noted in the previous chapter, this is exactly what some have done (e.g., Block 2007; Carruthers 2015).

A second, minor objection has to do with what I called the 'functionality condition'.[74] As you will recall, in §6 I noted that, according to GWM, the occurrence of global broadcasting is not a sufficient condition for cognitive accessibility. On the contrary, in addition to global broadcasting, it is also necessary that the consuming systems – systems of report, deliberation, etc. – be functional. What do I mean by 'functional'? As explained, to say that the consuming

[74] I would like to Stephen Stich for this objection.

systems are functional is to say that they aıre not impaired, at least as regards their capacity to take the content of a globally broadcast representation as input.

But suppose you are under the effect of some drug that impairs *some* of the consuming systems connected to the workspace – systems for the rational control of action, say. Would a globally broadcast representation still count as accessible? Or would it count as accessible, but only "in part"? And what would happen if you took a drug that impaired *all* the consuming systems *except one?* These are all good questions – questions that surely deserve careful consideration. For present purposes, however, these questions are relevant insofar as they reveal how the functionality condition and the motivations behind it – namely, that accessibility is accessibility *to* the consuming systems – are not at all uncontroversial.[75]

Fortunately, however, the unsolid ground upon which the functionality condition rests is not a real threat to the claim that AR and GWM are incompatible. For even though I have argued for that claim by showing that the two provide different sets of jointly sufficient conditions for accessibility, this is not the only argumentative strategy available. To be sure, such a strategy has the advantage of providing a clearer picture of how AR and GWM relate with one another. Strictly speaking, however, one could make do by noting that AR and GWM identify divergent necessary (but not sufficient) conditions for accessibility. That is, one could make do by noting that, while AR explains cognitive accessibility (partly) in terms of p-consciousness (and hence self-representation), GWM explains the same property (partly) in terms of global broadcasting. Accordingly, even if the

[75] In fact, one might even push further and question the very idea that the functionality of the consuming systems – or even their very presence – is a necessary condition for cognitive accessibility in the first place. Chalmers (1997) seems to do just that. In passing, he points out that even when the consuming systems "are mostly shut down", cognitive accessibility is still present (Ibid., 148). He does not, however, offer any justification for such a claim.

functionality condition turned out to be ungrounded, the claim that AR is incompatible with GWM would still be defendable.

As anticipated, there is also a third objection that I should consider. I have argued that GWM and AR are incompatible on the basis that they provide different sets of jointly sufficient conditions for a-consciousness. But one might object that that won't suffice to establish their incompatibility. To do that, the objection goes, I also need to consider a possible scenario in which Kriegel's conditions for a-consciousness are satisfied, but global broadcasting does not occur. For reasons of exposition, I will set this objection aside for now, and address it at the very end of the next section.

5.2. The Objection from Attention

This section argues against AF, on the ground that AR gets the relationship between a-consciousness and attention wrong. In the previous chapter, we saw that there is convincing evidence for the thesis that a-consciousness depends on attention. As I will argue, however, AR entails that a representation can be a-conscious even in the (almost complete) absence of attention. I should note that even though espousing GWM entails endorsing the claim that a-consciousness depends on attention – the latter being one of GWM's core claims – endorsing the latter claim does not entail espousing GWM. One can buy into the claim that attention is necessary for a-consciousness without having to endorse the claim that a-consciousness hinges on global broadcasting. For this reason, the objection I raise in this section will hold regardless of one's commitment to GWM.

Let us begin by considering the following scenario from Block, which we already encountered in the previous chapter:

> Suppose you are engaged in intense conversation when suddenly at noon you realize that right outside your window there is–and there has been for

> some time–a deafening pneumatic drill digging up the street. You were aware of the noise all along, but only at noon you are *consciously* aware of it. That is, you were P-conscious of the noise all along, but at noon you are both P-conscious *and* A-conscious of it (Block 1995, 234).

An implicit assumption in the above passage is that a-consciousness – but not p-consciousness – depends on attention (Block 2001, 203). While Block is engaged in intense conversation, all or almost all his attentional resources are focused on the conversation. This suggests that very few attentional resources, if any, are committed to the sound of the pneumatic drill. But if a-consciousness depends on attention, then no auditory representation of the pneumatic drill can be (and is) a-conscious. Still, as Block points out, the same auditory representation was p-conscious all along.

According to Kriegel, too, Block's experience of the pneumatic drill was p-conscious the whole time. As he notes, "while engrossed in conversation, Block has a very peripheral awareness of his auditory experience" (2009, 191). But now note that because of AR, Kriegel will also have to say that Block's auditory experience of the pneumatic drill was a-conscious all along as well. That is, the contents of Block's auditory experience were available to systems of report, the rational control of action, etc., even before noon.[76]

A first – if, admittedly, weak – point to be made about Kriegel's view is that, phenomenologically, it seems very implausible. Scenarios like the one described in Block's passage above are very familiar. I am engaged in doing something, when, all of a sudden, I realize that the rain is – and has been for some time – hitting a nearby window, or that the wind turbine just outside is – and has been for some time – making a lot of noise. But I find it hard to make sense of the idea that the contents of my auditory experience were directly available for me to report all

[76] The latter claim is also endorsed by (Chalmers 1997).

along, *even in the absence of attention*. They would become available, instead, were I to turn my attention to them (or, against the backdrop of Kriegel's account of a-consciousness, were I to turn my attention *to my experience of them*). Observations made on the basis of purely phenomenological intuitions should be trusted only to an extent, however. It is quite fortunate, then, that my point also fits nicely with the empirical data.

Consider the experiment by Simons and Chabris (1999) that we encountered in the previous chapter. Participants were instructed to watch a video where two teams, one in black, and one in white, are passing basketballs. They were also instructed to pay attention to either one team or the other, and to keep count of the basketball passes made by the attended team. At some point, a subject wearing a gorilla costume walked towards the middle of the scene, stopped, thumped its chest, and then continued walking across the scene. Surprisingly, when asked whether they had noticed the gorilla, a significant percentage of participants gave a negative response.

What would Kriegel's take on the experiment be? My best guess is that, given his interpretation of Block's passage above, he would say that the participants' experience of the gorilla was p-conscious. More precisely, as in the case of the pneumatic drill, he would claim that participants had a very peripheral awareness of their experience. According to AR, this entails that their experience of the gorilla was also a-conscious. But if, as I argued earlier, experiments like Simons and Chabris's experiments, along with conditions like hemineglect, lend support to the thesis that a-consciousness is strongly dependent on attention, then Kriegel is wrong. I conclude that AR gets the relation between attention and a-consciousness wrong. For this reason, AR, as well as AF, should be rejected.

But maybe my move was too quick. For I have been tacitly assuming that for a subject to be peripherally aware of her experience, she need not attend to it *at all*. According to Kriegel, however, even peripheral awareness requires attention. In his

own words, "it is impossible to be aware of something when 0% of one's attentional resources are dedicated to it. Awareness requires some minimal attention, though not focal attention" (2009, 257). In the present context, what this means is that, on Kriegel's account, it is not entirely true that Block was not attending to his experience of the pneumatic drill before noon. If he was peripherally aware of it, as Kriegel contends, then he was also attending to it, albeit only peripherally.

Does this mean that Kriegel is right to claim that Block's experience of the pneumatic drill was a-conscious all along? I do not think so. For consider Simons and Chabris's experiments again. On the assumption that participants were peripherally aware of, and hence peripherally attending to, their experience of the gorilla – as Kriegel would probably contend – their inability to report seeing it strongly suggests that peripherally attending to one's experience is not sufficient for a-consciousness. In this respect, then, it seems that peripheral attention is no better than the complete absence of attention.[77]

The considerations above bring to light an issue which I mention only to set aside: how do cognitive scientists think of attention? Do they think of it as an all or nothing phenomenon, or do they think it admits of degrees? I suspect it will be hard to find a *single* answer to these questions. At any rate, in connection to the Simons and Chabris's experiments, it is apparent that if, like Kriegel, we think of attention as admitting of degrees – or as admitting of a focus and a periphery – then a-consciousness requires focal attention.

One last point. In §5.1.1., I presented an objection to the argument from the incompatibility with GWM that, for reasons of exposition, I said I would only

[77] It should be noted that inferences from hemineglect can no longer help us counter Kriegel's claim that Block's experience of the pneumatic drill was a-conscious all along. For while Kriegel's take on Block's scenario strongly suggests that he would take participants in the Simons and Chabris experiments to be peripherally aware of their experience of the gorilla, it is hard to tell what Kriegel would say about hemineglect patients; i.e., would they be peripherally aware of their experience of what goes on in the neglected side of space?

address at the very end of this section. As you will recall, the objection said that, in order to establish that AR and GWM are really incompatible, I would also have to consider a possible scenario in which Kriegel's conditions for a-consciousness are satisfied, but global broadcasting does not occur. Since I find that to be a valid point, here is one such scenario.

You will recall that, according to KA, Block's experience of the pneumatic drill is p-conscious both before and after noon. According to GWM, however, the contents of the same experience are globally broadcast only at noon – that is, only once Block turns his attention to the noise. Even though, for the purposes of the present section, I have only been concerned with the claim that attention is necessary for a-consciousness, the reader will remember that GWM takes global broadcasting to depend on attention. We thus have a scenario in which p-consciousness is present but global broadcasting is not.

6. Concluding Remarks

According to Kriegel, if p-consciousness is the categorical basis of a-consciousness, then the former can be construed as the functional role of the latter. This chapter has rejected Kriegel's account of the functional role of p-consciousness by arguing against the former claim. In particular, I have put forward two related and yet independent arguments. First, I have brought to light KA's incompatibility with GWM. On the assumption that any account of the relation between p-consciousness and a-consciousness must be compatible with GWM, this entails that KA must be jettisoned. Second, I have argued that KA gets the relationship between a-consciousness and attention wrong. According to KA, a mental state can be a-conscious even in the (almost complete) absence of attention. But this flies directly in the face of the point made in the previous chapter, namely, that a-consciousness is strongly dependent on attention. This gives us an additional reason to reject KA, one that can be shared independently of one's commitment to GWM.

4

Closing in on the Relation Between Phenomenal Consciousness and Access Consciousness

1. Introduction

In the previous chapter, I introduced and argued against Kriegel's account of the functional role of p-consciousness. I suggested that the account is untenable for two reasons. First, it is incompatible with a well-established model of a-consciousness: the Global Workspace Model (GWM). Second, it gets the relation between a-consciousness and attention wrong. In search of a more promising account, in chapter 5, I will attempt to make sense of, and evaluate, Block's own view on the functional role of p-consciousness. To that end, however, I first need to say more on Block's positive characterization of p-consciousness, and introduce his most recent account of the relationship between p-consciousness and a-consciousness (e.g., 2007a; 2008). Doing so will also allow me to defend Block's account of the relationship between the two from a recent criticism advanced by Carruthers (2015b). Carruthers's criticism threatens to show that Block's view suffers from serious problems that would evidently make it unacceptable regardless of its potential as an account of the functional role of p-consciousness. My response to Carruthers will, by removing these potentially devastating obstacles to Block's view, allow me to focus precisely upon its relevance for this thesis's main concern; i.e., the question of what the functional role of p-consciousness is.

As explained in chapter 2, in his (1995), Block argues that p-consciousness and a-consciousness are numerically different properties on the basis that it is metaphysically possible for either to occur without the other. That, I also explained, is compatible with the two properties' being coextensive in the actual world. Block (2007a, 2007c; 2008; 2011) goes one step further, and offers his Overflow-Mesh Argument to make a much stronger claim: the neural basis of a-consciousness is *not included* in the neural basis of p-consciousness. One consequence of this is that, in this world, a mental state can be p-conscious without being a-conscious.[78] Carruthers (2015b) is happy to accept Block's distinction as drawn in the latter's (1995). At the same time, he argues that the conclusion of the Overflow-Mesh Argument is untenable: the neural basis of p-consciousness is *identical to* that of a-consciousness. Having reconstructed Block's and Carruthers's arguments, in this chapter I show that the conclusion of the Overflow-Mesh Argument can still be savaged. The reason, as we will see, is that Carruthers's argument against the latter rests on a view of attention that, in light of certain empirical results, is highly implausible.

Here is the plan. To begin with (in §2) I provide the reader with a general idea of Block's approach to p-consciousness, which, as we will see, can labelled as a *biological* approach. Second (in §3) I say something about why, according to GWM, attention is not only necessary, but, in conjunction with other conditions, also *sufficient* for global broadcasting. Third (in §4) I provide a detailed picture of the notion of *working memory*. As we will see, the latter plays a central role in Block's Overflow-Mesh Argument. Having introduced the notion of working memory, I then reconstruct (in §5) the first step of Block's argument, which we may just call the *Overflow Argument*. After that, (in §6) I turn to the second step of

[78] As opposed to the pneumatic drill scenario discussed earlier, which was meant to be an *intuitive* case of p-consciousness's *actually* occurring without a-consciousness, the Overflow-Mesh argument is Block's first, systematic attempt to establish that, in this world, the former can (and does) occur without the latter.

the Overflow-Mesh Argument, namely, the *Mesh Argument*. The reconstruction of the argument itself will be preceded by a discussion of the notion of *neural basis*, and of the neural and functional events surrounding global broadcasting. In the last part of the chapter, I present Carruthers's criticism against the Overflow-Mesh argument (in §7), and defend the latter's conclusion (in §8).

2. A First Look at the Biological Approach to Phenomenal Consciousness

Just what kind of property does Block take p-consciousness to be? We have seen what Block thinks p-consciousness is *not:* it is neither a functional nor a representational property. It is not a functional property in that what makes a mental state p-conscious is not the state's functional role. And it is not a representational property in that it is not identical to a mental state's representing the world as being a certain way. But what kind of property *is* it, then?

In many of his writings (e.g., 2001, 2007a, 2009)[79] Block has been leaning toward a *biological approach* to p-consciousness: p-consciousness is a *biological* property of the brain. In order to obtain a first, general idea of what that means, it will be useful considering a couple of theories that do *not* characterize p-consciousness in the same way. Theories like GWM, along with theories like Rosenthal's higher-order representationalism (HOT), for example, are sometimes called by Block *cognitive* theories of p-consciousness (2009, 2015). What these theories have in common is precisely that they take p-consciousness to be a *cognitive* phenomenon. (Importantly, for the purposes of this section only, I understand GWM as a model of *both* a-consciousness *and p*-consciousness. As you will recall from chapter 2, this is how its original advocates construe the model.)

Let's begin by considering GWM. According to the model, p-consciousness *just is* cognitive accessibility. A mental state's being p-conscious is just a matter of its content's being made available to a number of *cognitive* systems – systems of

[79] But see also Block (2007c, 2015).

thought, memory, reasoning, etc. As such, GWM counts as a *cognitive* account of p-consciousness. Consider HOT, now. As hinted at in the previous chapter, according to HOT, a mental state's being p-conscious is to be explained in terms of a subject's harbouring two suitably-related representations: a lower-order and a higher-order representation. The lower-order representation – the experience – represents the world as being a certain way. The higher-order, *thought-like* representation, instead, represents the lower-order one. According to HOT, it is only when an experience figures in the representational content of (is represented by) a higher-order thought, that it is p-conscious.

Qua cognitive theories, GWM and HOT account for p-consciousness at a level of explanation that abstracts away from the biology of the brain. The idea is that even though, in beings like us, p-consciousness is implemented in a certain biological system – the brain – in theory, it could also be implemented in a functionally isomorphic *non*-biological system – a super-computer, say. P-consciousness, accordingly, is not to be explained by appealing to biological properties of the brain – properties that are *exclusive to* neurons, or to the ways in which neurons interact with one another, for example.

Mental representations are often thought to be implementable in different systems with different properties – e.g., computers with electrical properties, and robots with electrical and hydraulic properties. Accordingly, understanding why, on the HOT account, p-consciousness does not depend on the biological makeup of the brain should be quite straightforward. Things, however, might not appear as straightforward in connection to GWM. As noted, this account, especially in the hands of Dehaene and his colleagues, makes explicit reference to the brain and its structure. The global workspace itself, for example, is thought to "consist" of a network of workspace neurons whose long axonal projections make it possible to transmit information from, say, the visual cortex, to those areas of the brain underpinning higher-level cognitive processes/capacities (report, deliberation,

etc.). How could the biology of the brain be irrelevant to p-consciousness, then? As hinted at earlier on, the answer is that, in spite of being couched in neuronal terms, a global workspace architecture could be implemented in many ways. According to GWM, what is essential to p-consciousness is just that a representation enters into certain suitable causal relations with the relevant components of the cognitive systems. Whether the representation, the causal relations it enters into, etc. are realized in a brain, rather than in a hydraulic machine, is completely irrelevant to p-consciousness. As long as the relevant functional relations are in place, p-consciousness is guaranteed to be present. Accordingly, GWM is as functional an account of p-consciousness as any other.

If cognitive theories abstract away from the biology of the brain, the same does not hold for the biological approach to p-consciousness. To put it with Block, cognitive theories

> leave out too many details of the actual working of the brain to be adequate theories of [phenomenal] consciousness. Information in the brain is coded electrically, then transformed to a chemical code, then back to an electrical code, and it would be foolish to assume that this transformation from one form to another is irrelevant to the physical basis of [phenomenal] consciousness (Block 2009, 1113).

According to cognitive theories, the relationship between p-consciousness and the biology of the brain is contingent. Although p-consciousness happens to sit in brains with a specific biological makeup, nothing about it requires *that* biological makeup, or a biological makeup *at all*. The biological approach, instead, holds that the biology of the brain is *essential* to p-consciousness. By purporting to account for p-consciousness in biological terms, the biological approach rejects the characterization of consciousness as a cognitive phenomenon. Cognitive theories

may be successful at explaining cognitive accessibility and surrounding phenomena, but not p-consciousness.

I will return to Block's biological approach to p-consciousness below, when discussing the Mesh Argument. In the next section, I come back to the relation between attention and global broadcasting.

3. Attention and Global Broadcasting Again

In chapter 2, I explained that, according to GWM, attention is *necessary* for global broadcasting. For the purposes of this and the next chapter, however, I now need to address the question of whether it is also thought to be *sufficient*. The advocates of GWM, as well as many philosophers/cognitive scientists (Prinz 2013; Carruthers 2015a) agree that, strictly speaking, it is not. As Carruthers (2015a, 57–58) notes, the studies typically cited in support of this claim (e.g., McCormick 1997; Bressan and Pizzighello 2008) purport to show that, under experimental conditions, attention can be directed to a stimulus without resulting in the subject's ability to use the relevant information for report, the rational control of action, etc.

