To be published in *Behavioral and Brain Sciences* (in press)

Consciousness, accessibility, and the mesh between psychology and neuroscience

Ned Block
New York University
Department of Philosophy
ned.block@nyu.edu

Abstract: How can we disentangle the neural basis of phenomenal consciousness from the neural machinery of the cognitive access that underlies reports of phenomenal consciousness? We see the problem in stark form if we ask how we can tell whether representations inside a Fodorian module are phenomenally conscious. The methodology would seem straightforward: Find the neural natural kinds that are the basis of phenomenal consciousness in clear cases – when subjects are completely confident and we have no reason to doubt their authority – and look to see whether those neural natural kinds exist within Fodorian modules. But a puzzle arises: Do we include the machinery underlying reportability within the neural natural kinds of the clear cases? If the answer is "yes," then there can be no phenomenally conscious representations in Fodorian modules. But how can we know if the answer is "yes"? The suggested methodology requires an answer to the question it was supposed to answer! This paper argues for an abstract solution to the problem and exhibits a source of empirical data that is relevant, data that show that in a certain sense phenomenal consciousness overflows cognitive accessibility. I argue that we can find a neural realizer of this overflow if we assume that the neural basis of phenomenal consciousness does not include the neural basis of cognitive accessibility and that this assumption is justified (other things being equal) by the explanations it allows.

Keywords: access consciousness; accessibility; change blindness; consciousness; mind/body problem; NCC; phenomenal consciousness; refrigerator light illusion; reportability; unconscious; vegetative state; working memory

1. Introduction

In *The Modularity of Mind*, Jerry Fodor argued that significant early portions of our perceptual systems are modular in a number of respects, including that we do not have cognitive access to their internal states and representations of a sort that would allow reportability (Fodor 1983; see also Pylylshyn 2003; Sperber 2001). For example, one representation that vision scientists tend to agree is computed by our visual systems is one which reflects sharp changes in luminosity; another is a representation of surfaces (Nakayama et al. 1995). Are the unreportable representations inside these modules phenomenally conscious? Presumably there is a fact of the matter. But since these representations are cognitively inaccessible and therefore utterly unreportable, how could we know whether they are conscious or not? It may seem that the appropriate methodology is clear in principle even if very difficult in practice: Determine the natural kind (Putnam 1975; Quine 1969) that constitutes the neural basis of phenomenal consciousness in completely clear cases – cases in which subjects are completely confident about their phenomenally conscious states and there is no reason to doubt their authority– and then determine whether those neural natural kinds exist inside Fodorian modules. If they do, there are conscious within-module representations; if not, there are not. But should we include the machinery underlying reportability *within* the natural kinds in the clear cases? Apparently, to decide whether cognitively inaccessible and therefore unreportable representations inside modules are phenomenally conscious, we have to have decided already whether phenomenal consciousness includes the cognitive accessibility underlying reportability. So it looks like the inquiry leads in a circle. I will be calling this problem "the methodological puzzle of consciousness research."

The first half of this paper is about the methodology of breaking out of this circle. The second half brings empirical evidence to bear on actually breaking out of it, using the principle that other things being equal, a mesh between psychology and neuroscience is a reason to believe the theory that leads to the mesh.

2. Two illustrations

Before giving a more precise statement of the methodological puzzle, I'll give two illustrations intended to give the reader a feel for it.

Nancy Kanwisher and her colleagues (Kanwisher 2001; Tong et al. 1998) have found impressively robust correlations between the experience of faces and activation at the bottom of the temporal lobe, usually in the subject's right hemisphere in what they call the "fusiform face area." One method used to investigate the neural basis of face perception exploits a phenomenon known as "binocular rivalry" (see Koch 2004, ch. 16). Presented with a face-stimulus to one eye and a house stimulus to the other, the subject experiences a face for a few seconds, then a house, then a face, and so on. Examination of the visual processing areas of the brain while the face/house perceptual alternation is ongoing found stronger shifts with the percept in the fusiform face area than in other areas. The fusiform face area lights up when subjects are experiencing seeing a face and not when subjects are experiencing seeing a house, despite the fact that the stimuli are unchanging. The fusiform face area also lights up when subjects imagine faces (O'Craven & Kanwisher 2000).

In highly constrained experimental situations, observers viewing fMRI recordings are 85% accurate in telling whether subjects in a scanner are seeing faces or houses (Haynes & Rees 2006). However, Rafi Malach and his colleagues (Hasson et al. 2004) have been able to get similar results from free viewing of movies by correlating activations in a number of subjects (see also Bartels & Zeki 2004).

There has been some dispute as as to what exactly the fusiform face area is specialized for, but these issues can be put aside here. (See Grill-Spector et al. 2006, 2007; Kanwisher 2006, 2007; Tsao et al. 2006)

No one would suppose that activation of the fusiform face area all by itself is sufficient for face-experience. I have never heard anyone advocate the view that if a fusiform face area were kept alive in a bottle, that activation of it would determine face-experience – or any experience at all (Kanwisher 2001). The *total* neural basis of a state with phenomenal character C is itself sufficient for the instantiation of C. The *core* neural basis of a state with phenomenal character C is the *part* of the total neural basis that distinguishes states with C from states with other phenomenal characters or phenomenal contents[i], for example the experience as of a face from the experience as of a house. (The core neural basis is similar to what Semir Zeki [Zeki 2001; Zeki & Bartels 1999] has called an *essential node.*) So activation of the fusiform face area is a candidate for the core neural basis – not the total neural basis – for experience as of a face (see Block 2005; Chalmers 2000; Shoemaker 1981).

For purposes of this paper, I adopt the physicalistic view (Edelman 2004) that consciousness is identical to its total neural basis rather than John Searle's view that consciousness is determined by but not identical to its neural basis (McLaughlin 1992; Searle 1992) . The issue of this paper is not physicalism versus dualism but rather whether consciousness includes the physical functions involved in the cognitive accessibility that underlies reportability.

What is the total minus core neural basis? That is, what is the neural background required to make a core neural basis sufficient for a phenomenally conscious experience? There is some evidence that (and I will be assuming that) there is a single neural background of all experience involving connections between the cortex and the upper brain stem including the thalamus (Churchland 2005; Laureys 2005; Llinás 2001; Llinás et al. 1998; Merker 2007; Tononi & Edelman 1998). This background can perhaps be identified with what Searle (2005) calls the "unified conscious field." Perhaps the most convincing evidence is that disabling connections to the thalamus seems the common core of what different general anesthetics do (Alkire & Miller 2005). Although Merker (2007) does not make the distinction between core and total, he presents evidence that children born pretty much without a cortex can have the conscious field with little or nothing in the way of any conscious contents: that is, they have the total without much in the way of core neural bases.

Nancy Kanwisher (2001) and Dan Pollen (2003, forthcoming) argue that activation of areas of the brain involved in spatio-temporal binding is required for perceptual phenomenology.; Of course some states that have phenomenology, for example, emotions and thoughts, are not experienced as spatially located. But Kanwisher and Pollen may be right about temporal aspects of experience. Further, Antonio Damasio (1999) and Pollen argue that all experience requires a sense of self, partly based in the posterior parietal lobe. If true, this would be part of the background.

