Linking phenomenal and access consciousness: a case for sparse representations

Abstract: In his recent article "Consciousness, accessibility, and the mesh between psychology and neuroscience" Ned Block uses the evidence of recent experimental psychology to defend the idea that phenomenal consciousness is richer than, or 'overflows', access consciousness. This enables us to speculate about whether subjects might have experiences which they cannot access at all. In chapter 1 of the thesis, I explain the main terminology and issues of the debate, and briefly outline Block's position. In chapters 2 and 3, I attack the idea that there may be phenomenally conscious states that are totally inaccessible, and argue that any theory of visual perception that admits this possibility is burdened with an implausible and theoretically problematic commitment. We should instead look for theories that do not force us to admit the possibility of inaccessible phenomenally conscious states whilst meshing plausibly with the empirical data. I go on to outline such a 'sparse representations' theory in Chapter 4, before applying it to the experimental data in Chapter 5. I conclude that sparse representations theory allows us to explain the relevant empirical data whilst avoiding the implausible possibility of inaccessible phenomenal consciousness, and should therefore be preferred to Block's account.

CONTENTS

LIST OF ABBREVIATIONS

A-consciousness = Access consciousness

IPC = Inaccessible phenomenal consciousness

P-conscious = Phenomenal consciousness

PRS = Partial Report Superiority

RLI = Refrigerator Light Illusion

SR = Sparse Representations

CHAPTER 1

INTRODUCING THE ISSUES

1.1 – Introduction

The history of philosophy is replete with radical ideas, many of which have long since been abandoned: few if any contemporary philosophers still talk in terms of Leibnizian monads, or have sympathies with Malebranche's Occasionalism. Other ideas once considered radical have become firmly lodged in the contemporary philosophical mindset; indeed, Physicalism itself was once considered an extreme viewpoint. To paraphrase Carl Sagan, however, extraordinary claims require extraordinary evidence, and the slow acceptance of Physicalism, under pressure from the natural sciences, is testament to how glacially philosophical viewpoints change.

This thesis deals with a radical idea that has been popularised by Ned Block in a recent important article[1], in which he argues that individuals may have conscious states that they *sincerely believe they do not have*. Although Block confines such states to individuals with severe neurological deficits, it is worth taking a moment to grasp just how radical his postulation is: it not only contradicts the idea that the subject is always the best judge of their own mental states, but contradicts the idea that a crucial part of what it is for a subject to have an experience is for it to be in some way accessible by their belief-formation and decision-making processes.

The relationship between a subject's having an experience and believing that they are having it has a venerable history; Immanuel Kant believed that without an experience being at least in principle self-ascribable, "something would be represented in me which could not be

1 Block 2007a, "Consciousness, accessibility, and the mesh between psychology and neuroscience"

thought at all, and that is equivalent to saying that the representation would be impossible, or at least it would be nothing to me." Such representations "would not then belong to any experience, consequently would be without an object, merely a blind play of representations, less even than a dream."[2]

I will argue that contrary to Block's argument, his claim does not have a sufficiently extraordinary evidential basis in contemporary psychology. More strongly, I will argue that the idea of experiences which are inaccessible to their subject is sufficiently problematic and implausible as to cast into doubt any theory of perception from which it follows as a conclusion. I will instead argue that a 'sparse representations' account of perception allows us to avoid the possibility of inaccessible experience, and provides plausible interpretations for the experimental paradigms Block advances. Given its fit with the experimental evidence, and the fact that inaccessible experience seems to follow as an unpalatable conclusion from alternative theories, we have good reasons for upholding some form of sparse representations theory.

The thesis falls into five chapters. In this chapter, I introduce the key concepts of the debate and provide a description and analysis of the target article, Block's 2007 paper. In chapters 2 and 3, I consider the idea of inaccessible P-conscious experience, and present arguments against it. Finally, in Chapters 4 and 5 I attempt relate philosophical theories of perception to the evidence of contemporary experimental psychology. Chapter 4 explain the sparse representations theory of perception I defend, and chapter 5 demonstrates how it can plausibly account for the empirical evidence that provides Block's basis for postulating the existence of inaccessible experience.

1.2 – P-consciousness and A-consciousness

2 Kant, Critique of Pure Reason B132, A112

A vital distinction that runs throughout this paper is the division between *phenomenal consciousness*, or 'P-consciousness', and access consciousness, or 'A-consciousness'. The distinction was popularised in a 1995 article[3] by Ned Block, and since then has attracted considerable attention.

In short, P-consciousness "is experience; what makes a state phenomenally conscious is that there is something "it is like" (Nagel, 1974) to be in that state."[4] I will use the adjective 'P-conscious' to describe states that there is something it is like to experience, and I will use the term 'P-consciousness' to refer to token P-conscious states as well as the phenomenon as a whole. I also speak of phenomenology, meaning the P-conscious contents of perception.

Whereas Block's definition of P-consciousness is descriptive, his definition of A-consciousness is stipulative. He states that it is a condition of a state's being A-conscious for a given perceiver that it be "poised to be used as a premise in reasoning, and... poised for [rational] control of action"[5]. A state's being A-conscious, then, is a matter of it relating to a subject in these specified ways. Block also sometimes talks of a subject's "A-consciousness", thereby referring all of a subject's states which are appropriately related to that subject, such as being suitably poised for action.

One point that may benefit from clarification is his requirement that a state is A-conscious only if suitably poised for *rational* behaviour. This does not exclude animals or pre-linguistic infants from A-consciousness[6], but instead excludes those cases in which information makes itself available to subjects without realising that it is at their disposal. For example, blindsight patients can sometimes make accurate guesses about the contents of visual information which plausibly is not P-conscious. Note that, in recent papers, Block has replaced talk of A-consciousness with talk of *cognitive* access, although I will use the two terms interchangeably in this thesis. I will clarify my precise interpretation of what I count as

3 Block, 1995, "On a confusion about a function of consciousness"
4 Ibid. 228
5 Ibid. 231
6 Ibid. 277

(cognitive) access in section 1.3 below.

One problem relating to A-consciousness is whether there is always one or may be multiple systems of access. We normally take it for granted that a subject capable of engaging in rational control of behaviour, report, and so on has a single integrated system of beliefs, intentions and so on. We normally assume, for example, that if a subject S is A-conscious of a perception P1 and a perception P2, they can report on having *both* perceptions. However, the famous cases with split brain patients[7] have shown that this is not always the case: a split brain patient who has two words presented to their left and right eye, such as "pen" and "knife" respectively, will report seeing only the word "knife". However, we cannot conclude that they are not A-conscious of the word "pen", since if asked to pick out the object they saw with their left hand, they will pick out a pen.

It is clearly over-simplistic to assume that we may always talk of a single set of A-conscious states. However, in any investigation, one has to control for certain factors, and Block generally treats the notion of "access" as referring to a single integrated system of intentional attitudes. I shall follow him in making this assumption in this thesis, but readers should be aware of this as an oversimplification.

1.3 – What counts as access?

I now wish to clarify what is meant by access, or, the term now preferred by Block, *cognitive* access. When I say a state is accessed or cognitively accessed, I mean that it influences that subject's total set of beliefs about the world, or their *doxastic state*, and does so non-inferentially. Seeing a red apple, for example, I acquire beliefs to the effect that there is an apple in front of me, that it is red, that I am having an experience of a red apple, and so on. I include the non-inferentiality clause to exclude cases like the following: a blindsight subject,

7 For a good discussion, see Tye 2003, Ch.5

who knows that her guesses about the contents of their blind field are usually correct, may realise she is about to guess that there is a red apple in front of her, and so come to believe that there *is* in fact a red apple in front of her. This is an inferentially acquired belief, however, based her knowledge that her guesses are usually accurate, and that in this particular instance, she is disposed to guess that there is a red apple in front of her. As a result, it is not sufficient for her perceptual experience to count as accessed that it play a merely *inferential* role in her acquisition of belief.

I do not mean to suggest that a subject must actually *think* "there is a red apple in front of me" in order for us to ascribe that belief to her. A subject's doxastic state, as I intend the term, allows of extremely finely-grained modifications. Looking at an arrangement of objects on a table, my visual experience may affect my doxastic state in a way that would have been different had the arrangement of objects been slightly different. The notion of doxastic state I am reaching for, then, is something akin to the total information that is poised for use in rational action by a subject.

I also allow that a subject may have cognitive access to an experience without being willing to express the belief that she has undergone the experience, though she must at least believe that she *may have had* the experience. Imagine that a word is presented to you tachistoscopically, for a fraction of a second. Below a certain exposure time, you may deny having seen any word presented to you at all. We cannot conclude from this that the experience did not alter your doxastic state. Different subjects may have different thresholds of confidence at which they feel comfortable in reporting having had a particular experience, and this may vary according to cultural or idiosyncratic factors (such as how intimidated a subject may be by the testing environment). One can imagine a subject in the above test being unsure if they saw a word, and yet still non-inferentially acquiring beliefs from that experience, perhaps thinking something along the lines of, "I cannot be sure if I saw a word,

but if I did, it may have been the word 'house"'[8].

If the subject did enjoy some degree of cognitive access, but not enough to make her feel confident to assert that she had the experience, then we would expect that, if asked to choose from a list of words she may or may not have seen, she would perform better than would be expected through mere chance. Unfortunately, this cannot serve as a clear test for cognitive access, since a subject may be *unconsciously* primed to be more likely to select the word she was exposed to without her doxastic state having been affected (that is, without the information being accessed).

An alternative test for cognitive access, one that can distinguish between consciously and unconsciously held beliefs, is required. One such test might be a *betting* paradigm. Imagine next that a subject is tachistoscopically presented with a word, and she denies having seen anything. She is then given a list of four words which may have been the word presented to her, and she is asked whether she would like to make a bet on whether she can guess the word correctly. If her doxastic state has not been influenced by the stimulus, then we would expect her, assuming she is rational, to be unwilling to make a bet on any odds shorter than 4-1, that is, the odds of guessing the word by blind chance. But if she thinks she *may* have seen a word, then again, assuming she is rational, she would make a bet even if the odds were slightly shorter, for example, 3-1. In this case, we could conclude that her experience did alter her doxastic state, since it has influenced her rational behaviour, in this case, her betting behaviour.

This is a conceptual rather than practical test, since there are two kinds of disruptive factor for which it would be unable to control. First, most people do not always act rationally, and might naturally fail to grasp the relationship between their knowledge and the odds they are facing. Second, a subject may realise that she was being unconsciously primed, so that her guesses were more likely than average to be correct. This might caused inferentially acquired

8 See Block 2005 for an in depth discussion on such tachistoscopic presentations and subject's cognitive responses to them.

beliefs to influence her betting behaviour. These factors would prevent the test from being a reliable indicator of whether a given experience was cognitively accessed. However, at a purely conceptual level, this 'betting test' allows us to describe what it means for a perceptual experience to affect a subject's doxastic state.

1.4 - Putative cases of divergence

A range of situations have been proposed in which a state may be P-conscious but not A-conscious. I will now consider three such types of case. It is vital to note here a distinction between the notions of *non-accessed* and *inaccessible* P-consciousness. The former terms includes P-conscious states that a subject *could* have accessed in the circumstances, had she so wished, while the latter term specifically indicates P-conscious states that a subject could not have accessed in the circumstances even if she had wished to.

(i) Failure of access

Example: *I am sitting in my room reading a book. I suddenly notice the noise of a ticking clock, and realise I have been hearing it all this time. I was not A-conscious of the noise of the clock, but I was P-conscious of it.*

This case demonstrates a failure of attention: the information alleged to have been P-conscious but not A-conscious is information that was *accessible* but *not accessed*: I could, had I so wished, directed my attention to the sound of the clock, but did not do so. I can subsequently recover memories of having heard the clock, lending further credence to the idea that I was P-conscious of it at the time.

(ii) Overflow

Example: [Sperling, 1960] *In a psychology experiment, a matrix consisting of three rows of four alphanumeric characters is flashed to me. After the matrix has been removed, I am cued to report the contents of one particular row. However, I cannot then go on to report on other rows, and this has nothing to do with where I was looking when the image was first presented. It is confirmed in subsequent trials that I can report accurately on whichever row I am first cued to report on. Before cueing, I am P-conscious of every row, but can access only one of them.*

This case is another example of *non-accessed* P-conscious information. Unlike the clock example, however, under no circumstances could I have accessed *all* the information: the P-conscious contents were too rich for me to access fully in the time available to me. One way of putting this is to say that P-consciousness *overflows* access (I use the term 'overflow' to describe such cases).

This is not just a case of my being unable to fully describe the contents of my perceptions. It may be impossible to describe the fine-grained details of a photograph via speech. However, looking at a photograph, I can access the data of my perceptions even if I cannot describe it, and I can make fine-grained discriminations which I may be unable to express. The problem in the above case is rather that there will inevitably be some data I cannot access at all. I could not say, for example, whether the two uncued rows in the second example are identical or non-identical to rows with which I am subsequently presented.

(iii) Inaccessibility

Example: [Block, 2007a] *I am a patient with a rare condition called visuospatial extinction. This means that when I am presented with a single object on my right or left, I can attend to them, but if objects are presented to both sides of my visual field simultaneously, I only see the one to the right. When presented with an image of a face to my left and an image of a house to my right, I deny seeing a face. Only the image of the house influences my doxastic state. However, an fMRI scan reveals that the fusiform region of my brain,*

strongly associated with experiences as of faces, lights up strongly. Could I be having totally inaccessible experience as of a face of which I am nonetheless P-conscious?

This case presents the most extreme kind of divergence of P- and A-consciousness, namely a case in which there is a state that is P-conscious but not even *possibly* A-conscious in the circumstances. In other words, it is both non-accessed and *inaccessible*.

All three kinds of case are considered in this thesis. Chapters 2 and 3, as already noted, are concerned with wholly inaccessible experience. In Chapters 4 and 5 I will provide a 'sparse representations' account that deals with the first and second sort of example whilst denying that P-consciousness is ever richer than A-consciousness, or that there are cases of P-consciousness without A-consciousness.

1.5 – A summary of Block's argument

I turn now to exposition of Block's argument in the target article[9]. Block is concerned with the question of whether we could find experimental mechanisms for establishing the occurrence of P-consciousness that we could use even in the absence of cognitive access. Since such consciousness would be inaccessible, the subject would sincerely believe they were not experiencing it (as defined in 1.3, they would make no difference to a subject's doxastic state). Hence Block's controversial move is suggesting that we might find an objective criterion for P-consciousness that would *trump* subjects' reports of which experiences they are having.

Block marshals experimental evidence that, he believes, shows that subjects can have experiences which are non-accessed (though accessible). This evidence does not quite demonstrate the possibility of subjects' having cognitively inaccessible P-conscious

9 Block, 2007a.

experience, but Block believes that it shows that information can be P-conscious without being A-conscious. Given this, he suggests, we should conclude that P-consciousness and A-consciousness have at least partially distinct neural correlates[10].

Block goes on to make a conjecture about the neural basis of phenomenology in general, namely that is consists in recurrent feedback loops in the back of the brain. If this is correct, the occurrence of such feedback loops might provide us with evidence of P-consciousness even where a subject is totally oblivious to the experience. Whereas until now, there has been a one-way transit from psychology to neurology in informing the search for neural correlates of consciousness, Block hopes that by providing an *objective* neurological criterion of the occurrence of conscious states that can operate even without report or access, he can provide a two-way 'mesh' between neurology and psychology.

I now present a more lengthy exposition of Block's argument, which I have divided into three sections.

1.6 – Block's argument: part one (sections 1-8)

In the first third of the paper, Block describes a fundamental methodological problem in psychology. Prima facie it might seem that the only evidence we can have for a subject's having a given experience is that the subject accesses that experience in some way, most obviously via report. Normally, to establish whether a subject can see a red dot on a card, we have simply to ask them. If this were impossible for some reason we might still attempt to establish whether they were having a conscious experience through noting whether they could rationally respond to the experience, for example, by using the perception in the formation of complex behaviours.

10 Block, 2007a: 494.

These evidential criteria, however, seem to rule out finding occurrences of non-accessed P-conscious experiences, for without a subject being able to access their experiences, they could not deploy them for rational action including report. There seems to be a fundamental methodological impediment: our only means of establishing whether an experience is occurring requires access, so we cannot investigate the possibility of non-accessed experiences. The question is analogous to the famous problem of whether a tree falling in the forest makes a sound. However, it has the added difficulty that, whereas whether or not an event is accessed by a conscious mind is largely irrelevant to our theory of the physical world, it seems very much an open question whether access by a subject might enable a brain state to become P-conscious.

Block gives two examples of such prima facie unanswerable questions, before attempting to provide a way out of the problem. The first example is that of Fodorian modules, the early processing units in perceptual systems[11]. Processing within Fodorian modules would not be accessible, but would play a causal role in the construction of more complicated representations higher up in subjects' perceptual systems, hence subjects could not possibly report the occurrence of P-consciousness in Fodorian modules. P-consciousness in Fodorian modules might therefore seem to lie outside the scope of investigation.

The example concerns an instance in which an fMRI scan shows that an area of a subject's brain thought to be correlated with a particular kind of experience is strongly activating, but where the subject in question denies having that experience. Normally, we would conclude that we had simply misidentified the neural correlate of that experience, taking the subject's word as definitive. However, if we knew that the patient had a neurological deficit which limited their access to certain representations, then the conclusion would be cast into doubt: what if the patient were having P-conscious experiences that were inaccessible owing to their neurological deficit? Again, however, the question is one that

11For more on Fodorian modules, see Fodor 1983.

prima facie could never be confirmed.

Such a case has arisen in empirical psychology, as was described in example (iii) of section 1.4 above. Block reports on a patient, GK, suffering from visuospatial extinction, a condition in which, owing to a lesion on one hemisphere, patients are sometimes incapable of accessing visual stimuli on their left and right sides simultaneously: one stimulus, associated with the unimpaired hemisphere, extinguishes the other, even though stimuli on both sides are perceptible when presented individually. When GK undergoes a binocular rivalry experiments, in which one eye is presented with the image of a house and the other an image of a face, the fusiform region of his brain lights up under an fMRI scanner. In normals, strong activations of the fusiform region are often associated with both A- and P- conscious experiences as of faces. GK, however, denies having any such experience. It is possible, Block suggests, that GK is having P-conscious experience of a face that is simply not A-conscious; yet given the methodological difficulties described, it seems prima facie impossible to establish whether this is the case.

Block terms the view that P-consciousness and A-consciousness are inseparable for investigative purposes 'epistemic correlationism', and firmly rejects it. Stronger than epistemic correlationism is the view that, not only are A-consciousness and P-consciousness inseparable for all purposes of investigation, but that a correlation between the two is metaphysically necessary. This view, which Block terms *metaphysical* correlationism, is independent of epistemic correlationism (one might arrive at it through an analytic functionalist theory of mind, for example) but is similarly opposed to Block's belief that we could have empirical reasons for thinking that P-consciousness diverges from A-consciousness.