In the studies at issue, however, attention's failure to trigger global broadcasting seems to result from one or a combination of the following factors: a) the presented stimulus is too weak; b) the stimulus is presented for an extremely short period of time; c) insufficient attentional resources are deployed. On the contrary, when none of the above factors are in play, attention regularly results in global broadcasting (Carruthers 2015a, 57–58). For this reason, Carruthers (Ibid.), as well as Dehaene and Naccache (2001), for example, conclude that, to put it with the former's own words, "attention is sufficient for global broadcasting in cases where *enough* attention is devoted to a stimulus that is *sufficiently intense and long-lasting*" (Ibid., 58).

I now turn to the notion of working memory.

4. Working Memory

4.1. A Temporary Storage Capacity

As anticipated, a notion that plays a key role in Block's Overflow-Mesh Argument is the one of *working memory*. What is working memory? This is not an easy question, as theorists use the term in slightly different ways (Cowan 2005b). Two minimal definitions that should be accepted by virtually all researchers, including Block, are Baddeley's and Cowan's, two of the most prominent scholars working on working memory. The former characterizes working memory as "a temporary storage system [...] that underpins our capacity for complex thought" (Baddeley 2007, 1). The latter, instead, defines it as "the retention of information in a temporarily accessible form" (Cowan 2005b, 155). The gist of both definitions is roughly the same: working memory is a temporary storage capacity – hence, 'memory' – whose contents can be easily accessed by the subject and used in complex cognitive tasks – hence, 'working'.

Below are some examples of the many activities that are thought to require working memory. Consider adding 44 to 37 in your head. Among other things, this requires that one hold 11 (7+4) in mind, while adding 4 to 3, and so on. Given the general characterization of working memory provided above, mental arithmetic requires working memory in that it requires holding in mind partial results, retrieving those results at a later stage, combining them with other partial results, etc. Another example would be keeping a long alphanumerical password in your head while writing it down on a piece of paper. Still, other examples would include figuring out the best route to go from the philosophy department to the best restaurant in town, or imagining the different ways in which a set of boxes might be organized into the trunk of a car (Carruthers 2015a, 12). All the activities above are believed to require working memory in that they involve keeping in mind certain pieces of information while working with them.

What are the cognitive mechanisms involved in working memory? Block does not say much about this. In more than one paper (2007c; 2008; 2007a), however, he notes that there is an *intimate relation* between working memory and the global workspace. The idea seems to be that the global broadcasting of a representation is somehow necessary for that representation being in working memory. In this section, I explain how working memory is thought to work, and how it is thought to be related to the workspace. In doing so, I will mostly – although not exclusively – follow the work of Peter Carruthers (2011, 2013, 2015a), who, unlike Ned Block, provides a very detailed account of how working memory works.

4.2. Working Memory and Long-Term Memory

Let me begin by pointing out that everyone agrees that working memory should be distinguished from long-term memory (Cowan 2008; Carruthers 2015a). First, the latter should be thought of as consisting of "stored representations that are no longer in a [heightened state of accessibility]" (Carruthers 2015, 12; but see also Shelton et al. 2008). In this respect, long-term memory differs from working memory in that the contents of the latter *are* in a heightened state of accessibility (what this means, exactly, will become clearer below).

Second, working memory and long-term memory differ in terms of capacity. While the former has a vast storage capacity, the storage capacity of the latter is thought to be very limited (Cowan 2008). In relation to working memory, the phrase 'storage capacity' is to be understood as "the amount that an individual can hold in mind at one time" (Cowan 2005a, 4). In general, psychologists believe that working memory capacity is limited to three or four "items" (Cowan 2008). The notion of item is slightly ambiguous, however. For it seems that an item may be either a separate entity, such as a letter, or a "chunk" (Ibid.). Intuitively, chunks can be said to correspond to larger, meaningful units. These might be "syllables, words, etc." (Cowan 1988, 166). Consider the string of letters IRSCIAFBI (Cowan

2008, 326). This can be chunked into three meaningful units: IRS, CIA, and FBI. Indeed, these are three acronyms for federal agencies in the United States (IRS stands for Internal Revenue Service, CIA for Central Intelligence Agency, and FBI for Federal Bureau of Investigation) (Ibid.). To say that working memory can hold up to three or four chunks of information is thus to say that it can hold up to three or four meaningful units like the ones above.[80]

A third aspect in relation to which working memory and long-term memory are thought to differ is what Carruthers (2013, 10372) calls *sensitivity to attentional interference*. Unlike long-term memory, the contents of working memory are quite feeble. That is to say that whatever the subject is holding in mind at any one time will be lost the moment she turns her attention fully to something else (Ibid.). This can be easily seen in those cases where one is trying to hold a phone number in mind, for example. Any distraction, even the smallest, has the potential to make the number (or part of it) slip out of one's mind. This connects to a fourth, fundamental difference between the two forms of memory at issue. Unlike long-term memory, the contents of working memory need to be *actively sustained* by the subject. In this sense, working memory is an *active* memory capacity. A good example, again, is holding a phone number in mind. This requires that one repeatedly rehearses the number in one's head, or else it will be lost.

4.3. Working Memory and the Global Workspace

There are several models of working memory. The most influential are Baddeley and Hitch's (1974) and Cowan's (1988, 2005a). Although the details are different, both models agree on a number of essential points. A picture of working memory that focuses on the shared tenets of these two models, while abstracting away from

[80] Chunking – the process of grouping together more items into larger, meaningful units – is a strategy that every one of us employs in order to "increase the efficiency of use" of the limited working memory capacity (Cowan 2008, 329) It should be borne in mind, however, that chunking takes time. For this reason, when subjects are required to hold in mind briefly presented perceptual items, this strategy is not available.

their particular differences, has recently been offered by Carruthers (2013, 2015a, 2011). In fact, Carruthers goes even further than Baddeley and Cowan: as noted, in line with Block, he contends that working memory is *intimately related to the global workspace*. Carruthers's picture of working memory is the one that I assume in this chapter. [81]

Working memory can be construed as a multi-component system that comprises the following three ingredients: a central executive, the sensory areas of the brain, and the global workspace.[82] The central executive is thought to be a system under voluntary control (Cowan 1988), tasked with the control of selective attention (Carruthers 2013). Selective attention, in turn, is thought to work by targeting representations in the sensory areas of the brain. The latter are the areas where sensory representations are processed (Postle 2006; Carruthers 2015a), but also where long-term memories are stored (Mayes and Roberts 2001; Carruthers 2015a). By targeting representations in the sensory areas, attention allows the former entry to, and sustains them in, the global workspace. There, those representations can be manipulated for purposes of reasoning, report, action-planning, etc.

Here is an instance of a perceptual representation's being sustained in working memory: you are presented with a row of letters, and are instructed to hold the letters in mind *after* stimulus offset. According to the picture offered by Carruthers, what happens is this. Upon presentation of the row of letters, attentional signals will target the areas of the brain where the letters are represented. In turn, this will result in the global broadcasting of the relevant information. In order for that information to gain entry into working memory,

[81] The reader should note that Carruthers draws extensively on Dehaene and colleagues' work (e.g., Dehaene and Naccache 2001; Dehaene et al. 2006). More specifically, Carruthers's understanding of the relation between working memory and the global workspace is almost identical to Dehaene and colleagues'.

[82] Carruthers does not put things exactly in these terms, but I believe he would agree with me on this.

however, it will have to be kept *in* the global workspace. To do this, it is necessary that attentional signals keep targeting the sensory areas even *after* stimulus offset. Having gained entry in working memory, information will be kept there on condition that attentional signals keep targeting the same areas.

But why would one want to think of the global workspace as a component of the working memory system? As should be apparent, the main reason has to do with the very notion of working memory (2011, 2015a). As we have seen, it is widely accepted that working memory is to be understood as the retention of information in an *accessible* state. But if accessibility is construed – in line with GWM – as a property of representations whose contents are broadcast in the workspace, then the contents of working memory representations will be globally broadcast contents themselves. This suggests that the global workspace is an integral part of the working memory system.

It is noteworthy that while all working memory representations are always cognitively accessible, typically, not all cognitively accessible representations are working memory representations. In other words, normally (although by no means always), the contents of working memory representations are only a *proper subset* of the contents of the global workspace (Carruthers 2015a, 83–84). Notice that the asymmetry in question is diachronic; that is, the amount of information that is cognitively accessible over a certain period of time may far exceed the amount of information that is in working memory over the same amount of time (Ibid., 84). The reason for this asymmetry lies in the fact that, as mentioned earlier, working memory representations are only those representations that are *actively sustained* by the subject. And, we may assume, only *some* of the broadcast representations are so sustained. "Put differently, since working memory is a form of actively sustained memory, much of what is [cognitively accessible] may never enter working memory because no steps are taken to keep it actively in mind" (Ibid., 83).

Information that is globally broadcast information *can* be held in working memory, "but much may not be" (Ibid.).

With the concept of working memory in hand, we can now turn to the first step of Block's Overflow-Mesh Argument, namely, the Overflow Argument.

5. Block's Overflow Argument

The Overflow Argument (Block 2007a, 2007c, 2011) purports to establish that more information can be (and is) p-conscious than is in working memory and, on that basis, that "the machinery of phenomenology is at least somewhat different from the machinery of cognitive accessibility" (Block 2007a, 489). The argument is strongly dependent upon Sperling's partial report paradigm (Sperling 1960).[83] For this reason, I begin with a few words on the latter.

5.1. Sperling's Partial Report Paradigm

Sperling ran a number of experiments to test how much information can be processed by the human visual system in a single brief exposure; that is, basically, how much one can *see* at a glance. In one experiment – call it the *whole report paradigm* – subjects were presented with arrays of letters for a very brief period – 50 msec. (Sperling 1960, 6). The arrays comprised a varying number of letters. Some consisted of 2 rows of 3 letters, while others consisted of 3 rows of 4 letters, for example. After stimulus offset, subjects were required to give a complete report of the letters. Although subjects believed that they had seen *all* the letters with *great clarity*, they were only able to report 4.3 letters on average. Sperling took this result as evidence for the claim that the "immediate memory span" – what we now call *working memory* – is limited to 4.3 letters (Ibid., 5).

[83] To make his case, Block also appeals to a couple of more recent experiments from Landman et al. (2003) and Sligte et al. (2008). Both experiments, however, are quite similar to Sperling's. In what follows, I will focus my discussion exclusively on the latter.

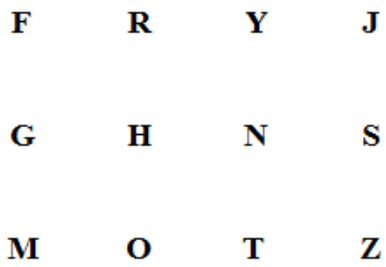

Figure 2. Example of array of letters used in Sperling's experiments.

Given the subjects' belief that they had seen much more than they could report, Sperling hypothesized that "the immediate memory-span sets a limit on a process that is otherwise rich in available information" (Ibid., 26). In other words, he suggested that although, due to working memory limitations, subjects could only report 4.3 letters, they had probably chosen those letters from an otherwise larger store of letters which had been briefly available to them.

To test his hypothesis, Sperling had to find a way to circumvent the limit imposed by working memory. That is, he had to find a way to test whether subjects had actually *seen* much more than they could report – a way that would be unaffected by the limited capacity of working memory. He thus devised an ingenious experiment where subjects were required to give only a *partial* report of the contents of the stimulus: t*he partial report paradigm*. In the experiment, subjects were presented with arrays of letters like the ones used in the previous experiment. 150 msec. after stimulus offset, a randomly chosen row was cued by sounding a tone lasting about 0.5 sec.: a high tone for the upper row, a medium tone for the middle row, and a low tone for the lower row. Depending on which tone was sounded, subjects had to report the letters from the relevant row.

Sperling found that, on each trial, subjects were able to accurately report almost all of the letters from the cued row (typically 3 out of 4 letters). He took this to confirm his initial hypothesis: subjects had actually seen much more than they

were able to report. The thought was roughly the following. If subjects were able to accurately report the letters of *any* of the rows, after this had been cued, then the amount of information available to them during the cueing was much larger than they could report. Differently put, if subjects were able to report the letters of any of the rows, after cueing, then the cueing acted as a lead for the subjects to select some letters (the letters of the cued row) from a larger store of letters which was shortly available to them. Sperling calculated that this larger store could contain an average of 9.1 letters – many more letters than subjects could actually report.

That subjects can see much more than they can report is one of the two major findings of Sperling's experiment. Another finding has to do with the persistence of the visual image after stimulus offset. As noted above, in the partial report paradigm, a tone was sounded 150 msec. *after* stimulus offset. The fact that subjects were able to accurately report almost all of the letters from the cued row suggests that what they had seen during the 50-msec. exposure, they continued seeing for a very short time right after stimulus offset – just enough time for the cue to attract their attention on the relevant row. In Sperling's own words, this suggests that "the stimulus information is thus "stored" for a fraction of a second as a persisting image of the objective stimulus" (Ibid., 20).[84] This "persisting image" is what Ulric Neisser later called *iconic memory* (1967, 20). As Neisser suggests, from a phenomenological perspective this persisting image is *perceptual.* That is, from the subjects' point of view, it appears no different than the visual image they had while the stimulus was being shown. Iconic memory, as Sperling (1960, 11) calculated, lasts about a second. This was clear from the fact that, if the instruction

[84] As Sperling suggested, there was also further evidence for this claim. There was, for example, phenomenological evidence. Apparently, the subjects of the experiment reported that an image of the display was still available the moment a tone was heard, namely, 150 msec. after stimulus offset. In addition, as he noted, it is very implausible to suppose that the subjects should stop seeing the stimulus the moment the display goes blank. As he writes, "the rise and fall of sensation may be rapid, but they are not instantaneous" (1960, 20).

to report a certain row was delayed until one second after stimulus offset, the accuracy of the reports drops down dramatically.

5.2. The Argument

We are now ready to introduce Block's Overflow Argument. This can be reconstructed as follows:

(1) Subjects have a p-conscious experience of/see all or almost all the characters in the array (below, I follow Block in using 'see' in the sense of p-conscious seeing);

(2) Subjects report much fewer characters than they are p-conscious of/see;

(3) Report of the characters in the array is underpinned by working memory; therefore,

(4) More information is p-conscious than is in working memory; therefore,

(5) More information is p-conscious than is cognitively accessible; therefore,

(6) "the machinery of phenomenology is somewhat different from the machinery of cognitive accessibility" (Block 2007a, 489).

Start with premise (1). In both the whole and partial report paradigm, subjects claim to see all or almost all the characters in the array very clearly. Differently put, they report a *rich* and *detailed p-conscious* experience of the array. As noted, the partial report paradigm appears to confirm this. When, right after stimulus offset, a tone corresponding to one of the rows is sounded, subjects accurately report the letters of the relevant row. From this, it is inferred that subjects really did, as they claim, have a rich and detailed p-conscious experience of the entire array.[85]

[85] One might wonder whether the view that one can have a p-conscious experience *of* the array (as opposed to an intransitive p-conscious experience) does not entail a view on which the p-conscious properties of a state are identical to the state's representational properties (e.g., Jacob 2007). Block (2007c) thinks that it does not. Indeed, as I explained in chapter 2, while he rejects the equation of p-conscious properties with representational

Granted the truth of premise (1), premise (2) seems quite uncontroversial. When, in the *whole report paradigm*, subjects are required to report the letters that they saw, they typically only report an average of 4.3 letters. But considering that, according to Sperling's calculations, subjects could actually *see* an average of 9.1 letters, it follows that the number of letters that subjects are able to report is much smaller than the number of letters they could (and did) actually see.[86]

As for premise (3), there is not much to say about this that has not been said before. The thought is just that, in order to report the letters that they saw, subjects need to actively sustain a representation of the letters in working memory. Report takes a considerable amount of time, so it is necessary that a representation of the letters be actively sustained in working memory, for at least as long as it takes to make a report. But, the thought goes, if working memory underpins report of the letters in the array, then the number of letters that subjects report will be identical to the number of letters represented in working memory. This number, as we have seen, Sperling calculated to be about 4.3. But, given that, upon stimulus presentation, subjects enjoy a p-conscious experience of about 9.1 letters, it follows that (4), more information is p-conscious than is in working memory. Now, on the assumption that all working memory information is cognitively accessible – as noted in §4.3 – Block can conclude that (5), more information is p-conscious than is cognitively accessible; that is, p-consciousness *overflows* cognitive accessibility.[87]

---

properties, he also thinks that p-conscious properties represent/have content. In virtue of being able to represent, p-consciousness can then also be *p-consciousness of*.

[86] One might wonder why Block's argument refers to the number of letters that subjects report in the *whole report paradigm*, rather than in the partial report paradigm. The answer is that, in the partial report paradigm, subjects are only required to report the letters of the cued row. In this sense, if we want to know how many letters subjects can report in relation to the entire array, then we need to refer to the whole report paradigm.

[87] In his (2011), Block points out that the Overflow Argument does not show that any items are cognitively inaccessible, but only that most items are unaccessed. *Prima facie*, this suggests that he has changed his mind. In fact, I believe he has not. For everyone accepts that most items are unaccessed in the sense that they are not reported – this is an undisputable empirical claim. What some of Block's opponents (e.g., Carruthers, 2015) challenge, instead, is the claim that any items are inaccessible. But why would Block (2011)

But (5) is just one of two claims that the Overflow Argument seeks to establish. The other claim, which Block takes to follow from (5), is that, (6), "the machinery of phenomenology is somewhat different from the machinery of cognitive accessibility" (2007a, 489). In the present context, we can take the term 'machinery' to broadly refer to whatever brain mechanisms/processes *suffice* for the instantiation of p-consciousness, on the one hand, and cognitive accessibility, on the other. With that in mind, the idea is the following. If more information can be (and is) p-conscious than is cognitively accessible, then whatever mechanisms/processes suffice for the instantiation of p-consciousness must be "at least somewhat different" from the ones that suffice for the instantiation of cognitive accessibility. But why the 'somewhat'?