At the risk of confusing the reader with yet another distinction, it is important to keep in mind the difference between a causal condition and a constitutive condition. For example, cerebral blood flow is causally necessary for consciousness, but activation of the upper brainstem is much more plausibly a constitutive condition, part of what it is to be conscious. (What does 'constitutive' mean? Among other things, *constituent*: Hydrogen is partially constitutive of water since water is composed of hydrogen and oxygen.) The main issue of this paper is whether the cognitive access underlying reportability is a constitutive condition of phenomenal consciousness.

Here is the illustration I have been leading up to. There is a type of brain injury which causes a syndrome known as *visuo-spatial extinction.* If the patient sees a single object on either side, the patient can identify it, but if there are objects on both sides, the patient can identify only the one on the right and claims not to see the one on the left (Aimola Davies 2004). With competition from the right, the subject cannot attend to the left. However, as Geraint Rees has shown in two fMRI studies of a patient identified as "GK", when GK claims not to see a face on the left, his fusiform face area (on the right, fed strongly by the left side of space) lights up almost as as when he reports seeing the face (Driver & Vuilleumier 2001; Rees et al. 2000, 2002b). Should we conclude that GK has face experience that – because of lack of attention – he does not know about? Or that the fusiform face area is not the whole of the core neural basis for the experience as of a face? Or that activation of the fusiform face area is the core neural basis for the experience as of a face but that some other aspect of the total neural basis is missing? How are we to answer these questions, given that all these possibilities predict the same thing: no face report?

I will use the phrase "core neural basis of the experience" instead of Frances Crick's and Christof Koch's "NCC," for neural correlate of consciousness. Mere correlation is too weak. At a minimum, one wants a match of content between the mental and neural state (Chalmers 1996a; Noë & Thompson 2004).

3. The puzzle

The following is a principle that will be appealing to many, though not to me: Whatever it is about a state that makes it unreportable would also preclude its being phenomenally conscious. We can call this the Phenomenally Conscious ➔ Reportable Principle, or for short, the Phenomenal➔ Reportable Principle. But how could we test the Phenomenal➔ Reportable Principle? If what we mean by a "direct" test is that we elicit reports from subjects about unreportable states, then a direct test will always be negative. And it might seem that there could not be an indirect test either, for an indirect test would have to be based on *some* direct method, a method of investigating whether a state is phenomenally conscious independently of whether it is reportable, a method that apparently does not exist.

A brain-oriented version of the point: Suppose empirical investigation finds a neural state that obtains in all cases in which a phenomenally conscious state is reportable. Such a neural state would be a candidate for a core neural basis. Suppose in addition, that we find that the putative core neural basis is present sometimes when the state is unreportable because mechanisms of cognitive access are damaged or blocked. Would that show the existence of unreportable phenomenal consciousness? No, because there is an alternative possibility, that we were *too quick* to identify the core neural basis. Perhaps the supposed core neural basis that we identified is necessary for phenomenal consciousness but not quite sufficient. It may be that whatever it is that makes the state unreportable also makes it unconscious. Perhaps the cognitive accessibility mechanisms underlying reportability are a constitutive part of the core neural basis, so that without them, there cannot be a phenomenally conscious state. It does not seem that we could find any evidence that would decide one way or the other because any evidence would inevitably derive from the reportability of a phenomenally conscious state, and so it could not tell us about the phenomenal consciousness of a state that cannot be reported. So there seems a fundamental epistemic (that is, having to do with our knowledge of the facts rather than the facts themselves) limitation in our ability to get a complete empirical theory of phenomenal consciousness. This is the methodological puzzle that is the topic of this paper.

Note that the problem cannot be solved by giving a definition of ‘conscious’. Whatever definition one offers of this and other terms, the puzzle can be put in still other terms – there would still be the question, does what it is like to have an experience include whatever cognitive processes underlie our ability to report the experience?

The problem does not arise in the study of, for example, water. On the basis of the study of the nature of accessible water, we can know the properties of water in environments outside our light cone – that is, in environments that are too far away in space and time for signals traveling at the speed of light to reach us. We have no problem extrapolating from the observed to the unobserved and even unobservable in the case of water because we are antecedently certain that our cognitive access to water molecules is not part of the constitutive scientific nature of water itself. In homing in on a core neural basis on the basis of reportable episodes of phenomenal consciousness, we have a choice about whether to include the aspects of those neurological states that underlie reportability *within* the core neural basis. If we do, then unreportable phenomenally conscious states are ruled out; if we do not, unreportable phenomenally conscious states are allowed. Few scientifically minded people in the twenty-first century would suppose that water molecules are partly constituted by our cognitive access to them (Boghossian 2006), but few would be sure whether phenomenal consciousness is or is not partly constituted by cognitive access to it. It is this asymmetry that is at the root of the Methodological Puzzle of phenomenal consciousness.

This issue – whether the machinery of cognitive accessibility is a constitutive part of the nature of phenomenal consciousness – is the focus of this paper. I will not mention evidence concerning inaccessible states within Fodorian modules or whether GK has face experience, but I do claim to show that the issue of whether the cognitive accessibility underlying reportability is part of the constitutive nature of phenomenal consciousness can be resolved empirically and that we already have evidence for a negative answer.

I will turn to a consideration of reportability, but first I want to mention one issue that will not be part of my discussion. Readers will no doubt be familiar with the "explanatory gap" (Levine 1983; Nagel 1974), and the corresponding "Hard Problem" of phenomenal consciousness (Chalmers 1996b), the problem of explaining why the neural basis of a given phenomenal quality is the neural basis of that phenomenal quality rather than some other phenomenal quality or none at all. No one has any idea what an answer would be, even a highly speculative answer. Is the explanatory gap an inevitable feature of our relation to our own phenomenology? Opinions differ (Churchland 1994; McGinn 1991). I will argue that we can make at least some progress on solving the Methodological Puzzle even without progress in closing the explanatory gap.

I have been talking about consciousness versus reportability, but reportability is not the best concept to use in thinking about the puzzle.

4. Cognitive accessibility versus reportability

Empirical evidence about the Phenomenal➔ Reportable Principle *seems* unobtainable, but that is an illusion: *that* principle is clearly false even though another closely related principle is problematic. If a locked-in subject loses control of the last twitch, all mental states can become unreportable. There has been progress in using electrodes implanted in the brain, and less intrusively, EEG technology to enable patients to communicate with the outside world. But if the patient is not trained with these technologies before the total loss of control of the body, these technologies may not work. (See the articles on this topic in the July, 2006 issue of *Nature*.)