Block hopes to present experimental evidence to rebut both positions and show how we might have good reason to believe that subjects were having non-accessed experiences. If he could prove this point, he would have scored a decisive blow against both kinds of

correlationism, since it would show that it was possible to have good reason to ascribe an experience E to a subject S without S accessing the contents of E. However, he also believes that by proving this, he will demonstrate that P- and A-consciousness must have distinct neural correlates, and this, he believes, licenses the possibility that individuals could have P-conscious states that they could not, in the circumstances, cognitively access, as in GK's case. This would provide a "mesh" between psychology and neurology, that is, provide a means of studying the occurrence of consciousness from the purely neurological level, and might thereby operate independently of subjects' reports.

1.7 – Block's argument: part two (sections 8-12)

The second section of Block's paper is devoted to providing experimental evidence in support of the existence of non-accessed phenomenology. The three experiments that figure most prominently in his argument are the Sperling[12], Landman et al.[13], and Sligte et al.[14] Paradigms (hence, Landman and Sligte paradigms). I will consider these experiments in depth in Chapter 5, but I will now provide a brief description of their common features.

Fig.1 – a matrix like that in the Sperling test

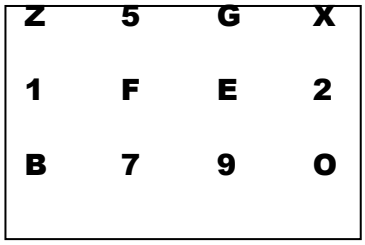
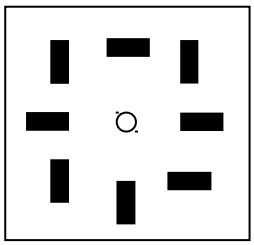


Fig.2 – an array like that in the Landman and Sligte tests



12 Sperling 1960.
13 Landman et al. 2003.
14 Sligte et al. 2006.

All of the experiments are *partial report procedures*, and involve presenting a subject with more information than they can access in the short period of time it is present, before requiring them to demonstrate knowledge of some subset of the total information. This information may be in the form of alphanumeric characters, as in the Sperling test (see Fig.1), or rectangles of various sizes and orientations, as in the Landman and Sligte experiments (see Fig.2). Subjects in these paradigms demonstrate the ability to recall proportionately more of a subset of the stimulus rather than on the whole of the stimulus even when cued subsequent to the removal of the stimulus.

For example, in the Sperling test, subjects were briefly shown a grid of twelve characters. If they were asked, after the removal of the stimulus, to name as many characters as they could, they could name only four or five. However, if they were cued up to 1000m *after* the removal of the stimulus to report on just the top, middle, or bottom row, they could report an average of three characters from that row. In other words, when making reports on a subset of the grid, they were more proportionately accurate (recognising on average 3/4 items) than when making reports on the whole of the grid (where they recognise 4-5/12 items). This is termed *partial report superiority* (hence, PRS).

The Landman and Sligte paradigms are also partial report procedures in which subjects are successively shown two images of rectangles at various orientations with a grey screen in between the presentations. Subjects are required to report on whether a given rectangle has changed orientation, with the rectangle to be reported on being cued during the grey screen interval. Subjects display PRS in detecting changes in *any* of the cued rectangles. I consider all of these arguments in more depth in Chapter 5.

The existence of PRS seems to demonstrate that subjects, in making partial reports, were drawing on some kind of visual short-term memory (VSTM). VSTM seems to degrade rapidly, hence rendering subjects unable to report on more than a small number of items, but

apparently has a higher capacity than *working memory*. Working memory is the memory function which allows for the intelligent deployment of information in behaviour, and its contents are identified by many (including Block) with the *accessed* contents of perception.

In all of these experiments, subjects claimed to have *seen* more data than they had time to access, and moreover, demonstrate that they had retained more data than they were able to report. Block believes the natural interpretation of this is that subjects were P-conscious of the rich data of VSTM. If the contents of VSTM are indeed P-conscious, then we have found a clear case of P-consciousness overflowing A-consciousness, and an instance in which we can ascribe experiences to subjects even where they do not access the contents of those experiences. This demonstrates the overflow of A-consciousness by P-consciousness, and hence shows that the capacities of P-consciousness and A-consciousness are different.

Block holds that these experiments show that there are P-conscious states that are *accessible* but contingently *non-accessed*. Block holds that a state could be P-conscious but not A-conscious only if the neural machinery of the two did not entirely overlap, that is, if the neural correlates of P-consciousness and A-consciousness are non-identical. Hence Block concludes that P-consciousness and A-consciousness are "based at least partly on different systems with different properties"[15].

If we accept that P-consciousness and A-consciousness have different neural correlates, then it is hard to see how the mere *accessibility* of states to working memory should make a difference to whether or not the state is P-conscious. After all, P-consciousness is plausibly an intrinsic rather than relational phenomenon: the question of whether a P-conscious state can be subsequently taken up by working memory seems irrelevant to rendering that state P-conscious in the first place.

15 Block 2007a: 494

Many philosophers would question this step in Block's argument, and hence deny an entailment from the kind of *non-accessed* P-consciousness alleged to occur in the Sperling test to the possibility of *inaccessible* P-consciousness, as in GK's case. I will not consider this step in the argument in great detail, as it would require me to provide a broader theory of the nature of conscious content than is possible within this paper. Moreover, I feel that the possibility of inaccessible experience can be adequately challenged on a priori grounds without attacking this claim. I do, however, say a little bit more about this claim in section 2.4.

Block's argument for the possibility of inaccessible P-consciousness, then, is as follows. We have empirical evidence that P-consciousness has a higher capacity than A-consciousness, and hence that not all P-conscious states are accessed. If some P-conscious states are contingently non-accessed, then there could be P-conscious states which *could* not be accessed, since the difference between the two situations consists in the presence or absence of relational properties irrelevant to the occurrence of P-consciousness. GK might provide us with one such case of inaccessible P-consciousness: the fusiform region of his brain is generating P-consciousness as of a face, but because of his neurological deficit, his fusiform region cannot pass on its data to working memory, so preventing the P-conscious representation as of a face from becoming A-conscious.

1.8 – Block's argument: part three (sections 12-15)

The conclusion Block draws from this is that phenomenology overflows cognitive access. The former is based in the high capacity representations of VSTM and the later in working memory. He is required to give evidence, however, that VSTM and working memories have different neural bases. Block notes first that working memory has relatively limited storage capacities, rarely being able to retain data of more than eight items, though there is

considerable variation dependent on the task (in the partial report paradigms, it seems limited to four or five items), and it has been suggested that there may be different working memor*ies* based on different systems. Moreover, what counts as an item varies greatly. Subjects might struggle to remember the sequence "AGQXCTPF", but would have no difficulty in remembering the sequence "ABCDEFGH", for example, even if it were displayed to them only briefly, because they can bring it all under a single concept sufficient for its recall (namely the first eight letters of the alphabet)[16].

The location in the brain of the working memory processes used by subjects in the Sperling, Landman, and Sligte experiments, Block argues, is somewhere in the prefrontal cortex. Block quotes Curtis and D'Esposito in saying that this area "aids in the maintenance of information by directing attention to internal representations of sensory stimuli and motor plans that are stored in more posterior regions"[17].

Assuming that working memory is located somewhere in the prefrontal cortex, Block next claims that "arguably, the core neural basis of visual phenomenology is in the back of the head"[18]. His main evidence for this is a wide series of studies that have shown perturbation to subjects' experience of motion associated with stimulation of area V5 in the back of the head. In order for this stimulation to generate an experience of motion, however, a recurrent feedback loop between area V1 and V5 needs to be established. Block concludes from this that a particular V1-V5-V1 loop may provide the *core* neural basis of some experiences as of motion.

The problem, Block notes, with the argument given in the preceding section is that apparently the only way we can be sure that such cases generate phenomenology is through demonstration via access, and this requires that these feedback loops induce activity in the front of the brain. Block now sets out to show how we might demonstrate whether these

16 Block gives a similar example in 2007a: 495
17 Curtis and D'Esposito 2003: 415
18 Block 2007a: 496

frontal activations characteristic are part of the *total* neural basis of phenomenology.

Before moving any further, I wish to very briefly explain the language of the "global workspace", since Block makes his claims in these terms. Global workspace originated with Bernard Baars[19], and holds, roughly, that there are many competing coalitions (patterns of firing) of data active in the brain at any one time. What determines whether these coalitions are accessed by the subject is whether they trigger similar activations in the frontal lobes. Once a coalition in the back of the head triggers similar activity in the frontal lobes via 'feed-forward' processes, this frontal activity reinforces this pattern in the back of the head via 'feed-back' processes, causing this particular pattern to come to dominate and drown out 'losing coalitions'. Hence the frontal lobes are called the 'global workspace', since data that arrives there is subsequently broadcast throughout the system. The idea is, roughly, that competing perceptual stimuli generate competing patterns of activity in the brain, and that whichever pattern 'wins out' and is broadcast throughout the brain is the stimulus to which we consciously attend. Hence access is a kind of 'fame in the brain'[20], that is, a given firing pattern's coming to predominate. This 'winner takes all' model of neuronal firings is supported, as Block notes, by recent experiments, notably Sergent and Dehaene[21].

It is Block's view is that it is not only the winning coalitions that are P-conscious, and that activity at the back of the brain does not require feed-forward processing to become P-conscious. Block notes that there are some very strong coalitions in the back of the head which very narrowly 'lose out' to stronger coalitions and thus fail to trigger a winning coalition in the front of the head. Considering a further experiment[22] in which subjects were confronted with brief stimuli, either just noticeable by the subject or entirely subliminal, Block notes that the latter stimuli strongly activated visual areas in the back of the head, but

19 See, e.g., Baars 1988.
20 See, e.g., Dennett 1996.
21 Sergent and Dehaene 2004.
22 Kouider et al. 2007.

failed to activate frontal coalitions; they did, however, "*modulate* frontal activity"[23].

This leads to Block's final point. Kouider and Dehaene argue that losing coalitions are neural bases of *preconscious* states. They assume this on the grounds of their unreportability, which, Block notes, unfairly assumes some form of correlationism: "[a] better way of proceeding would be to ask whether a phenomenal state might be present even when it loses out in the competition to trigger a winning frontal coalition."[24] However, if we assume that strong but still losing recurrent feedback loops in the back of the head constitute the neural basis of phenomenal states, they provide an explanation for the differential in capacity between phenomenology and the global workspace. If Block's explanation of the partial report experiments is correct, we must find a high capacity neural correlate for the rich, non-accessed P-consciousness that subjects experience in the Sperling, Landman, and Sligte paradigms. The many losing coalitions in the back of the brain are just such a correlate, and so provide a plausible reason for making the assumption that the core neural bases of P- and A-consciousness are distinct. Hence psychological experiments enable us to pin down the neural correlate of P-consciousness, and once found, we can use the presence of such activity to establish if a subject is be having P-conscious experience even if it is inaccessible. This is the promised mesh between neurology and psychology.

Block concludes, then, that the degree of overlap of the machinery of cognition and phenomenology is open to empirical investigation. Second, he holds, there is evidence to think that the machinery of the latter does not include the machinery of the former. One experimental prediction he makes is that the unaccessed representations involved in the Sperling, Sligte, and Landman experiments involve recurrent feedback loops in the back of the brain. Most importantly, he argues these may provide a way of settling the question of whether GK has P-conscious experience as of a face even if it were inaccessible.

23 Block 2007a: 497
24 Block 2007a: 498

1.9 – Conclusion, and the relevance of the issue

For the purposes of this thesis, I treat Block's core argument as having the following form.

(1) Postulating unaccessed P-conscious states is the best way to account for the empirical data.
(2) If unaccessed P-consciousness is possible, then *inaccessible* P-consciousness is possible.

Hence,

(C) Inaccessible P-consciousness is possible.

I attempt to rebut Block's argument by tackling it in reverse, to whit:

(1) Inaccessible P-consciousness is at best implausible, at worst impossible. (Chapters 2 and 3)
(2) Unaccessed P-conscious would entail the possibility of inaccessible P-consciousness. (Section 2.4)

Hence,

(C) We have a strong motivation for finding plausible interpretations of the experimental data that does not rely on unaccessed P-consciousness. (Chapters 4 and 5)

The question of whether inaccessible P-conscious is possible is not merely a philosophical curiosity. It has a broader importance for the philosophy of mind for four reasons, which I will now provide.

The first is that investigation of the nature of inaccessible experience compels us to think about the self and its relationship to experience, and what it means to say that an experience E belongs to a subject S, an issue which I consider in chapters 2 and 3. Similarly, if, as I argue, inaccessible P-consciousness is problematic, this has important consequences for us in formulating our theories of perception. In chapters 4 and 5, I attempt to show how a

'sparse representations' theory of perception can allow us to avoid the possibility of inaccessible P-consciousness.

Third, consideration of inaccessible P-consciousness is intimately related to the relationship of philosophy to psychology and neuroscience. Although I do not focus on this issue in detail in this thesis, I deny that the specific evidence raised by Block provides us with empirical grounds for making inferences about the occurrence of P-consciousness in the absence of access.

Finally, it is my hope that long term investigation of these issues may strengthen our framework for tackling the hard problem of consciousness, that is, how activations of neurons in the brain can give rise to the phenomenon of experience. Direct investigation of the hard problem has yielded much material of interest, but little in the way of conclusions. I will not directly engage with this debate, but it is my hope that by developing our general understanding of different aspects of consciousness and perception, we help to establish a firmer theoretical basis for tackling the hard problem itself.

CHAPTER 2

INACCESSIBLE PHENOMENAL CONSCIOUSNESS – SOME WORRIES

2.1 – Introduction

As noted in the previous chapter, Block's most striking conclusion in the target article is that there may be cases where subjects have experiences which they honestly believe they are not having, and he suggests the patient GK may provide an example of this. I term this phenomenon inaccessible phenomenal consciousness (hence, IPC), and it should be distinguished from the accessible but non-accessed P-consciousness that Block takes to be present in the Sperling, Landman, and Sligte paradigms. Subjects in such tests *could have*

become A-conscious of *any* of their P-conscious states had they been appropriately cued. Conversely, IPC is not accessible in the conditions in which it occurs. Of course, were circumstances different, there may be cases in which otherwise inaccessible P-conscious states might become accessible: for example, if GK were not faced with two competing stimuli, his inaccessible P-consciousness of a face would have been P- and A-conscious.

I think we have good philosophical reasons for rejecting the idea of IPC, stronger even than those which we have for rejecting the kind of contingently non-accessed P-consciousness that Block believes occurs in the Sperling test, for example (though as noted, both Block and I believe that once we allow for unaccessed by accessible P-consciousness, we have good reasons for also admitting the possibility of IPC). In this chapter and that which follows, I criticise the notion of IPC. This chapter presents some initial worries about IPC, showing how it is unrelated to our normal concept of experience, and forces upon us implausible theoretical commitments. Chapter 3 directly attacks the idea of IPC, arguing that we could not have empirical grounds for ascribing it to a subject, and providing a priori arguments against its coherency. The arguments in this section, then, are weaker than those in Chapter 3, insofar as they do not question the possibility of IPC, but I hope now to begin the process of convincing the reader that it is a tenuous and problematic concept.

This chapter consists of two arguments. I first argue that IPC is entirely removed from our ordinary concept of what it is for a subject to have an experience, with the result that ascriptions of IPC to a subject would be meaningless for all practical purposes. Second, I argue that IPC commits us to the implausible view that the neural correlates of P-consciousness could still generate P-consciousness even if removed from a complex psychological system. Note that this same argument provides evidence for the view that a commitment to accessible but non-accessed P-consciousness forces us to also commit to the possibility of IPC (step (2) as described in 1.9). I conclude that allowing for the possibility of

IPC should be regarded as a serious demerit of any theory of consciousness.

2.2 – What is it like to have inaccessible experience?

Before considering any arguments against IPC, I wish first to clarify precisely what IPC may and may not consist in. In the absence of a subject's being able to form any thoughts about their experiences or access data within it, it is questionable whether we can even understand the notion of a subject *having* these experiences. Block appeals to the notion of *awareness*, which he admits is involved in anything that deserves to be called an experience[25]. The kind of awareness that a subject has of an inaccessible experience is course not any kind of *cognitive* awareness. The notion at play seems rather to be some kind of raw phenomenological acquaintance relation, though Block does not spell this out. Hence GK is directly acquainted with his inaccessible experience as of a face; he just cannot make *use* of this experience in cognition.

Prima facie this notion of direct acquaintance may seem straightforward: when I look at a red apple, there is a sense in which I seem to be directly acquainted with its redness, or when feeling a pain, acquainted with how the pain feels. In both of these cases, however, we also have *access* to some representational content; for example, we know that the apple's surface has a distinctive kind of property, namely its colour. When we introspect on an instance of P-consciousness, it is also a case of A-consciousness, that is, it informs us about the world or our experiences in some way. Whilst I do not deny that there is a distinctive phenomenological aspect to experience, I doubt whether we can distinguish from the notion of having information presented to us in a way that would allow us to clearly conceive of phenomenology in the absence of all access.

25"We may suppose that it is platitudinous that when one has a phenomenally conscious experience, one is in some way aware of having it." (Block 2007a: 484)

Despite my reservations, I will endeavour to explicate this notion of awareness in a manner maximally sympathetic to Block's position. Let us assume as a starting point that some kind of subjective acquaintance is what Block has in mind when talking of the kind of awareness involved in IPC. What more can we say about the kind of awareness at stake? Without access, a subject could not become aware of an experience's representational content, but we should not thereby conclude that it altogether lacks such content. IPC may be subject to some form of categorisation. Consider the phenomenon reported by Christopher Mole[26] in which mothers with young children are more likely to awaken to the sound of a child crying than to other equally shrill and noisy stimuli. Presumably, some sort of categorisation process is occurring even in the mother's sleeping mind, according to which the sound is categorised as that of a crying child, and so prompts the mother to wake.

Indeed, there is some evidence of complex categorical structure even in unconscious experience. In one small group study[27], doctors read a prewritten script during an operation with patients under general anaesthetic, in which they expressed shock and worry that something had gone wrong. Though patients later claimed to be unaware as to whether anything untoward had occurred during the operation, under hypnosis they were able to reproduce very closely the semantic content of the script. Even if we balk at saying the patients actually *understood* the script while unconscious, what we cannot doubt is their brains had subjected it to some process of interpretation which allowed them to remember its content appropriately, despite massively diminished or absent consciousness at the time of reading.

I conclude, then, that if wholly unconscious perceptions can be subjected to complex forms of categorisation, we should *not* conclude from the mere fact of the inaccessibility of IPC that it may not have a complex representational structure.

---

26 Mole 2008.

27 Levinson 1965.

2.3 – Is being P-consciousness of something enough for it to count as *my* experience?

I have doubts about whether IPC is a comprehensible notion and whether it could be ascribed to subjects, as I argue in the following chapter, but I wish now to argue that, these qualms aside, the kind of minimal consciousness involved in IPC is far removed from what we normally mean when we talk of an experience's *belonging* to a subject. Indeed, I think that in all cases where a subject's having an experience *matters,* it is in fact that subject's *access* to their experience that is relevant. Hence even if subjects can have IPC, they cannot have that experience in any rich sense, thereby rendering IPC at best a tenuous concept.