For even though participants are only able to report very few letters, they also claim to have seen them all in detail. In this sense, subjects are p-conscious of

---

make the above claim, then? The answer, I believe, is that what he means by 'inaccessible' and 'unaccessed' in his (2011) is simply different from what he means by the same terms in his (2007). In order to appreciate what Block (2011) means by 'inaccessible' and 'unaccessed', I need to say something about Block's (2007a, 492) distinction between *broad* and *narrow* cognitive accessibility. A representation is said by Block to be cognitively accessible in the *narrow* sense if it is actually globally broadcast in the workspace, and thus directly available to report, the rational control of action, etc. 'Directly' here means that, once globally broadcast, the representation is available to the consuming systems "without further processing" (Ibid.). As for broad cognitive accessibility, a representation is said by Block to cognitively accessible in the *broad* sense if it is not actually, but only *potentially*, broadcast in the workspace. In other words, a representation's being accessible in the broad sense is a matter of its being *poised* to be globally broadcast. Now, in all likelihood, what Block (2011) means by 'inaccessible' is broad inaccessibility. For any one item would be globally broadcast if properly cued (Block 2007a, 489). Thus, the Overflow Argument does *not* show that any items are inaccessible in the broad sense. Still, most items are not globally broadcast. Thus, what the Overflow Argument *does* show is that most items are unaccessed in the broad sense. At this point, one thing that I should note is that, as should be clear, accessibility in the narrow sense is the notion of accessibility that I have been assuming so far. Moreover, unless otherwise specified, this is also the notion that I will assume in the remainder of this work. There are two different reasons for focusing on narrow accessibility. First, this is what Block (and Kriegel) have in mind when using the term 'cognitive accessibility': cognitively accessible representations are (actually) globally broadcast representations. Second, as we have seen, according to Dehaene and colleagues, it is globally broadcast representations that are conscious – both in the access and phenomenal sense, if you will.

all the letters in detail and a-conscious of their own p-conscious experience. More precisely, what is a-conscious is a conceptual representation of their own p-conscious experience, a representation with the "generic content" that they "[have] had specific phenomenology" (Block 2007c, 539). But, for all we know, this might mean that cognitive access to a representation of our own p-conscious states is a necessary condition of p-consciousness.[88] It may be that a representation R cannot be p-conscious unless a representation of R, R_1, is cognitively accessible. Still, the thought goes, the machinery of p-consciousness has to be at least partly different from the machinery of cognitive access. For while subjects are p-conscious of all the letters in the array – and *know* that they are – they can only report very few of them.

6. Overflow Meets Neuroscience: The Mesh Argument

This section introduces the second step of the Overflow-Mesh Argument, namely, the Mesh Argument. The latter proceeds from the claim that (4) p-conscious information overflows what is in working memory, to the conclusion that the neural basis of a-consciousness is not included in the neural basis of p-consciousness. In a nutshell, Block's reasoning is that if we assume that the neural basis of a-consciousness is not included in the neural basis of p-consciousness, then we will have a mechanism that explains why p-conscious information overflows working memory (Block 2007, 498). This, in Block's view, is a strong enough reason to make the above assumption.

Clarifications are already in order. In particular, I first need to explain what Block means by 'neural basis'. This is what I will do immediately below.

[88] Block is not explicit about this. That is, although he notes that the 'somewhat' in (6) is due, in part, to the fact that subjects report that they have seen all the characters, he is silent as for why that is the case. The idea that, because subjects know that they have seen all the letters, access to a representation of the phenomenal character of the experience may be a necessary condition of p-consciousness, however, is made explicit by Katalin Balog (2007).

6.1. The Neural Basis of Phenomenal Consciousness and Access Consciousness

According to a standard definition, the neural basis of consciousness (*tout court*) is "the minimal set of neuronal events and mechanisms jointly sufficient for a specific conscious percept" (Koch 2004, 16).[89] We can tweak Koch's definition so as to make it more general by replacing 'percept' with '(mental) state'. Importantly, in the present context, we also need to distinguish between the neural basis of p-consciousness and the neural basis of a-consciousness. The former can be defined as the minimal set of neuronal events and mechanisms jointly sufficient for a specific p-conscious state. The latter, instead, can be defined as the minimal set of neuronal events and mechanisms jointly sufficient for a specific a-conscious state.

Koch's definition of neural basis concerns *specific* types of p-conscious/a-conscious states; e.g., the neural basis of a p-conscious experience with a reddish phenomenal character, and the neural basis of the a-conscious experience of a horizontal line. But this does not mean that one might not also be interested in the neural basis of *broader* types of p-conscious/a-conscious states; e.g., the neural basis of a p-conscious visual experience, the neural basis of an a-conscious auditory experience. Crucially, it also does not mean that the core of Koch's definition cannot work as a definition of 'neural basis' in relation to these broader types of states. Thus, for example, we can easily define the neural basis of a p-conscious *visual* experience as the minimal set of neuronal events and mechanisms jointly sufficient for that experience, and the neural basis of an a-conscious *auditory* experience as the minimal set of neuronal events and mechanisms jointly sufficient for that experience.

Let us now have a closer look at the neural basis of p-consciousness and a-consciousness, respectively. Echoing Chalmers (2000), Block (2005, 2007a)

[89] Strictly speaking, this is the definition that Koch (2004) provides of 'neural correlate of consciousness'. As I understand it, however, this definition captures Block's notion of neural basis just as well.

distinguishes between the *total* and *core* neural basis of p-consciousness.[90] The former, he defines as follows: "[t]he total neural basis of a state with phenomenal character C is itself sufficient for the instantiation of C" (Block 2007a, 482). Block's definition can be construed as concerning both specific and broader types of p-conscious states. Thus, for example, the total neural basis of a mental state with a reddish phenomenal character is sufficient for the instantiation of that reddish phenomenal character. The total neural basis of a mental state with visual phenomenal character, instead, is sufficient for the instantiation of visual phenomenal character. As for the core neural basis, Block suggests that "The core neural basis of a state with phenomenal character C is the part of the total neural basis that distinguishes states with C from states with other phenomenal characters or phenomenal contents" (Ibid.). Again, this definition can be understood as concerning specific and broader types of p-conscious states. The core neural basis of a state with a reddish phenomenal character, for example, is the part of the total neural basis that distinguishes states with reddish phenomenal character from states with bluish phenomenal character. The core neural basis of an experience with visual phenomenal character, instead, is the part of the total neural basis that distinguishes experiences with visual phenomenal character from experiences with auditory phenomenal character. (Note that, in light of the distinction between total and core neural basis, Koch's definition above counts as a definition of *total* – rather than core – neural basis).

But what is the "total minus core neural basis" (Block 2007a, 482) of p-consciousness? That is, what is it that needs to be added to the core neural basis of a p-conscious state to get its total neural basis? Block defines the total minus core neural basis of a p-conscious state as the set of "enabling conditions" (2005, 47) for

[90] Koch (2004), too, draws a similar distinction between *enabling* and *specific* factors. As he notes, "enabling factors are tonic conditions and systems that are needed for any form of consciousness to occur at all, while specific factors are required for any one particular conscious percept, such as seeing the glorious, star-studded alpine night sky" (Koch 2004, 88).

a p-conscious state. More precisely, it is what makes a (potentially p-conscious) state p-conscious (2009, 1112). (More on this below).[91] Thus, while the core and total neural basis change from one type of p-conscious experience to another, the total minus core neural basis is the same for *all* types of p-conscious experiences.

For later purposes, it will be worth saying something about the neural mechanisms and events that Block identifies as potential candidates for the core neural basis of p-consciousness, on the one hand, and its total minus core neural basis, on the other. (Perhaps it is best to say *visual perceptual p-consciousness*, rather than p-consciousness *simpliciter*, as the former is what Block is primarily concerned with. Unless otherwise specified, below I assume that p-consciousness is always visual perceptual p-consciousness). The most plausible candidate for the core neural basis of visual p-consciousness, Block thinks is *recurrent activation* in the *occipito-temporal* areas of the brain. It is recurrent activation in those areas that Block thinks determines how the phenomenal character of p-conscious visual experiences differs from the phenomenal character of p-conscious auditory experiences, gustatory experiences, and so on.[92] His chief example concerns the experience as of motion. Block cites different lines of evidence which all point to the idea that the core neural basis of the phenomenal experience as of motion is recurrent activation between V1, the primary visual cortex, and V5, in the extrastriate visual cortex (Block 2007a, 496).

Now for the total minus core neural basis. Citing, for example, Alkire and Miller (2005), Block suggests that the "single neural background of all experience" involves the activation of connections between the cortex and the thalamus (Block 2007a, 482). Alkire and Miller (2005) review several studies on the relationship between general anaesthesia and consciousness which all converge on one point: disabling certain connections between the cortex and the thalamus what general

[91] I will also come back to this in chapter 5.

[92] I discuss the notion of recurrent activation at length in the next chapter.

anaesthetics do. On the assumption that consciousness (in any sense of the term 'consciousness') is absent when a subject is under general anaesthesia, the authors infer that the activation of the aforementioned connections may be necessary for consciousness (in any sense of the term of the term 'consciousness').[93] On the basis of Alkire and Miller's study, Block hypothesises that activation of those connections may be the total minus core neural basis of p-consciousness.

But what, exactly, is the relationship between p-consciousness (in general; i.e., visual, auditory, and maybe even non-perceptual) and its total neural basis, according to Block? Block suggests that "[p-]consciousness is *identical to* its total neural basis" (2007a, 482 emphasis mine).[94] Let us see if we can shed some light on this statement. We have seen that the total neural basis of a kind of conscious state is defined as "the minimal set of neuronal events and mechanisms jointly sufficient for [it]" (Koch 2004, 16). This definition, however, is compatible with a wide range of views about the relationship between p-consciousness and the neuronal events and mechanisms that constitute its total neural basis. For example, it is compatible with the view according to which p-consciousness is *determined* by said neuronal events and mechanisms. A view of this sort is endorsed by Searle (1992). According to Searle, p-consciousness *supervenes* on its total neural basis. Roughly, this means that any difference in p-conscious properties guarantees a difference in the neuronal events and mechanisms that constitute its total neural basis, but not the other way around. Looking at the notion of supervenience from a different angle,

[93] I should note that Block's reasons to claim that the activation of certain connections between the cortex and the thalamus is the total minus core neural basis of p-consciousness can be easily called into question. Alkire and Miller (2005) write that, in the reviewed studies, "the 'unconscious' endpoint referred to in many anaesthetic studies is that point at which an anaesthetic dose is given in sufficient quantity that it causes a subject to be unable to respond to a verbal command or to a rousing shake" (2005, 232). But, if anything, this is evidence for the claim that the activation of certain connections between the cortex and the thalamus is necessary for *a-consciousness*. While one's inability "to respond to a verbal command or to a rousing shake" suggests that one's experience of the command/shake is not a-conscious, it is compatible with that experience's being p-conscious.

[94] But see also Block (2009).

"sameness of neurophysiology guarantees sameness of mentality; but sameness of mentality does not guarantee sameness of neurophysiology" (Searle 1992, 125).

But the above definition of neural basis is also compatible with the claim that p-consciousness is *identical with* the neuronal events and mechanisms that constitute its total neural basis. This is Block's own view. To say that p-consciousness is identical with its total neural basis is the same thing as saying that p-consciousness *just is* a certain brain state, or that the property of being p-conscious *just is* a certain biological property. Consider visual p-consciousness again. To say that visual p-consciousness just is a certain brain state is to say that visual p-consciousness just is the brain's having recurrent activation in the occipito-temporal areas of the brain, plus activation of certain neural connections between the cortex and the upper brain stem. To say that visual p-consciousness just is a biological property, instead, is to say that the property of being visually p-conscious just is the property of being recurrent activation in the occipito-temporal areas, plus activation of certain neural connections between the cortex and the upper brain stem.

What about the neural basis of a-consciousness? For our purposes, identifying the core and total neural basis of a-consciousness is of no relevance. What needs pointing out, however, is that a crucial component of the neural basis of a-consciousness is activation in the frontal, prefrontal, and cingulate areas of the brain. As you will remember from chapter 2, workspace neurons are particularly dense in those areas, which is where the consuming systems are located. Indeed, one might even see those areas as the "seat" of the global workspace. It is those very areas that, in his argument, Block refers to as 'the neural basis of a-consciousness', and this is how I will use the same phrase below.

One last brief point. What is the nature of the relation between a-consciousness and its total neural basis? Above we saw that Block construes the relation between p-consciousness and its total neural basis as an identity relation.

But the same cannot be said with respect to the relation between a-consciousness and its total neural basis. As you will remember, a-consciousness is a *functional* property. This entails that, unlike p-consciousness, the relation between a-consciousness and its total neural basis is one of *realization*. As it happens, in humans as well as other animals, a-consciousness is *realized* by its total neural basis.

Before introducing Block's argument, I need to take care of one last thing. In particular, I need to say something about the brain mechanisms and processes surrounding global broadcasting.

6.3. The Neuronal and Functional Events Surrounding Global Broadcasting

Consider visual perception. Here is a rough picture of what, according to GWM, happens upon presentation of a stimulus. Within a few msec. from the presentation of the stimulus, bottom-up (i.e., sensory) stimulation results in the activation of the occipito-temporal areas in the "back of the head" (e.g., Dehaene and Changeux 2004; Dehaene et al. 2006). In general, if the stimulus is sufficiently strong and long-lasting, activation in the above areas will be strong, too. According to GWM, if the stimulus is targeted by attention, the already strong neuronal activation in the relevant areas will be amplified to the point that it will spread to the parietal, prefrontal, and cingulate regions. These areas of the brain are normally referred to as *higher association cortices* (Dehaene et al., 2006) and, as already noted, are thought to be the areas where systems of report, deliberation, etc. are located.

Importantly, according to GWM, the activation of the higher association cortices by the occipito-temporal areas hinges on the presence of the long-distance axons of workspace neurons. These connect the occipito-temporal areas to parietal, prefrontal, and cingulate areas, and the latter areas among themselves – indeed, as noted on several occasions, workspace neurons are particularly dense precisely in the areas in question. Dehaene et al. (2006) refer to the information processing

state corresponding to this state of widespread activation – the state in which activation has extended from the occipito-temporal areas to the parietal, prefrontal, and cingulate areas – as *conscious*. According to Block's distinction between p-consciousness and a-consciousness, however, this information processing state would only qualify as *a*-conscious. In the language of GWM, the spreading of activation from the occipito-temporal areas to the higher association cortices corresponds to the global broadcasting of information about the relevant stimulus, and thus to such information becoming accessible to the relevant consuming systems.[95]

But what happens when the presented stimulus is sufficiently strong and long-lasting, and yet it is *not* targeted by attention? In that case, brain activation, although intense – indeed, almost as intense as it would have been had the stimulus been targeted by attention – will remain confined to the occipito-temporal areas. Dehaene *et al.* (2006) and Dehaene and Changeux (2011) refer to the information processing state corresponding to this state of confined activation as *preconscious*. Information carried by the strong activation of the occipito-temporal areas is only *potentially* accessible information: it would become (*actually*) accessible if the relevant stimulus were targeted by attention, which would in turn trigger the global broadcasting of the relevant representation.[96]

It might also happen that a visual stimulus is too weak, short-lasting, or both. In cases such as this, although presentation of the stimulus will still cause activation in the occipito-temporal areas of the brain, this activation will be very weak, and will be barred from spreading to more frontal areas. Dehaene *et al.*

[95] Here and in the remainder of this chapter I (unharmfully) assume that if the functionality condition that I argued for in chapter 3 has to indeed be satisfied for a-consciousness to be present, then it is *normally* satisfied. This means that the occurrence of global broadcasting *normally* guarantees the presence of a-consciousness.

[96] Consider the phenomenon of inattentional blindness, encountered in chapter 2. The visual stimulus is quite strong and yet, because the subjects' attention is directed elsewhere the relevant representation is inaccessible. It would become accessible, however, were the subjects to attend to the stimulus. This is a clear case of preconscious processing.

(2006) and Dehaene and Changeux (2011) refer to the information processing state corresponding to this state of weak and confined activation as 'subliminal'. As in the case of preconscious processing, information involved in subliminal processing is *in*accessible. Unlike the case of preconscious processing, however, subliminal information is not even potentially accessible. That is, it could not be made available to the brain's consuming systems even if the subject's attention were directed to the relevant stimuli.[97]

But a visual scene will normally contain many objects. Accordingly, the number of stimuli concurrently causing bottom-up activation in the occipito-temporal areas of the brain will normally be quite large. So while neuronal activation caused by one or more strong and long-lasting stimuli will, in virtue of being amplified by attention, extend to parietal, prefrontal, and cingulate areas, neuronal activation caused by the remainder of the stimuli in the visual field, regardless of their initial strength, will not. Crick and Koch (2003), Koch (2004), and Block (2007a, 2009) speak of *neural coalitions*. The notion of neural coalition is somewhat flexible. In the present context, a neural coalition may be thought of as a temporary assembly of neurons whose overall activation represents a specific stimulus/group of stimuli (Fahrenfort and Lamme 2012). At any one time, different coalitions feature different levels of activation, depending, among other things, on the initial strength of the stimuli they carry information about, but also on the attentional boost they (do or do not) receive.[98]

[97] Subliminal processing can also be induced through *masking* in laboratory settings. Masking consists in the reduction or elimination of the visibility of a stimulus S through the presentation of other stimuli S_n, in close temporal or spatial contiguity with S. As an example, consider the visual presentation of a word for about 33ms. That word is fully visible (and reportable) when presented in isolation, but stops being so when it is immediately preceded or followed by the presentation of a geometrical shape (Dehaene & Changeux, 2011, p. 201).

[98] What is the relationship between neural coalitions and the processors discussed in earlier? In chapter 1, I defined processors as highly specialized group of neurons dedicated to the processing of very specific kinds of information. In light of this, neural coalitions

Following Block's terminology (Block 2007a, 2009), we can thus distinguish between "winning" or "dominant" and "losing" neural coalitions. Winning coalitions are assemblies of neurons whose (already strong) activation has been amplified by attention and has thus resulted in global broadcasting. Normally, these coalitions are selected because of their relevance to the organism's current goals (Dehaene & Changeux, 2004). As for losing coalitions, these come in two kinds. On the one hand, there are *losing and weak* coalitions. These are assemblies of neurons whose activation level is too low to *ever* cause global broadcasting – in the language of Dehaene et al. (2006), the information carried by activation of these coalitions would count as subliminal. On the other hand, there are *losing but strong* coalitions. These are assemblies of neurons whose activation is strong, albeit *not sufficiently* strong to have resulted in global broadcasting – in the language of Dehaene et al. (2006), the information carried by the activation of these coalitions would count as preconscious. Losing but strong coalitions, however, are also *potentially* winning coalitions. What that means is that, were the relevant stimulus/stimuli be targeted by attention, the activation of these coalitions would become strong enough to trigger global broadcasting.

We are now ready to introduce the Mesh Argument.

6.4. The Argument

As noted, the Mesh Argument proceeds from the claim that (4) p-conscious information overflows what is in working memory, to the conclusion that the neural basis of a-consciousness is not included in the neural basis of p-consciousness. More precisely, in light of the distinction between core and total neural basis, the argument concludes that the neural basis of a-consciousness is not included in *either* the core *or* total neural basis of p-consciousness. The

might also be seen as groups of processors whose activity codes for the properties of the relevant stimulus/groups of stimuli.

argument, which takes the form of an inference to the best explanation, can be reconstructed as follows:

(4) More information is p-conscious than is in working memory;

(5_1) If we assume that the neural basis of a-consciousness is not included in either the core or total neural basis of p-consciousness, then we have a neural mechanism that explains (4);

(6_1) The assumption in (5_1) is the best explanation for (4) in that it leads to a mesh between psychological and neuroscientific results;

(7) Therefore, the neural basis of a-consciousness is not included in either the core or total neural basis of p-consciousness.