There is a distinct problem with the Phenomenal➔ Reportable Principle, namely that a person who is not paralyzed may lose all ability to produce or understand language, and so not have the language capacity required for reporting. In some forms of this syndrome (profound global aphasia), subjects clearly have phenomenal states – they can see, they have pain, and they can make clear what they want and don't want in the manner of a prelinguistic child – but they are totally without the ability to report in any nonextended sense of the term. (Come to think of it, the same point applies to prelinguistic children and animals.) And if an aphasic *also* had locked-in syndrome, the unfortunate conjunctively disabled person would be doubly unable to report conscious states. But there is no reason to think that conscious states would magically disappear. Indeed, given that aphasia is fairly common and locked-in syndrome, though infrequent, is not rare, no doubt there have been such conjunctive cases.

Of course there can be nonverbal reports. Giving a thumbs-up and shaking one's head come to mind. But not every behavioral manifestation of cognitive access to a phenomenal state is a report except in an uninterestingly stretched version of the term. Reportability is legacy of behaviorism that is less interesting than it has seemed. The more interesting issue in the vicinity is not the relation between the phenomenal and the reportable, but rather the relation between the phenomenal and the cognitively accessible.

Adrian Owen and colleagues (Owen et al. 2006) report that a patient who at the time of testing satisfied the criteria for a vegetative state responded to requests to imagine a certain activity in a way indistinguishable from normal patients on an fMRI scan. Her premotor cortex was activated upon being asked to imagine playing tennis, and her parahippocampal place area was activated on being asked to imagine walking through rooms in her house. Paul Matthews objected that the brain activity could have been an associative response to the word *tennis,* but Owen counters that her response lasted 30 seconds – until he asked her to stop (Hopkin 2006). In an accompanying article in *Science*, Lionel Naccache insists on behavioral criteria for consciousness. He says, "Consciousness is univocally probed in humans through the subject's report of his or her own mental states" and notes that Owen and colleagues "did not directly collect such a subjective report" (Naccache 2006). But the evidence is that the patient is capable of an intentional act, namely the act of imagining something described. That should be considered no less an indication – though of course a fallible indication – of consciousness than an external behavioral act. As an editorial in *Nature* suggests, instead of "vegetative state" we should say "outwardly unresponsive." (Editorial 2006)

In the rest of this paper, I will be talking about cognitive accessibility instead of reportability. Reportability is a behavioristic ladder that we can throw away.

In previous papers (Block 1995b, 2001, 2005), I have argued that there can be phenomenally conscious states that are not cognitively accessible. (I put it in terms of phenomenal consciousness without access consciousness.) But I am mainly arguing for something weaker here. Cognitive accessibility could be a causally necessary condition of phenomenal consciousness without being a constitutive part of it. Bananas constitutively include CH_2O molecules but not air and light. Still, without air and light, there could be no bananas—they are causally necessary. The focus here is on whether accessibility is constitutively necessary to phenomenal consciousness, not whether it is causally necessary.

5. Why the Methodological Puzzle matters

I will mention two ways in which it matters whether we can find out whether phenomenal consciousness includes cognitive accessibility. First, if we cannot get evidence about this, we face a fundamental limit in empirical investigation of the neural basis of phenomenal consciousness – we cannot tell whether the putative core neural basis we have found is the neural basis of phenomenal consciousness itself or the neural basis of *phenomenal consciousness wrapped together with the cognitive machinery of access to phenomenal consciousness*.

Second, there is a practical and moral issue having to do with assessing the value of the lives of persons who are in persistent vegetative states. Many people feel that the lives of patients in the persistent vegetative state are not worth living. But do these patients have experiences that they do not have cognitive access to? It is not irrational to regard a rich experiential life – independent of cognitive access to it – as relevant to whether one would want the feeding tube of a loved one removed.

6. Phenomenal consciousness and awareness

We may suppose that it is platitudinous that when one has a phenomenally conscious experience, one is in some way aware of having it. Let us call the fact stated by this claim – without committing ourselves on what exactly that fact is – the fact that phenomenal consciousness requires Awareness. Sometimes people say Awareness is a matter of having a state whose content is in some sense "presented" to the self or having a state that is "for me" or that comes with a sense of ownership or that has "meishness" (as I have called it; Block 1995a).

Very briefly, three classes of accounts of the relation between phenomenal consciousness and Awareness have been offered. Ernest Sosa (2002) argues that all there is to the idea that in having an experience one is necessarily aware of it is the triviality that in having an experience, one experiences one's experience just as one smiles one's smile or dances one's dance. Sosa distinguishes this minimal sense in which one is automatically aware of one's experiences from noticing one's experiences, which is not required for phenomenally conscious experience. At the opposite extreme, David Rosenthal (2005) has pursued a cognitive account in which a phenomenally conscious state requires a higher order thought to the effect that one is in the state. That is, a token experience (one that can be located in time) is a phenomenally conscious experience only in virtue of *another* token state that is about the first state. (See also Armstrong 1977, 1978; Carruthers 2000; Lycan 1996) for other varieties of higher order accounts.) A third view, the "Same Order" view says that the consciousness-of relation can hold between a token experience and *itself*. A conscious experience is reflexive in that it consists in part in an awareness of itself. (This view is discussed in Brentano 1874/1924; Burge 2006; Byrne 2004; Caston 2002; Kriegel 2005; Kriegel & Williford 2006; Levine 2001, 2006; Metzinger 2003; Ross 1961; Smith 1986).

The same order view fits both science and common sense better than the higher order view. As Tyler Burge (2006) notes, to say that one is necessarily aware of one's phenomenally conscious states should not be taken to imply that every phenomenally conscious state is one that the subject notices or attends to or perceives or thinks about. Noticing, attending, perceiving, and thinking about are all cognitive relations that need not be involved when a phenomenal character is present to a subject. The mouse may be conscious of the cheese that the mouse sees, but that is not to say that the mouse is conscious of the visual sensations in the visual field that represent the cheese or that the mouse notices or attends to or thinks about any part of the visual field. The ratio of synapses in sensory areas to synapses in frontal areas peaks in early infancy, and likewise for relative glucose metabolism. (Gazzaniga et al. 2002, p. 642–43). Since frontal areas are likely to govern higher-order thought, low frontal activity in newborns may well indicate lack of higher-order thoughts about genuine sensory experiences.

The relevance of these points to the project of the paper is this: the fact of Awareness can be accommodated by either the same order view or the view in which Awareness is automatic, or so I will assume. So, there is no need to postulate that phenomenal consciousness requires cognitive accessibility of the phenomenally conscious state. Something worth calling "accessibility" may be intrinsic to any phenomenally conscious state, but it is not the cognitive accessibility that underlies reporting.

The highly ambiguous term 'conscious' causes more trouble than it is worth in my view. Some use the term so as to trivially include cognitive accessibility. To avoid any such suggestion I will be abandoning the term 'phenomenal consciousness' (which I think I introduced [Block 1990, 1992]) in favor of 'phenomenology'.

In the next section, I will discuss the assumption underlying the Methodological Puzzle and in the section after how to proceed if we drop that assumption.

7. Correlationism

Correlationism says that the ultimate database for phenomenology research is reports that let us find correlations between phenomenal states and features on the one hand, and scientifically specifiable states and features, viz. neural states and features on the other. These reports can be mistaken, but they can be shown to be mistaken only on the basis of other reports with which they do not cohere. There is no going beyond reports.