Experiences matter to subjects in three main ways, which I will now consider: they are informative; they have normative force; and through these two mechanisms play a role in agency.

I will consider first how experiences are informative. Prima facie, IPC could never inform a subject about anything: without having access to the experience, she could not form beliefs from it, and belief is widely held to be a condition for knowledge. One objection to this might be the case of a subject for whom IPC generated unconscious capacities. For example, a subject with inaccessible visual experience might be primed by her experiences to make accurate guesses about the content of the blind field in her vision. If she knew this fact, she might come to *believe* that her guesses were reliable. Via her inaccessible experiences, then, she might come to have knowledge about the world.

The problem with this view is that it is not such a subject's IPC that provides the conscious basis for her knowledge about the world, but her awareness of her guesses. She is A-conscious of her having guessed that an apple is present, and trusts her guess to be reliable, and so also knows that an apple is present. Whilst her IPC may be causally responsible for her guesses, this is disanalogous to perceptual knowledge which involves awareness of a stimulus

whose content generates first-order beliefs. In this case, however, the subject has no access to the content of her perception, nor even needs to believe that there is any IPC going on. Hence there is no compelling argument here that IPC generates her knowledge *qua* perceptual experience.

There is a further kind of knowledge that a subject of experience might be held to possess in virtue of having IPC, namely knowledge of phenomenal character. For example, it might be sufficient to know what red looks like that I just have the appropriate P-conscious experience, whether or not this experience is accessible.

On closer inspection, however, I find this argument implausible. Consider Marla, a variant on Jackson's famous achromatopic colour scientist[28]. Marla's colour vision is normal, but the areas of her brain responsible for colour vision are modified such that all her colour experience is *inaccessible*. Looking around the world, she is convinced that she sees only in grey-scale. If asked to say whether two objects of the same brightness and saturation but of different hue look the same or different to her, she reports, and believes, that they look the same. One day, scientists repair the damage to Marla's brain, allowing her to access the information in her brain relating to colour. "At last," she announces after the operation, "I know what red looks like!" Plausibly, Marla did *not* know what red looked like before the operation, and for those philosophers inclined to explain this example in terms of possession of phenomenal concepts, Marla did not possess the phenomenal concept 'red'. Hence, I conclude, not merely 'knowing that', but 'knowing what it is like' requires access, and hence IPC is irrelevant to such knowledge.

One way around this objection might be to claim that Marla *did* know what red looked like insofar as she had P-conscious experiences of red, but that she did not have a higher order thought to the effect that she was having these experiences; in other words, she does not *know* that she knows what red looks like. However, even if it is possible to know

28 Jackson 1982.

something without thereby knowing that one knows, I do not believe this applies in the Marla case insofar as she fails to bring her experience under any concept at all.

By way of analogy, consider the following example of three ornithologists on the Scottish moors, listening for the call of a Ptarmigan. The first has heard a Ptarmigan before, and knows what it sounds like, whereas the second and third have not. One night when in their tent, they all hear a strange sound, which is in fact a Ptarmigan. The first recognises the noise as belonging to a Ptarmigan. The second hears the noise but does not know what it is. The third does not notice the cry at all, being engrossed in a telephone call. The next morning, the second and third ornithologist complain to the first that despite having been on the moors for almost a week, they still do not know what a Ptarmigan sounds like. The first turns the second, and corrects him, saying, "no, *you* do know what a Ptarmigan sounds like; do you remember that strange sound we heard last night?" In this case, the second ornithologist knows what a Ptarmigan sounds like, even though he does not know that knows what a Ptarmigan sounds like, in virtue of having conceptualised the experience in *some way or another*, specifically as a strange noise. Conversely, the third ornithologist does not know what a Ptarmigan sounds like because he failed to bring the noise under any concept at all: he could not be similarly corrected by the first. In other words, even if our ordinary attributions of knowledge of what some *x* looks or sounds like do not depend on knowing that one knows, they are dependant upon having brought that sensation under *some* concept. This would be impossible in Marla's case, since she could not bring her P-conscious experience of red under *any* concept, not even a bare demonstrative one as "that hue"[29].

A second important feature of our mental states is the normative force they possess. They *matter* to us with an immediacy that other people's perceptions and feelings do not. If I say, for example, that I am unhappy that some person is in pain, it is reasonable to ask who

29 Tye offers an account of P-consciousness according to which P-consciousness of an object or feature requires minimally bringing it under the concept "what is that?"; see Tye 2009.

she is such that I feel such empathy for her. Conversely, if I myself am in pain, and say that I am unhappy about this, it is unreasonable to ask why I care: the experience is mine, and that is justification enough. That good experiences are *immediately* good for the person who has them, and *bad* experiences are immediately bad for the person who has them, is part of our notion conception of what it is to have an experience.

This normativity does not apply to inaccessible experiences. If someone offers you a hundred pounds in exchange for undergoing an inaccessible but P-conscious experience of pain, then it seems rational to accept the offer: after all, you will not think you are in pain, nor will you flinch, scream, or have the thought that anything unpleasant is happening to you. As far as you are concerned, when the experience happens, you will not notice it at all. Perhaps intuitions will be divided on this issue, but I would gladly be willing to undergo the experience. IPC of pain could not be good or bad for me, at least directly; perhaps I would find myself suffering subconsciously acquired post-traumatic stress disorder from an inaccessible pain, but this would not be bad in virtue of having been causing by *pain*; it would be bad in and of itself. Indeed, arguably we should not count an IPC experience with the phenomenal character of pain as being *painful*: what we mean by finding something painful is in part having an appropriate normative response to it[30].

Against this, one might cite the various reports of patients who have taken morphine and who report that their pains are not painful[31]. I will say two things about such cases. First, in such a case, we have another reason to think that pain is occurring, namely that patients report it, and so are A-conscious of the representational content of their pain experience. This A-consciousness may be a further element of what we mean by pain, but one which is of course missing in IPC, hence again debarring it from counting as pain. Moreover, 'painless

30 However, see Lycan's (1996; Ch.2, n. 6) example from a John Grisham novel of a commonsense use of the idea of "unfelt pain": "Every step was painful, but the pain was not felt. He moved at a controlled jog down the escalators and out of the building." This seems best interpreted, however, in one of two ways: either the pain was felt, but he was so determined, brave, or frightened that he did not respond to it; or, he was so focused on escape that he did not feel any pain until later.

31 See, e.g., Dennett 1978: 431

pain' plausibly would have a different phenomenology from ordinary pain, since the normative content of pain seems to contribute to its phenomenology: it is partly constitutive to how the ache in my foot feels at this moment that it feels *unpleasant*. If this is the case, then the 'painless pains' of morphine patients will not count against my criterion that for all sensations with ordinary pain phenomenology, the pain must be *bad* for the subject.

The defender of IPC might grant that inaccessible experiences lack immediate normative force, but argue that they could matter to the subject in a more general way. For example, if someone was asked to choose between either death or a life consisting of wholly inaccessible P-conscious experience, the latter might be thought preferable. If IPC matters in this sense, however, it cannot matter very much, and I am doubtful as to whether it matters at all. Imagine that you are a prisoner on death row, about to be executed. A jury, having considered and rejected your appeal, nonetheless offers you a choice of two further reprieves. Either you can have one more day of life in your present state, able to spend a few last hours with your family and eat a final meal, or you can live out the whole of your natural life with all higher level processing disconnected. You will be looked after and presented with a wide range of pleasant phenomenology through appropriate stimulation of lower level processing areas of the brain, but none of these will give rise to thoughts and you will remain wholly access unconscious.

In such a case, I am drawn significantly more to the former option, for all its ghastliness, simply because the alternative seems utterly irrelevant to me as a subject. I could not think that any of the situations I was undergoing were pleasant, or indeed, have any *thoughts* at all. In order to truly count as mine, then, for the purposes of my welfare, I hold that experiences must be accessible. Raw phenomenology without access is not something that could make a difference to me in any normative sense.

Tying these two threads together, I wish to point finally to the fact that, without the ability to inform our beliefs or motivate us in any way, IPC could have no bearing on a subject's behaviour as an agent. Someone could not be held praised or blamed for failing to respond to perceptions which were inaccessible to them, for example. Thus fenced off from agential, and by extension, moral considerations, inaccessible experience once again seems an irrelevance[32].

I hold that experience, insofar as it can *matter* to us, must involve access as well as phenomenology. Even if we accepted that GK was P-conscious of a face, without access the experience would not belong to him in any rich sense, and asserting of a given subject that she had IPC would lack the significance which the notion of *having an experience* carries in ordinary discourse. The only sense of experiential *possession* remaining to it would be one of raw phenomenological acquaintance, hence making the notion of IPC almost wholly irrelevant for all but purely philosophical considerations[33].

2.4 – What is it like to be a fusiform region?

The previous objection charged that there is no rich sense in which we could ascribe inaccessible experiences to a subject. A second objection I wish to raise is that IPC leads to implausible consequences. Insofar as IPC is possible only if the neural correlates of P- and A-conscious are distinct, it might seem to force us to accept the implausible scenario in which a tiny sample of brain tissue might generate P-consciousness all by itself, even if kept alive in a bottle.

Consider G.K.'s fusiform region as it activates during a binocular rivalry experiment: here,

32 This is the objection of Clark and Kiverstein to the possibility of IPC. See Clark and Kiverstein (2007).
33 There is a question here about whether, if a patient in a vegetative state could be shown to be having IPC but no A-conscious experiences, this would provide a motivation for keeping them alive. This is a very delicate issue on which I do not wish to enforce an opinion. However, as should be clear from my death row example, for my part I regard it as clear that inaccessible experiences cannot positively or negatively influence my welfare.

Block alleges, the fusiform region generates IPC. However, Block also explicitly states that no-one takes seriously the idea that a fusiform region kept alive in a bottle might generate P-consciousness all by itself[34].

I agree with Block that such a scenario is implausible. However, I find it hard to see how, given his commitments, he can help but acknowledge its possibility. For if the neural correlates of P-consciousness and A-consciousness are distinct, why should the former (or some part of the former) not continue to generate P-consciousness even if removed from an organism? Block might wish to argue that some kind of integration into a broader neural system is required for individual neural correlates of P-conscious states to generate P-consciousness. The kind of integration at stake cannot be integration into a system consisting just of *other* P-conscious states. Block argues that "there are different core neural bases for different phenomenal characters"[35], and there seems no reason why the total neural bases of these phenomenal states should include one another. After all, subjects can have severely or totally impaired P-consciousness in some modalities whilst preserving others.

An alternative kind of neural integration that Block might appeal to, in explaining why a fusiform region kept alive in a bottle would not be P-conscious, would be integration into a system at least capable of access[36]. This would not rule out IPC, but would require that IPC occurred in a creature capable of also having accessible P-conscious states. The problem for this prima facie plausible view lies in the fact, on Block's picture, the uptake of data by working memory is *subsequent* to its becoming P-conscious.

Consider, for example, a visual experience VE_1 generated by brain processes at the back of the head at time t_1. This data has not yet been accessed by working memory systems; hence if we, with Block, equate A-consciousness with actual access by working memory, this

34 Block, 2007a: 482

35 Block 2007a: 496

36 Note that this view would be incompatible with a statement made by Block (1995: 233), that an animal which had had all of its higher order processing centres removed might still have P-conscious without have A-conscious states.

data is not as yet A-conscious. A *subset* of this data is subsequently accessed by working memory, resulting in an A-conscious experience AE_1 at time t_2. Given that VE_1 occurs prior to AE_1, how can it be the case that whether or not VE_1 is conscious depends on whether or not that data is *going* to be accessed?

It might be replied by Block that a brain region's generating P-consciousness at time *t* is dependant on it bearing some suitable *relations* to other regions, such that that its P-consciousness could in different circumstances be taken up by working memory. After all, the significance of a given event may be determined by extrinsic factors; whether or not my driving at 70 miles per hour counts as a case of law breaking depends on factors external to the act itself. Similarly, it may be the case that a fusiform region can generate P-consciousness only if it is suitably structurally integrated into a system which in principle allows for the information represented by these regions to be subsequently taken up by higher level systems.

The problem with this view is that it is not clear how the occurrence of P-consciousness in a given region at a given time could be determined by relational properties, since P-consciousness, unlike A-consciousness, is plausibly an intrinsic phenomenon. Tyler Burge summarises the point powerfully. "Consciousness is an occurrent, not a dispositional, condition. We have no good idea how mere dispositional accessibility to working memory could be causally necessary to occurrence of consciousness before working memory operates. Why should the door's being open matter to the occurrence of something that does not use the door until after it already occurs?"[37]

The problem can be brought out by the following example. Imagine that my fusiform region activates at t_1 in response to an external stimulus. At this moment, my fusiform region *does* bear the suitable relations of interconnectivity to other regions, so according to the position outlined, my fusiform region generates P-consciousness at t_1. However, at t_2, before

37 Burge, 2007: 501

the information from the fusiform region gets passed on to my working memory, all of the parts of my brain responsible for access are totally disintegrated. The fusiform region itself is unaffected by this disintegration process, and is still firing in exactly the same way as it was at t_1. I find it deeply implausible that this should cause my fusiform region to *stop* generating P-consciousness. More abstractly, it seems implausible that some brain region R should be generating P-consciousness at a time t_1 without the direct involvement of any other brain regions, and that the destruction of those uninvolved brain regions an instant later, without any impact on R itself, could cause R to *stop* generating P-consciousness. Consider now my fusiform region. Perhaps it is still integrated into parts of my brain responsible for other kinds of P-consciousness. However, as argued above, it seems implausible that the P-consciousness of a region should depend on other kinds of P-consciousness going on around it. Yet if we allow these to be destroyed, what we are left with is equivalent to Block's case of a fusiform region kept alive in a bottle.

A final way around the problem would be to say that in order to be P-conscious, a brain event must possess certain *intrinsic* properties which ensure that it is accessible. The problem, however, is that accessibility is determined partly by *extrinsic* properties of a brain state, such as whether there are connections present which link that brain region to *another* brain region capable of deploying transferred information in rational action. Whatever intrinsic properties we assign to the P-conscious region in question, it seems possible that these properties be present whilst the connections to the neural bases of A-consciousness, or indeed these neural bases themselves, are not.

If it is accepted that accessibility to working memory (and higher order systems more generally) cannot be a condition of whether a brain region can generate P-consciousness, then it seems hard to see how Block can avoid endorsing the conclusion that an isolated fusiform module could generate P-consciousness all on its own, even if it is simply being kept

alive inside a bottle. Even if we assume that the fusiform region had to be connected to some other (non P- or A-conscious) regions in order to generate P-consciousness, we are left with the improbable conclusion that there is something it is like to be a fusiform region. There seems no easy way for Block to avoid this, given his assertion that the neural bases of A- and P-consciousness do not overlap. This seems a good reason in itself to look for theories of consciousness which link P- and A-consciousness directly together, both at the explanatory and neurological level.

I have provided this argument for two reasons. The first is simply to demonstrate an implausible consequence of Block's position. Secondly, however, it serves to bolster the broader aims of this thesis. As noted in section 1.9, my argument relies in part on using the implausibility of inaccessible P-consciousness also to rule out accessible but non-accessed P-consciousness. If, as I have argued, dispositional properties cannot determine whether or not a given brain state gives rise to P-consciousness, then we lack any way of showing how contingently non-accessed P-conscious states (such as those alleged by Block to occur in the Sperling test) could be possible but *inaccessible* states, such as those of GK could not; for the only difference between the two is a dispositional one, namely whether or not the state *could* have been accessed in other circumstances. Hence if we have good reasons for rejecting the possibility of inaccessible P-consciousness, then we also have good reason for rejecting any theory which allows for contingently non-accessed P-consciousness, since the possibility of the latter would entail the possibility of the former.

2.5 – Conclusion

In this chapter, I have attempted to show some implausible features of IPC in general. First I argued that even if we *could* establish that GK did have IPC as of a face, the notion of

experiential possession we would be left with would be an utterly minimal one, removed from all kinds of normative, epistemic, and hence agential considerations. Though this does not directly challenge Block's claim that GK might have IPC, I hope that it shows that there is no *rich* sense in which a subject might have IPC, and renders IPC a highly bizarre and rarefied concept.

Secondly I argued that there is no clear way to argue for the independence of P-consciousness from access without also admitting that isolated brain regions kept alive in laboratory situations could also be P-conscious. Again, this is an implausible commitment of Block's theory, and one which ideally we would rule out via a theory that linked P-consciousness specifically to A-consciousness. This stage in the argument was also intended to provide some limited support for the move that Block and I both endorse, namely that a theory allowing of non-accessed P-conscious states must also allow for IPC.

In the next chapter, I move from arguments of plausibility to stronger attacks. First, I attempt to show that there is no empirically derivable criterion which would allow us to ascribe IPC to a particular subject. Secondly, I present arguments relating to the representation of time and space, and attempt to show that there is no easy way for Block's theory to allow for the temporal or spatial integration of IPC with other experiences.

CHAPTER 3

INACCESSIBLE PHENOMENAL CONSCIOUSNESS: THE BIG PROBLEMS

3.1 – Introduction

In the previous chapter, I raised some basic worries about IPC. In this chapter, I turn directly to more damning arguments against the notion.

There are three main arguments in this chapter. First, I question Block's claim that we could ever have good reasons, informed by empirical evidence, for attributing IPC to a particular subject. I consider a range of putative conditions which would give us grounds for asserting that a given subject was experiencing IPC, noting the flaws in each, and concluding that any criterion we might use to make such an attribution would depend on our properly *philosophical* commitments. Secondly, I consider Block's own grounds for asserting that GK himself is the subject of the IPC as of a face, namely that the experience occurs within GK's visual field, and I conclude that without access, the kind of subjective spatial integration Block has in mind is impossible. I also develop from this some extremely improbable conclusions to which a theory endorsing IPC seems committed. My third argument holds that problems similar to those which prevent IPC from being spatially integrated with a subject's experience also suggest that it might not be able to be temporally integrated. These latter two arguments in particular, I conclude, constitute good a priori reasons for rejecting the possibility of IPC.

3.2 – Assigning IPC to subjects

Let us accept for the moment Block's claim that we could have empirical evidence for the occurrence of IPC in a person's brain. Imagine that we detect a inaccessible P-conscious event in the brain of a subject, Susan. What would be required for us to make an ascription of IPC to Susan? The question I am concerned with here is not how we could know that inaccessible P-conscious experience was occurring in Susan's brain, but rather how we could know that it belonged to Susan. Another way of putting this would be to ask what could suffice to make the IPC in Susan's brain *subject unified* with her accessed experiences.