Let us have a closer look at the argument. As noted, Sperling's partial report paradigm, as well as other experiments in the same vein, can be understood as warranting the claim that (4) more information is p-conscious than is in working memory. How can we explain this overflow of working memory by p-consciousness? Suppose we assume that the neural basis of a-consciousness is not included in either the core or total neural basis of p-consciousness. In other words, suppose we assume that *all* strong – both winning and losing – neural coalitions of activation in the occipito-temporal areas of the brain can be p-conscious in the absence of activation of the more frontal areas that underpin global workspace activation. Then we have a neural mechanism that explains why more information is p-conscious than is later stored in working memory. Even though, at any one time, only one (or some) neural coalitions trigger global broadcasting – which results in the information they carry becoming a-conscious – many other losing, albeit strong, coalitions will still be p-conscious.[99]

[99] One tacit assumption here is that, as we have seen, there is a strong relationship between a-consciousness and working memory. As noted, Block is not entirely clear about

The assumption that the neural basis of a-consciousness is not included in either the core or total neural basis of p-consciousness leads to a mesh between psychology and neuroscience. That is, it leads to a mesh between the psychological result that more information can be (and is) p-conscious than is in working memory and the result from neuroscience that some coalitions of activation in the occipito-temporal areas of the brain can be almost as strong as the coalitions that trigger activation in more frontal areas. In virtue of leading to the mesh, that assumption is the best explanation for (4).

7. Carruthers's Criticism of the Overflow-Mesh Argument

This section introduces Carruthers's (2015b) criticism of the Overflow-Mesh Argument. To this end, I begin by providing a more concise reconstruction of the Overflow Argument, and leave out (5) and (6) above:

(I) Subjects are p-conscious of all or almost all the letters in detail;
(II) Subjects only report about 4 letters in the whole report paradigm;
(III) More information is p-conscious than is in working memory.

As we already know, the second step of the Overflow-Mesh Argument – the Mesh Argument – argues for the conclusion that the neural basis of a-consciousness is not included in either the core or the total neural basis of p-consciousness, on the grounds that this is a powerful explanation for (III).

The Overflow-Mesh argument has been challenged in multiple ways. Many critics have attempted to deny the claim that (III) more information is p-conscious than is in working memory, by rejecting the claim that, upon presentation of the

what this relationship is. There is little doubt, however, that he thinks that a representation's being a-conscious – and thus being globally broadcast – is a necessary condition for its being a working memory representation. Now, if we assume that neural coalitions in the "back of the head" can be p-conscious in the absence of global broadcasting – in the absence of a-consciousness – it follows that information can be p-conscious without being in working memory.

array, (I) subjects are p-conscious of all or almost all the letters in detail. Consistently with the outcome of Sperling's partial report paradigm, Cohen and Dennett (2011), for example, claim that upon presentation of the array subjects "are conscious only of the few letters they attend to and the impression that there are other items on the display whose identities they do not know" (2011, 359). Similarly, Naccache and Dehaene (2007) suggest that, upon presentation of the array, subjects distribute their attention evenly over it. Because attentional resources are limited, this results in only a gist representation of the array being globally broadcast, and thus conscious. But on this and the previous account, how is the participants' ability to report all (or almost all) the letters of any cued row to be explained, then? The thought seems to be that the effect of cueing is to attract the subjects' attention to part of an *un*conscious representation of the array. This results in the relevant representation's being globally broadcast in the workspace, and thus in its being available for report. The reason why subjects claim to have seen every letter in the array in great detail is just that they have fallen victim to some kind of illusion.

To appreciate what this means, consider Dehaene and Naccache's proposal (2001). According to the authors, in the Sperling experiments, subjects have "the intuition of a rich phenomenological world" (Ibid., 30) because, were they to focus their attention on any of the letters, they would actually be able to see them in great detail. In this sense, to use Block's terminology, subjects confuse *potential* with *actual* phenomenology (2007a, 491). Block (2001, 2007a) dubs the alleged phenomenon the 'refrigerator light illusion': "[t]he allusion being to the possibility that a technologically naïve person might have the illusion that the refrigerator light is always on because it is always on when he looks" (2001, 209).[100]

Unlike the above critics, Carruthers (2015b) wishes to grant the truth of (III). He believes that it may indeed be possible for subjects to have a rich p-conscious

[100] Cohen and Dennett (2011) espouse a view that is similar to Dehaene and Naccache's.

experience of the array and that, given that (II) they only report 4 letters, this suggests that (III) more information is p-conscious than in working memory. On the other hand, Carruthers aims to challenge Block's idea that (III) is best explained by the claim that the neural basis of a-consciousness is not included in either the core or total neural basis of p-consciousness. In his view, there is a better explanation for this fact – one that supports Dehaene *et al.*'s claim that p-consciousness and a-consciousness are identical with the property of being globally broadcast, and thus with each other. To make his point, Carruthers (2015b, 4) puts forward five related theses, which he suggests are all widely accepted among cognitive scientists:

i. "Attentional signals directed at representations in sensory regions of the brain are a necessary (and, with other factors, sufficient) condition for those representations to be globally broadcast";
ii. "attention is a limited resource: only so much information can be attended to at any one time";
iii. "the effect of attentional signals is to boost the neural activity underlying the targeted representations (while also suppressing competing activity, and perhaps also sharpening the representations in question)";
iv. "working memory uses the same attentional network to sustain previously-presented sensory representations in the global workspace";
v. "global broadcasting takes place when some sort of threshold of neural activity is reached".

With the above theses in place, Carruthers notes that the reason why Block's argument fails is fairly straightforward. The thought is that the amount of attention needed to actively sustain a representation in working memory is much larger than the amount required to globally broadcast the same representation. In the latter case, "the incoming signal will involve exogenously caused neural activity that is already above baseline, needing less of an attentional boost in order to reach the

threshold for global broadcasting" (2015b, 4). [101] The same cannot be said with respect to sustaining a representation in working memory, however. For in that case attention will have to do its work in the absence of exogenously caused neural activity. Attention will thus have to boost the relevant neural activity all by itself. In virtue of that, it is reasonable to suppose, more attention will be needed to sustain a representation in working memory than to broadcast the same representation in the global workspace. It follows that a richer and more detailed representation of a stimulus will be globally broadcast (and hence a-conscious) than will later be held in working memory.

In this way, Carruthers can grant that (I) subjects in the Sperling experiments are conscious "of most of [the letters] in identity-defining detail" (2015b, 4); explain why (II) subjects can only name about 4 letters; and vindicate the claim that (III) more information is p-conscious than is in working memory. At the same time, however, he can reject the conclusion of Block's argument, namely, that the neural basis of a-consciousness is not included in either the core or total neural basis of p-consciousness. When, in the whole report paradigm, subjects are presented with the array of letters, they distribute their attention "evenly over the entire display" (2015b, 5). Because exogenously caused neural activity is already high, attention may be sufficient to cause the global broadcasting of a detailed representation of the entire array. At this stage – call it *T1* – information about the array will thus be broadcast in the workspace, but not yet retained in working memory.

After stimulus offset – call this stage *T2* – exogenously caused neural activity will gradually drop. In order to report the letters that they have seen, subjects will now have to hold the relevant information in working memory. That is, they will now have to actively sustain the relevant neural activity, "holding [it] far enough

[101] What does 'baseline' mean, in this context? The idea is that, even in the absence of exogenously caused activity – activity caused by the presence of a stimulus – there is still some spontaneous activity in the relevant brain areas.

above baseline for global broadcasting to continue to take place" (Ibid.). Crucially, however, in the absence of bottom-up stimulation, attentional resources alone won't suffice to sustain as much neural activity as is needed to keep a detailed representation of the array in a broadcast state. Instead, all attentional resources will now have to be focused on a much smaller number of characters – three or four – and withdrawn from the others. As a result, three or four is the number of letters that subjects can normally report.

But if global broadcasting is all that is needed to explain (III), namely, that more information is p-conscious than is in working memory, then Carruthers can conclude that there is only one neural basis of consciousness (tout court), and that p-consciousness and a-consciousness are thus one and the same thing.

8. In Defence of the Overflow-Mesh Argument

What should we think of Carruthers's reply? In what follows, I argue that it does not undermine the Overflow-Mesh Argument. Let me begin by laying out three assumptions. First, let us assume the truth of all five theses introduced in §7. As Carruthers notes, although not uncontroversial, all five theses are widely accepted among cognitive scientists. Second, let us grant Carruthers that, at T1, subjects distribute their attention evenly over the entire display. This seems a reasonable assumption, for in the absence of reasons for doing otherwise – e.g., the experimenter's instruction to attend to one specific row – the even distribution of attention over the entire display is what one would probably expect. Finally, let us also grant Carruthers that, at T1, a representation of the array is broadcast in the global workspace.

My argument will proceed in two steps. First, I will challenge the claim that, because of "the exogenously caused neural activity that is already above baseline" (Carruthers 2015b, 4), the even distribution of attention over the entire display results in the global broadcasting of a detailed representation of the array. Having

established that no detailed representation of the array is broadcast in the workspace, I will then argue that if, in line with Block and Carruthers, we wish to grant that (III) more information is p-conscious than is in working memory, then we will have to assume with Block that all strong coalitions – both winning and losing – in the occipito-temporal areas of the brain can be (and are) p-conscious in the absence of global broadcasting. In other words, we will have to assume that the neural basis of a-consciousness is not included in either the core or total neural basis of p-consciousness.

In their commentary to Block (2007), Naccache and Dehaene write:

> When subjects report seeing "all the letters," we suggest that they distribute their attention globally over the array, and thus are only able to determine its approximate numerosity and "letterhood"; [...] only this approximate content, not the detailed letter identities, accesses a fronto-parietal global neuronal workspace. When subjects report individual letters, they change to a focused attention mode that allows them to enhance one location at a time, to let its specific content access the global workspace, and, therefore, to name it (Naccache and Dehaene 2007, 519).

Upon presentation of the array, subjects in the Sperling experiments distribute their attention evenly over the entire display. But evenly distributed attention over the entire display results in the global broadcasting of only an *approximate* representation of the array. If Naccache and Dehaene are right, attention is therefore too much of a limited resource to trigger the global broadcasting of a detailed representation of the array.[102] But are they right?

Naccache and Dehaene justify their claim that only an approximate representation of the array will be broadcast in the global workspace by appealing

[102] Note that, as one would expect, Naccache and Dehaene's goal is antithetic to mine – they offer those considerations in the context of arguing *against* Block. Still, this does not mean that we can't use the same considerations to argue *in favour* of Block.

to the work of Sabine Kastner et al. (1998) on attention.[103] Using functional magnetic resonance imaging (fMRI), Kastner et al. have provided evidence suggesting that when attention is distributed evenly over a scene consisting of four or more objects, representations of the objects in the occipito-temporal areas of the brain "interact in a competitive, mutually suppressive fashion" (Ibid., 108). One plausible consequence of the representations' mutually suppressive interaction is that only *some* information about each of the objects in the subject's visual field will be able to reach further stages of processing. In the language of GWM, this means that only *some* information about each individual object will be broadcast in the workspace. On the other hand, the focusing of attention upon a specific location/object in the same scene has the effect of cancelling out "the suppressive influence of nearby stimuli, thereby enhancing information processing at the attended location" (Ibid.). Again, in the language of GWM, the focusing of attention will result in detailed information about the attended stimulus being broadcast in the workspace, just as would have happened had the stimulus been presented alone.

Below is a concise account of the evidence that Kastner and colleagues have provided for the claim that evenly distributed attention over a scene consisting of four or more objects results in the mutually suppressive interaction of the relevant extrastriate representations. Eight subjects were presented with complex visual images in the *absence* of spatially directed attention. The images were shown in random order within a quadrant under sequential and simultaneous presentation conditions. "In the sequential condition (SEQ), each of the stimuli was shown alone in one of the four locations. In the simultaneous condition (SIM), the stimuli appeared together in all four locations" (Ibid.). On the basis of results from previous experiments in the same vein conducted on monkeys (e.g., Treue and Maunsell 1996; Miller et al. 1993), Kastner and colleagues hypothesised that, if

[103] But see also Kastner *et al.* (1999), Kastner *et al.* (2001), and Beck & Kastner (2005).

representations of the stimuli in the occipito-temporal areas did really interact in a mutual suppressive fashion in the simultaneous condition, fMRI signals concerning those very areas would be smaller than in the sequential condition. This is exactly what they found. As they suggest, fMRI signals during the presentation of single stimuli were "significantly greater" than the signals during the presentation of the same stimulus plus three other stimuli (Ibid., 109).

On the assumption that Kastner and colleagues are right, Carruthers's claim that evenly distributed attention over the entire display will result in the global broadcasting of a detailed representation of the array is rejected. For given the mutually suppressive effect of the simultaneous presence of multiple objects, evenly distributed attention over the Sperling array is unlikely to suffice for the broadcasting of a detailed representation of the same. As Naccache and Dehaene point out (2007, 519), subjects will only be "able to determine [the array's] approximate numerosity and 'letterhood' [...] only this approximate content, not the detailed letter identities, accesses a fronto-parietal global neuronal workspace".

But if we still wish to grant, in line with Block and Carruthers, that (III) more information is p-conscious than is in working memory, then we will also have to grant – again, in line with both authors – that (I) subjects are p-conscious of all or almost all the letters in detail. But having established that only an approximate representation of the array is broadcast in the workspace, it seems clear that 'p-conscious' cannot be understood in terms of global broadcasting, but *a la* Block. In other words, in order to account for the richness of the experience, we will have to assume with Block that all strong neural coalitions – both winning and losing – in the occipito-temporal regions can be (and are) p-conscious in the absence of global broadcasting.

I would now like to anticipate a possible response. The second step of my argument has focused on Carruthers's inability to account for (I) – namely, that subjects are p-conscious of all or almost all the letters in detail – by appealing to

global broadcasting. But in light of the evidence on attention discussed above, Carruthers might now give up his attempt to vindicate (I), while noting that a weaker claim will suffice to establish (III) just as well. For all that is needed to establish (III) is that *more* information is globally broadcast than is later encoded in working memory. And although the empirical evidence is incompatible with the global broadcasting of a detailed representation of the array, it *might* still be compatible with the broadcasting of an amount of information that is just a bit larger than the amount later held in working memory.

Were Carruthers to pursue this line of response, the burden of proof would be on him to show that the evidence is indeed compatible with the above. And even if he were able to do that, part of my point would still stand. Even though, clearly, I would no longer be able to salvage Block's conclusion, I will have at least shown two things. For one thing, the claim that a rich representation of the array can be globally broadcast is almost certainly false. For another, the claim that (I) subjects are p-conscious of all or almost all the letters in detail can only be vindicated by assuming that all neural coalitions – both winning and losing – in the occipito temporal areas of the brain can be (and are) p-conscious in the absence of global broadcasting. In other words, (I) can only be vindicated by assuming that the neural basis of a-consciousness is not included in either the core or total neural basis of p-consciousness.

Carruthers has offered a story of the cognitive processes at play in the whole report paradigm. Is it possible to offer an alternative story, this time against the background of my defence of the distinctness of p-consciousness and a-consciousness? And is it possible to offer such a story while preserving the five assumptions put in place by Carruthers? I believe so. Start with the presentation of the array – T1. Suppose that subjects enjoy a *p*-conscious experience of the entire array in detail. At the same time, the even distribution of attention on the array results in the global broadcasting of an approximate representation of it. In other

words, the even distribution of attention on the array results in an a-conscious, approximate representation of it. Right after stimulus offset – T2 – subjects continue enjoying a p-conscious representation of the array for a brief period of time. The existence of a persisting image after stimulus offset – *iconic memory* (Neisser, 1967) – is, as we have seen, taken to be shown by Sperling's partial report paradigm. How could subjects report almost all the letters from the cued row? Sperling's answer was that what subjects had seen during the 50msec. exposure, they continued seeing for a very short time right after stimulus offset – just enough time for the cue to attract their attention on the relevant row. Given the assumption that a detailed representation of the entire array is p-conscious at T1, we can also assume that the same representation will be p-conscious for as long the neural activity underlying it stays the same – about 1 sec., to be exact (Sperling 1960, 11). We will now expect that, in order to make a report, all attentional resources will be focused on three or four characters. This will result in the global broadcasting of a representation of those characters. But given that report takes a considerable amount of time, those characters will have to be sustained in working memory. That is, a representation of those characters will have to be kept in a broadcast state for as long as it takes the subject to make a report.

9. Concluding Remarks

In order to understand and evaluate Block's position on the functional role of p-consciousness, we first need to appreciate his positive account of the latter, as well as close in on his understanding of the relation between p-consciousness and a-consciousness. In §2, I provided a general picture of the biological approach to p-consciousness, which Block favours. Having explained that, unlike the defenders of cognitive theories of consciousness, Block thinks that an account of p-consciousness cannot abstract away from the biological/neurological details, I then moved onto analysing the Overflow-Mesh Argument. In its ultimate attempt to disentangle the neural machinery of p-consciousness from the machinery of a-

consciousness, the Overflow-Mesh Argument has worked as a powerful tool to further our understanding of Block's take on p-consciousness, and the latter's relation with a-consciousness.

In short, the picture that we have obtained is the following. First, p-consciousness is identical with its total neural basis. In the specific, this is to say that p-consciousness is identical with activation of a certain kind in the occipito-temporal areas, plus activation of certain connections between the cortex and the brain stem. Second, the neural basis of a-consciousness is not included in either the core or total neural basis of p-consciousness. This means that, in effect, p-consciousness can (and does) occur in the absence of a-consciousness.

Importantly, discussing the Overflow-Mesh has also allowed me to defend the last two claims from a recent criticism advanced by Carruthers (2015b). By appealing to the empirical work of Kastner et al.'s (1998), I have argued that first, Carruthers's argument rests on a view of distributed attention that is empirically implausible and second, that Kastner et al.'s empirical results can be used in direct support of Block's conclusion.

In the next chapter, I put many of the notions introduced in this chapter to work, and attempt to make sense of, and assess, Block's account of the functional role of p-consciousness.

5

Block on the Functional Role of Phenomenal Consciousness

1. Introduction

In this chapter I consider, and assess, Block's take on the functional role of p-consciousness. Block has touched on the question of what the functional role of p-consciousness is on several occasions (e.g., 2009, 2015). To a first approximation, his view is that p-consciousness figures in the causal explanation of a-consciousness. To a second approximation, his view is that (the core neural basis of) p-consciousness *contributes to the triggering of global broadcasting* (personal communication, April 21, 2017).[104] Unfortunately, this is as detailed as Block's view on the matter gets.