One version of correlationism is stated in David Papineau's (2002) *Thinking about Consciousness*, in which he says:

> If the phenomenal property is to be *identical* with some material property, then this material property must be both necessary and sufficient for the phenomenal property. In order for this requirement to be satisfied, the material property needs to be present in all cases where the human subjects report the phenomenal property – otherwise it cannot be necessary. And it needs to be absent in all cases where the human subjects report the absence of the phenomenal property – otherwise it cannot be sufficient. The aim of standard consciousness research is to use these two constraints to pin down unique material referents for phenomenal concepts. (Papineau 2002, p.187)

Consider, for example, what an adherent of this methodology would say about patient GK. One kind of correlationist says we have misidentified the neural basis of face experience and so some aspect of the neural basis of face experience is missing. That is, either activation of the fusiform face area is not the core neural basis for face experience, or if it is, then in extinction patients some aspect of the total neural basis outside the core is missing. Another kind of correlationist does not take a stand on whether GK is having face experience, saying that we cannot get scientific evidence about it.

So there are two versions of correlationism. *Metaphysical* correlationism – the first version just mentioned says that there is (or can be) an answer to the sort of question I have raised about GK and that answer is no. The metaphysical correlationist thinks that the cognitive access relations that underlie the subject's ability to report are a part of what constitutes phenomenology, so there could not be phenomenology without cognitive accessibility (Papineau 1998).

Epistemic correlationism says that GK might be having face experience without cognitive accessibility, but that the issue is not scientifically tractable. According to epistemic correlationism, cognitive accessibility is intrinsic to our knowledge of phenomenology but not necessarily to the phenomenal facts themselves. Epistemic correlationism is more squarely the target of this paper, but I will say a word about what is wrong with metaphysical correlationism.

Why does the metaphysical correlationist think GK cannot be having face experience? Perhaps it is supposed to be a conceptual point – that the very concepts of phenomenology and cognitive accessibility make it incoherent to suppose that the first could occur without the second. Or it could be an empirical point: The evidence (allegedly) shows that the machinery of cognitive accessibility is part of the machinery of phenomenology. I have discussed the conceptual view elsewhere (Block 1978, 1980).

The neuroscientists Stanislas Dehaene and Jean-Pierre Changeux (2004) appear to advocate epistemic correlationism. They say the following. (References are theirs but in this and other quotations to follow are listed in the style of this journal.)

> We shall deliberately limit ourselves, in this review, to only one aspect of consciousness, the notion of *conscious access*… Like others (Weiskrantz 1997), we emphasize *reportability* as a key property of conscious representations. This discussion will aim at characterizing the crucial differences between those aspects of neural activity that can be reported by a subject, and those that cannot. According to some philosophers, this constitutes an "easy problem" and is irrelevant to the more central issues of phenomenology and self-awareness (e.g. Block 1995b). Our view, however, is that conscious access is one of the few empirically tractable problems presently accessible to an authentic scientific investigation.

Kouider and colleagues say: "Given the lack of scientific criterion, at this stage at least, for defining conscious processing without reportability, the dissociation between access and phenomenal consciousness remains largely speculative and even possibly immune to scientific investigation.” (Kouider et al. 2006) (*Access-consciousness* was my term for approximately what I am calling “cognitive accessibility” here.)

In a series of famous papers, Crick and Koch (1995) make use of what appears to be metaphysical correlationism. They argue that the first cortical area that processes visual information, V1, is not part of the neural correlate of phenomenology because V1 does not *directly* project to the frontal cortex. They argue that visual representations must be sent to the frontal cortex to be reported and for reasoning or decision-making to make use of those visual representations. Their argument in effect makes use of the hidden premise that part of the function of visual phenomenology is to harness visual information in the service of the *direct* control of reasoning and decision-making that controls behavior. Jesse Prinz (2000) argues for the AIR theory, for *attended intermediate representations.* The idea is that “consciousness arises when intermediate-level perception representations are made available to working memory via attention.” Because of the requirement of connection to working memory, this is a form of metaphysical correlationism

David Chalmers endorses epistemic correlationism. He says,

Given the very methodology that comes into play here, we have no way of definitely establishing a given NCC as an independent test for consciousness. The primary criterion for consciousness will always remain the functional property we started with: global availability, or verbal report, or whatever. That’s how we discovered the correlations in the first place. 40-hertz oscillations (or whatever) are relevant *only* because of the role they play in satisfying this criterion. True, in cases where we know that this association between the NCC and the functional property is present, the NCC might itself function as a sort of “signature” of consciousness; but once we dissociate the NCC from the functional property, all bets are off. (Chalmers 1996a)

Victor Lamme (2006) gives the example of the split-brain patient who says he does not see something presented on the left, but nonetheless can draw it with his left hand. There is a conflict between normal criteria for conscious states. Lamme says “preconceived notions about the role of language in consciousness” will determine our reaction and there is no objective truth about which view is right. He argues for “letting arguments from neuroscience override our intuitive and introspective notion of consciousness,” using neuroscientific considerations to motivate us to define ‘consciousness’ as recurrent processing, in which higher areas feed back to lower areas, which in turn feed forward to the higher areas again, thereby amplifying the signal. He doesn’t claim the definition is correct, just that it is the only way to put the study of consciousness on a scientific footing. Although Lamme does not advocate correlationism in either its metaphysical or epistemic forms, his view depends on the idea that the only alternative to epistemic correlationism is neurally based postulation.

Often philosophers – Hilary Putnam (1981) and Dan Dennett (1988, 1991) come to mind – argue that two views of the facts about consciousness are “empirically indistinguishable” – and then they in effect conclude that it is better to say that there are no such facts than to adopt epistemic correlationism. One example is Putnam’s thought experiment: We find a core neural basis for some visual experience, but then note that if it occurs in the right hemisphere of a split-brain patient, the patient will say he doesn't see anything. If we restore the corpus callosum, the patient may then say he remembers seeing something. But we are still left with two "empirically indistinguishable" hypotheses, that the hypothesis of the core neural basis is correct so the memory is veridical and, alternatively, that the memory is false.

I will give an empirical argument that we can achieve a better fit between psychology and neuroscience if we assume that phenomenology does not include cognitive accessibility and hence that epistemic correlationism is wrong.

8. An alternative to epistemic correlationism

The alternative I have in mind is just the familiar default "method" of inference to the best explanation, that is the approach of looking for the framework that makes the most sense of all the data, not just reports (Harman 1965; Peirce 1903, vol. V, p. 171).