The easiest way to show that a subject is having a particular experience is, of course, for them to access it. I am undergoing at this moment P-conscious experiences of being in a warm environment, sitting on a soft surface, and so on, and I can report this. Even if report were impossible, it might be inferred from my behaviour that I was undergoing a particular experience: when a thirsty person reaches for a glass of water, we can infer that they have seen it[38]. However, this criterion would not allow us to settle whether a particular subject was having *inaccessible* P-conscious experiences. We require some other sufficient condition for ascription of IPC to a subject.

I can foresee one immediate objection to this line of argument. If we accept that there is P-conscious activity somewhere in Susan's brain, then even if it were inaccessible, provided we are not confronted with a case where we plausibly have two distinct subjects in a single brain, can we not just assume that it belongs to her? After all, who else could it belong to?

This view makes two major assumptions. First, it assumes that P-consciousness must be ascribable to a subject at all, and this assumption can be challenged. Perhaps there are instances of P-consciousness that are not bound to a subject. Some may find the very idea of 'something it is like' without 'someone for whom it is like' incoherent; but there are views of

38 At a more abstract level, there is a problem of how we can justify ascribing P-consciousness to other minds, but for present purposes it seems reasonable to treat such ascriptions are justified.

the self, such as the Humean bundle theory, according to which experiences might exist without thereby implying the existence of a distinct subject of those experiences. Second, it assumes that there could not be numerous transitory 'micro-subjects' present in S's brain which existed as momentary bearers of P-conscious states. While I do not wish to debate either assumption, neither is trivially true, or even uncontroversially plausible. Moreover, both seem straightforwardly philosophical questions unable to be resolved by empirical investigation. Hence the question of whether the IPC in Susan's brain belongs to her *qua* subject is an open one.

I will now describe some putative criteria we might use to ascribe IPC to Susan *qua* subject. The first criterion I wish to consider is the view that the subject unity of two experiences is secured by their occurring in the same organism. This does not commit us to the view that subjects are identical to organisms, nor to the view that *only* organisms can be subjects. However, it does commit us to saying that if two P-conscious events occur within a given organism, the two events are subject unified.

In daily life, we encounter a one-to-one correspondence of subjects to organisms, making this view seem plausible. However, more exotic scenarios cast it into doubt. It seems conceivable, for example, that an individual could survive the loss of an entire cerebral hemisphere whilst preserving enough functions to make attributions of P-consciousness to their remaining hemisphere plausible. Imagine, then, if two human organisms, A and B, both lose an entire cerebral hemisphere, and moreover suffer devastating damage to their bodies. Doctors decide to transplant their remaining hemispheres into new bodies, but unfortunately, only *one* donor body is available. Rather than choose whether A or B is to survive, doctors transplant both hemispheres into a single body, resulting in a single organism; call him Janus. To preserve the autonomy of each hemisphere, doctors do not connect the two hemispheres;

perhaps they even insert a non-conducting plate between them to prevent any unwanted neuronal interactions between them. Finally, in the interests of fairness, they divide control of Janus' body between the hemispheres in an equitable way; one hemisphere might have control of the mouth and auditory systems, for example, and the other control of the hands and eyes.

Janus would be very strange, but he seems conceivable, and I think that in such a case we could have reasons for believing there were two subjects of experience in his body. For example, the hemisphere that controlled speech centres and audition might complain verbally about blindness, while the hemisphere that controlled the hands and vision might use sign language to complain about its deafness, and they might both complain about being stuck in the same body with an unwanted companion. In such a case, I suggest, we would have good reason for thinking that two subjects of experience were present in a single organism.

One reply would to assert that an *organism* is essentially based in its brain, and that the situation can be redescribed as one in which we have cut away the extremities of two organisms and placed them in a single shell. Even thus modified, however, problems can be raised. Imagine that a parasite capable of P-consciousness, such as a highly complex worm, burrows into the brain of some unfortunate subject, and integrates itself comfortably with that subject's biological systems so as to keep itself happily supplied with oxygen and nutrients. After many years, as tissues grew around the worm, the worm becomes increasingly integrated with its host, unable to survive removal from its environment. It may be extremely difficult for scientists to differentiate the worm from the host's brain. We might even say that, by this point, the worm has become part of the brain. Yet plausibly, we are still left with two subjects of experience: the worm, and the individual it has infested. Alternatively, consider the example given by Block himself of a creature whose very neurons were composed of tiny microscopic P-conscious creature. These creatures would be part of the creature's brain, yet we would not naturally think that their P-consciousness would thereby belong to the same subject.

We are owed at the very least a non-*ad hoc* account of why, in the above examples, the worm or the tiny creatures are *not* part of the subject's brain. We might say that they are not really a single organism insofar as they do not have the same genetic code, or insofar as they are not appropriated neurally integrated with the organism's brain. However, the conditions under which biological systems constitute a single organism seem open to debate: there are symbiotic organisms whose interdependence is so close as to make it a matter of scientific stipulation as to whether or not they are one being or two. Besides, if nature does not provide us with truly borderline cases, it is easy for the philosopher to conjure them up.

I do intend the above account to count decisively against the view that all P-conscious states shared by an organism are thereby shared by a single subject. However, I hope I have done enough to show that this is not obviously the case, and that there are problematic features of 'one organism, one subject' accounts.

An alternative sufficient condition of subject unity would be some kind of psychological integration criterion. On this view, it would be sufficient for two P-conscious states to belong to the same subject that they belonged to an integrated psychological system. We might say, for example, that two P-conscious visual experiences were psychologically unified and hence subject-unified if information could be shared between them. There is a real problem in applying such a psychological unity condition to IPC, however, insofar as it be difficult to find a way of framing psychological integrity that does not make recourse to access.

Moreover, the psychological integrity condition also seems prone to challenge by thought experiments. Referring back to the Janus example given earlier, imagine that some neural functions were shared between the two hemispheres, contrary to the example. Scientists might graft a few neurons between the two hemispheres to allow the hemisphere

with access to vision to pass on visual information at will to the other hemisphere. In a case where both hemispheres allowed the transfer of sensory data to the other, they might be jointly conscious of the same two visual and auditory stimuli and able to form beliefs based on the two. Even so, given that the two subjects would consider themselves distinct subjects, with different beliefs and desires, it seems plausible that there are two subjects present.

A last redoubt in seeking a sufficient condition of subject unity might be as follows. If a given brain region has at some time produced P-conscious experience that we know belonged to a particular subject (for example, via introspective report), then we might say that any future events in that region which we know to be P-conscious would also belong to that subject, even if inaccessible. Hence, considering Block's example of GK, since GK's fusiform region normally produces accessed P-conscious experience as of a face, we might assume that the P-consciousness of a face belongs to him also on those occasions when it is inaccessible.

There is a major problem with this, however, since in formulating this sufficient condition of subject unity, we must include some constraint regarding what kind of alterations that brain region could undergo before its P-conscious activations would cease to be subject-unified. For example, my fusiform region may be producing certain sensations now that are P-unified with my other experiences, but if it were excised and stimulated in a bottle, assuming it were still to generate P-consciousness, that P-consciousness would of course no longer belong to me. We have no idea what kind of alterations a brain region might be able to undergo whilst allowing its P-conscious activations to remain subject unified with activations elsewhere in my brain. Perhaps whatever change suffices to prevent access in GK's case also causes a breakdown in subject unity. Hence we cannot conclude from the fact that activations in a given brain region were previously subjected unified when accessed that they will also be subject unified when inaccessible.

Any of the above conditions of subject unity might be turn out to be correct, but none of them are *obviously* correct, and in many cases provide contradictory accounts. Of course, if we could establish that a subject was P-conscious of the IPC in her brain, then we could infer that the experience belonged to her. Without cognitive access, however, I see no way that this could be empirically demonstrated.

As a result, even if we allow for the possibility of IPC, there is a more restricted kind of scepticism we can engage in, namely that in order to answer questions about whether the P-consciousness associated with activations in a given brain region is subject unified, we ultimately require recourse to report or some other clear indication of access. Hence, *contra* Block, even if we knew that experience as of a face was occurring in GK's brain, we would not have grounds via the method of inference to the best explanation for thereby concluding that GK himself was having P-conscious experience as of a face.

3.3 – Does subjective spatial unity require access?

Block does not in fact adopt any of the criteria outlined above in seeking to show how we might attribute IPC to a subject. Instead, Block makes the following claim:

...we can understand the face experience as [GK's] experience by noting that it is in his visual field. One could meaningfully ask, for example, whether it is the same half of his visual field in the vertical dimension as his experience on the right, or which is closer to the periphery, the one on the left or the one on the right.[39]

This is best interpreted, I will argue, as the claim that all of GK's perceptions, including his IPC, are *subjectively spatially integrated*. If Block could demonstrate this, naturally he would have

39 Block, 2008: 291

demonstrated that the IPC belonged to GK. However, I will argue IPC does not admit of subjective spatial integration with accessed experiences, and that as a result, Block cannot appeal to it as a way of showing that GK's experience might belong to him. Of course, this does not rule out that GK's experiences might be P-unified in some other way, but it obstructs Block's preferred route to this conclusion. I also claim that IPC generates such implausibilities at the level of subjective spatial integration that we have good reasons for thinking that access is a necessary condition of visuospatial representation generally.

Before moving on to this argument, I wish to clear up an ambiguity in Block's claim that GK's experiences may be his in virtue of forming part of his visual field.

There are at least three ways we can construe 'visual field'. One would be a biological notion of a visual field: we might say that any two P-conscious visual experiences arising from, saying, stimulation of the same *retina* would thereby belong to the same subject. However, this does not seem a plausible interpretation. Imagine a version of the Graeae of Greek mythology, two organisms A and B with only a single eye between them. Although located in only one of the organisms, through the use of radio transmitters, signals from the optic nerve can be distributed between the two of them. Organism A might be in receipt of the signals from one half of the eye's retina and B in receipt of the other. In such a case, we would not imagine there was subject unity; hence constituting a same visual field, in this biological sense, is not sufficient for subject unity.

A second way we might construe visual field is in terms of its being a depiction of an actually continuous expanse of space. Even if part of our visual field is completely occluded, as when we wear a mask that blocks much of our vision, we know that the occluded parts of our vision are in reality juxtaposed spatially to the non-occluded parts. However, actual spatial continuity cannot be a sufficient condition of the unity of visual phenomenology. For

consider again the Graeae case: where individual A sees the left half of a room and individual B sees the right half of a room, even though the two images depict a single expanse of space, they are not subject-unified.

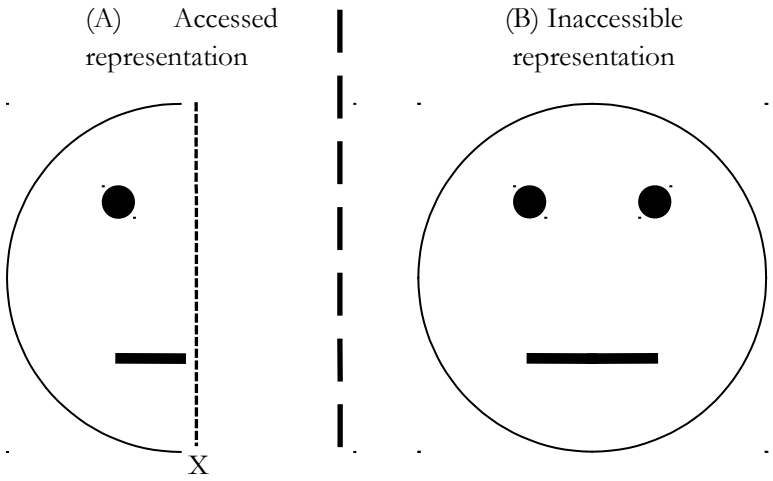
The most charitable way to interpret Block's reference to a visual field is via the notion of *subjective spatial unity*, that is, that all of my P-conscious visual representations seem to have visuospatial relations with one another. What I mean by this is that, for any of my visual representations, there is a matter of fact about whether their contents appear higher or lower or to the left or the right of all others. This is how we should interpret Block's reference to a visual field, and how I will use the term from now on.

If the IPC as of a face in GK's case were subjectively spatially unified with his other visual experiences, that would certainly *suffice* to ensure that the IPC was also subject unified with his visual experiences. Prima facie, this suggests a relatively straightforward way of establishing whether the P-conscious activations as of a face in GK's brain truly were actually experienced by him. If we could show that these activations had undergone visuospatial binding with his other experiences, then we might conclude that they formed a P-conscious continuum. This conclusion would be premature, however, since we cannot know whether what some degree of access might be required for the P-conscious representation of spatial relations. The fact that information has undergone a visuospatial binding process does not demonstrate that the subject is thereby P-conscious of it as visuospatially bound with their other visual experiences. As I will now argue, spatial representation is best understood as a process in which some degree of access is required.

Consider a subject, Anna, who is presented with an image as of a face. Part of her visual experience, specifically of the right half of the face, is inaccessible. Being able only to access the left half of her representation of a face, Anna fails to realise that she is looking at

a face, or even half a face: she just reports seeing a circle and a line inside a non-closed semi-circle (see Fig.3, below).

Fig.3 – Anna's experience



Let us assume that Anna does in fact have P-consciousness as of a whole face, though part of it is inaccessible. There are two arguments I wish to raise at this point.

The first is that, if Anna is asked where the *limit* of her visual field is, she will reply that it is the line marked "X" in the above diagram. Plausibly, she says this in virtue of how things seem to her. I do not think that it is merely the *absence* of accessible items to the right of X that is responsible for this judgement. This would be possible only insofar as the phenomenological representation of the location of the contents of P-consciousness was a feature of those contents themselves, in other words, analogous to properties like redness. As it is, it seems more plausible to me that the representation of subjective spatial location is not a representational function of individual contents of consciousness. I suggest rather that the representation of an object's location requires representing its location relative to *other* items,

and a visual field is an area in which objects are completely interrelated. Hence a subjective spatial map that relates A as being to left of B and C, C as being to the right of A and B, and B as intermediate between A and C is distinct from one with *four* elements, A, B, C, and D, which links A, B, and C as being to the left of D, and so on.

Given that Anna claims that she does have one complete visual field, this creates a problem for Block. For if we wish to assert that she also has a P-conscious representation of her visual field in which items located to the right of X, then, given my above argument, this would suggest that she has two rather than one P-conscious representations of her visual field, one in which items to the left of X are subjectively spatially interrelated only with each other, and one in which items to the left and right of X are mutually subjectively spatially interrelated. At one level, for example, the edge of her visual field is intersected by the lines of a semicircle, but at another it is not. Not only is 'two visual fields' view implausible, but it counters Block's point that IPC might belong to a subject in virtue of forming a subjective single visual field: if Anna does have a visual field including her IPC, it is not the visual field that it *seems* to her she is seeing.

I will now provide a second argument against Block's position. I wish the reader to note that seeing a *face* is a very different experience from that of seeing a set of shapes. The P-conscious representational content involved in an experience of (B) is not just the content of (A) plus a horizontally flipped duplicate of (A). There is a distinctive 'face phenomenology'. Block certainly seems committed to this view, given that he asserts that GK has experience as of a face, and this in virtue of firings in the fusiform region, an area associated with experience of faces.

I also suggest that in seeing (B) the phenomenology present in (A) of a selection of shapes is *not* present. When we look at (B), it does not feel like we see lots of shapes in a

seemingly arbitrary arrangement, as we might when looking at (A)[40]. Rather, we see the shapes in (B) as an eye, a mouth, and so on. However, one cannot see a circle as an eye unless one sees the overall image as a face, and one cannot see the overall image as a face unless one sees both left and right sides of the figure. The point I wish to make, then, is that Anna's accessible experience of (A) is phenomenologically very different from her inaccessible experience of (B). Even if they have common elements (such as representations of colour), they also have elements which are not present in one another: whereas in (A), there are P-conscious representations as of random shapes, in (B), there are P-conscious representations of eyes and a mouth.

We have now to ask why Anna reports seeing a set of shapes rather than a face. The plausible answer is that she reports this in virtue of how things look to her; that is, in virtue of her phenomenology. Certainly, it does not seem as if anything *but* her phenomenology can influence this judgement. But if her phenomenology does determine her judgement about (A), then, assuming she also has inaccessible experience as of a face, then she must have two distinct phenomenological representations of the image, since her phenomenology as of a set of shapes is not part of her phenomenology of (B).

The first conclusion I draw from this argument is that by allowing that an experience like Anna's is possible, we are again pushed towards the improbable position that Anna has not one but two distinct phenomenological representations. Insofar as the left half of the face features in both of them, with different phenomenological features in each, it seems highly improbable that the two images form a single visual field. For that would seem to require that the left half of the face was represented *twice* in Anna's visual field, once as part of image (A) and once as part of image (B). But where would these two be located relative to one another? The only non-arbitrary answer is that they would have to be located in the same

40 Note that, given the nearby presence of (B), readers will likely see (A) as "half a face" rather than merely a collection of shapes. The phenomenology distinctive of seeing a set of shapes can be readily restored, however, if one looks at (A) upside down.

place; but given that each half-face fully occupies the place in our vision where it is located, I struggle to even coherently conceive of how the two could fully overlap in Anna's visual field whilst having distinct phenomenological properties.

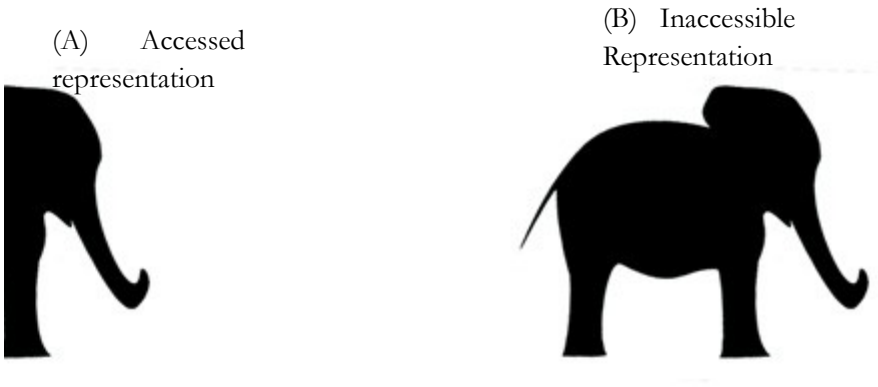
If Block cannot advert to Anna's experience forming a single visual field as a criterion for ascribing her IPC to her, then it seems entirely ad hoc for him to use in GK's case. Moreover, it seems that *any* explanation that Block could give of Anna's experience would be implausible, insofar as it commits us either to deny her introspective judgement that she sees shapes or accept that she has two quite distinct phenomenological representations of the same object at the same time. It seems far more plausible to assume that visuospatial representation requires access, and Anna does not have experience as of a face at all.

A second problem for Block arising from this example is that it suggests that the neural correlate of Anna's experience as of a set of shapes includes working memory. For imagine that Anna's entire experience of the image were inaccessible. In that case, we uld just conclude that she was having a single inaccessible P-conscious experience as of a face. It is only because a subset of the total visual data relating to the image reaches her working memory that she has experience as of a set of shapes at all. Yet this would suggest that uptake of data by working memory is partly constitutive of her phenomenology as of as face, hence suggesting that P- and A-consciousness are not based on distinct systems.