Against the background of, first, Block's view on p-consciousness, and, second, the definition of functional role of p-consciousness provided in chapter 1, can we do the philosophical work for Block, and explain just in what sense contributing to the triggering of global broadcasting may be said to be (part of) the

[104] As explained earlier, Block distinguishes between the total and core neural basis of p-consciousness: "[t]he *total* neural basis of a state with phenomenal character C is itself sufficient for the instantiation of C. The *core* neural basis of a state with phenomenal character C is the *part* of the total neural basis that distinguishes states with C from states with other phenomenal characters or phenomenal contents" (2007a, 482). I will come back to this below.

functional role of (the core neural basis of) p-consciousness?[105] Below I attempt to do just that, and argue for the following construal of Block's view: potential a-consciousness is the functional role of (the core neural basis of) p-consciousness.

With a clearly-defined account of the functional role of (the core neural basis of) p-consciousness in hand, I then also argue, contra Block, that the account is untenable. In a nutshell, the reason is that potential a-consciousness is best understood as the functional role of a property *other than* (the core neural basis of) p-consciousness.

Here is the plan. In §2, I provide a careful explanation of the notion of *local recurrent activation.* Doing so is a necessary step towards making sense of Block's view. For as I anticipated in chapter 4, Block thinks that the core neural basis of p-consciousness is exactly local recurrent activation. In the remainder of the chapter, I put to work many of the notions introduced in §2 as well as in chapter 4. In §3, I argue that Block's view should be understood as the view that potential a-consciousness is the functional role of a mental state's property of being a local recurrent activation. Finally, in §4, I argue that, if my construal of Block's view is right, that view is unacceptable. Note that, as in the previous chapter, I will focus exclusively on *visual perceptual* p-consciousness.

2. Lamme and Block on Local Recurrent Activation and Phenomenal Consciousness

In this section, I explain what local recurrent activation is, and how, exactly, Block thinks it is relevant to p-consciousness. What we know about local recurrent activation we know mostly thanks to Victor Lamme, a cognitive neuroscientist. As a consequence, in what follows, I will mainly – although not exclusively – draw on

[105] I assume here that Block's proposal does not concern the *entire* functional role of p-consciousness, but only *part* of it. For the purposes of exposition, below I am going to omit the 'part of' qualification, however.

his work. As it happens, as I will briefly explain, Lamme, too, thinks that this kind of activation is crucial to p-consciousness.

This section is divided into several chunks, so it will be useful outlining its structure. I begin with a few words on the structure of the visual system, with a specific focus on its *hierarchical* structure. After that, I introduce the *feedforward sweep*, a form of neural activation that is thought to temporally precede the onset of local recurrent activation itself. Having introduced the feedforward sweep, I then move onto recurrent activation *tout court*, and distinguish between *local* and *widespread* recurrent activation. Last, I close in on Block's take on the relationship between local recurrent activation and p-consciousness.

2.1. The Parallel and Hierarchical Structure of the Human Visual System

Neuroscientists often speak of the visual system (or parts thereof) as having a *parallel* as well as *hierarchical* structure (Lamme and Roelfsema 2000, 572; Gazzaniga et al., 2014, 189–90; Postle 2015, 139–40). I take the term 'visual system' to refer to all the areas, connections, and structures of the nervous system that underpin the processing of visual information. These would certainly include the retina, the optic nerve, the visual cortex, and a variety of areas in the parietal and temporal lobes. As already mentioned in chapter 2 in the context of explaining the phenomenon of blindsight, in humans and other primates the optic nerve leaving the eye projects directly into an area of the thalamus known as *lateral geniculate nucleus* (Weiskrantz 1990, 3-6). From there, new fibres originate which terminate into an area of the occipital lobe known as *primary visual cortex*, *striate cortex*, or *V1*.

Two main parallel pathways emerge from within the primary visual cortex: a *dorsal* pathway and a *ventral* pathway. The former, which flows to the parietal lobe, is involved in the processing of information for online, real-time action. The latter, which flows into the temporal lobe, is involved in the processing of

information for perception (Lamme and Roelfsema 2000, 572). (I will come back to this distinction in §2.2.1. below). "Within this parallel flow", neuroscientists distinguish a number of functionally (as well as anatomically) individuated areas, among which are V2, V3, V3a, V4, V5 (also known as MT), TEO, and TE (Ibid.). Along with a number of other areas, the areas just mentioned constitute the so-called *extrastriate cortex*. The primary visual cortex plus the extrastriate cortex, instead, are typically referred to as the *visual cortex simpliciter*.

While the parallel aspect of the visual system has to do, *partly*, with the ventral/dorsal distinction, its hierarchical aspect – which is where our main interest lies – has to do mainly with how, starting from the striate cortex, or V1, "progressively higher levels of stimulus representation are constructed at progressively higher levels of the [visual] system" (Postle 2015, 139). In order to make things clearer, the idea that the visual system has a hierarchical structure can be analysed as the conjunction of the following two claims: first, the system is thought to be organized in a multi-level fashion, one level differing from another on the basis of the stimulus features it is tasked with processing. Second, level L_n is said to be *higher up in the hierarchy than* L_{n-1} iff, "[elaborating] on the representation derived by processing" in L_{n-1}, L_n is dedicated to the processing of more complex stimulus features than the latter (Gazzaniga et al. 2014, 189).

Neurons in V1 – which is thought to be at the very bottom of the hierarchy – for example, are dedicated to the detection of fairly simple visual features: edges (Ibid.). Neurons in V2, instead, use the information processed by V1 "to represent corners and edge terminations" (Ibid.). And so on, all the way up to the highest levels of the hierarchy. As concerns the *ventral stream*, the highest levels of the hierarchy can be found in the *temporal lobe*. As concerns the *dorsal stream*, instead, the top is harder to accurately identify (Lamme and Roelfsema 2000, 571). Relatedly, a second reason why the visual system is also said to be organized in a

parallel fashion is that there are multiple parallel connections from one level to the next even *within the same stream* (dorsal or ventral).

2.1. The Feedforward Sweep

With the above picture of the organization of the visual system in hand, let us now consider what happens upon presentation of a visual scene. By means of multiple parallel neural connections "activation [will spread] from low-level to high-level areas of the visual cortical hierarchy (Ibid.). This spreading of activation, is typically known as the *fast feedforward sweep* (Ibid.). The feedforward sweep is said to be *fast* because it only takes about 100 msec. for activation to spread from the lowest to the highest areas of the hierarchy in the ventral stream, for example (Ibid.). Importantly, the earlier the latencies, the larger the number of stimuli that are represented at each stage of the sweep (Lamme 2004, 868). At a latency of 40ms, for example – which is when activation has spread to the whole of V1 – all the stimuli in the visual scene are represented.[106] At a latency of about 60-80ms, instead – which is when activation has spread to some of the extrastriate areas – a smaller number of stimuli is processed in full. In general, however, *most* stimuli will be processed to a good extent even at the highest areas of the hierarchy. Construing, as in the previous chapter, neural coalitions as temporary assemblies of neurons whose overall activation carries information about/represents a specific stimulus/group of stimuli (e.g., Crick and Koch 2003; Cohen and Dennett 2011), we may say that, during the feedforward sweep, there is relatively little competition among different neural coalitions.

Before continuing, let me pause for a moment to point out that although the story that I am currently telling is, in part, different from the one I told in §6.3. of chapter 4, the two are not to be seen as alternative. In that section, I looked at some of the neural events occurring in the "back of the head" following the

[106] More precisely, once feedforward activation has spread to the whole of V1, *the edges of all stimuli* are represented.

presentation of a visual scene. Given my purposes in that section and chapter more generally, however, I mainly focused on distinguishing weak vs strong, and winning vs losing, neural coalitions, and on identifying the different relations that, according to the Global Workspace Model (GWM), exist between those coalitions and global broadcasting. Here my focus is different – I want to understand what local recurrent activation is. For this reason, I need to make some fine-grained distinctions that were absent in §6.3. But one might still wonder whether those distinctions are compatible with what I said in the previous chapter, and GWM more generally. The answer is that they are. Indeed, except for Lamme's take on the relationship between local recurrent processing and p-consciousness, which we will briefly look at in §2.2.1., the advocates of GWM do explicitly endorse Lamme's characterization of the feedforward sweep, recurrent processing, etc. I will come back to this in a couple of pages.

In going back to the feedforward sweep, there are still a couple of things to note about it. First, in spite of being so fast, the sweep is thought to involve quite a great deal of processing. It enables, for example, the detection of features such as edges, colour, orientation, and motion (Lamme n.d., 5). Second, the sweep is thought to be responsible for enabling "reflex-like" behavioural responses, such as the "grasping of moving objects" (Lamme et al. 1998, 531). Third, feedforward activation is *wholly unconscious*, in both the phenomenal and access sense (Lamme 2004, 869). This is inferred from the fact that feedforward activation can also be detected in anesthetised animals, and in cases where a stimulus has been made invisible by backward masking,[107] for example (Lamme 2004, 867–868; Lamme n.d., 7).

[107] Backward masking involves the presentation of a stimulus shortly after the presentation another stimulus. In this way, the second stimulus renders the former invisible (see, for example, Lamme, 2004, 867).

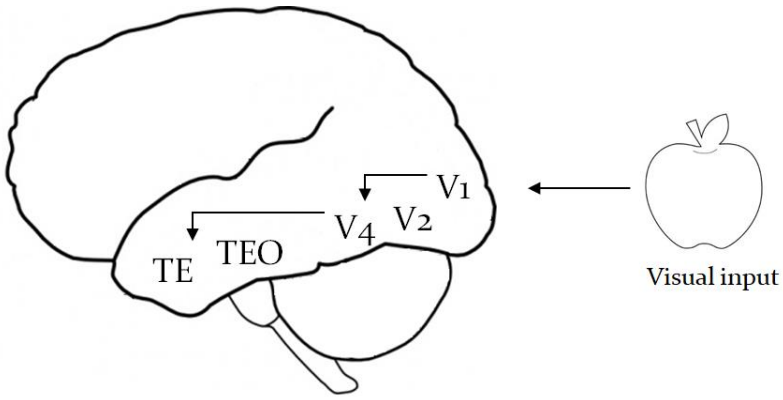

Figure 3. The feedforward sweep in the ventral stream. Upon presentation of a stimulus, activation spreads rapidly from the bottom to the top of the visual cortical hierarchy.[108]

2.2. Recurrent Activation

If a number of conditions are met – e.g., a stimulus has not been made invisible by backward masking or it is not too weak and short-lasting – the fast feedforward sweep is followed, at each stage of processing, by *recurrent* (or *re-entrant*) *activation*. Better, as Lamme explains, "as soon as the feedforward sweep has reached an area [of the visual system hierarchy], recurrent interactions between neurons within that area and neurons that have been activated earlier at lower levels may start" (2004, 867). Thanks to feedback and feedforward connections, information from higher areas of the hierarchy is thus fed back to lower areas, then fed forward again to higher areas, and so on (Ibid.).

A distinction of central importance here is that between two kinds of recurrent activation: *local* and *widespread* (e.g., Lamme 2004; 2006). Before expanding on such a distinction, it will be worth, especially for later purposes, clarifying the notion of recurrent activation itself. On the basis of what I have said so far, we may characterize recurrent activation as follows: recurrent activation is

[108] This illustration is similar to Lamme's (2009, 556).

a kind of neural activation involving bi-directional (feedforward and feedback) interactions among neurons from different levels of the visual system hierarchy.

2.2.1. Local Recurrent Activation

A few words on *local* recurrent activation. Local recurrent activation is a kind of activation involving feedforward and feedback interactions *between neurons in V1 and neurons in the extrastriate areas of the visual cortex only* (Lamme 2004, 867). (A terminological note. V1 + the extrastriate areas are typically referred to by the name of 'sensory areas'. In what follows, I will sometimes use that term, too).

Local recurrent interactions are thought to be responsible for a level of perceptual processing that is more advanced than the one occurring during the feedforward sweep. As explained, the sweep enables, for example, the detection of features such as edges, colour, orientation, and motion. But thanks to local recurrent interactions, many of those features start to get *bound* together, to produce a unitary, coherent representation of an object (Lamme 2003, 17; 2004, 870). More precisely, local recurrent processing is thought to be responsible for the generation of what are sometimes known as *intermediate level representations* (Lamme 2003, 17). To put it with Jesse Prinz (2013, 50), an intermediate level representation provides "a coherent representation of an object's boundaries. It represents surface textures, separates figure from ground, and [captures] information about depth".

I mentioned that Lamme, too, thinks that local recurrent activation has something to do with p-consciousness. Indeed, according to Lamme, local recurrent activation is both *necessary* and *sufficient* for (visual) p-consciousness. More strongly, in his view, (visual) p-consciousness *just is* local recurrent activation.[109] At a more fine-grained level, recurrent activation involving neurons in V1 and neurons in one (or more) specific extrastriate area(s) can be equated with

[109] I shall say something about Lamme's argument(s) below.

specific kinds of phenomenally conscious visual experiences. Consider the experience as of a kind of motion, for example. As already noted in the previous chapter, Block (2005, 2007a) cites several studies that seem to suggest that the (core) neural basis of the experience as of (a certain kind of) motion is recurrent activation from MT to V1, the former being an area in the so-called 'temporoparietal-occipital junction' (Heywood and Kentridge 2009). On Lamme's account, what these studies would probably suggest is that the experience as of (a certain kind of) motion *just is* recurrent activation from MT to V1.

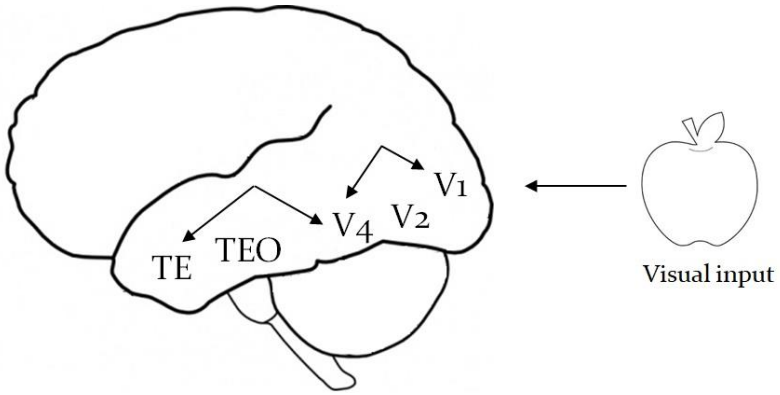

Figure 4. Local recurrent activation in the ventral stream. Once feedforward activation has spread from V1 to the whole of V4, for example, bi-directional interactions from V4 to V1 begin.[110]

Before explaining how *local* recurrent activation differs from *widespread* recurrent activation, I would like to say something more about the *ventral stream*. As we have seen, Lamme thinks that p-consciousness is identical to local recurrent activation. In fact, however, Lamme does not think p-consciousness is identical to local recurrent activation *tout court*, but to local recurrent activation *in the ventral stream*.[111] Briefly expanding on the ventral stream will also help the reader understand why, as noted on several occasions, the areas of the brain where p-

[110] This illustration, too, is similar to Lamme's (2009, 556).

[111] An exception here seems to be recurrent activation involving MT. For this area – which, as we have seen, Block (and plausibly also Lamme) think is part of the neural basis of the visual experience as of (a certain kind of) motion – is actually located between the ventral and the dorsal streams.

conscious and/or a-conscious visual perceptual representations are processed are sometimes called *occipito-temporal* areas.

At the very beginning of §2, I mentioned that two major parallel pathways originate from within V1: the ventral and the dorsal streams. The ventral stream flows from V1 – which is in the *occipital* lobe – into the *temporal* lobe; hence, the relevant areas are called *occipito-temporal.* The dorsal stream, on the other hand, flows from V1 into the parietal lobe. Milner and Goodale (e.g., 1995, 2003) have argued for a thesis that is now accepted by several researchers in vision science and beyond: the ventral stream is involved in processing information "for perception" (2003, 45), whereas the dorsal stream is involved in processing information "for action" (Ibid., 46). Berit Brogaard (2011) summarizes the distinction nicely. Start with the ventral stream.

Activation in the ventral stream, as Brogaard puts it, "is responsible for object recognition and classification" (2011, 452). Information processed by activation in the ventral stream can be temporarily stored in different forms of memory (e.g., working memory and episodic memory) and used in the planning of "offline" action (non-real-time action). Activation in the dorsal stream, on the other hand, is involved in the processing of information that "guides programming and unfolding of on-the-fly actions needed when delayed action is counterproductive" (Ibid.). Examples would be grasping a pen without giving it much thought, catching a ball mid-air, or dodging a punch. When movement is delayed, however, "dorsal stream representations decay" and action becomes guided by information in the ventral stream (Ibid.).

For many, the moral of the story just outlined is that it is only activation in the ventral stream (and hence the *occipito-temporal* areas) that is involved in the processing of information that is relevant to perceptual visual consciousness (*simpliciter*). This claim is endorsed by authors with completely different views about consciousness (*simpliciter*). On the one hand, it is endorsed by those who,

like Lamme (and, as we will see in more detail below, Block), think that local recurrent activation plays a key role in accounting for *p-consciousness*. On the other hand, it is also endorsed by those who, like the advocates of GWM, identify p-consciousness with a-consciousness, and think that there is no consciousness (*simpliciter*) without global broadcasting – and hence without the extension of neural activation to more frontal areas.[112]

2.2.2. Widespread Recurrent Activation

Now for *widespread* recurrent activation. If targeted by attention, some local recurrent coalitions in the ventral system – cross-level local coalitions whose neurons are engaged in recurrent interactions – grow stronger and spread from the sensory areas (V1 + extrastriate areas) to the higher association cortices (parietal, prefrontal, and cingulate areas).[113] It is precisely when it reaches the interface between sensory and higher association areas that recurrent activation is amplified (or suppressed) by the intervention of attentional signals (Lamme 2004, 869). While selective attention plays little to no role during the feedforward sweep, it plays a key role in connection to the extension of recurrent activation to the higher association areas. Thus, whereas at a local level, the number of recurrent coalitions – and hence of stimuli being represented – is relatively high (2004, 869), that number drops dramatically the moment one goes from the sensory areas to the higher association cortices.[114]

---

[112] Relatedly, as one would expect from what I said in the two paragraphs above, it is only activation in the ventral stream that can be targeted by attention. Together with the fact that, as noted, information processed by the ventral stream can be temporarily stored in different forms of memory (e.g., working memory and episodic memory), that should make it apparent that it is only information in the ventral stream that can be broadcast in the global workspace (Dehaene et al., 1998, 14530; Carruthers 2006, 84).

[113] We already encountered the higher association cortices in earlier chapters. As explained, these are thought to be those areas where systems of report, deliberation, and so on, are located.