The reader may feel that I have already canvassed inference to the best explanation and that it did not help. Recall that I mentioned that the best explanation of all the data about observed water can give us knowledge of unobserved – even unobservable – water. I said that this approach does not apply straightforwardly to phenomenology. The reasoning that leads to the Methodological Puzzle says that inevitably there will be a choice about whether to include the neural basis of cognitive access within the neural basis of phenomenology. And that choice – according to this reasoning – cannot be made without some way of measuring or detecting phenomenology independent of cognitive access to it. But we don't have any such independent measure. As I noted, there is a disanalogy with the case of water, since we are antecedently certain that our access to information about water molecules is not part of the natural kind that underlies water molecules themselves. But we are not certain (antecedently or otherwise) about whether our cognitive access to our own phenomenology is partly constitutive of the phenomenology. Without antecedent knowledge of this – according to the reasoning that leads to the Methodological Puzzle – we cannot know whether whatever makes a phenomenal state cognitively inaccessible also renders it nonphenomenal.

Here is the fallacy in that argument: The best theory of *all* the data may be one that lumps phenomenology with water molecules as things whose constitutive nature does not include cognitive access to it. To hold otherwise is to suppose – mistakenly – that there are antecedent views – or uncertainties in this case – that are not up for grabs.

Perhaps an analogy will help. It might seem, offhand, that it is impossible to know the extent of errors of measurement, for any measurement of errors of measurement would have to be derived from measurement itself. But we can build models of the sources of measurement error and test them, and if necessary we can build models of the error in the first level models, and so on, stopping when we get a good predictive fit. For example, the diameter of the moon can be measured repeatedly by a number of different techniques, the results of which will inevitably vary about a mean. But perhaps the diameter of the moon is itself varying? The issue can be pursued by simultaneously building models of source of variation in the diameter itself and models of error in the various methods of measurement. Those models contain assumptions which can themselves be further tested.

The puzzle of how it is possible to use measurement itself to understand errors of measurement is not deep. As soon as one sees the answer, the problem of principle falls away, although it may be difficult to build the models in practice. I do not believe the same is true for the Methodological Puzzle. One reason is the famous "explanatory gap" I mentioned earlier. There may be reasonable doubt whether the method of inference to the best explanation can apply in the face of the explanatory gap. A second point is that with the demise of verificationism (Uebel 2006), few would think that the nature of a physical magnitude such as length or mass is constitutively tied to our measurement procedures. The mass of the moon is what it is independent of our methods of ascertaining what it is. But verificationism in the case of consciousness is much more tempting – see Dan Dennett's "first person operationism" (Dennett 1991) for a case in point. Lingering remnants of verificationism about phenomenology do not fall away just because someone speaks its name.

The remainder of this article will describe evidence that phenomenology overflows cognitive accessibility and a neural mechanism for this overflow. The argument is that this mesh between psychology and neuroscience is a reason to believe the theory that allows the mesh. The upshot is, there are distinct mechanisms of phenomenology and cognitive accessibility that can be empirically investigated.

Figure 1. Compare this with Figure 4 without looking at the two figures side by side. There is a difference between the two pictures that can be hard to be aware of, a fact that motivates the appellation (a misnomer in my view) "Change Blindness."

9. Phenomenology overflows accessibility

George Sperling (1960) showed subjects arrays of alphanumeric characters, for example three rows of four characters, for 50 ms, followed by a blank field. Subjects said they could see all or almost all of the characters. This has also been reported in replications of the experiment (Baars 1988, p. 15). The phenomenology of a version of the experiment was described by William James in his *Principles of Psychology*: "If we open our eyes instantaneously upon a scene, and then shroud them in complete darkness, it will be as if we saw the scene in ghostly light through the dark screen. We can read off details in it which were unnoticed whilst the eyes were open," (James 1890) and may be what Aristotle was talking about when he said, "even when the external object of perception has departed, the impressions it has made persist, and are themselves objects of perception" (Aristotle in Ross 1955, 460b).

When Sperling asked subjects to say what letters they had seen, subjects were able to report only about 4 or 5 of the letters, fewer than half the letters they said they'd seen. (This result was first reported by Cattell [1885] – I am indebted to Patrick Wilken [2001]). Did the subjects really see all or almost all the shapes as they said? Sperling's clever idea was to test whether people really did see all or almost all of the characters and whether the phenomenology persists after the stimulus is turned off by playing a tone soon after the array was replaced by a blank. Subjects were to report the top row if the tone was high, the bottom row if the tone was low, and the middle row in case of an intermediate tone. The result? Subjects could report all or almost all the characters in any given row. Versions of this type of experiment have been done with as many as 32 alphanumeric characters with similar results (Sligte et al. 2006). An attractive picture of what is going on here – and one that I think makes the most sense of the data – is that although one can distinctly see all or almost all of the 9–12 objects in an array, the processes that allow one to conceptualize and identify the specific shapes are limited by the capacity of "working memory," allowing reports of only about 4 of them. That is, the subject has experiences as of specific alphanumeric shapes, but cannot bring very many of them under specific shape or alphanumeric concepts (representations) of the sort required to report or make comparisons. The subject can bring them under a general concept – "alphanumeric character" – which is why the subjects can report that they have seen an array of alphanumeric characters but not under the more specific concepts required to identify which alphanumeric character. Interestingly, Sperling found the same results whether he made the exposure of the grid as short as 15 ms or as long as 500 ms.

Sperling's experiment is often described as showing that a "visual icon" persists after the stimulus is turned off. However as Max Coltheart (1980) notes, this term is used ambiguously. In my terms, the ambiguity is between (1) phenomenal persistence and (2) persistence of accessible information concerning the stimulus. Because these are the very notions whose empirical separation is the topic of this paper, the term *icon* is especially unfortunate and I will not be using it further.[ii]

The idea that one does in fact phenomenally register many more items than are (in a sense) accessible and that that phenomenology persists beyond the stimulus is further tested in a combination of a change "blindness" paradigm with a Sperling-like paradigm (Landman et al. 2003).

First, I will sketch the change "blindness" paradigm. In these experiments, a photograph is presented briefly to subjects, followed by a blank, followed sometimes by an identical photograph but other times by a similar but not identical photograph, followed by another blank. Then the cycle starts over. When the two photographs differ, they usually differ in one object that changes color, shape, or position, or appears or disappears. The surprising result is that subjects are often unaware of the difference between the two pictures, even when the changed region takes up a good deal of the photographic real estate. Even with 50 repetitions of the same change over and over again, people are often unaware of the change. It is widely agreed that the phenomenon is an attentional one. The items that change without detection have been shown to be items that the subjects do not attend to. But the controversial question – to be discussed later – is whether the phenomenon is one of inattentional blindness or inattentional inaccessibility.[iii]