To avoid this argument, Block has to assert in some way that Anna's accessible phenomenology is not a subset of her inaccessible phenomenology. One way to do this would for Block to assert that (B) includes all the phenomenology of (A). Block might claim that, in seeing a face, Anna also has the phenomenology characteristic of seeing a selection of shapes. However, I find this highly implausible. Anna does not see the shapes in (B) as just shapes, but as features of a face; and she does not see the shapes in (A) as features of a face, because she does not know she is seeing a face.

To emphasise this problem, consider Fig.4, below. Here I find it plausible that Anna sees the left half of image (B) as 'the rear half of an elephant'. She sees image (A), conversely, as merely a strange shape, perhaps a bit like 'a hair dryer with a tail'. If the defender of IPC is to claim that (A) and (B) have parts in common, and hence that they form a single phenomenal representation, she must assert either that Anna sees the left half of image (B) as a strange shape, which seem implausible, or image (A) as the left half of an elephant, which seems more implausible still: how can Anna see (A) as the left of an elephant when she cannot access all of image (B), and hence *realise* she is seeing an elephant?

Fig.4 – An Alternative Presentation



I will now briefly summarise the argument given. First, if part of a subject's visual field is inaccessible but P-conscious, we are forced to the implausible conclusion that she has two P-conscious representations of where her visual field begins and ends, contrary to Block's view that GK's IPC might be subjectively spatially integrated with his other representations. Secondly, in Anna's case, we have good reason for thinking that her accessible P-conscious representation has elements that are not held in common with her inaccessible P-conscious representation. This suggests, again, that she has two distinct P-conscious representations of what she is seeing, which cannot be subjectively spatially unified, and second, that access by working memory plays a role in generating some P-conscious representations. I find the idea that individuals with IPC would have multiple distinct representations of their visual fields

sufficiently implausible as to count against the notion of IPC in general. We would have a far less implausible and far more straightforward theory if we held that an individual had visual P-conscious experience only of accessed representations.

In passing, I wish to note a similar issue raised in section 2.3 above by my earlier example of Marla, the colour scientist without access to her P-conscious perceptions of colour. Assuming she is P-conscious but not A-conscious of the colours around her, what does the world *look like* to her? Recall that she behaves in exactly the same way as a true achromatope, being sensitive only to brightness and saturation but not hue, and claims to see things in greyscale. It is natural to suppose that in some way the world *looks* greyscale to her. Yet, like the case of Anna above, this would suggest that there are in fact *two ways* the world looks to her at any one time, in one of access was involved at a constitutive level. Not only does this seem implausible – an object's looking grey all over seems to preclude it also looking red all over – but it again provides a plausible case that access may play a role in the generation of P-conscious representations.

3.4 – Does representation of temporal relations require access?

I raised the problem that in certain cases it is plausible to suppose that working memory plays a role in determining visuospatial phenomenology, and that a commitment to IPC forces us to admit that individuals may have distinct P-conscious representations of the same objects. I wish to suggest now that similar arguments can be applied to phenomenology via the role played by time in our experience.

Imagine a person, William, whose brain has been badly damaged. In particular, the regions of his brain responsible for transferring P-conscious visual data into working memory have failed completely, hence rendering all his P-conscious visual experience inaccessible.

Doctors replace this machinery with a surgical implant, which replicates the functions of his original machinery, but with one disadvantage: the materials used are much slower conductors than the originals which they replace. This has the consequence that there is a delay of a few hundred milliseconds in the transfer of P-conscious visual data into William's working memory. Now, imagine that at a time t_1 William sees a bright flash and hears a loud noise, and that the two events cause simultaneous P-conscious neural events in William's brain. However, owing to the poor performance of the implant, William becomes A-conscious of the noise at t_2 and of the flash at t_3. If asked to describe his experience he says he heard a noise and then saw a flash (we are assuming his brain does not 'backdate' the flash).

We have first to ask whether the subset of P-conscious data that reaches working memory at least partly determines William's experience of the temporal orderings of the two events. This seems plausible to me: William *says* that he experiences a noise and then a flash, and he says this purely in virtue of how things seem to him. But if we accept that the subjective temporal ordering of events has a phenomenal character (what it is like to experience E_1 and E_2 as successive is different from what it is like to experience them as simultaneous), and we allow working memory a role in determining the subjective ordering of events, then again we allow working memory to play a role in determining phenomenal character.

Although not congenial to Block's position, the defender of IPC could admit this conclusion. However, if they wish to allow that William's experience of the noise and the flash is also at some level independent of working memory, then they have to make one of two claims: either that the subjective temporal relation between two events can be determined by something other than working memory, or that it is possible for a subject to experience two events without their experiencing them as having any temporal relation between them whatsoever.

Let us first consider the first option. We could say that even prior to the P-conscious representations of the noise and flash reaching William's working memory, there is be a matter of fact about their temporal ordering. One option would be, for example, to say that because the two P-conscious events occurred simultaneously at t_1, there is an inaccessible level at which William experiences them as simultaneous. But if accept that there is also something it is like for William to experience the two events as successive, we are forced to admit that William has *two* subjective temporal orderings of events. In one of these, the noise seems simultaneous with the flash, and in another, the noise seems subsequent to the flash. It is highly implausible that a single subject could have two contradictory P-conscious interpretations of the temporal relations holding between the same two events.

The second option would be to assert that William might experience the noise and the flash without experiencing them as simultaneous or as successive. He might just experience a noise and a flash at t_1, without thereby having an experience of a noise and a flash together, and then have a further experience of them as successive once the two reached working memory. However, this would also commit us to the possibility of a radical kind of subjective temporal disunity in experience. It seems plausible that all experience is necessarily P-consciously represented as having a subjective temporal order. What I mean by subjective temporal order is this: for any two experiences E_1 and E_2 that occur to a subject S in an uninterrupted period of consciousness, it must *either* be the case that S has an experience of E_1 and E_2 as simultaneous, *or* S has an experience of E_1 and E_2 as linked by relations of succession. That is, even if E_2 does not seem to directly follow E_1, S's experience must constitute a chain of experiences which *do* seem to follow one upon the other, and E_2 must have some definite place in this chain. I will call this claim that experience must be subjectively temporally unified.

One putative counterexample this claim may be the experiences had by some split brain patients. If a split brain patient's left hemisphere is A-conscious only of the word "pen", and the patient's right hemisphere is A-conscious only of the word "knife", there is something intuitively plausible in the idea that each hemisphere is also P-conscious only of one stimulus. If this were true, however, a subject in this case would have an experience as of the word "pen" and an experience as of the word "knife" such that she did not experience them as simultaneous, nor as linked to one another through orderly succession of experiences[41].

We might attempt to say that something similar might occur in William's case. Prior to P-conscious representations arriving in working memory, there is a kind of temporal disunity in his P-consciousness, his representation of a noise and his representation of a flash seeming neither successive nor simultaneous. However, I find this highly counterintuitive. For if Block is right, and P-conscious representations occur at the back of the head, then there would be some delay between a stimulus' becoming P-conscious and its being represented as having temporal interrelations with other stimuli, such as those from other sense modalities. Hence we would face the prospect that all of us routinely undergo P-conscious experiences that are temporally disunified with our other experiences. Given that experience never normally seems temporally disunified, this seems highly implausible.

Let me summarise the arguments given. First, I argued that there was something it was like for William to experience a noise and a flash as successive, and this in virtue of some subset of his representations reaching working memory. Hence I suggest we have good reason for believing that working memory plays a constitutive role in subjective temporal unity. If we are to allow for the possibility of IPC, we have to allow either that experience is

41 Tye (2003:126ff) is one philosopher who is committed to this claim.

not necessarily subjectively temporally unified, or that subjective temporal unity can also be secured at a level below working memory. The first assumption would require us to assume that there is radical subjective temporal disunity in every day experience. The second assumption would require us either to reject the point given above, that there is something it is like for William to experience the noise and flash as successive, or to assert that a subject may have contradictory P-conscious representations of the temporal order of events. I find both options highly implausible.

Conversely, if we accept that rendering representations subjectively temporal unified is a function of working memory, then we have a very straightforward account of what it is like to be William, namely that he *only* experiences the noise and flash as successive. Moreover, if we make the further plausible assumption that experience is necessarily subjectively unified, then this rules out IPC all together, since without access, subjective temporal ordering and hence experience in general would be ruled out. Again, I think above arguments give us good a priori motivations for ruling out IPC. At the very least, I hope they have shown that the defender of IPC is forced to make a range of extremely problematic commitments in order to render their theory coherent.

3.5 – Conclusion

In this chapter I have provided three arguments against IPC. First I argued that we lack a clear criterion for assigning IPC to a particular subject. Rather, we have multiple philosophical theories of the self, and there is no clear way, and certainly no empirical way, of adjudicating between these. The second and third arguments concerned representations of time and space, and aimed to show that the defender of IPC is committed to some extremely problematic principles in explaining how IPC may spatially and temporally relate to our other

representations. Moreover, I suggested, if we take subjects' statements at face value in the above thought experiments, we have good reason for thinking that working memory may play a constitutive role in the generation of P-consciousness.

I hope I have shown in this chapter and that which preceded it that IPC is at best highly implausible and problematic. This provides us with good grounds for seeking a theory which does not commit us to its possibility. In the next two chapters, I go on to outline just such a theory, and to defend it against several putative cases of P-consciousness without A-consciousness. I also review the experimental evidence mustered by Block to this effect, and show how it is compatible with a theory that denies the possibility of IPC. Given the dubiousness of the notion of IPC, then, I argue that any theory that comports equally well with the evidence should be adopted, and that just such a theory is available to us.

CHAPTER 4

SPARSE REPRESENTATIONS THEORY

4.1 – Introduction

In the previous two chapters, I attempted to provide some philosophical motivations for rejecting IPC. Given these, I argued, we have good reasons for seeking a theory of consciousness which links P-consciousness to access.

However, regardless of the philosophical merits in seeking such a theory, if there is empirical evidence that directly supports the possibility of IPC, then such a theory will have to be discarded in favour of one which incorporates IPC. As noted in the second chapter, Block regards the Sperling, Landman, and Sligte experiments as providing evidence that P-consciousness *overflows* access, hence showing that divergence of P-consciousness from A-consciousness is possible. Theories which are committed to this view I term *overflow theories*, in contrast to my own approach, outlined below, which denies that P-consciousness ever overflows A-consciousness.

In this chapter, I will describe the challenged posed by Block's experiments to theories which reject the possibility of overflow, and develop and defend a theory of experience which I believe overcomes these challenges. In the next chapter, I will apply this theory directly to these experimental paradigms.

4.2 – Introducing Sparse Representations Theory

As noted in the second chapter, Block holds that our best interpretation of the Sperling, Landman, and Sligte paradigms is that subjects have more phenomenology than they access.

This principal basis of this conclusion is that subjects *claim* to have seen more than they access, and we have evidence to the effect that their short-term memory encoded more than they accessed. For example, subjects in the Sperling test say that they have seen all twelve alphanumeric characters, and they can report on *any* row in the grid, but not on *every* row. It is natural to imagine that the two phenomena are connected: that the reason subjects demonstrate the ability to report on any row is because, as they say, they see every row. If correct, this demonstrates a disparity between the informational capacity of P-consciousness and A-consciousness which entails that subjects fail to access some of their P-conscious representations.

There is a very strong challenge here for theorists like myself who wish to assert that P-consciousness does not diverge from A-consciousness. It is one thing to assert that subjects' ability to report on any row in the grid is based on a visual short term memory which is not P-conscious. This leaves the problem, however, of explaining why, if subjects have P-conscious visual experiences just of the accessed items, they *say* they saw the whole grid.

I now wish to introduce my own view of the relationship between P-consciousness and A-consciousness, which I term *sparse representations* (hence, SR) theory. On this view, the P-conscious contents of any experience are without exception A-conscious. More specifically, they are cognitively accessed, in the sense that they non-inferentially determine a subject's doxastic state[42]. Roughly, SR theory holds that it is a necessary and sufficient condition for a given perceptual experience's being P-conscious that it generate *non-inferentially acquired* beliefs in the subject of that experience[43]. Further it holds that the richness of P-consciousness

42 I refer reader back to 1.5 for more on cognitive access.

43 In asserting that triggering non-inferentially acquired belief is sufficient for the occurrence of P-consciousness, I might seem to deny the possibility of zombies, since on the zombie hypothesis, zombies could have all sorts of perceptual beliefs without being P-consciousness. I do not wish to rule out the possibility of zombies in other possible worlds, hence I confine this condition to goings on in *this* world, leaving open the issue of whether or not there may be possible worlds where a given stimulus generates a non-inferentially acquired belief but does not result in P-conscious experience.

exactly mirrors the richness of information that is cognitively accessed, and that there is consequently no overflow.

One problem with this account that I wish to deal with at once is what I will term the problem of 'passive belief'. Assume that there is a background noise, for example, the ticking of a clock, that a subject does not attend to. Nonetheless, we may ascribe to the subject the belief that no sounds in the room were louder than 10db. Given that the ticking clock was one of the noises in the room, this might seem to have the consequence that the noise of the ticking clock itself influenced the subject's doxastic state. In that case, we should be forced to admit that the subject *was* P-conscious of the noise of the ticking clock.

There are two routes we could take at this point. We could either admit that the subject's doxastic state *was* influenced by the noise of the ticking clock, because she believes it was not louder than 10db. Hence it was to some degree cognitively accessed, so there is no problem in imagining that it was also P-conscious.

Whilst the strategy is compatible with the denial of overflow, I think it is misguided. I question whether the subject's belief that the noise of the ticking clock was not louder than 10db was in fact acquired non-inferentially. Rather, she may have acquired these beliefs through knowing that, had she heard any extremely loud noises, she would have noticed them. As it is, she can recall her behaviour and recall that she did not exhibit signs of having noticed any such thing.

Consider the following thought experiment. I take a powerful narcotic, which puts me into a state of bliss under which I am completely apathetic towards sensory experiences. In such a state, I would not notice loud auditory stimuli. This thought experiment hence blocks a possible *inferential* route to knowledge of whether unusual auditory stimuli occurred: I cannot know on the basis of my failure to exhibit reactions to unusual stimuli that there were no such stimuli present.

If there is a viable non-inferential route by which I might acquire this knowledge, however, namely how things sounded to me, then I ought to be confident in my belief that there were no unusual auditory stimuli, even though I would not have reacted to them. However, it is far from clear that I can claim such knowledge. Hence the 'passive belief' challenge can be met, and hence we need not extend non-inferential influence on a subject's doxastic states to vast numbers of unattended to environmental phenomena.

This leaves us with two remaining problems. The first is the problem that subjects in the Sperling test believe that they saw a matrix of twelve characters, but are not aware of the identities of all individual characters. On the SR account, subjects have the former belief in virtue of their phenomenology, but did not have phenomenology of individual characters. Granted that a subject cognitively accesses the fact that they are seeing a three by five matrix of alphanumeric characters, how can it not also be the case that they see the specific identities of each these characters, most of which they do not cognitively access?

The second problem, which I term the 'missing phenomenology' problem, is as follows. All of us have undergo many experiences which are not cognitively accessed, such as background noises, objects in the periphery of my field of vision, and the pressure of the clothes I am wearing on my body. Surely we would notice if these experiences had literally zero phenomenology? I intend to answer these objections in turn over the next two sections, making use of two additional ideas, namely *generic phenomenology* and the *refrigerator light illusion*.

4.3 – Generic and Specific Phenomenology

Imagine looking down an optician's chart. You can recognise the larger letters, but moving down the chart you will come to some letters which you are less certain of. You may wonder, for example, whether a given character is a B or a P. Near the bottom of the chart, there will

be some letters you cannot identify at all; at best, you can confirm that they are some kind of alphanumeric character. There may be some letters which you cannot be sure are even letters, but appear simply as black dots.

This example is designed to show that there are many cases in ordinary life in which we might say that our visual phenomenology has various degrees of *determinacy*. By determinacy, I mean that our phenomenological representations vary in degree of detail. A representation of a letter which is clearly identifiable as a B is a *more* determinate representation than a representation which may be a B or may be a P, which is in turn more determinate that a representation which may be a B, a P, or a D. It is not the case that we have maximally detailed P-conscious visual phenomenology of all these letters which we just cannot access. Rather, I suggest, we have distinct phenomenology in each case, of greater or lesser degrees of determinacy, and the letters literally look different.

It is not just our representations of small phenomena that can have various degrees of determinacy. If you are shown an optician's eyechart in the *periphery* of your vision, you will find that your ability to identify letters significantly decreases. Large letters which would be immediately identifiable were you to foveate them become less easy to recognise. Again, however, this is not just a case of being unable to *access* all the detail in our representations: our P-conscious phenomenology is itself less rich and less informative.

Representations that are relatively indeterminate and provide less data about the world to cognition I will term *generic*[44]. I now wish to claim that the determinacy of phenomenology does not vary simply according to whether an item appears large or small, or is in the middle or periphery of our vision, but according to how closely we *attend* to and thereby access it. Consider the following quotation from Tye:

44 Following Grush 2007: 504. Note that the terms 'generic' and 'specific' are relative to a level of description.

Focus on an object – a watch, say – lying on a magazine in the middle of your field of view... Now switch your attention slightly from the watch to the letters... there will be no marked or sudden change in the phenomenology... Still, the phenomenology does change a little. At a conscious level, there is a detail in the letters that simply was not there before – a detail that now enables you to identify them.[45]

It should be clear why an SR theorist is committed to this view. Changing the object of our attention changes what we cognitively access, and if the richness of phenomenology correlates precisely with the richness of cognitively accessed information, then items to which we attend less closely must literally look different, appearing less determinate.

I now wish to consider some objections to the idea of generic phenomenology. The most important objection is why, if so much of our phenomenology is generic, we do not notice it as such. I will return to this point in the next section, but one point to note here is that the genericity of a representation is not part of the first-order content of that representation. Genericity is not a quality like redness or squareness. I term generic a representation which is *merely* generic, but there are generic elements present even when I have highly specific phenomenology. If I see a dog out of the corner of my eye, I may have generic phenomenology which informs me that there is a large, black, dog shape moving towards me. Were I to look at the dog in more detail, I would still have this phenomenology, but I would *also* have more specific phenomenology representing its appearance in more detail. Generic phenomenology does not represent to me that any information is missing, but merely informs me about the general features of an object. If I introspect on my experience and develop higher order thoughts about it, I may realise that little information is present and that my representation is generic; indeed, I think that we are clearly aware of this in cases like that of the optician's chart.

45 Tye 2009: 172-3

A second objection to generic phenomenology is as follows. Imagine that a picture is tachistoscopically presented to you. The presentation is so brief that, whilst you are sure you saw something, you have no idea what it was. You would confidently bet that you saw something, but would expect to perform no better than chance when asked to guess the colour, shape, and other features of the thing you saw. In such a case, the SR view is committed to the idea that your phenomenology is extremely indeterminate, consisting just in the P-conscious representation 'something'.