[114] Let me say something about Lamme's general approach to the study of consciousness (*tout court*), as well as his reasoning for the claim that local recurrent activation is *necessary and sufficient for p-consciousness*. Lamme's point of departure is the observation

According to Lamme, recurrent activation that has spread to more frontal areas is (at least part of) the neural underpinning of a-consciousness. At this stage of processing, information is thus made available "for conscious access and can be reported about" (Ibid.). It is noteworthy that, unlike GWM, Lamme does not

---

that current scientific research on consciousness almost exclusively relies on gauging the presence of the latter from behavioural and introspective data (Lamme 2006, 494). But this approach, he claims, is problematic for at least two different reasons. First of all, one has to choose what, exactly, would count as evidence for the presence of consciousness. For example, does verbal report count as evidence of a subject's having a conscious experience? And what about the pushing of a button? Secondly – and most importantly – by gauging the presence of consciousness from behavioural and introspective data, current scientific research automatically excludes the possibility of unreportable – or, more generally, cognitively inaccessible – consciousness. Is there a viable alternative? Lamme argues for "letting arguments from neuroscience override our intuitive and introspective notion of consciousness" (2006, 499). On the basis of purely neuroscientific considerations, he then proceeds to argue that local recurrent activation is *necessary* for consciousness. This claim is not deeply controversial, as it is also accepted by the major opponents of Lamme's view: the advocates of GWM. What is controversial, instead, is the claim that local recurrent activation is also *sufficient* for consciousness. Lamme has a number of arguments for the latter claim. Here I will sketch only one (Lamme 2007). Recall Dehaene et al.'s (2006) tripartite classification of processing states: *subliminal* (I_1), *preconscious* (I_2), and *conscious* (I_3). Information processed by neural activation at I_1 is neither actually nor potentially accessible. Information processed by neural activation at I_2 is potentially, but not actually accessible. Information processed by neural activation at I_3 is actually accessible. Lamme notes that there is no question as to whether information processed by activation at I_1 is really unconscious. Everyone agrees that information processed at I_1 is information processed during the feedforward sweep. And there is also no question as to whether information at I_3 is really conscious. If subjects report that they are conscious of a stimulus x, as is typically required in Dehaene et al.'s experiments, then they are probably really conscious of x. What is far from settled, however, is whether information at I_2 – which both Lamme and Dehaene agree is processed by recurrent activation – is really unconscious, as Dehaene et al. contend. According to Lamme, this issue can be addressed by asking ourselves "whether I_2 is more like I_1 (i.e., unconscious) or like I_3 (conscious)" (Lamme 2007, 512). If it turns out that I_2 is more like I_3, then it is likely that local recurrent activation is also sufficient for consciousness – and that, as a consequence, there is a distinction to be made between p-consciousness as a kind of consciousness that does not require accessibility, and a-consciousness. One strategy that could be adopted to establish whether I_2 is more like I_1, or more like I_3 is to asking whether the properties we normally associate with conscious perception have already been processed at (I_2). Lamme, as well as many others (e.g., Prinz 2013; Carruthers 2015a), would agree that they have. Another strategy is to ask "what the critical neural differences are between I_1, I_2, and I_3 states" (Ibid.). While information at both I_2 and I_3 is processed by recurrent activation, information at I_1 is not. Hence, "the critical neural dichotomy [lies] between I_1 and I_2/I_3" (Ibid.). For these reasons, Lamme concludes that I_2 is not just necessary, but also sufficient for p-consciousness.

provide a detailed account of a-consciousness. However, what little he says about a-consciousness is entirely compatible with the former. Thus, for example, both Lamme and the advocates of GWM agree that, first, local recurrent activation is necessary, but not sufficient for a-consciousness, and, second, that attention is necessary (and, together with other factors, sufficient) for the spreading of activation from the "back" to the "front of the head" (Dehaene et al. 2006; Dehaene and Changeux 2011).

I now close in on Block's take on the relationship between local recurrent activation and (visual) p-consciousness.

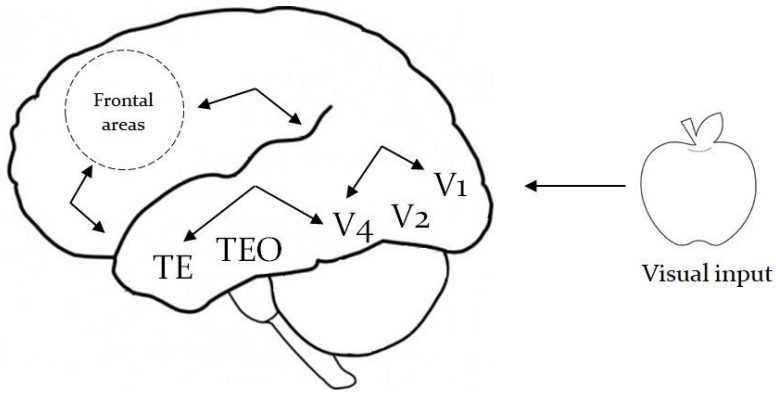

Figure 5. Widespread recurrent processing. Recurrent activation has spread beyond the extrastriate cortex to more frontal areas.[115]

2.3. Block on Local Recurrent Activation

In personal communication (April 21, 2017), Block has suggested that, on his view, local recurrent activation in the occipito-temporal areas of the brain (in the ventral stream) is sufficient for p-consciousness. This claim needs refining.

For one thing, as explained, Block leans towards a *biological account* of p-consciousness. On such an account, p-consciousness is *identical to* a brain state/property. Accordingly, what Block has to say is that local recurrent activation

[115] This illustration is similar to Lamme's (2004, 497).

just is p-consciousness. Mere sufficiency is too weak. For the claim that local recurrent activation in the occipito-temporal areas is sufficient for p-consciousness is compatible, for example, with the claim that the latter empirically correlates with, but is not identical to, the former.[116] (A terminological note. From now on, I will always speak of local recurrent activation *tout court*, rather than local recurrent activation in the occipito-temporal areas/ventral stream).

But not only is Block not allowed to say that local recurrent activation is sufficient for p-consciousness. He is also not allowed to say that local recurrent activation is, *all by itself*, sufficient for p-consciousness. For as I anticipated in chapter 4, Block contends that recurrent activation in the occipito-temporal areas of the brain is only the *core* neural basis of p-consciousness; that is, it is the part of the total neural basis that distinguishes states with *this* phenomenal character from states with *that* phenomenal character (Block 2007a, 482). But what reasons does Block provide for the claim that local recurrent activation is *only* the core neural basis of p-consciousness?

Block does not have a proper argument for this claim. Instead, partly by appealing to the reader's own intuition, and partly by citing Nancy Kanwisher (2001), he simply limits himself to note that, for example, "no one would take recurrent activation of MT/V5 + V1 all by itself in a bottle as sufficient for experience of motion" (2005, 47). Something else must be added to local recurrent activation in order to get the total neural basis of p-consciousness. The total minus core neural basis of p-consciousness, as we have seen, Block hypothesises is the activation of certain connections between the thalamus and the cortex.

[116] To a first approximation, we may say that *a property A empirically correlates with a property B* when, and only when, relative to our experience of A and B, any instance of B is accompanied by an instance of A. (Correlation may be synchronic or diachronic. In the present context, however, it is only the former that seems relevant.) But although empirical correlation suggests sufficiency – any instance of B may be said to suffice for any instance of A – it says nothing about identity: as far as empirical correlation is concerned, A and B may be different properties.

In light of the above, a more accurate description of Block's is that local recurrent activation, *together with the activation of certain connections between the thalamus and the cortex* – the total minus core neural basis – *is identical to* p-consciousness.[117] We are now ready to look at Block's proposal about the functional role of p-consciousness.

3. Block's Take on the Functional Role of Phenomenal Consciousness

Block has touched on the question of what the functional role of p-consciousness is on multiple occasions. In two papers written twenty years apart, for example, he claims that p-consciousness "greases the wheels of [a-consciousness]" (1995, 242; 2015, 157). In his (2009), he notes that "according to the biological account, global broadcasting [is] what [p-]consciousness *does* rather than what consciousness *is*. That is, one function of [p-]consciousness on the biological view is to promote global broadcasting" (2009, 1113). More recently, in personal communication (April 21, 2017), Block suggests that "local recurrent processing is one factor that contributes to the triggering [of global broadcasting]", the other factor being attention.

Block's position on the functional role of p-consciousness appears to have remained consistent over the years, the very last of the abovementioned passages providing a seemingly more precise expression of it than the first two.[118] That

[117] Thus, as anticipated, Block's view differs from Lamme's in one main respect. While Lamme takes local recurrent activation to be identical to be p-consciousness, Block takes it to be identical to p-consciousness's core neural basis only.

[118] In the context of discussing the difference between GWM (as a theory of both a-consciousness and a-consciousness) and the biological approach to p-consciousness, Block (2001) writes: "[t]he theory that consciousness is ventral stream activation plus, for example, neural synchrony, and the theory that consciousness is broadcasting in the global neuronal workspace are instances of the two major rival approaches to consciousness in the philosophical literature, physicalism and functionalism. The key to the difference is that functionalism identifies consciousness with a role, whereas physicalism identifies consciousness with a physical or biological property that fills or implements or realizes that role in humans" (Ibid., 203). (Note that Block uses the term "functionalism" to refer to GWM, and "physicalism" to refer to the biological approach/identity theory). The claim

passage is the one that I will focus in what follows. One thing that stands out immediately is that, strictly speaking, what Block has in mind is the functional role of *a proper part of* p-consciousness (of its core neural basis) rather than that of p-consciousness "as a whole", so to speak (of its total neural basis). For as I noted in several places, Block thinks that local recurrent activation is only "the part of the total neural basis that distinguishes states with [phenomenal character] C from states with other phenomenal characters or phenomenal contents" (2007a, 482). With that in mind, Block's idea is that local recurrent activation figures in the causal explanation of global broadcasting. More accurately, local recurrent activation, which is part of p-consciousness, *contributes to the triggering of global broadcasting*.

Alas, as a picture of the functional role of part of p-consciousness this is still quite vague. For what does it mean to say that local recurrent activation *contributes to the triggering of global broadcasting*, exactly? As far as my knowledge of Block's work goes, the answer is nowhere to be found. Below, I am thus going to do some of the philosophical work for him, and attempt to develop his proposal further. Having argued for what I believe is a plausible, more informative way of understanding such a proposal, I then also argue that, in effect, Block's attempt to identify the functional role of part of p-consciousness fails.

I wish to draw the reader's attention here is Block's claim that, according to the biological approach, p-consciousness realizes a-consciousness or, in other words, that a-consciousness is (part of) the functional role of p-consciousness (call this claim C_1). Should our discussion of Block's view on the functional role of p-consciousness take C_1 into account? I suggest not. For a moment's reflection reveals that C_1 is incompatible with the (2007a) claim that the neural basis of a-consciousness is not included in the neural basis of p-consciousness (call this claim C_2). For if a-consciousness were the functional role of p-consciousness – if p-consciousness realized a-consciousness – the neural basis of p-consciousness and a-consciousness would coincide. Given the plain inconsistency between C_1 and C_2, and given how dear C_2 is to Block, I will assume that C_1 is the result of some sort of confusion on Block's side and, accordingly, disregard it.

3.1. Potential Access Consciousness as the Functional Role of Phenomenal Consciousness

We can start making some progress by clarifying the relationship that I believe Block (e.g., 2007a, 2007b) assumes between *local recurrent activation* and *(visual) perceptual mental states*. Consider the notion of *neural coalition* again. As noted on several occasions, in the present context, we can think of a neural coalition as a temporary assembly of neurons whose overall activation carries information about/represents a specific stimulus/group of stimuli. In effect, then, the overall activation of a neural assembly *is a representation*. That is, roughly, it is an item that carries information about something. In the present context, it carries information about a *visual* stimulus/group of *visual* stimuli.

Let us refer to the overall activation of a neural assembly by the name of 'neural activation'. Some neural activations in the occipito-temporal areas of the brain are recurrent activations. As I explained towards the end of §2.2., this is to say that some neural activations in the occipito-temporal areas involve feedforward and feedback interactions among neurons from different levels of the visual cortical hierarchy. *Qua* neural activations in the sense just specified, local recurrent activations carry information. The information they carry depends on the recurrent interactions that they involve. For example, local recurrent activations involving interactions among neurons from V1 to V5 (or MT) can be thought of as carrying information about a (kind of) motion. In short, then, local recurrent activations can be construed as representations. More strongly, we should be able to construe *all* local recurrent activations as (visual) *perceptual mental states*.

I am not quite sure what a thorough argument for the claim that all local recurrent activations are perceptual mental states would look like. Nonetheless, I think we may be able to at least establish its *prima facie* plausibility. For notice that given Lamme's and Block's commitments, local recurrent activations have at least some of the features that we would happily attribute to perceptual mental states. Here are two examples that come to mind. For one thing, you will recall that local

recurrent activation is thought to enable the binding of different features into a unitary, coherent representation of an object (Lamme 2004, 870). For another, local recurrent activations carry information about the world that is *potentially a-conscious*. I will come back to this below, but you will recall from §2.2.2. that local recurrent activations can, if targeted by attention, spread to more frontal areas. In this sense, they can (potentially) become a-conscious, and hence available for deliberation, report, etc.

Having clarified the relationship that Block thinks there is between local recurrent activations and perceptual mental states, we should now consider the question of what, exactly, it is that we wish to know by asking what the functional role of local recurrent activation is. The reasoning behind this claim goes as follows. Block addresses the question of what the functional role of p-consciousness is by telling us that local recurrent activation (which, remember, is part of p-consciousness) contributes to the triggering of global broadcasting. As already noted, as an answer to the above question, however, this is quite vague. So, what might we do to obtain a more informative answer? One natural suggestion is that we formulate a more specific question. As should be apparent, though, such a question must be constrained by Block's commitments. That is, it must be true to Block's own view.

On the basis of a number of claims that, to my knowledge, are espoused by many philosophers – Block included – in chapter 1 I argued that the phrase 'the functional role of p-consciousness' is to be understood as follows:

> *Functional Role of P-Consciousness Relative to a Mental State M: the fund of appropriate dispositional properties that p-consciousness endows M with.*

If p-consciousness has a functional role at all, then it endows the perceptual mental states (mental states, for short) that have it with a fund of appropriate dispositional properties, where a dispositional property counts as appropriate relative to a psychological theory (see chapter 1, §3). To ask about the functional role of p-

consciousness, then, is to ask with which appropriate dispositions p-consciousness endows the mental states that have it.

Against the background of what I have said thus far – and stressing that, on Block's view, local recurrent activation is *part of p-consciousness* – I believe that the question we are looking for might be formulated as follows: which appropriate dispositions does the property of *being a local recurrent activation* endow the (p-conscious) mental states that have that property with? Or, equally, which appropriate dispositions does a (p-conscious) mental state have in virtue of *being a local recurrent activation*? Assuming that we are on the right track, the next major task is to address these very questions.[119]

At any one time, some mental states are actually a-conscious, others are only potentially a-conscious, and still other mental states are neither actually nor potentially a-conscious. Let's focus on the *potentially a-conscious* ones. What is it to say that a mental state is potentially a-conscious? Well, if a mental state is *actually* a-conscious if its content is available to systems of report, deliberation, and so on, then a mental state is *potentially* a-conscious if its content is *potentially* available to the same systems. Consider a visual perceptual experience as of a tomato on the counter. That experience is potentially a-conscious if the content *that there is a tomato on the counter* is potentially available to the consuming systems. Assuming GWM, we may say that *a mental state is potentially a-conscious if its content can be broadcast in the workspace*. Crucially, like a-consciousness, potential a-consciousness is a dispositional property. That is to say that a mental state's being potentially a-consciousness is just a matter of its being *poised* to become a-conscious.

[119] It might be objected that the right question to ask is this: which appropriate dispositions does a p-conscious mental state have in virtue of *having a local recurrent activation?* Given the relationship between (visual perceptual) mental states and local recurrent activation, however, such a question does not really make much sense.

Block's general answer to the general question of what the functional role of p-consciousness is, is that local recurrent activation (which is part of p-consciousness) contributes to the triggering of global broadcasting. Armed with i) the more specific question of which dispositional properties p-consciousness endows the mental states that have it with, and ii) the considerations offered in the paragraph immediately above, a natural suggestion now be the following: on Block's view, a mental state's *property of being a local recurrent activation is the categorical basis of its being potentially broadcastable in the workspace* – and, thus, the categorical basis of the state's being potentially a-conscious.[120] Potential a-consciousness, in other words, seems a good candidate for being the functional role of being a local recurrent activation (and hence of part of p-consciousness). But is it, really? Below I argue for a negative answer.

4. Potential A-Consciousness Is Not the Functional Role of Phenomenal Consciousness

As with any dispositional notion, one can offer what is sometimes called a *simple conditional analysis* of potential a-consciousness (Lewis 1997, 143). By that, I mean an analysis along the following lines: "[s]omething *x* is disposed at time *t* to give response *r* to stimulus *s* iff, if *x* were to undergo stimulus *s* at time *t, x* would give response *r*" (Ibid.).[121] Consider, as in chapter 3, the fragility of wine glass. In general, to say that wine glass is fragile is to say that it is disposed to break under certain circumstances. On the simple conditional analysis of wine glass's fragility, wine glass is disposed to break iff, if it were struck, it would break (Ibid.). Having

[120] Here, again, I am making the (unharmful) assumption that the functionality condition that I argued for in chapter 3 is satisfied. For if it weren't, being potentially broadcastable would not be sufficient for being potentially a-conscious.

[121] For my purposes, I assume the truth of the simple conditional analysis. Note, however, that such an assumption is not uncontroversial. Lewis (1997) himself, for example, argues that such an analysis is inadequate.

outlined the simple conditional analysis, we can now ask what a simple conditional analysis of potential a-consciousness would look like.

As I have noted on several occasions, global broadcasting is thought to occur "when some sort of threshold of neural [activation] is reached" (Carruthers 2015b, 4). In this context, attention plays a crucial role: it amplifies (some) activations in the occipito-temporal areas of the brain until their content becomes globally broadcast (e.g., Dehaene and Changeux 2004, 1147; Carruthers 2015b, 4). As with the striking of a glass, or the pouring of some salt in a glass of water, the targeting of an activation by attention "triggers", to use Kriegel's (2015, n/a) terminology, the "manifestation" of the disposition. That is, attentional amplification makes it so that we go from *potential* a-consciousness to *actual* a-consciousness. With that in mind, a conditional analysis of potential a-consciousness will plausibly look something like the following: to say that an *activation/representational mental state* is potentially a-conscious is to say that the state *would* become (actually) access conscious – or, equally, that its content *would* be broadcast in the workspace – *were* it targeted by attention.