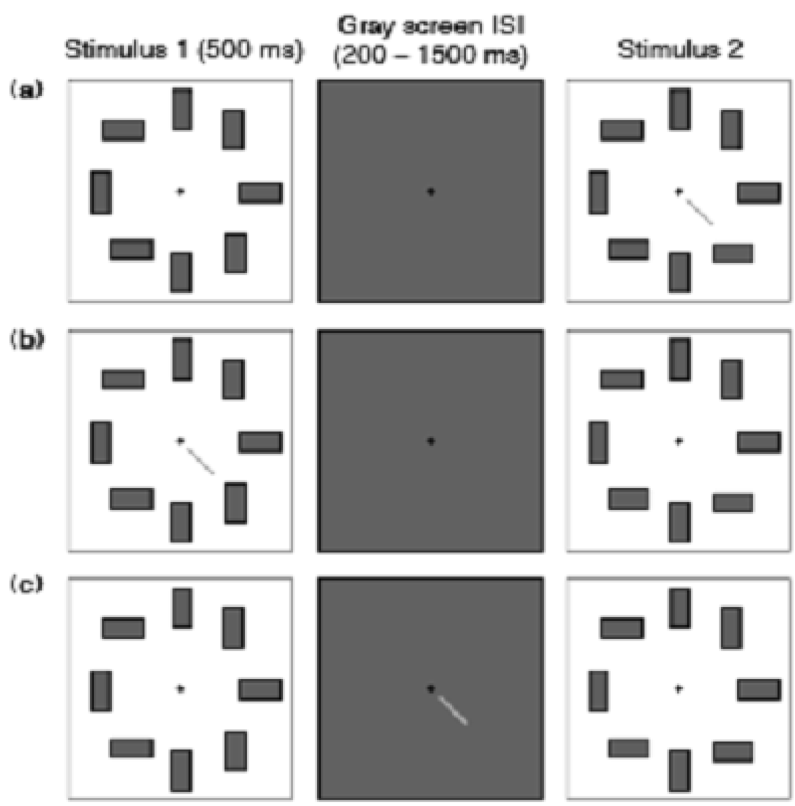

Figure 2. Landman et.al.'s paradigm combining change "blindness" with Sperling's experiments on iconic memory. The rectangles are displayed here as line drawings but the actual stimuli were defined by textures. (From Lamme 2003.)

Now for the experiment by Landman et al. (2003). The subject is shown 8 rectangles for half a second as in (a) of Figure 2. There is a dot in the middle which the subject is supposed to keep looking at. (This is a common instruction in visual perception experiments and it has been found using eye-tracking that subjects have little trouble maintaining fixation.) The array is replaced by a blank screen for a variable period. Then another array appears in which a line points to one of the objects which may or may not have changed orientation. In the example shown in Figure 2, there is an orientation change. Using statistical procedures that correct for guessing, Landman et al. computed a standard capacity measure (Cowan's K – see Cowan 2001) showing how many rectangles the subject is able to track. In (a), subjects show a capacity of 4 items. Thus, the subjects are able to deploy working memory so as to access only half of the rectangles despite the fact that in this as in Sperling's similar task, subjects' reported phenomenology is of seeing all or almost all of the rectangles. This is a classic "change blindness" result. In (b), the indicator of the rectangle that may or may not change comes on in the first panel. Not surprisingly, subjects can get almost all right: their capacity measure is almost 8. The crucial manipulation is the last one: the indicator comes on during the blank after the original rectangles have gone off. If the subjects are continuing to maintain a visual representation of the whole array – as subjects say they are doing – the difference between (c) and (b) will be small, and that is in fact what is observed. The capacity measure in (c) is between 6 and 7 for up to 1.5 seconds after the first stimulus has been turned off, suggesting that subjects are able to maintain a visual representation of the rectangles. This supports what the subjects say and what William James said about the phenomenology involved in this kind of case. What is both phenomenal and accessible is that there is a circle of rectangles. What is phenomenal but in a sense not accessible, is all the specific shapes of the rectangles. I am taking what subjects say at face value (though of course I am prepared to reject what subjects say if there is evidence to that effect). Whether that is right will be taken up in the section after next.

Subjects are apparently able to hold the visual experience for up to 1.5 seconds – at least "partial report superiority" (as it is called) lasts this long – considerably longer than in the Sperling type experiments in which the production of 3–4 letters for each row appears to last at most half a second. The difference (Landman et al. 2003) is that the Sperling type of experiment requires a good enough representation for the subjects to actually volunteer what the letters were, whereas the Landman et. al. methodology only requires a "same/different" judgment. (Yang 1999) found times comparable to Landman using similar stimuli.

In one variation, Landman and colleagues did the same experiment as before but changed the size of the rectangles rather than the orientation, and then in a final experiment, changed either the size or the orientation. The interesting result is that subjects were no worse at detecting changes in either orientation or size than they were at detecting changes in size alone. That 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 supporting the subjects' reports.

There is some reason to think the longest lasting visual representations of this sort come with practice and when subjects learn to "see (and not look)". Sligte and colleagues found long persistence, up to 4 seconds in a paradigm similar to that of Landman with lots of practice (Sligte et al. 2006). Others (Long 1980; Yang 1999) have noted that practice in partial report paradigms makes a big difference in subjects' ability to make the visual experience last. These experiments are hard for the subjects and there are large differences among subjects in ability to complete the experiment (Long 1985, Yang 1999); in some cases experimenters have dismissed subjects who simply could not perform the tasks (Treisman et al. 1975). Snodgrass and Shevrin (2006) also find a difference (in a different paradigm) between "poppers" who like to relax and just see and "lookers" who are more active visual seekers.

The main upshot of the Landman and Sligte experiments (at least on the surface – we will consider debunking explanations later) is along the same lines as that of the Sperling experiment: The subject has persisting experiences as of more specific shapes than can be brought under the concepts required to report or compare those specific shapes with others. They can all be brought under the concept "rectangle" in the Landman experiment or "letter" in the Sperling experiment, but not the specific orientation-concepts which would be required to make the comparisons in Landman or to report the letters in Sperling.

Why are subjects able to gain access to so few of the items they see in the first condition of the Landman experiment (as described in (a) of Figure 2) and in the Sperling phenomenon without the tones? I am suggesting that the explanation is that the "capacity" of phenomenology, or at least the visual phenomenal memory system, is greater than that of the working memory buffer that governs reporting. The capacity of visual phenomenal memory could be said to be at least 8–32 objects – at any rate for stimuli of the sort used in the described experiments. This is suggested by subjects' reports that they can see all or almost all of the 8–12 items in the presented arrays and by the experimental manipulations just mentioned in which subjects can give reports which exhibit the subjects apprehension of all or almost all of the items. In contrast, there are many lines of evidence that suggest that the "working memory" system – the "global workspace" – has a capacity of about 4 items (or fewer) in adult humans and monkeys and 3 (or fewer) in infants.

When some phenomenal items are accessed, something about the process erases or overwrites others, so in that sense the identities of the items are not all accessible. However, any one of the phenomenal items is accessible if properly cued, and so in that sense all are accessible. Another sense in which they are all accessible is that the subject knows that he sees them all (or almost all). The upshot is that there is phenomenology without accessibility (Block 1995a) in one sense of the term but not another (Chalmers 1997; Kobes 1995). Of course, there is no point in arguing about which sense of the word 'accessibility' to use.

The argument of this paper is thus importantly different from that in earlier works (Block 1995b, 1997) in which I argued that the Sperling experiment directly shows the existence of phenomenal states that are not cognitively accessible, a conclusion that is not argued for here. In this paper, I use the fact of overflow to argue for the conclusion that the machinery of phenomenology is at least somewhat different from the machinery of cognitive accessibility. I will then argue that there is a neural realization of the fact of phenomenological overflow – if we assume that the neural basis of phenomenology does not even include the neural basis of cognitive access to it, and that is a reason to accept that assumption. The neural argument suggests that machinery of cognitive access is not included in machinery of phenomenology.