It might be thought simply incoherent to imagine that, as in this example, we could see something without thereby seeing its colour or shape. For we see *by* seeing colour and *by* seeing shape. It is certainly true that light of a particular wavelength hits our retinas, and it hits our retinas with a certain dispersal pattern, and hence that some information relating to colour and shape reaches our eyes. However, we need not assume that this information reaches the level of P-consciousness. If the exposure of our retinas to these effects is sufficiently brief, there may be insufficient data to generate a P-conscious representation of colour and shape.

I suspect that this objection is wedded to what might be termed the 'snapshot' conception of visual experience. Just as a photograph captures all of the detail in a scene, and cannot represent anything without representing it as having definite colour and shape, the same seems naturally true of vision. Just as, when we watch a recording of a given scene, although we do not notice all of the background detail, all of the detail is present, and could have been accessed had we attended to it.

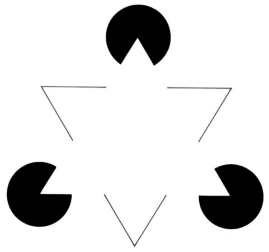
If we accept this 'snapshot' view of visual experience, then we will find the idea that unattended to phenomenology may be generic utterly perplexing. However, even the snapshot theorist has to admit that visual experience is not really like a video recording. Almost all our cone-cells (responsible for tracking colour) lie outside the fovea, hence our

brain cannot accurately encode the colour of items in the periphery of our vision. Hence we might judge that something on the periphery of our vision is some kind of reddish shade, for example, without enough information being available to our brain to determine whether it was scarlet, vermilion, or magenta. This kind of submaximal determinacy cannot be captured in the snapshot perspective, so we should abandon the position, together with our reservations about the conceivability of varying degrees of determinacy in P-conscious representations.

It is worth noting also that various neuropathies demonstrate a breakdown of the usual connection between features such as colour and motion. . Different areas of the brain have been linked to the generation of representations of contrast and motion (V1 and V5 respectively), and Hulme and Whitely in their reply to Block (2007a) note that patient G.K. seems to undergoing an experience of V5 without V1 when he exhibits sensitivity to motion but not to colour, an experience he describes as "black moving on black" (Zeki & ffytche 1998). Likewise, in certain images such as the Kanizsa triangle (Fig.5), even normals can experience object boundaries without colour differences.

If we accept the idea of generic phenomenology, we are left with a conception of vision according to which visual perception is more like a series of highly detailed symbolic representations, or a sketch, than a snapshot. An artist sketching a picture does not need to make all aspects of the representation equally detailed, and may omit some parts all together. A sketch of a woman sitting on a chair may show the woman's face in fine detail, her body only roughly, and the chair may be incomplete. Another analogy w say that our vision represents to us in ways comparable to *descriptions*: may refer to a car without mentioning the colour, size, shape, or other ar, so we might come to be aware of a car passing us in our periphera having visual phenomenology as of specific colours, sizes, and shapes. The description

Fig.5 – The Kanizsa Triangle



analogy is imperfect, however, insofar as it may misleadingly suggest that the sparse representations theorist is committed to a view according to which all perception is conceptual. I have deliberately avoided adopting a stance on this issue, and the SR theorist is not committed to one view or another. She might require that perception be representational whilst allowing that this may be in a non-conceptual form.

I regard it as an open question whether all features of experience admit of degrees of determinacy, and hence degrees of phenomenological genericity. SR theory must allow for genericity, however, in any cases where a subject's doxastic state is sensitive to a stimulus only insofar as that that stimulus falls under some more general category (or determinable) rather than more determinate features of that stimulus.

We might give the following SR account of when generic phenomenology occurs. A subject S is presented with a stimulus with feature F, where F is some determinate D of a determinable *d*. S's doxastic state is affected by the stimulus only insofar as that stimulus was *d*, such that the stimulus could have been *any other* determinate D_2 of determinable *d* without this having affected S's doxastic state. In this case, we must allow that S's phenomenology consisted just in the representation of *d* rather than of D. Otherwise, we allow for overflow and the separation of P-consciousness from A-consciousness. In a case where a letter 'B' is presented to me, for example, but I am wholly unable to say anything about it other than that it is *some alphanumeric character*; I have phenomenology only as of an alphanumeric character.

A final point I wish to make is that there is no reason why the overflow theorist cannot also make use of generic phenomenology. Indeed, given our insensitivity to colour and fine detail outside the region of our fovea, generic phenomenology seems almost indispensable in accounting for phenomenology in peripheral vision. It is particularly useful

for the SR theorist, however, in allowing her to explain away putative cases of overflow. Moreover, having admitted generic phenomenology into our philosophical toolkit, overflow theories are less strongly motivated insofar as we have such a ready explanation of putative overflow cases to hand.

4.4 – The Refrigerator Light Illusion

As noted above, a second problem for SR accounts concerns 'missing phenomenology': can we really believe that, when a subject does not notice a stimulus, that stimulus contributes nothing to their phenomenology? By way of illustration, imagine suddenly hearing an air conditioning system turn off, and realising that one has been hearing it all along. If we assume that the noise of the air conditioning system was not cognitively accessed and did not affect your doxastic state, should we assume that one had literally zero P-consciousness of it until it turned off?

One problem here concerns how I obtained the memory of the sound of the air conditioning unit if I was not P-conscious of it. However, there is no difficulty in supposing that subjects sometimes lay down memories of events which they are not P-conscious of at the time. Consider the case, mentioned in section 3.2, of subjects who were able to recall events which occurred under general anaesthetic. We need not assume, then, that just because you have a memory of the noise of the air conditioner, you heard it all along.

More difficult is rendering plausible the view that I had literally no P-consciousness of the air conditioner. To do this we can appeal to something called the 'refrigerator light illusion' (hence, RLI). This attempts to explain why we are naturally deceived into thinking that phenomenology is present when it is in fact absent. We can attend at a moment's notice to anything in any of our sensory modalities; normally if we *check* to see if we are having, for

example, auditory experiences, we find that we are. Insofar as, on the SR model, cognitive access enables phenomenology, and any instance of checking for the presence of phenomenology will involve access, we never notice the absence of phenomenology. It is like someone who comes to believe that the refrigerator light is always on because, whenever she opens the door to check, the light is on. Whenever we check if we are perceiving something, we thereby focus our attention on that source, and hence the situation never arises that we find ourselves not conscious of some item under investigation.

The illusion is completed by the fact that we regularly become conscious of intrinsically salient items without deliberately attending to them. For example, whilst engrossed in reading philosophy, my attention may be suddenly caught by a book falling off a shelf. Insofar as we become conscious of such intrinsically salient stimuli, we naturally conclude that we have been consciously monitoring them all along. Some form of monitoring certainly does occur: our attention is immediately drawn to moving objects, bright colours, strong smells, or loud noises. The SR theorist denies, however, that this monitoring is P-conscious, considering it an unconscious process.

The RLI enables SR theory to explain how we fail to notice the absence of phenomenology of unaccessed items. However, it is only loosely speaking an *illusion*. Normally illusions are cases where experience represents the world as being one way whereas in fact it is another. However, the RLI would be an illusion that applied not to our judgements about the world but to judgements about our own experience. If it were a true illusion, it would be a particularly extreme form, one in which the way things *seem* to seem to me is not the way they *actually* seem to me. Some commentators, notably Block, have questioned the very existence of such "hyper-illusions"[46].

For this reason, it is better to characterise the RLI as a faulty inference than a true

46 Block 2007a: 493

illusion. Compare the example of the observed motion of the sun around the Earth, and the natural ensuing conclusion that the sun orbits the Earth. This is not exactly an illusion, but more a case of the most *obvious* interpretation of observed phenomena not being the correct one. The RLI charges us with making a similar kind of mistake in perception.

Moreover, it is not clear that most people really are subject to the illusion. We do not naturally think of ourselves as being blind to unattended to items, as in the air conditioning example above; but, equally, when we fail to notice something it is natural to say that we did not perceive it. The question of whether unattended to items have phenomenology is not necessarily one on which common sense delivers a clear verdict, and the very concept of phenomenology is sufficiently rarefied that we need not convict the majority of humankind of being under the spell of an illusion.

I wish finally to note that something like the RLI helps to explain why we tend not to notice that those contents of consciousness to which we do not closely attend possess only generic phenomenology[47]. If I attend carefully to a watch lying on top of a magazine, the text of the magazine will not be clearly defined in my vision, but given that I my attention is largely on the watch, I may not notice this. If I carefully introspect on the quality of my experience, I may become aware that the items to which I am attending carefully are more richly visually represented than those to which I maintain a more limited attention, but this is not pre-reflectively obvious.

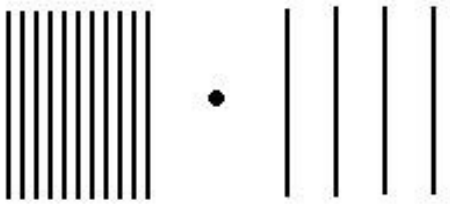
I hope I have given the reader a basic account of SR theory. In the next two sections, I will apply it to two philosophical thought experiments, both to aid explication of the theory and to forestall potential difficulties.

47 Grush (2007: 504) distinguishes this from the RLI and calls it the "waveform collapse illusion".

4.5 – SR theories at work: speckled hens

The first problem I wish to consider concerns experience of objects with highly complex features. When confronted with highly complex patterns and images, we are unable to attend to all of their many features simultaneously. One example of this is shown by Fig.6 below[48]. Here, when focusing on the central dot, most observers cannot attend *individually* to the lines on the left, being unable, for example, to count them. This is to be contrasted with our ability to attend to and count the lines on the right while thus focusing.

Fig.6 – a demonstration of the limitations of attention



A famous illustration of these difficulties arises in Chisholm's case of the speckled hen[49]. Seeing a hen with many speckles on its back, we know straight away that it has many speckles, but we do not need to attend to all of the speckles in order to know there are many of them. Plausibly, we see all of the speckles: we certainly see some of them, and there seems no non-arbitrary way of deciding which we do not see. However, if we see, for example, forty two speckles, but can only access a limited number of them, then it seems we have a prima facie case for the overflow of access by phenomenology.

The natural answer for the SR theorist is to argue that we have a merely generic representation of the number of speckles in the hen. We have a representation that there are many speckles, but our phenomenology does not represent exactly how many there are. If we counted the speckles, we would break them down into groups of twos or threes to which we

48 Taken from Tye (2009: 15)
49 Chisholm 1942.

would attend successively. In such cases, specific phenomenology is added to our generic representation, but our phenomenology does not remain specific when we go on to count and thereby attend to a different cluster of speckles. At no point does our phenomenology specifically represent the total number of speckles; at best, it represents there as being three speckles here, and many elsewhere.

This may strike some readers as counterintuitive. It might be objected that it is incoherent to suppose that our phenomenology represents the hen as having no definite number of speckles, equivalent to saying that a major earthquake will occur in 2009, but not in January, February, March, or any other one month. This is a valid criticism if we regard our phenomenology as itself having some particular feature for every speckle it represents, but SR theories would deny this. In representing that there are *n* speckles on the hen, our phenomenology does not need to consist of *n* items of phenomenology. If this were the case, then it would indeed be impossible for our phenomenology to represent there being indefinitely many speckles, since this would require that our phenomenology consisted of indefinitely many items. In fact, our phenomenology consists of a single representation of numerosity, and there is nothing incoherent about this. It is equivalent to someone asserting that a major Californian earthquake will occur in 2009 without specifying or intending to specify that it will come in one particular month in 2009.

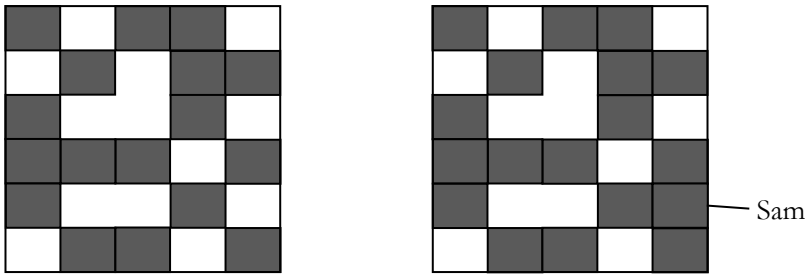
If the reader remains unconvinced that phenomenology might represent there being *many* speckles on the hen, one final example might serve to make the suggestion more plausible. Try to generate a mental image of a speckled hen in your mind's eye, as clearly as possible. It seems plausible that there is something akin to phenomenology in our image of the hen; after all, creating a mental image of a hen is a very different thing from merely thinking about a hen. Yet would we really wish to say that the hen has a definite number of speckles in our mental image of it? For, without intending to imagine a hen with a definite

number of speckles, what could determine this number? We cannot advert to any *actual* number of speckles, as we can when dealing with a real hen. If this line of thought is correct, then we have found a compelling example of how phenomenal representations of numerosity may also be generic.

4.6 – SR theories at work: Another brick in the wall

I wish now to use a problem from Dretske[50] as the basis for considering a further problem for generic phenomenology. He uses an example of two images of walls full of yellow bricks, in the second of which an additional brick, which he calls 'Sam', is present. Even without noticing Sam, he suggests, we still *see* Sam. We see him because we know of the second image that none of the bricks there were tilted or blue. Hence, we see Sam, since we know of each individual brick that it was not tilted or blue (see Fig.7 below).

Fig.7 – Brick walls like those used by Dretske



SR theory has a direct reply to this which I provided in section 4.2, namely that we know that Sam is not tilted or blue because we know we would have *noticed* if he had been tilted or blue. Hence our knowledge is based on inferential rather than non-inferential considerations, and hence non-perceptual.

The more subtle problem in this example that I wish to consider is that, given that

50 Dretske (2007)

there is a large number of bricks, and we can only attend to and thereby access a small number of these at a time, it might seem as if we are forced to assert that we do not see every brick, or at least, have only a kind of generic awareness of the wall as a whole. Yet this seems extremely counterintuitive.

One point to make is that even a generic awareness of the image as a whole may allow for me to 'see' every brick, in the sense that, for example, I have a representation of the shape of the wall, which is a consequence of actual arrangements of various bricks. Moreover, I am sensitive to every brick insofar as, if one brick were to change colour, I would notice it. The meaning of the English word 'see' is sufficiently ambiguous to allow, then, that I may in a sense see each brick even if I do not have specific phenomenology of it.

However, a further factor is at play, namely bundling. The amount of information of which we can be A-conscious at any one time varies depending on how effectively it can be bundled. For example, compare the ease of recall of character string (A) compared with character string (B):

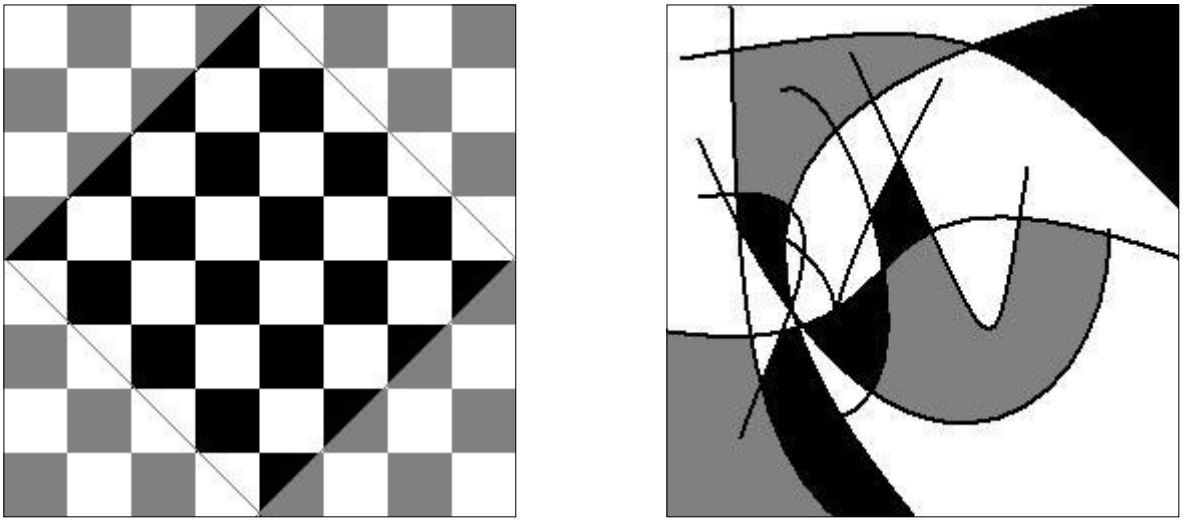
(A) TB NS AM SS MS AA **(B)** CIA FBI NSA IRS[51]

Assuming the reader notices that the character strings in (B) are all US government departments, they will find it easy to remember all of the characters in (B), whereas recalling (A), an entirely random string of characters, may take careful study while we attend to letters successively. Moreover, it feels to me as if I can attend to the character strings in (B) simultaneously whilst being aware of their individual identities, something I find difficult in the case of (A). Consider now the two images in Fig.8 below.

51 This example is similar to one given in Block 2007a: 495.

Fig.8 – an example of bundling in images

(C) **(D)**



Most readers, I imagine, would find it easy to recall or reproduce image (C), but impossible to recall or reproduce image (D). The answer lies in the fact that we can 'bundle' (C)'s components into a much more manageable set of regularly interrelated parts, thus allowing us to have a single accessed specific P-conscious representation. Bundling (D), by contrast is much harder, and it seems to me that I cannot have highly specific phenomenology of all of (D) simultaneously; the relations between its different elements are simply too irregular. Note also that whereas it is relatively easy to attend to the whole of image (C) at once, including all of its component parts, in looking at image (D), our attention either 'hovers' over the image as a whole without providing detailed awareness of its component parts, or 'zeroes in' on one or another element within it.

The point I wish to make is this. Attending to (C), we cannot simultaneously split our attention between every square, but because we bundle it together, we *do* have specific phenomenology of the whole image[52]. In the case of highly irregular images, such as (B), above, such bundling is impossible, so we either have a generic representation of the whole (as when I look at the whole image without zeroing in on its component parts), or specific

52 Cf. Dennett's account of the Marilyn Monroe wallpaper, 1991: 354-5.

representations of some but not all individual elements within it. The same may be true of Dretske's bricks, if the wall image is fairly regular: I can have specific phenomenology of the entire wall as a single item, and hence I *do* see every brick in the wall. On the other hand, if the image is too irregular, I will not be able to bundle it into a single specific representation, and I will have merely a generic P-conscious representation of the image, 'hovering over' it in my mind's eye or 'zeroing in' on individual elements. As noted, however, even in this case, there will still be a sense in which see every brick, being, for example, sensitive to changes in every part of the wall.

4.7 Conclusion

In this chapter, I have endeavoured to explicate SR theory as an alternative to overflow approaches. Explaining these theories required introducing the concepts of generic phenomenology and the RLI. Note that whereas some sparse theories of perception have grounded the appeal of their theories partly on experimental evidence such as change blindness[53], my commitment to SR comes rather from a combination of its general plausibility and the problematic consequences that would accompany adopting an overflow alternative, that is, the prospect of admitting IPC and its unpleasant philosophical ramifications. It remains to be shown, however, whether SR theory can avoid the most challenging evidence in favour of overflow theories, namely the Sperling, Landman, and Sligte experiments, and this will be the concern of the next chapter.