According to GWM, when a local activation – an activation that is confined to the occipito-temporal areas in the "back of the head" – is weak, the global broadcasting of its content *cannot* occur (such an activation, as we know, Dehaene et al. (2006) would call *subliminal*). When a local activation is sufficiently strong, however, it *can* spread to more frontal areas (such an activation, Dehaene et al. (2006) would call *preconscious*).[122] It is thus *because* a mental state is a sufficiently strong local activation that its content is poised to be broadcast in the workspace. The relation between being a sufficiently strong local activation and potential a-consciousness can be made clearer with the aid of the conditional analysis offered in the previous paragraph: it is *in virtue of* the fact that a mental state is a

[122] The notion of sufficiency here is obviously vague. How strong counts as *sufficiently strong* for global broadcasting is an empirical matter.

sufficiently strong local activation that, were the state targeted by attention, it would become a-conscious – or, equally, that its content would be broadcast in the workspace. A mental state's property of *being a sufficiently strong local activation* is thus the categorical basis of its being potentially a-conscious.

But if a mental state's property of being a sufficiently strong local activation is the categorical basis of potential a-consciousness, then we can readily construe potential a-consciousness as the *functional role of being a sufficiently strong local activation*. That is, we can construe potential a-consciousness as the appropriate disposition that a mental state's property of being a sufficiently strong local activation endows the state with. In the present context, an important consequence of this is that there seems to be no role left for the property of being a local recurrent activation to play vis-à-vis potential a-consciousness. Differently put, it seems that what makes a mental state potentially a-conscious is not its being a local recurrent activation, but its *being a sufficiently strong local activation.* From this, we may then conclude that potential a-consciousness is *not* the functional role of local recurrent activation (and hence of part of p-consciousness).

To sum up, I started by speculating that, according to Block, asking what the functional role of local recurrent activation is, is to ask how the property of *being a local recurrent activation* disposes a representational mental state. I then suggested that Block's claim that local recurrent activation contributes to the triggering of global broadcasting should be understood as the claim that a mental state's being a local recurrent activation is the categorical basis of its being potentially a-conscious. The latter claim turned out to be wrong, however. For potential a-consciousness is not the functional role of a mental state's being a local recurrent activation, but of its being a sufficiently strong local activation. In the next section, I introduce and respond to three objections.

4.1. Objections and Replies

Block might concede that potential a-consciousness is not the functional role of being a local recurrent activation. Yet, he might also insist that the latter can still be said to play a role vis-à-vis the former. If he is right, he might then still be able to salvage the more general claim that local recurrent activation contributes to the triggering of global broadcasting. But in what sense could being local recurrent activation be said to play a role vis-à-vis potential a-consciousness? By accounting for *part* of the categorical basis of potential a-consciousness. That is, more precisely, by accounting for a local activation's property of *being sufficiently strong*.

Before explaining what exactly it is that I have in mind, let me immediately note that even if such a response were successful, my argument would be defeated only in part. For as I hinted at above, that response is only an attempt at salvaging the idea that being a local recurrent activation figures in the causal explanation of potential a-consciousness, not that the latter is the functional role of the former. That being said, let me develop the response in question.

Recall the characterization of recurrent activation that I offered in §2.2.: recurrent activation – be it local or widespread – is a kind of neural activation involving bi-directional (feedforward and feedback) interactions between neurons from different levels of the visual system hierarchy. Lamme (2004, 870, 2009, 559) as well as Block (2007a, 486) note that, because it involves cross-level feedforward and feedback interactions, recurrent activation is typically a *strong* form of activation. Neurons from level L_n of the visual system hierarchy interact with neurons of level L_{n+1} L_{n-1}, for example, and, in so doing, they mutually excite each other, leading to an overall high level of activation (Lamme 2009, 559). In light of the above, Block might plausibly suggest that an activation's property – be the activation local or widespread – of *being recurrent* (*recurrency*, for short) accounts

for its being *strong*, and from then proceed to argue that recurrency accounts for a local activation's *being sufficiently strong*.[123]

But I doubt such a move would work. For notice that we are now being told that what accounts for a local activation's property of being sufficiently strong is not its being a local recurrent activation, but *recurrency*. In other words, we are being told that the reason why a local activation is sufficiently strong is just that it is recurrent. But the objection had it that a local activation's being sufficiently strong was to be explained by *its being a local recurrent activation*. As a consequence, it is unsuccessful.

The second objection that I would like to consider is that my development of Block's view rests on a fundamental misunderstanding. I have been taking Block's claim that "local recurrent processing is one factor that contributes to the triggering [of global broadcasting]" (the other factor being attention) as representative of his view. I have also relied on the claim that, on Block's view on p-consciousness, local recurrent activation is *part* of what p-consciousness *is*, the other part being its total minus core neural basis. But the objection has it that the claim that local recurrent activation is *part* of p-consciousness is mistaken. On a correct understanding of Block's take on p-consciousness, local recurrent activations are just states that are *made p-conscious* by something else. In this sense, although local recurrent activation plays a key role in his account of p-consciousness, it is just wrong to say that it is a *part* of p-consciousness.

[123] I deliberately leave the nature of the relation between *recurrency* and *being sufficiently strong* vague. The reason is simply that I am not sure what that is. Intuitively, being sufficiently strong is an occurrent, rather than dispositional property. As a consequence, it seems that the relationship between the properties in question cannot be an instance of the categorical basis-disposition relation. That, however, does not rule out that one property – recurrency – may be said to *ground* the other property – sufficient strength. Still, I am not certain that the relation between the two may be seen as one of grounding.

In effect, the view just sketched seems to be the view endorsed by Block in a passage of a recent paper (2009, 1112). There, Block suggests that local recurrent activations are only *potentially* p-conscious representations. And to the question of what *makes those representations p-conscious*, he answers by making reference to what, as explained, he takes to be the total minus core neural basis of p-consciousness; i.e., the activation of certain connections between the cortex and the thalamus. According to this picture, it thus seems wrong to say that local recurrent activation is *part of* p-consciousness. In fact, local recurrent activations are just *made p-conscious* by the activation of the cortex-thalamus connections.

In my defence, I think that the passage just mentioned does not accurately reflect Block's take on p-consciousness. This is for three reasons. First, there is evidence suggesting that, under general anaesthesia, recurrent activation in general is either partly or fully suppressed (Lamme 2004, 868). Now, as I explained in the previous chapter, general anaesthetics are thought to work by disabling – suppressing the activation of – the very cortex-thalamus connections mentioned in the paragraph above. This suggests that there can be no recurrent activation *without* the activation of those connections. The latter claim, in turn, suggests that there can be no such things as *potentially p-conscious* local recurrent activations, as Block claims in the passage above.

But maybe Block is unaware of that.[124] Or maybe the claim that there can be no such things as potentially p-conscious local recurrent activations is still compatible with the claim that the activation of the cortex-thalamus connections is what makes local recurrent activations p-conscious. Luckily, as mentioned, there are two further points in my support. First of all, if local recurrent activation were

[124] Interestingly, if Block *is* aware that there is evidence suggesting that when the cortex-thalamus connections are disabled, recurrent activation is suppressed, that might explain why he takes the studies reviewed by Alkire and Miller (2005) to show that there can be no p-consciousness without the activation of the cortex-thalamus connections (see footnote n.93).

not part and parcel of p-consciousness, then Block's claim that p-consciousness often *has representational content* would be unintelligible. In chapter 2, I explained that Block thinks that p-consciousness is often like the paint that represents Marat in the painting *La Mort de Marat*. In other words, p-consciousness is often what does (at least part of) the representing in a mental state. But the only way I can make sense of that claim is if local recurrent activation is *part* of p-consciousness. For unless Block wishes to say that it is the activation of the cortex-thalamus connections that do (part of) the representing, then the only candidate left for the job is local recurrent activations.

Secondly, as the reader may have noticed, the claim that local recurrent activation is not part of p-consciousness flies right in the face of Block's claims that (i) local recurrent activation is the core neural basis of p-consciousness, and (ii), that p-consciousness is identical to its total neural basis (i.e., the core, plus total minus core neural basis). The conjunction of (i) and (ii), as noted in several occasions, I take to be the core of Block's take on p-consciousness. But why would the claim that local recurrent activation is not part of p-consciousness fly in the face of the conjunction of those two claims? For the latter make it apparent that local recurrent activation *is* part and parcel of p-consciousness. Taken together, the points that I have just made should suffice to show that, on a correct understanding of Block's account of p-consciousness, local recurrent activation is *part* of p-consciousness, the other part being its total minus core neural basis.

I conclude that my argument is safe, and that potential a-consciousness is not the functional role of being a local recurrent activation.

4.2. Is Block's Question the Right Question to Begin With?

If my development of Block's view on the functional role of p-consciousness is correct, I suspect that there may be a further – and maybe even bigger – issue for that view. Here is what I have in mind. I have argued that we should understand

Block's proposal as follows: *potential a-consciousness is the functional role of a perceptual mental state's being a local recurrent activation*. But if this is really how Block's view should be understood, such a view tells us nothing about how a mental state's *being p-conscious* appropriately disposes the state. What it tells us, instead, is how a mental state's *being (part of) p-consciousness* disposes the state.

Whether this is an issue for Block's view boils down to what we think an account of the functional role of p-consciousness should tell us. Differently put, it boils down to how we understand the phrase 'the functional role of p-consciousness'. According to the definition that I put forward in chapter 1, here is how that phrase is to be understood:

> *Functional Role of P-Consciousness Relative to a Mental State M: the fund of appropriate dispositional properties that p-consciousness endows M with.*

For the purposes of my development of Block's proposal, in §3.1., I implicitly assumed that the above definition was compatible with construing the functional role of p-consciousness as the fund of appropriate dispositional properties that a mental state has in virtue of *being part of p-consciousness*. The reason I did so was that I saw no other way of making sense of Block's proposal. In the specific, I saw no way of understanding how Block might attribute a functional role to local recurrent activation, other than by suggesting that *being a local recurrent activation* (and, hence, being part of p-consciousness) is that property that endows a (p-conscious) mental state with certain appropriate dispositions.

But the reader may have noticed that construing the functional role of p-consciousness in this way is, in fact, *incompatible* with the definition in question. For as I explained in chapter 1, to say that the functional role of p-consciousness is the fund of appropriate dispositional properties that p-consciousness endows M with, is to say that the functional role of p-consciousness is *the fund of appropriate dispositional properties that M's being p-conscious endows M with*.

That definition, as I argued in the same chapter, should capture what most philosophers wish to know when they ask about what the functional role of p-consciousness is. On top of that, it should be in line with what we have in mind when we ask about the functional role of some non-mental property in some non-mental system.

But if both my definition and my development of Block's view are correct, then it follows that the latter rests on a fundamental misunderstanding of the notion of functional role of p-consciousness. In this sense, the question that I have attributed to Block – the question of which appropriate dispositions the property of being a local recurrent activation endows a mental state with – is not the right question to ask if what one is seeking to understand is what the functional role of p-consciousness is.

5. Concluding Remarks

In this chapter, I have considered, and assessed, Block's view on the functional role of p-consciousness. As we have seen, although Block has touched on the question of what that functional role is on several occasions, he has never provided a detailed account. Armed with the notions introduced in chapter 4, along with the notion of local recurrent activation introduced in §2, I have thus made an attempt at developing Block's proposal further. First, I have argued that, on Block's view, the question of what part of the functional role of p-consciousness is, boils down to the following question: which appropriate dispositions does the property of *being a local recurrent activation* endow the p-conscious mental states that have that property with? After that, I have argued that Block's answer to that question involves *potential a-consciousness*. In the specific, according to Block, the property of being a local recurrent activation endows the relevant states with the disposition to become a-conscious. That view, however, turns out to be untenable. Why? For what disposes a mental state to become a-conscious is not the state's being a local recurrent activation, but its being a sufficiently strong activation. Last, I have

argued that there may be an even bigger issue for Block. If my development of his view is correct, the issue is that that view rests on a fundamental misunderstanding of the notion of functional role of p-consciousness.

Concluding Remarks

My main aim in this thesis has been to explore what in the introduction I called the *functional question*: assuming that p-consciousness has a functional role *at all*, *what* is it? Against the background of Block's distinction between p-consciousness and a-consciousness, I have considered two main views: Kriegel's and Block's very own. Both authors contend that p-consciousness figures, somehow, in the causal explanation of a-consciousness. According to Kriegel, a mental state's being p-conscious is that property in virtue of which that state is also (*actually*) a-conscious. According to Block, a mental state's being part of p-consciousness is that property in virtue of which that state is *potentially* a-conscious.

As I have argued, both accounts are untenable. As concerns Kriegel's, I have made two main points. First, the account is incompatible with GWM (the Global Workspace Model). Given that the GWM is a well-established model of a-consciousness, this gives us reason to reject the account. Second, the account entails a view of the relation between a-consciousness and attention that, in light of certain empirical discussed in chapter 2, is highly implausible.

Things are a bit more complicated as concerns Block's proposal. As we have seen, although Block has touched on the functional question on several occasions, he has never really offered what one would usually think of as a "proper" account. I have thus further refined and developed his view, and I have argued that it should be further articulated as the view that being a local recurrent activation is the categorical basis of potential a-consciousness. In other words, the idea is that a

(visual perceptual) mental state's being a local recurrent activation is that property in virtue of which the state is also potentially a-conscious. Alas, this view, too, has turned out to be unsatisfactory. This is for two main reasons. First, the categorical basis of potential a-consciousness is not a state's property of being a local recurrent activation, but a state's property of being a sufficiently strong local activation. Second, given the definition of the notion of functional role of p-consciousness that I provided in chapter 1, the claim that potential a-consciousness is the functional role of being a local recurrent activation is not an answer to the functional question *at all*.

Where does that leave us? Well, although Kriegel's and Block's accounts may not be the accounts to look at for a promising answer to the functional question, future exploration might profit greatly by continuing to look for the functional role of p-consciousness at the interface between the latter and a-conciousness. This is for two reasons. First, the noted continuity between Kriegel's and Block's accounts seems to be more than just an accident. That is, the fact that both accounts take (or claim to take) p-consciousness to figure, somehow, in the causal explanation of a-consciousness seems to reveal an underlying, shared intuition. The intuition is that what p-consciousness contributes to the functioning of the mind, it must contribute vis-à-vis our capacity to access our mental states, and use their content in reasoning, deliberation, etc. Of course, intuitions are just that: intuitions. But it would be silly to abandon them at the current stage of research – which, as far as the functional role of p-consciousness is concerned, is still a very early stage.

Second, to look for the functional role of p-consciousness at the interface between the latter and a-consciousness is, ideally, to look for that functional role at the interface between p-consciousness and a-consciousness as accounted for by GWM. For if there is one positive thing that this thesis has hopefully been able to show is that although we still know very little about p-consciousness, we do know

a great deal about a-consciousness. In this sense, GWM provides us with a theoretically and empirically solid background against which we may keep testing our hypotheses about the functional role of p-consciousness. What I mean can be easily understood from my treatment of Kriegel's account: if our hypotheses about the functional role of p-consciousness turn out to be incompatible with GWM, that should probably sound a note of warning.

References

Alkire, Michael T., and Jason Miller. 2005. "General Anesthesia and the Neural Correlates of Consciousness." *Progress in Brain Research* 150: 229–597.

Armstrong, David Malet. 1968. *A Materialist Theory of the Mind*. London: Routledge and Kegan Paul.

Atkinson, Anthony P., and Martin Davies. 1995. "Consciousness without Conflation." *Behavioral and Brain Sciences* 18 (2): 248–49.

Baars, Bernard J. 1988. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

———. 2005. "Global Workspace Theory of Consciousness: Toward a Cognitive Neuroscience of Human Experience." *Progress in Brain Research* 150: 45–53.

Baddeley, Alan D. 2007. *Working Memory, Thought, and Action*. Oxford: Oxford University Press.

Baddeley, Alan, and Graham J. Hitch. 1974. "Working Memory." In *Recent Advances in Learning and Motivation*, 8:47–90. New York: Academic Press.

Balog, Katalin. 2007. "Access to Phenomenality: A Necessary Condition of Phenomenality?" *Behavioral and Brain Sciences* 30 (5–6): 499–500.

Bayne, Tim. 2009. "Functions of Consciousness." In *The Oxford Companion to Consciousness*, 315–18. Oxford: Oxford University Press.

Beck, Diane M, and Sabine Kastner. 2005. "Stimulus Context Modulates Competition in Human Extrastriate Cortex." *Nature Neuroscience* 8 (8): 1110–16.

Bickle, John. 2016. "Multiple Realizability." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/spr2016/entries/multiple-realizability/.

Block, Ned. 1995. "On a Confusion about a Function of Consciousness." *Behavioral and Brain Sciences* 18 (2): 227–47.

———. 1996. “Mental Paint and Mental Latex.” *Philosophical Issues* 7: 19–49.

———. 2001. “Paradox and Cross Purposes in Recent Work on Consciousness.” *Cognition* 79 (1): 197–219.

———. 2002. “Concepts of Consciousness.” In *Philosophy of Mind: Classical and Contemporary Readings*, edited by David Chalmers, 206–2018. Oxford: Oxford University Press.

———. 2005. “Two Neural Correlates of Consciousness.” *Trends in Cognitive Sciences* 9 (2): 46–52.

———. 2007a. “Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience.” *Behavioral and Brain Sciences* 30 (5–6): 481–99.

———. 2007b. *Consciousness, Function, and Representation*. Cambridge, Mass.: The MIT Press.

———. 2007c. “Overflow, Access, and Attention.” *Behavioral and Brain Sciences* 30 (5–6): 530–48.

———. 2009. “Comparing the Major Theories of Consciousness.” In *The Cognitive Neurosciences*, edited by Michael S. Gazzaniga, 4th ed., 1111–22. Cambridge, Mass.: The MIT Press.

———. 2011. “Perceptual Consciousness Overflows Cognitive Access.” *Trends in Cognitive Sciences* 15 (12): 567–75.

———. 2015. “Consciousness, Big Science, and Conceptual Clarity.” In *The Future of the Brain: Essays by the World’s Leading Neuroscientists*, edited by Jeremy Freeman and Gary Marcus, 161–76. Princeton: Princeton University Press.

Block, Ned, and Cynthia MacDonald. 2008. “Consciousness and Cognitive Access.” *Proceedings of the Aristotelian Society* 108: 289–317.

Bourget, David, and Angela Mendelovici. 2014. “Tracking Representationalism.” In *Philosophy of Mind: The Key Thinkers*, edited by Andrew Bayley, 209–35. London: Bloomsbury.

Bressan, Paola, and Silvia Pizzighello. 2008. “The Attentional Cost of Inattentional Blindness.” *Cognition* 106 (1): 370–83.

Broadbent, Donald. 1958. *Perception and Communication*. London: Pergamon Press.