What does it mean to speak of the representational capacity of a system as a certain number of objects? Working memory capacity is often understood in terms of "slots" that are set by the cognitive architecture. One capacity measure that is relevant to phenomenology is the one mentioned above in connection with the Landman and Sligte experiments – Cowan's (and Pashler's) K , which I will not discuss further. Another capacity measure relevant to phenomenology is what subjects say about seeing all of a number of items.

There are two significant remaining issues:

1. How do we know that the Sperling, Landman, and Sligte effects are not retinal or otherwise prephenomenal?

2. How do we know we can believe subjects' reports to the effect that they experience all or almost all of the objects in the Sperling and Landman experiments? Perhaps subjects confuse potential phenomenology with actual phenomenology just as someone may feel that the refrigerator light is always on because it is on when he looks.

10. Is the effect retinal or otherwise prephenomenal?

The persistence of phenomenology is based in the persistence of neural signals. But some neural persistence may *feed* phenomenology rather than *constitute* it, and that creates a difficulty for the view that the capacity of phenomenology is larger than the capacity of working memory. It is no news that the "representational capacity" of the retina is greater than 4! Activations of the retina are certainly not part of the minimal neural basis of visual phenomenology. Rather, activations of the retina are part of the causal process by which that minimal supervenience base is activated. A minimal neural basis is a necessary part of a sufficient condition for conscious experience. We know the retina is not part of a minimal core neural basis because, for one thing, retinal activation stays the same even though the percept in the binocular rivalry experiments mentioned earlier shift back and forth. It may be said that phenomenological capacity is no greater than that of working memory, the appearance to the contrary deriving from tapping large capacity preconscious representations.

The experimental literature points to activations at all levels of the visual system during phenomenal persistence (Coltheart 1980; Di Lollo 1980). However, there are clear differences between the effects of retinal persistence and persistence higher in the cortex.

One of the neatest dissections of phenomenal persistence in low level-vision from that in high-level vision comes from an experiment by G. Engel (1970). (The discussion of this experiment in Coltheart [1980] is especially useful.) Engel used as stimuli a pair of "random dot stereograms" of the sort invented by Bela Julesz in 1959. I can't show you an example that would allow you to experience the effect because the kind Engel used requires a stereo viewer. I can best explain what they are by telling you how they are made.

Figure 3. Random-dot stereograms (Thanks to Júlio M. Otuyama)

You start with a grid of say 100 by 200 tiny squares. You place a dot in a random location within each square in the grid. The result is something that looks bit like snow on an old black and white TV – like one of the rectangles in Figure 3. Then you copy the grid dot by dot, but you make a certain change in the copy. You pick some region of it and move every dot in that region slightly to the left, leaving the rest of the dots alone. The right rectangle in the figure is the result of moving a square-shaped region in the middle horizontally. The resulting figure looks to the untrained naked eye just like the first rectangle, but since the visual system is very sensitive to slight disparities, if each eye is presented with one of the two rectangles in a stereo viewer (or if you can "free fuse" the images), the viewer sees a protruding square.

The illusion of depth requires the two rectangles to be presented to the two eyes, but if the presentations to the two eyes are at different times, there will be no loss of the experience of depth so long as the two presentations are not separated by too much time. Suppose the left stereogram is presented to the left eye for 10 ms, and then the right stereogram is presented to the right eye 50 ms later. No problem: There is an experience of depth. Indeed, one can present the second stereogram up to 80 ms later and still get the experience of depth. Monocular persistence, persistence of the signal to a single eye, lasts 80 milliseconds. Think about the left eye getting its stereogram and the right eye then getting its stereogram 50 ms later. If there is no independent stereo persistence, the depth experience would expire in another 30 ms, when the monocular persistence to the left eye runs out. But that does not happen. Depth experience goes on much longer. Engel considered the question of how long one can wait to present another pair of stereograms before the subject loses the experience of depth. He presented sequences of the left stimulus, then the right, then the left, then the right, and so on. If the initial left was followed by a right within 80 ms, he found that the next left had to come within 300 ms in order for the subject's experience of depth to be continuous. That is, the experience of depth lasts 300 ms. The retina is of course completely monocular; each retinal activation depends on input to just one eye. Indeed, substantial numbers of binocular cells are not found in early vision. The conclusion: Depth requires processing in areas of the visual system higher than early vision.

So this experiment shows two kinds of persistence, monocular persistence lasting 80 ms and binocular persistence lasting 300 ms; and the binocular persistence is clearly phenomenal because it is a matter of the subject continuing to see depth.

Here is another item of evidence for the same conclusion. There is phenomenal persistence for visual motion which cannot be due merely to persistence of neural signals in the retina or in early visual areas. Anne Treisman (Treisman et al. 1975) used a display of 6 dots, each of which moved in a circular pathway, either clockwise or counterclockwise. Subjects were asked to report whether the motion was clockwise or counterclockwise. Treisman and colleagues found a partial report advantage much like that in Sperling's experiment. See also (Demkiw & Michaels 1976).

Why can't this phenomenon be accounted for by neural persistence in the retina or early vision? The point can be understood by imagining a moving dot that goes from left to right on a TV screen. Suppose the screen is phosphorescent so that the images leave a ghost that lasts 1.5 seconds (inspired by the Landman et. al. experiment) and suppose that the initial moving dot moves across the screen in 100 milliseconds. The viewer will see a dot on the left that expands into a line towards the right over a 100 ms period. The line will remain for 1,300 ms and then it will shrink towards the right to a dot on the right over another 100 ms. The idea of the analogy is to point to the fact that retinal persistence of the signals of a moving object cannot be expected to create phenomenal persistence of the experience of that moving object. Phenomenal persistence has to be based in neural persistence that is a good deal higher in the visual system. As will be discussed later, there is an area in the visual system (V5) that is an excellent candidate for the neural basis of the visual experience of motion.

Perhaps the strongest evidence for cortical persistence comes from the Sligte et al. paper mentioned earlier. There is evidence that the persisting visual experience can be divided into two phases. In the first phase, it is indistinguishable from visual perception. This phase typically lasts at most a few hundred ms (unless subjects are dark-adapted, in which case it lasts longer), and often less than 100 ms. The persistence of the experience can be tested by many methods, for example asking subjects to adjust the timing of a second visual stimulus so that what the subject experiences is a seamless uninterrupted visual stimulus. (See Coltheart, 1980 for a description of a number of converging experimental paradigms that measure visible persistence.) In the second phase, the subject has a fading but still distinctly visual experience. The first two phases are of high capacity and disturbed if the test stimulus is moved slightly, and easily "masked" (Phillips 1974, Sligte et al. 2006) by stimuli that overlap in contours with the original stimulus. (Such a mask, if presented at the right lag, makes the stimulus hard to see.)