53 For example, O'Regan and Noë 2001.

CHAPTER 5

DOING WITHOUT OVERFLOW

5.1 - Introduction

Let us review the argument thus far. In chapters 2 and 3, I presented the case against IPC, while in chapter 4, I outlined a theory, sparse representations, designed to allow us to do without it. It should be clear that I regard IPC as a dubious and theoretically inelegant consequence of overflow theories, and that the presumption should be in favour of theories that do without it. An analogy may be drawn with radical positions such as panpsychism: although it is not an incoherent view that inanimate matter might generate P-consciousness, it is sufficiently implausible and theoretically awkward as to warrant a presumption in favour of theories that can make do without it. The only thing that might motivate us to adopt panpsychism would be the existence of other strong empirical or philosophical reasons which made it our only reasonable option.

Block adopts precisely this strategy to demonstrate the possibility of IPC: he suggests that there are experiments which lack any plausible explanation save for the overflow of access by the phenomenal, and from there argues that overflow entails the possibility of IPC. In refuting Block's arguments against IPC, then, the burden of proof I shoulder is simply to show that there are no arguments or empirical evidence that overflow theories are *uniquely* well positioned to deal with. To that end, I will now attempt to demonstrate that SR theories can plausibly account for the Sperling, Landman, and Sligte results, thereby removing the motivation for overflow theories and their consequent commitment to IPC.

5.2 –The Sperling Test

I wish first to show how SR theory interprets the Sperling test. Readers will recall from 1.7 that the Sperling test involves the tachistoscopic presentation to subjects of a matrix of alphanumeric characters, and that subjects display the ability to report more relatively more accurately on subsets of the data than on the whole (this is partial report superiority, or PRS). Because subjects display PRS consistently no matter which row was cued, the Sperling test shows that subjects retain memory not of four to five but of *twelve* characters. The Sperling test first demonstrated, then, the existence of some form of visual short-term memory (or VSTM) which decays rapidly but with a higher capacity than working memory.

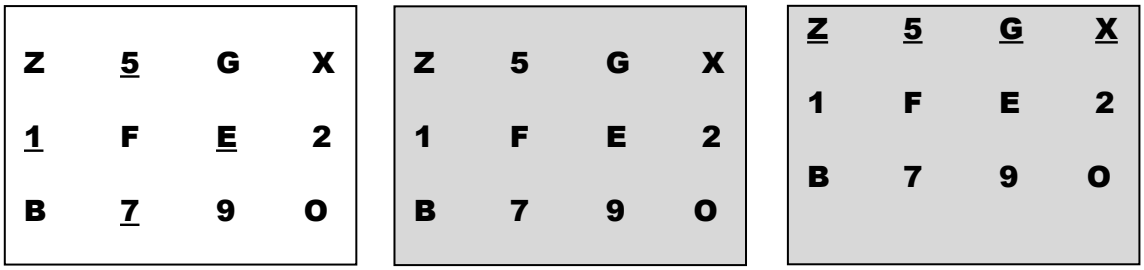
Undeniably, the total information retained in some form by subjects in the short term after the removal of the visual stimulus in the Sperling test is greater than the information that they can access. The total number of items that subjects can actually report is four or five, but they retain information in VSTM relating to around twelve items. The real question is whether PRS is a result of subjects' VSTM containing P-conscious representations. If so, then we have an empirical demonstration of overflow, and SR theories are in trouble. Moreover, as a subject in the test, it certainly feels like I have seen every one of the characters concerned. Taking subjects' introspective accounts at face value, we might have good reason to believe that phenomenology overflows access.

Overflow theories can provide a straightforward account of the Sperling test: I am P-conscious of all characters in the matrix when the stimulus is present (and perhaps also A-conscious of some), and retain this phenomenology prior to cuing. This phenomenology is degraded only *after* cuing in the time taken to make my report (see Fig.9 below).

However, there is also a plausible SR interpretation of the Sperling test[54], which makes use of the idea of generic P-consciousness outlined in the previous chapter. While the stimulus is being presented, subjects have a generic phenomenological representation of the matrix, that is, they have a visual representation of a four by three grid of alphanumeric characters without having specific P-conscious representations of the identities of these characters. They may also have specific phenomenology of some subset of the matrix, as their attention roves from one character to another. They may be A-conscious, then, of some of the characters while the stimulus is being presented, but this A-consciousness will be so fleeting that it is not likely to be recalled, particularly once they are given the subsequent, attention-grabbing task of reporting cued items. Following removal of the stimulus, they retain generic phenomenology representing a matrix of sixteen alphanumeric characters, but if they are not attending to specific characters in their mental image (but merely trying to keep an image of the matrix as a whole) they will not have specific P-consciousness of the identities of individual letters. Finally, when cued, they trigger a transfer of some of the contents of their VSTM into working memory, thereby generating specific phenomenology that represents to them the identity of the cued characters. A representation of this SR interpretation of the Sperling test is given below (see Fig.10).

54 The SR story as given here is influenced heavily by the replies made to Block by Papineau (2007) and Grush (2007), as well as by discussions held with Tim Bayne and Nick Shea.

Fig. 9 – an overflow interpretation of the Sperling test (**NB** underline indicates access)

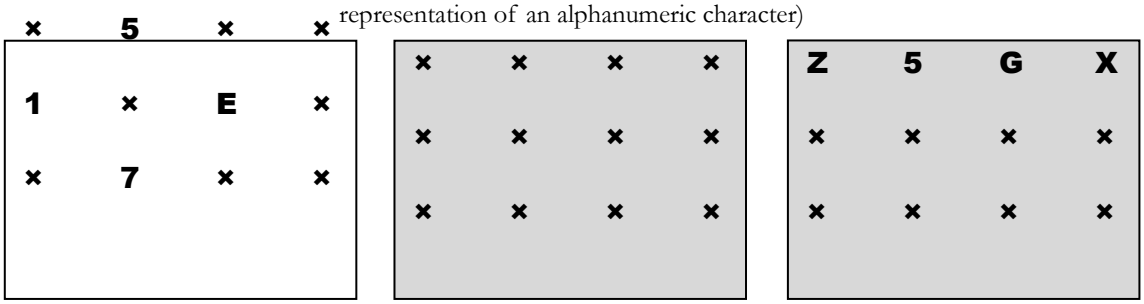


1: Stimulus presentation I have specific phenomenology of all items; I may also be A-conscious of some.

2: Stimulus removal I still have specific phenomenology of all items.

3: Cueing/report In response to cueing, I access some section of my phenomenology for report.

Fig. 10 – an SR interpretation of the Sperling test (**NB '×'** indicates a generic phenomenological representation of an alphanumeric character)



1: Stimulus presentation I have specific phenomenology of some items (whichever I happen to attend to) and generic phenomenology of others; all phenomenology is A-conscious

2: Stimulus removal I have A-conscious generic phenomenology of all items.

3: Cueing/report In response to cueing, I render some section of my generic phenomenology specific.

5.3 – The Landman et al. and Sligte et al. paradigms

I turn now to the Landman et al. and Sligte et al. paradigms, briefly described in Chapter 2, which form a key part of Block's case for overflow theories. Both are partial sampling paradigms in which subjects are presented with arrays of rectangles at various orientations.

Subjects in the Landman paradigm demonstrate a PRS lasting for up to 1.5 seconds, 500ms longer than in the Sperling case. The selection criteria were the *orientation* of the

rectangle in the one trial, and both orientation and size in another. No significant difference in PRS was found when subjects were required to report on changes in both selection criteria.

The Sligte experiment used the Landman paradigm with slight variations, displaying up to thirty-two rectangles instead of the Landman paradigm's four, and found that, with practice and directions to 'relax and let it happen', subjects demonstrated PRS lasting up to four seconds, far greater than the 1000ms recorded in the Sperling test and 1.5s recorded in the Landman paradigm. The Sligte paradigm also made use of tests involving bright stimuli and dark-adapted patients to test whether PRS might be based on mere retinal persistence rather than persistence in VSTM. It was found that dark-adapted patients showed an enhanced PRS when asked to report on black and white stimuli rather than isoluminant red and green stimuli, but that this differential lasted only for the first 1000ms. After the first 1000ms, subjects retained a partial report advantage for both sorts of stimuli, but there was no differential in the degree of superiority between the two sets of stimuli. Retinal persistence would be stronger for the black and white stimuli than for the red and green stimuli, and the fact that the differential in PRS between the two kinds of stimuli disappeared after 1000ms, but was still significantly more accurate than whole report, suggests that even if the early stages of PRS are determined by retinal persistence, this has ceased to be a factor after 1000ms. This conclusion was bolstered by the fact that the heightened PRS in the first 1000ms was significantly diminished by a flash, which one would expect to overwrite any retinal persistence, but not by a pattern mask, whereas superiority after this was abolished by a pattern mask but not a flash, suggesting the PRS effect in the second stage was not retinal persistence but based on the retention of some kind of mental image liable to be disrupted by a pattern mask.

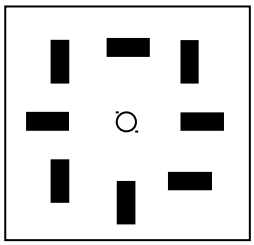
The overflow theorist's interpretation of the Landman and Sligte paradigms should be clear: subjects, presented with a complex stimulus, have rich specific P-consciousness of

every item, but their ability to *access* items is limited by working memory. Their rich P-consciousness persists for up to 1.5 or 4s and provides the basis for subjects' reports as to whether an individual rectangle has changed orientation or size, though in the course of making report, the rich P-consciousness, based in VSTM, degrades. A representation of this interpretation is given below (Fig.11).

Likewise, SR theorist's interpretation should now be clear. Subjects, presented with the stimulus, have generic P-consciousness of the whole array, and specific P-consciousness of any rectangles they happen to attend to. Once the stimulus is removed, subjects have generic A- and P-conscious representation of the array as a whole, including all eight rectangles. Upon cueing, subjects trigger a transfer of some of the contents of their non-conscious VSTM into working memory, thereby rendering its phenomenology specific, and during report, the rest of their VSTM degrades (see Fig.12).

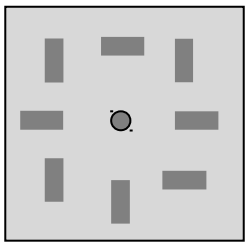
Note that the SR theory could allow that subjects had specific P-consciousness of the whole array in the Landman and Sligte tests were it sufficiently regular (cf. Dretske's brick in the wall case in 4.7). In typical presentations, the rectangles, having random orientations, are too irregular to admit of easy bundling. However, compare a case in which, as it happens, the rectangles in the test were either *all* vertical or *all* horizontal. In such circumstances, the array would easily be bundled into a single *specific* P- and A-conscious representation, and we would expect subjects to be able to report accurately on whether *each and every* one of the rectangles had switched orientation.

Fig. 11 – an overflow interpretation of the Landman and Sligte experiments



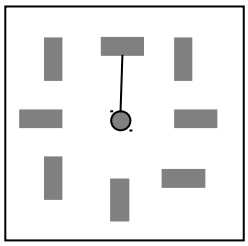
1: Stimulus presentation

I have specific phenomenology of all rectangles, but access only some subset of them.



2: Stimulus removal and gray interval

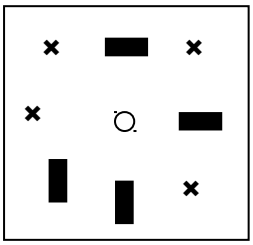
I still have specific phenomenology of all items.



3: Cueing/report

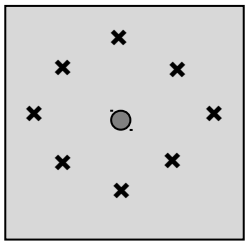
In response to cueing, I access some section of my phenomenology for report.

Fig. 12 – an SR interpretation of the Landman and Sligte experiments



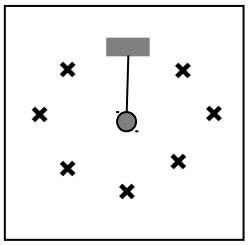
1: Stimulus presentation

I have specific phenomenology of some items (whichever I happen to attend to) and generic phenomenology of all items; all phenomenology is A-conscious



2: Stimulus removal and gray interval

I have A-conscious generic phenomenology of all items.



3: Cueing/report

In response to cueing, I render some section of my generic phenomenology specific.

5.4 – Does the SR view impute illusion to subjects?

In the next two sections, I will consider objections to SR accounts. The most serious is that an SR interpretation requires that we do not take subjects' reports quite at face value. After all, subjects do not report seeing anything they would describe as generic, nor do they report seeing any shifts from generic to specific phenomenology upon being cued. Block's view seems to be that subjects genuinely believe their experience is determinate. Perhaps this is true, insofar as subjects do not report anything akin to what goes on when we look at a

dentist's eye chart. The letters and rectangles in the above paradigms all seem clear and distinct: subjects just cannot report what they all were.

The first point I wish to make is that, as stated in 4.3, we do not normally notice generic phenomenology, since its genericity is not a first order property of the our phenomenology, but instead the *absence* of specific first order properties. It is only when we reflectively introspect on generic phenomenology that we can infer that it is generic. Secondly, as also stated earlier, we do not normally notice transitions from generic to specific phenomenology because these transitions are accompanied by a change in the focus of attention. Through a kind of low-level attention, we may become aware that things which were clearly defined while we were attending to them have become less well defined now that we are looking at something else. The third point concerns the RLI: subjects might easily fail to spot that the rectangles they were less closely attending to were less richly represented, precisely because the main focus of their attention was elsewhere.

A further crucial factor affecting subjects' reports is that in reporting that they saw all of the presented items, we cannot assume that they thereby indicate that they believe their phenomenology to have been specific. Phenomenology is a relatively rarefied philosophical notion, and asking subjects to report on what they saw is not exactly equivalent to asking them about the contents of their phenomenology. In particular, I suspect that subjects did not naturally distinguish between what they *believed* themselves to be seeing and the phenomenology itself.

Consider the following two situations. If someone is asked about a letter low down on the optician's chart, they may say that it looks "like a P or a B", thereby implying that their phenomenology is to some degree indeterminate. What they do not consider, yet could not rule out based on their phenomenology, is that the character they are seeing is not a P or a B, but some hybrid of the two, like the character shown in Fig.13. They do not entertain this

possibility, however, because their *expectations* influence their reports. Likewise, if someone is asked to hold their hand first in front of their face, and then in the periphery of their vision, and is asked whether it looks the same, a natural, pre-reflective judgement would be that it *does* look the same, even though, plausibly, the phenomenology in the second case is less determinate. Subjects naturally report that it looked the same, however, because they know the represented object has not changed: it is their hand, and they know because they are holding it there.

Fig.13 – a distorted character

I suggest, then, that subjects' reports of their own phenomenology frequently fail to distinguish between phenomenology proper and what they believe to be the object represented by their experience, where this latter belief is influenced by inferential (and hence non-phenomenological) considerations. Given that, in the above experiments, subjects attend to *some* items while the stimulus is present, they cognitively access the fact that (at least some of) the items before them are alphanumeric characters (or rectangles). It would not occur to subjects to report that, although they believe that all twelve of the items in the grid were specific alphanumeric characters or rectangles with specific orientations, their phenomenology was insufficient to ground this judgement; or to admit that *some* of the characters may not in fact have been alphanumeric characters or rectangles, but distorted forms. We routinely fail to peel apart inferentially and non-inferentially acquired data when reporting on the contents of perception, and hence we cannot use subjects' reports that they definitely saw items clearly as a source of information about their phenomenology.

Further support for this view can be drawn from the experiments by McConkie et al.[55]. In these experiments, subjects were presented with a screen of text, and told to attend to a particular area of it. Using an eye-tracking camera, the area foveated by subjects' was

55 See, e.g., McConkie et al. (1975, 1979)

measured, and areas of text around it were replaced with distorted text. However, because the distorted areas of text were never those to which subjects attended, they believed that all the text that they saw was normal. Speaking to a subject of such a test, he reported it as quite an astonishing experience: one is verbally assured that the text one is not attending to is distorted, but of course it is extremely difficult to *catch* the fact that the text is distorted, insofar as the eye-tracking camera changes which areas of the text are distorted as quickly as the eye can move.

Without being told otherwise, subjects report the whole page as seeming like normal text. We must ask where this belief comes from. We cannot suppose that their brains *assume* the whole page of text is normal, and so generate *specific* phenomenology as of normal text across their entire representation of the page. The SR theorist has a ready explanation. First is that subjects' representation of the text outside the point of foveation is merely generic and fails to represent the specific identities of individual (distorted) letters. Subjects still feel confident in reporting that their visual experience was of clear, undistorted text, however, because of the influence of *inferential* considerations: they have no reason to expect the text to look anything but normal, and indeed, they have no reason to imagine that the normal text they have already read might become distorted once their eye has moved on. For this reason, they quite naturally report that their experience was as of entirely normal text, and *not* that their experience was of *indeterminate* text. Only after being informed that the text is distorted may subjects admit that their experience of the text was sufficiently unclear as to prevent them from being *sure* that the text was undistorted.

Two ways of exploring this hypothesis further would be as follows. The first would be to combine the McConkie et al. and Sperling paradigms. If my interpretation of Block's experimental paradigms is correct, then subjects, despite reporting that they had clear and determinate representations of every character, would fail to notice that some of these

characters were distorted. This experiment is suggested by Kouider et al. in their reply to Block, and indeed, Block in his reply seems to admit that this it is plausible[56]. A second way of exploring the hypothesis would be to conduct the Sperling test with subjects who had some philosophical background. Speaking as a subject in the Sperling test myself, I highly doubt that I do have determinate P-conscious representations of the whole matrix. However, as Spener notes, philosophers are very good at drawing quite contradictory conclusions about the nature of visual experience based purely on 'infallible' introspective evidence, hence such evidence can scarcely be considered cogent.[57]

5.5 – Additional problems raised by the Landman and Sligte paradigms

I now move to two features of the Landman and Sligte paradigms that Block believes favour overflow interpretations. Block holds that the fact that, in the Landman and Sligte trials, subjects were able to attend to the stimulus for an extended period (500-1000ms rather than 50ms in the Sperling test) counts in favour of overflow interpretations. He states that "subjects are attending to arrays in full view in good viewing conditions to stimuli that last between half a second and a second, more than enough time for specific phenomenal content"[58]. Block's claim is that subjects attending to the stimulus while it is present have specific phenomenal content. However, the SR theorist can accept this: subjects would likely have time to attend to one or more rectangles prior to the removal of the stimulus, hence enjoying brief specific P-consciousness which they may well lose once the stimulus is removed (but need not, if they continue to focus on that part of their mental image).