Bronfman, Zohar Z., Noam Brezis, Hilla Jacobson, and Marius Usher. 2014. “We See More than We Can report: ‘Cost free’ Color Phenomenality Outside Focal Attention.” *Psychological Science* 25 (7): 1394–1403.

Broogard, Berit. 2011. “Are There Unconscious Perceptual Processes?” *Consciousness and Cognition* 20 (2): 449–63.

Burge, Tyler. 2007. *Foundations of Mind*. Oxford: Clarendon Press.

Carruthers, Peter. 2002. *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge University Press.

———. 2006. *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought*. Oxford: Oxford University Press.

———. 2011. *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.

———. 2013. “Evolution of Working Memory.” *Proceedings of the National Academy of Sciences of the United States of America* 110: 10371–78.

———. 2014. “On Central Cognition.” *Philosophical Studies* 170 (1): 143–62.

———. 2015a. *Centered Mind: What the Science of Working Memory Shows Us About the Nature of Human Thought*. Oxford: Oxford University Press.

———. 2015b. “Block’s Overflow Argument.” *Pacific Philosophical Quarterly*, January, n/a-n/a.

———. 2016. “Higher-Order Theories of Consciousness.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/fall2016/entries/consciousness-higher/.

Chalmers, David J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

Chalmers, David J. 1997. “Availability: The Cognitive Basis of Experience.” *Behavioral and Brain Sciences* 20 (1): 148–49.

———. 2000. “What Is a Neural Correlate of Consciousness?” In *Neural Correlates of Consciousness*, edited by Thomas Metzinger, 17–39. Cambridge, Mass.: The MIT Press.

———. 2010. *The Character of Consciousness*. Oxford: Oxford University Press.

Chudnoff, Elijah. 2012. “Presentational Phenomenology.” In *Consciousness and Subjectivity*, edited by Sofia Miguens and Gerhard Preyer, 51–72. Frankfurt: Ontos.

Coates, Paul, and Sam Coleman. 2015. *Phenomenal Qualities: Sense, Perception, and Consciousness*. Oxford: Oxford University Press.

Cohen, Michael A., and Daniel C. Dennett. 2011. “Consciousness Cannot Be

Separated from Function." *Trends in Cognitive Sciences* 15: 358–64.

Cowan, Nelson. 1988. "Evolving Conceptions of Memory Storage, Selective Attention, and Their Mutual Constraints Within the Human Information-Processing System." *Psychological Bulletin* 104 (2): 163–91.

———. 2005a. *Working Memory Capacity*. Hove: Psychology Press.

———. 2005b. "Working Memory Capacity Limits in a Theoretical Context." In *Human Learning and Memory: Advances in Theory and Applications*, edited by C. Izawa and N. Ohta, 155–75. Mahwah: Lawrence Erlbaum Associates.

———. 2008. "What Are the Differences between Long-Term, Short-Term, and Working Memory?" *Progress in Brain Research* 169: 323–38.

Crick, Francis, and Christof Koch. 2003. "A Framework for Consciousness." *Nature Neuroscience* 6: 119–26.

Cummins, Robert. 1975. "Functional Analysis." *The Journal of Philosophy* 72: 741–65.

Dehaene, Stanislas, and Jean-Pierre Changeux. 2004. "Neural Mechanisms for Access to Consciousness." In *The Cognitive Neurosciences*, edited by Michael S Gazzaniga, 1145–57. Cambridge, Mass.: The MIT Press.

———. 2011. "Experimental and Theoretical Approaches to Conscious Processing." *Neuron* 70 (2): 200–227.

Dehaene, Stanislas, Jean-Pierre Changeux, and Lionel Naccache. 2011. "The Global Neuronal Workspace Model of Conscious Access: From Neuronal Architectures to Clinical Applications." In *Characterizing Consciousness: From Cognition to the Clinic?*, edited by Stanislas Dehaene and Yves Christen, 55–84. Berlin: Springer.

Dehaene, Stanislas, Jean-Pierre Changeux, Lionel Naccache, Jérôme Sackur, and Claire Sergent. 2006. "Conscious, Preconscious, and Subliminal Processing: A Testable Taxonomy." *Trends in Cognitive Sciences* 10 (5): 204–11.

Dehaene, Stanislas, Michel Kerszberg, and Jean-Pierre Changeux. 1998. "A Neuronal Model of a Global Workspace in Effortful Cognitive Tasks." *Procedings of the National Academy of Sciences* 95 (24): 14529–34.

Dehaene, Stanislas, and Lionel Naccache. 2001. "Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework." *Cognition* 79 (1): 1–37.

Dennett, Daniel C. 1991. *Consciousness Explained*. New York: Little, Brown & Company.

———. 2001. “Are We Explaining Consciousness Yet?” *Cognition* 79 (1–2): 221–37.

Driver, Jon, and Jason B. Mattingley. 1998. “Parietal Neglect and Visual Awareness.” *Nature Neuroscience* 1 (1): 17–22.

Driver, Jon, and Patrik Vuilleumier. 2001. “Perceptual Awareness and Its Loss in Unilateral Neglect and Extinction.” *Cognition* 79 (1): 39–88.

Fahrenfort, Johannes J., and Victor A.F. Lamme. 2012. “A True Science of Consciousness Explains Phenomenology: Comment on Cohen and Dennett.” *Trends in Cognitive Sciences* 16 (3): 138–39.

Flanagan, Owen J. 1992. *Consciousness Reconsidered*. Cambridge, Mass.: The MIT Press.

Fodor, Jerry A. 1974. “Special Sciences (or: The Disunity of Science as a Working Hypothesis).” *Synthese*, 97–115.

———. 1981. *The Modularity of Mind*. Cambridge, Mass.: The MIT Press.

Gazzaniga, Michael S., Richard B. Ivry, and George R. Mangun. 2014. *Cognitive Neuroscience: The Biology of the Mind*. New York: W. W. Norton & Company.

Godfrey-Smith, Peter. 1996. *Complexity and the Function of Mind in Nature*. Cambridge, Mass.: The MIT Press.

Gulick, Robert Van. 1994. “Deficit Studies and the Function of Phenomenal Consciousness.” In *Philosophical Psychopathology*, 25–49. Cambridge, Mass: The MIT Press.

———. 2009. “Concepts of Consciousness.” In *The Oxford Companion to Consciousness*, edited by Tim Bayne, Axel Cleeremans, and Patrick Wilken, 163–66. Oxford: Oxford University Press.

Heywood, Charles, and Robert W. Kentridge. 2009. “Akinetopsia.” In *Oxford Companion to Consciousness*, edited by Tim Bayne, Axel Cleeremans, and Patrick Wilken, 24–25. Oxford: Oxford University Press.

Jacob, Pierre. 2007. “What is ‘Cognitive Accessibility’ Accessibility To?” *Behavioral and Brain Sciences* 30 (5/6): 508.

Kanwisher, Nancy. 2001. “Neural Events and Perceptual Awareness.” *Cognition* 79 (1): 89–113.

Kastner, Sabine, Mark A. Pinsk, Peter De Weerd, Robert Desimone, and Leslie G. Ungerleider. 1999. “Increased Activity in Human Visual Cortex during Directed Attention in the Absence of Visual Stimulation.” *Neuron* 22 (4): 751–61.

Kastner, Sabine, Peter De Weerd, Robert Desimone, and Leslie G Ungerleider. 1998. "Mechanisms of Directed Attention in the Human Extrastriate Cortex as Revealed by Functional MRI." *Science* 282 (5386): 108 LP-111.

Kastner, Sabine, Peter De Weerd, Mark A Pinsk, M Idette Elizondo, Robert Desimone, and Leslie G Ungerleider. 2001. "Modulation of Sensory Suppression: Implications for Receptive Field Sizes in the Human Visual Cortex." *Journal of Neurophysiology* 86 (3): 1398–1411.

Kim, Jaegwon. 1996. *Philosophy of Mind*. Boulder: Westview Press.

———. 2010. *Essays in the Metaphysics of Mind*. Oxford: Oxford University Press.

Koch, Christof. 2004. *The Quest for Consciousness : A Neurobiological Approach*. Englewood: Roberts & Company.

Kouider, Sid, and Jérôme Sackur. 2014. "Consciousness and Its Access Mechanisms." In *The Cognitive Neurosciences*, edited by Michael S. Gazzaniga and George R. Mangun, 855–63. Cambridge, Mass.: The MIT Press.

Kriegel, Uriah. 2004a. "Consciousness and Self-Consciousness." *The Monist* 87 (2): 182–205.

———. 2004b. "The Functional Role of Consciousness: A Phenomenological Approach." *Phenomenology and the Cognitive Sciences* 3 (2): 171–93.

———. 2007. "Consciousness: Phenomenal Consciousness, Access Consciousness, and Scientific Practice." In *Handbook of the Philosophy of Psychology and Cognitive Science*, edited by Paul R Thagard, 195–217. Amsterdam: Elsevier.

———. 2009. *Subjective Consciousness: A Self-Representational Theory*. Oxford: Oxford University Press.

———. 2011. "Self-Representation and the Explanatory Gap." In *Consciousness and the Self: New Essays*, edited by JeeLoo Liu and John Perry, 51–74. Cambridge: Cambridge University Press.

———. 2012. *Précis of Subjective Consciousness: A Self-Representational Theory*. *Philosophical Studies*. Vol. 159.

———. 2015. "Phenomenality as the Categorical Basis of Accessibility." Paper presented at the *Consciousness and Accessibility Conference, 20-21 November*. Paris.

Kripke, Saul. 1980. *Naming and Necessity*. Cambrdige, Mass.: Harvard University Press.

Lamme, Victor A.F. n.d. "Can Neuroscience Reveal the True Nature of

Consciousness?" *Unpublished Article*, 1–12.

———. 2004. "Separate Neural Definitions of Visual Consciousness and Visual Attention: A Case for Phenomenal Awareness." *Neural Networks* 17 (5–6): 861–72.

———. 2006. "Towards a True Neural Stance on Consciousness." *Trends in Cognitive Sciences* 10 (11): 494–501.

———. 2007. "Sue Ned Block! Making a Better Case for P-Consciousness." *Behavioral and Brain Sciences* 30 (5–6): 511–12.

———. 2009. "Re-Entrant Processing." In *The Oxford Companion to Consciousness*, edited by Tim Bayne, Alex Cleeremans, and Patric Wilken, 556–60. Oxford: Oxford University Press.

Lamme, Victor A.F., and Pieter R. Roelfsema. 2000. "The Distinct Modes of Vision Offered by Feedforward and Recurrent Processing." *Trends in Neurosciences* 23 (11): 571–79.

Lamme, Victor A F. 2003. "Why Visual Attention and Awareness Are Different." *Trends in Cognitive Sciences* 7 (1): 12–18.

Lamme, Victor A F, Hans Supèr, and Henk Spekreijse. 1998. "Feedforward, Horizontal, and Feedback Processing in the Visual Cortex." *Current Opinion in Neurobiology* 8 (4): 529–35.

Landman, Rogier, Henk Spekreijse, and Victor A. F. Lamme. 2003. "Large Capacity Storage of Integrated Objects before Change Blindness." *Vision Research* 43 (2): 149–64.

Leibniz, Gottfried W., and Robert Latta. 1898. *The Monadology and Other Philosophical Writings*. Oxford: Clarendon Press.

Levin, Janet. 2016. "Functionalism." In *Stanford Encyclopedia of Philosophy*, edited by Edward N Zalta. https://plato.stanford.edu/entries/functionalism/.

Levine, Joseph. 2001. *Purple Haze: The Puzzle of Consciousness*. *Purple Haze: The Puzzle of Consciousness*. Oxford: Oxford University Press.

———. 2003. "Experience and Representation." In *Consciousness: New Philosophical Perspectives*, edited by Quentin Smith and Alexandar Jokic, 57–76. Oxford: Oxford University Press.

Lewis, David K. 1972. "Psychophysical and Theoretical Identifications." *Australasian Journal of Philosophy* 50 (3): 249–58.

———. 1980. "Mad Pain and Martian Pain." In *Readings in the Philosophy of*

Psychology, edited by Ned Block, 216–22. Cambrdige, Mass.: Harvard University Press.

———. 1997. "Finkish Dispositions." *The Philosophical Quarterly* 47 (187): 143–58.

Locke, John. 1975. *An Essay Concerning Human Understanding*. Oxford: Clarendon Press.

Lycan, William G. 1996. *Consciousness and Experience*. Cambridge, Mass.: The MIT Press.

Mack, Arien, and Irvin Rock. 1998. *Inattentional Blindness*. Cambridge, Mass.: The MIT Press.

Marcel, Anthony J. 1986. "Consciousness and Processing: Choosing and Testing a Null Hypothesis." *Behavioral and Brain Sciences* 9 (1): 40–41.

Marcel, Anthony J. 1988. "Phenomenal Experience and Functionalism." In *Consciousness in Contemporary Science*, edited by Edoardo Bisiach, 121–58. Oxford: Clarendon Press.

Maudlin, Tim. 2007. *The Metaphysics Within Physics*. Oxford: Clarendon Press.

Mayes, Andrew R., and Neil Roberts. 2001. "Theories of Episodic Memory." *Pilosophical Transactions of the Royal Society of London* 356: 1395–1408.

McCormick, Peter A. 1997. "Orienting Attention without Awareness." *Journal of Experimental Psychology* 23 (1): 168–80.

Menzel, Christopher. 2016. "Possible Worlds." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N Zalta. https://plato.stanford.edu/entries/possible-worlds/.

Miller, Earl K., Paul M. Gochin, and Charles G. Gross. 1993. "Suppression of Visual Responses of Neurons in Inferior Temporal Cortex of the Awake Macaque by Addition of a Second Stimulus." *Brain Research* 616 (1–2): 25–29.

Milner, David, and Mel Goodale. 1995. *The Visual Brain in Action*. Oxford: Oxford University Press.

———. 2003. *Sight Unseen: An Exploration of Conscious and Unconscious Vision*. Oxford: Oxford University Press.

Mole, Christopher. 2011. "The Metaphysics of Attention." In *Attention: Philosophical and Psychological Essays*, 60–77. Oxford: Oxford University Press.

Naccache, Lionel, and Stanislas Dehaene. 2007. "Reportability and Illusions of

Phenomenality in the Light of the Global Neuronal Workspace Model." *Behavioral and Brain Sciences* 30 (5–6): 518–20.

Nagel, Thomas. 1974. "What Is It like to Be a Bat?" *The Philosophical Review* 83 (4): 435–50.

Neander, Karen. 1991. "Functions as Selected Effects: The Conceptual Analyst's Defense." *Philosophy of Science* 58 (2): 168–84.

Neisser, Ulric. 1967. *Cognitive Psychology*. Upper Saddle River: Prentice-Hall.

Piccinini, Gualtiero. 2007. "The Ontology of Creature Consciousness: A Challenge for Philosophy." *Behavioral and Brain Sciences* 30 (1): 103–4.

———. 2010. "The Mind as Neural Software? Understanding Functionalism, Computationalism, and Computational Functionalism." *Philosophy and Phenomenological Research* 81 (2): 269–311.

Polger, Thomas W. 2004. *Natural Minds*. Cambridge, Mass., Mass.: The MIT Press.

———. 2012. "Functionalism as a Philosophical Theory of the Cognitive Sciences." *Wiley Interdisciplinary Reviews: Cognitive Science* 3 (3): 337–48.

Polger, Thomas W., and Owen J. Flanagan. 2002. "Consciousness, Adaptation and Epiphenomenalism." In *Consciousness Evolving*, edited by James H. Fetzer, 21–42. Amsterdam: John Benjamin.

Postle, Bradley R. 2006. "Working Memory as an Emergent Property of the Mind and Brain." *Neuroscience* 139 (1): 23–38.

———. 2015. *Essentials of Cognitive Neuroscience*. Oxford: Wiley Blackwell.

Prinz, Jesse. 2013. *The Conscious Brain*. *The Conscious Brain*. Oxford University Press.

Pryor, Jim. 2002. "Lewis vs. Orthodox Functionalism." http://www.jimpryor.net/teaching/courses/mind/notes/lewisfunctionalism.html.

Putnam, Hilary. 1975. "The Nature of Mental States." In *Mind, Language and Reality. Philosophical Papers. Volume 2.*, edited by Hilary Putnam, 429–40. Cambridge: Cambridge University Press.

Rosenthal, David. 1986. "Two Concepts of Consciousness." *Philosophical Studies* 49 (3): 329–59.

———. 1997. "A Theory of Consciousness." In *The Nature of Consciousness:*

Philosophical Debates, edited by Ned Block, Owen Flanagan, and Güven Güzeldere, 729–54. Cambridge, Mass.: The MIT Press.

———. 2005. *Consciousness and Mind*. Oxford: Clarendon Press.

Searle, John R. 1992. *The Rediscovery of the Mind*. Cambrdige, Mass.: The MIT Press.

Shelton, Jill T., Emily M. Elliott, and Nelson Cowan. 2008. "Attention and Working Memory: Tools for Understanding Consciousness." *Psyche* 14: 1–6.

Simons, Daniel J., and Christopher F. Chabris. 1999. "Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events." *Perception* 28 (9): 1059–74.

Sligte, Ilja G., Steven H. Scholte, and Victor A. F. Lamme. 2008. "Are There Multiple Visual Short-Term Memory Stores?" *PLoS ONE* 3 (2). Public Library of Science: e1699.

Smart, John J. 1959. "Sensations and Brain Processes." *The Philosophical Review* 68 (2): 141–56.

Sperling, George. 1960. "The Information Available in Brief Visual Presentations." *Psychological Monographs: General and Applied* 74 (11): 1–29.

Treue, Stefan, and John H R Maunsell. 1996. "Attentional Modulation of Visual Motion Processing in Cortical Areas MT and MST." *Nature* 382 (6591): 539–41.

Tye, Michael. 1995. *Ten Problems of Consciousness*. Cambridge, Mass.: The MIT Press.

———. 2000. *Consciousness, Color, and Content*. Cambridge, Mass.: The MIT Press.

Watzl, Sebastian. 2017. *Structuring Mind: The Nature of Attention and How It Shapes Consciousness*. Oxford: Oxford University Press.

Weiskrantz, Lawrence. 1990. *Blindsight: A Case Study and Implications*. Oxford: Oxford University Press.

———. 1995. "Blindsight: Not an Island Unto Itself." *Current Directions in Psychological Science* 4 (5): 146–51.

———. 1998. *Consciousness Lost and Found*. Oxford: Oxford University Press.

———. 2007. "The Case of Blindsight." In *The Blackwell Companion to Consciousness*, edited by Max Velmans and Susan Schneider. Oxford: Blackwell.

Wright, Larry. 1973. “Functions.” *The Philosophical Review* 82: 139–68.

Wu, Wayne. 2014. *Attention*. New York: Routledge.