56 See Kouider et al., 2007 p511, and Block, 2007b p532: "Kouider et al. predict, plausibly, that in Sperling experiments that include some letter-like symbols which are not letters, subjects would treat false letters as similar to real letters."
57 See Spener (forthcoming)
58 Block 2008: 307.

Block's argument may be that, having had some specific P-conscious content during the display phase, subjects would *notice* if the content of their experience suddenly became generic. As he notes, "[i]n the Sligte version of the Landman experiment, the visual experiences last for four to five seconds, and it is plausible that subjects would be likely to be accurate about something they see for such a long period"[59]. Block seems to be making the assumptions, firstly, that subjects believe their phenomenology was determinate, and that this belief is credible given that subjects held a mental image for a long period.

As stated, I dispute Block's claim that we can interpret subjects' judgements as suggesting that they had determinate phenomenology; first, because it would be quite natural for subjects to fail to make the higher order judgement that there phenomenology was generic, and second, that their claims about their phenomenology may be tainted by subjects beliefs about what their mental image represented, namely the array of specific characters they had just seen. Subjects' reports are not just about their phenomenology, and so the fact that they retained their phenomenology for a prolonged period does not make their claims more useful in determining whether they had specific phenomenology of all items. The SR theorist need not claim that subjects *wrongly* report the nature of their mental image, then, but rather that their response cannot be viewed as being just about their phenomenology. Rather, subjects' claims will relate to what they think their mental image represented, namely the specific array that they had just seen, whether or not features of the actual array were precisely mirrored in their phenomenology.

Moreover, Block's idea that we should take subjects' claims at face value as suggesting that they had more phenomenology than they accessed runs a very real risk of incoherence. In order for them to know that their phenomenology was specific, they must have known that every rectangle was simultaneously represented as having some specific orientation. But

59 Block 2008: 307.

how could they be sure whether this was the case? The only way we have detailed knowledge of the contents of our phenomenology is via attending to it, but Block cannot claim that subjects *access* each individual rectangle simultaneously; not only would this tell against the scientific evidence that we can attend to only four to five items at any one time, but it would defeat his whole intention to show that phenomenology overflows access.

Rather, Block must claim that subjects know that their mental image consists of eight rectangles with specific orientations without knowing what each such orientation is. Yet, without checking individual rectangles, how could subjects be sure that every rectangle simultaneously had a specific orientation? They could at best be certain a particular rectangle or rectangles to which they were carefully attending had specific orientations. In particular, without access to detailed information about rectangles to which they were not attending, they could not rule out such rectangles were merely *generically* represented as 'rectangles at various orientations'.

The challenge to Block, then, is why we should take subjects' claims about the determinacy of their phenomenology as authoritative, when the evidence they *do* have is compatible with their having only generic phenomenology. If he admits, as it seems he must, that subjects could *not* simultaneously check whether every rectangle was represented as having a specific orientation, then I can see no grounds taking subjects' testimony to this effect as definitive.

The second of Block's arguments concerns the fact that subjects' PRS holds almost as well when they have to report on whether a rectangle has changed size *or* orientation[60]. Indeed, in studies by Luck et al.[61] And Woodman and et al.[62], it was found that subjects were almost as accurate at reporting on *four* changes, in orientation, size, colour, and whether or

60 See Block 2008: 308-9
61 Luck et al. 1997.
62 Woodman et al. 2003.

not there was a gap in a figure, as at reporting on just one change. Block holds that this supports the view that PRS arises from subjects' retaining a mental image of the rectangles. As he puts it, "[t]hat suggests that the subjects have a representation of the rectangles that combines size and orientation from which either one can be recovered with no loss due to the dual task, again backing up subjects' reports that indicate a kind of visual imagery"[63].

That subjects underwent visual imagery is perfectly compatible with an SR account. However, Block may be arguing that one natural explanation for the fact that subjects display the same degree of PRS whether reporting on one feature or several is that in the representations of the rectangles in VSTM, these features have all been bound together. One feature of uncontroversially unconscious representations in the visual system is that they process individual features of perceived items which are not yet bound together. Hence V1 processes for contrast, while V2 encodes for colour and orientation. At the conscious level, these features have been bound together. Hence, as many have argued[64], there does seem to be some kind of correlation between feature-binding and consciousness.

Although this is certainly a piece of evidence in favour of Block's theory, it cannot count as decisive. There is no reason that the SR theorist cannot, on the basis of other considerations, argue that VSTM allows for the binding of complex features whilst remaining unconscious. One might point, moreover, to the fact that in VSTM, feature binding is not complete, as is shown by the fact that no PRS is displayed in the Sperling test if the cued criterion is to report either on just letters or just numbers[65], suggesting that at the level of VSTM, alphanumeric characters have not been bound with the concepts *number* and *letter*.

5.6 – Another piece of evidence for overflow?

63 Block 2008: 309.
64 See, e.g., Crick and Koch 1990.
65 Sperling 1960.

Block has one more piece of evidence he adduces in favour of overflow. He argues that two experiments by Di Lollo et al.[66] and Loftus and Irwin[67] cannot be explained by SR theories using the same methods used to explain the Landman and Sligte experiments. "The upshot," Block contends, "is that there is a completely different paradigm in which the evidence favours high capacity specific phenomenal consciousness"[68].

The Di Lollo and Loftus and Irwin paradigms used a five by five grid in which all but one square was filled with dots. This grid was broken into two groups of twelve dots which were presented to subjects sequentially with variable delays. Subjects were challenged to work out the location of the missing dot by combining the two images in their mind's eye. In a second experiment, subjects were asked to state whether or not it had seemed to them as if they were seeing a *complete* grid of dots, or whether the pause between the two presentations had led to them seeing two completely distinct images. Crucially, Block argues, "subjects' ability to do the task correlates nearly perfectly with their phenomenological judgements of whether there appears to be a whole matrix rather than two partial matrices"[69]. However, the SR theorist can plausibly advert to retinal persistence to explain this correlation: if subjects literally had a combined afterimage of the two grids, their ability to perform the task would naturally be high.

A similar experiment by Brockmole, Wang and Irwin[70] showed the interesting result that subjects' performance in the task is good in the short term but tails off, bottoming out at around 100ms, and then actually increases, peaking at around 1.5s but remaining very high for up to five seconds. Pointing to work by on the generation of mental imagery by Kosslyn[71], Block argues that 1.5s is the delay one would expect if subjects were generating mental

66 Di Lollo 1980.
67 Loftus and Irwin 1998.
68 Block 2008: 310.
69 Ibid. 309
70 Brockmole et al. 2002.
71 Kosslyn 2006.

images of the first dot matrix which they integrate "in their mind's eye" with the second dot matrix.

The suggestion on Block's part seems to be that in these partial grid paradigms, subjects' ability in the short term (up to 100ms) is based on retinal persistence or something akin to it: they quite literally see the two grids as integrated. Their sustained PRS, however, is a result of their generating a complete image in their mind's eye on the basis of the two images already presented (the idea that this second stage involves a mental image is suggested by the Brockmole, Wang and Irwin paradigm), which allows them to judge the location of the missing dot. It is not immediately obvious, however, why subject's use of mind's eye imagery to complete the task should debar SR explanations.

The suggestion may be that subjects' mental images contain specific phenomenology for each dot. However, there seems no reason to assume that subjects' mental imagery *must* be thus specific. Plausibly, mental imagery can have generic elements: one can imagine a spotted leopard without there being a matter of fact as to how many spots it has on its body[72]. Subjects may, for example, generate mental images consisting of generic phenomenal representations of the twelve dots insofar as they form a single *shape*, and via awareness of this shape be able to locate the gap in the image. This would be compatible with their reports that, when they were best able to work out the location of the gap, they seemed to see a single combined dot-grid in their mind's eye. Moreover, the dot image might be sufficiently regular as to amenable to *bundling* into a single accessed specific P- conscious representation. The dot paradigms, then, seem susceptible to several interpretations compatible with SR accounts.

72 Cf. Schwitzgebel's (forthcoming: Ch.3) excellent discussion of mental imagery. I have modified his example of a striped tiger to a spotted leopard, which is more closely analogous to the DiLollo dot paradigm!

5.7 – Conclusion

In this chapter, I have endeavoured to show that the Sperling, Landman, and Sligte paradigms do compel us to allow for the overflow of access by phenomenology, and that SR theories have a way of explaining subjects' reports that is not implausible. I argued that subjects could conclusively verify that they had specific phenomenology only by attending closely to it, and both SR and overflow accounts make the same prediction regarding carefully attended to items, namely that they are specific. The point of dispute concerns items which subjects do not attend to closely, but here again, SR theories have a plausible explanation for why subjects report seeing them as specific. Generic phenomenology never represents anything *as generic*, and we might naturally expect subjects not to notice the genericity of representations which they do not attend to closely, given that the focus of their attention is elsewhere, and that genericity is a higher order property which is not pre-reflectively obvious. Moreover, as noted, their reports about what they saw are liable to influence from their awareness that their representations refer to a fully specific image which they have just seen.

Though overflow theories may comport slightly more straightforwardly with subjects' accounts, SR accounts can also plausibly account for them. Moreover, the SR approach makes use of philosophical tools, such as generic visual phenomenology and mind's eye imagery, which the overflow theorist is plausibly required to refer to in certain circumstances. Conversely, the SR theorist must introduce the notion of non-accessed P-consciousness, which, as an *addition* to philosophical theory, requires cogent evidence in its defence. More decisively, however, I agree with Block's assumption that, once we admit of non-accessed P-consciousness, we are required to admit of *inaccessible* P-consciousness, which, as I argued in Chapters 2 and 3, I consider a strikingly implausible notion. The SR theorist should be open to *future* experiments that would settle the question one way or another, but in the absence of compelling evidence, I submit that their position is the more plausible.

CONCLUSION

This thesis has attempted to describe and rebut Block's seminal 2007 article, "Consciousness, accessibility, and the mesh between psychology and neuroscience". This article brought to the forefront of contemporary philosophy of mind the relationship between P- and A-consciousness, and forcefully made two radical suggestions: that empirical psychology provides evidence that P-consciousness overflows A-consciousness; and second, that this would allow for the possibility of individuals' having wholly inaccessible P-conscious states. I have aimed in this thesis to rebut these two arguments, and provide the foundations for a theory of perception according to which P- consciousness is inextricably bound to access.

For the purposes of this thesis, I have treated Block's argument as having the following form.

(3) Postulating unaccessed P-conscious states is the best way to account for the empirical data.
(4) If unaccessed P-consciousness is possible, then *inaccessible* P-consciousness is possible.

Hence,

(C) Inaccessible P-consciousness is possible.

I have attempted to rebut Block's argument by tackling it in reverse, to whit:

(3) Inaccessible P-consciousness is at best implausible, at worst impossible. (Chapters 2 and 3)
(4) Unaccessed P-conscious would entail the possibility of inaccessible P-consciousness.

Hence,

(C) We have a strong motivation for finding plausible interpretations of the experimental data that do not rely on overflow.

It was the burden of chapters 4 and 5 to present such an interpretation and to deploy it to the relevant philosophical paradigms.

The one step in the argument that I not been able to dwell upon as much as I would have wished step (2). Insofar as I have been concerned primarily to rebut Block's argument, I have been happy to grant this second step, though I have also delivered some arguments that may be felt to support it (in section 2.4). Even if the reader rejects this second step, however, I hope that they will find my arguments against the plausibility of IPC convincing, and moreover, that SR theory provides us with a viable alternative strategy for explaining putative evidence of overflow.

A major unspoken influence running throughout this thesis has, of course, been that of Immanuel Kant. Like Kant, I have been concerned to show that experience wholly unconditioned by cognitive faculties, to the point of being not even *potentially* self-ascribable by its subject, is impossible; in other words, that bringing experience under some kind of cognitive framework is a necessary condition of experience in general. Again, like Kant, my most direct arguments in support of this thesis have concerned the problem of the subjective *binding* of experiences in time and space.

My aim in this thesis has not been a reductionist one. Even if P-conscious states are necessarily A-conscious, as I have argued, I fully accept a *conceptual* division between the two. The P-conscious aspect of experience is certainly the more challenging to explain in any kind of physicalist framework, and it is certainly the harder problem of consciousness.

I consider our familiar experience, in the form of states which are both A- and P-conscious, to be problematic enough as it stands. The postulation of a further kind of experience, utterly and necessarily removed from anything which recognise in our own case, is a radical move, and one which, as I have argued, is not motivated by a substantial base of philosophical or empirical evidence. Pending such evidence, we should turn our attention to formulating a theory of perception that seeks to explain experience as we know it, and I hope that in this thesis I have taken some early steps to doing precisely that.

BIBLIOGRAPHY

Baars, B. (1988), *A Cognitive Theory of Consciousness*, Cambridge, MA: Cambridge University Press.

Bayne, T., and Chalmers, D. (2003), 'What is the unity of consciousness?', in A. Cleeremans (ed.), *The Unity of Consciousness: Binding, Integration, Dissociation*, Oxford: Oxford University Press.

Block, N. (1978). 'Troubles With Functionalism', in. C.W. Savage (ed.), *Perception and Cognition: Issues in the Foundations of Psychology, Minnesota Studies in the Philosophy of Science*, vol. 9. Minneapolis: University of Minnesota Press.

Block, N. (1995), 'On a confusion about a function of consciousness', *Behavioral and Brain Sciences* 18: 227-47.

Block, N. (2005), 'Two neural correlates of consciousness', *Trends in Cognitive Sciences* 9 (2): 46-52.

Block, N. (2007a), 'Consciousness, accessibility, and the mesh between psychology and neuroscience', *Behavioral and Brain Sciences* 30: 481-99.

Block, N. (2007b), 'Author's response', *Behavioral and Brain Sciences* 30: 530-48.

Block, N. (2008), 'Consciousness and Cognitive Access', *Proceedings of the Aristotelian Society,* 105(3): 289-317.

Brockmole, J. R., Wang, R. F. & Irwin, D. E. (2002), 'Temporal integration between visual images and visual percepts', *Journal of Experimental Psychology: Human Perception and Performance* 28(2):315–34.

Burge, T. (2007), 'Psychology supports independence of phenomenal consciousness', *Behavioral and Brain Sciences* 30: 500-1.

Chisholm, R. (1942), 'The Problem of the Speckled Hen' *Mind* 51(204): 368–373.

Clark, A., and Kiverstein, J. (2007), 'Experience and agency: Slipping the mesh', *Behavioral and Brain Sciences* 30: 502-3.

Chalmers, D. (1996), *The conscious mind: In search of a fundamental theory*, Oxford: Oxford University Press.

Crick F., and Koch C. (1990), 'Towards a Neurobiological Theory of Consciousness', *Seminars in the Neurosciences* 2: 263-275.

Curtis, C., and D'Esposito, M. (2003), 'Persistent activity in the prefrontal cortex during working memory', *Trends in Cognitive Sciences* 7(9): 415-23.

Dennett, D.C. (1978), 'Why you can't make a computer that feels pain', *Synthese* 38: 415-56

Dennett, D.C. (1991), *Consciousness Explained*, Boston, MA: Little, Brown.

Di Lollo, V., (1980), 'Temporal integration in visual memory', *Journal of Experimental Psychology: General*, 109: 75-97.

Dretske, F. (2004), 'Change blindness', *Philosophical Studies* 120: 1-18.

Dretske, F. (2007), 'What change blindness teaches about consciousness', *Philosophical Perspectives* 21: 216-30.

Fodor, J. (1983), *Modularity of Mind: An Essay on Faculty Psychology*, Cambridge, MA: MIT Press.

Grush, R. (2007), 'A plug for generic phenomenology', *Behavioral and Brain Sciences* 30: 504-5.

Hulme, O., and Whitely L. (2007), 'The "mesh" as evidence – model comparison and alternative interpretations of feedback', *Behavioral and Brain Sciences* 30: 505-6.

Jackson, F. (1982), 'Epiphenomenal Qualia', Philosophical Quarterly 32: 127-136.

Kant, I. (1929), *Critique of Pure Reason*, trans. N. Kemp Smith, New York: St. Martin's Press

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

Kouider, S., de Gardelle, V., and Dupoux, E. (2007), 'Partial awareness and the illusion of phenomenal consciousness', *Behavioral and Brain Sciences* 30: 510-11.

Landman, R., Spekreijse, H., and Lamme, V. (2003), 'Large capacity storage of integrated objects before change blindness', *Vision Research* 43: 149-64.

Levinson, B.W. (1965), 'States of awareness during general anaesthesia', British Journal of Anaesthesia 37: 544-6.

Loftus, G., and Irwin, D. (1998), 'On the relations among different measures of visible and informational persistence', Cognitive Psychology 35: 135-99.

Luck, S. J. & Vogel, E. K. (1997), 'The capacity of visual working memory for features and conjunctions', *Nature* 390: 279–81.

Lycan, W.G. (1996), *Consciousness and Experience*, Cambridge, MA: Bradford Books/MIT Press.

McConkie, G. W., and Rayner, K. (1975), 'The span of the effective stimulus during a fixation in reading', *Perception and Psychophysics* 17: 578-86.

McConkie, G.W. and Zola, D. (1979), 'Is visual information integrated across successive fixations in reading?', *Perception and Psychophysics* 25: 221-24.

Mole, C. (2008), 'Attention and consciousness', Journal of Consciousness Studies 15 (4):86-104.

Nagel, T. (1974), 'What is it like to be a bat?', *Philosophical Review* 82: 435-56.

O'Regan, J. K., and Noë, A.(2001), 'A sensorimotor approach to vision and visual consciousness', *Behavioral and Brain Sciences* 24: 883–975.

Papineau, D. (2007), 'Reuniting (scene) phenomenology with (scene) access', *Behavioral and Brain Sciences* 30: 521.

Prinz, J. (2007), 'Accessed, accessible, and inaccessible: where to draw the phenomenal line', *Behavioral and Brain Sciences* 30: 521-2.

Schwitzgebel, E. (forthcoming), *Perplexities of Consciousness*.

Sergent, C., and Dehaene, S. (2004), 'Is consciousness a gradual Phenomenon? Evidence for an all-or-none bifurcation during the attentional blink', *Psychological Science* 15:720-28

Sligte, I., Lamme V., and Scholte H. (2006), 'Capacity Limits to Awareness', paper presented at the Association for the Scientific Study of Consciousness, Oxford, UK.

Spener, M. (forthcoming), 'Phenomenal adequacy and introspective evidence'

Sperling, G. (1960), 'The information available in brief visual presentaitons', *Psychological Monographs* 74(11): 1-29.

Tye, M. (2004), *Consciousness and Persons: Unity and Identity*, Cambridge, MA: MIT Press

Tye, M. (2009), *Consciousness Revisited: Materialism without Phenomenal Concepts,* Cambridge, MA: Bradford Books/MIT Press.

Woodman, G. F. & Luck, S. J. (2003), 'Dissociations among attention, perception, and awareness during object-substitution masking', *Psychological Science* 14: 605–11.

Zeki, S., and ffytche, D. H. (1998), 'The Ridoch syndrome: Insights into the neurobiology of conscious vision', *Brain* 121(Pt. 1): 25-45.