Sligte and colleagues used dark adaptation to increase the strength of the first phase, producing what could be described as a positive afterimage. They also introduced a further variable, two kinds of stimuli: a black/white stimulus and a red/gray isoluminant stimulus in which the foreground and background have the same level of luminance. The idea was to exploit two well-known differences between rods and cones in the retina. Rods are color blind and also have an extended response to stimulation whereas cones have a brief burst of activity. Rods react to isoluminant stimuli as to a uniform field. The black and white stimulus in dark adaptation will however maximize rod stimulation, producing longer visible persistence without affecting the later working memory representation (Adelson 1978). Sligte found, not surprisingly, that the black and white stimuli produced very strong visible persistences, much stronger than the isoluminant red and gray stimuli when the cue was given just after the array of figures was turned off. (In arrays with 16 items, the subjects had a capacity of 15 for the black and white stimuli but only 11 for the red and gray stimuli.) Here is the very significant result for the issue of retinal persistence versus cortical persistence. A brief flash of light just after the array wiped out this difference. However, when the flash of light was given later after about 1,000 ms after the array stimulus, it had no effect. Further, a pattern mask did have a huge effect at 1,000 ms, lowering the capacity to the level of working memory. The flash of light right after the stimulus interferes with retinal persistence, whereas the pattern mask after 1,000 ms interfered with cortical persistence.

As I mentioned, Sligte used as many as 32 items instead of the 8 of Landman. The capacity for the black/white stimulus was close to 32 for the early cue, the capacity of the red/gray stimulus was about 17 and both fell to about 16 for the cue late in the blank space. And both fell further to somewhat over 4 – as in Landman – once the new stimulus came on. If the cue was presented 10 ms after the first stimulus (the analog of (c) in Figure 2), the black/white stimulus produced greater retention, but if the cue was presented late in the blank period (or once the new stimulus came on as in (a)), the black/white and red/grey stimuli were equivalent. The upshot is that the first phase is very high capacity and is over by 1,000 ms, that the second phase is high capacity and lasts up to 4 seconds and that the third phase has a similar capacity to the working memory phase in Sperling and Landman.

The results mentioned earlier in connection with the Sperling and Landman experiments are likely to be based in central parts of the visual system, and so not due to something analogous to "looking again" as in the imaginary dialog presented earlier. However, the question of exactly which central neural activations constitute phenomenology as opposed to constituting input to phenomenology is just the question of what phenomenology is in the brain, and of course the topic of this paper is whether that can be empirically investigated. So it may seem that I have unwittingly shown the opposite of what I am trying to show, namely that every attempt to give an empirical answer ends up presupposing an answer. So how can my argument avoid begging the question?

I have three responses. First, the evidence suggests neural persistence at all levels in the visual system. There is no reason to think the phenomenal level is an exception. Second, as mentioned earlier, there is evidence to come that a certain kind of activation of V5 is the core neural basis of the experience of motion. We can see how experimental evidence from phenomenal persistence could dovetail with the evidence outside of memory for V5 as the neural basis for the visual experience of motion. If some version of Treisman's experiment were done in a scanner, my point of view would predict persisting V5 activations of the aforementioned kind. So the issue is not beyond investigation. Third, this paper is about the question of whether the machinery of cognitive access is part of the neural basis of phenomenology. That issue is orthogonal to the question of whether a given activation in early visual cortex is part of the neural basis of phenomenology or only an input to that neural basis. The issue of this paper is about the interface of phenomenology with cognitive accessibility, not the interface of phenomenology with retinal and other early visual processing.

11. The Refrigerator Light Illusion

The argument of this paper depends on the claim that subjects in the Sperling and Landman experiments have phenomenal experiences of all or almost all of the shapes in the presented array. One objection is that subjects' judgments to that effect are the result of an illusion in which they confuse potential phenomenology with actual phenomenology. In order to explain this allegation and defend against it, I will first have to say more about cognitive accessibility.

The dominant model of cognitive accessibility in discussions of consciousness – and one that is assumed both in this paper and by Stan Dehaene and his colleagues, the critics who I will be talking about in this section – is a model of broadcasting in a global workspace that started with the work of Bernard Baars (1988, 1997) The idea is closely related to my notion of access consciousness and Dan Dennett's (1993, 2001) notion of "cerebral celebrity" or fame in the brain.[iv] Think of perceptual mechanisms as suppliers of representations to consuming mechanisms which include mechanisms of reporting, reasoning, evaluating, deciding and remembering. There is empirical evidence that it is reasonable to think of perceptual systems as sending representations to a global active storage system, which is closely connected to the consuming systems. Those representations are available to all cognitive mechanisms without further processing. (That's why blindsight "guesses" don't count as cognitively accessible in this sense; further processing in the form of guessing is required to access the representations.) This workspace is also called "working" memory – the word "memory" being a bit misleading because after all, one can report an experience while it is happening without having to remember it in any ordinary sense of the term.

Dehaene and colleagues (Dehaene et al. 1998; Dehaene & Nacchache 2001) have given impressive evidence that our ability to report our phenomenal states hinges on such a global workspace and that the connection between perception and the workspace lies in long-range neurons in sensory areas in the back of the head which feed forward to the workspace areas in the front of the head.

In past publications, I argued for phenomenology without cognitive accessibility (Block 1995a, 1995b, 2001) on the basis of the Sperling experiment. Dehaene and Naccache replied, making use of the global workspace model.

> Some information encoded in the nervous system is permanently inaccessible (set I_1). Other information is in contact with the workspace and could be consciously amplified if it was attended to (set I_2). However, at any given time, only a subset of the latter is mobilized into the workspace (set I_3). We wonder whether these distinctions may suffice to capture the intuitions behind Ned Block's ((Block 1995b); see also (Block 2001)) definitions of phenomenal (P) and access (A) consciousness. What Block sees as a difference in essence could merely be a qualitative difference due to the discrepancy between the size of the potentially accessible information (I_2) and the paucity of information that can actually be reported at any given time (I_3). Think, for instance, of Sperling's experiment in which a large visual array of letters seems to be fully visible, yet only a very small subset can be reported. The former may give rise to the intuition of a rich phenomenological world – Block's P-consciousness – while the latter corresponds to what can be selected, amplified, and passed on to other processes (A-consciousness). Both, however, would be facets of the same underlying phenomenon. (Dehaene & Naccache 2001, p. 30)

The distinction between I_1, I_2, and I_3 is certainly useful, but its import depends on which one or more of these categories is supposed to be phenomenal. One option is that representations in both categories I_2 (potentially in the workspace) and I_3 (in the workspace) are phenomenal. That is not what Dehaene and Naccache have in mind. Their view (see especially section 3.3.1 of their paper) is that only the representations in I_3 are phenomenal. They think that representations in the middle category (I_2) of potentially in the workspace *seem* to the subject to be phenomenal but that this is an illusion. The only phenomenal representations are those that are *actually* in the workspace. But in circumstances in which the merely potential workspace representations can be accessed at will, they *seem* to us to be phenomenal. That is, the subjects allegedly mistake merely *potential* phenomenology for actual phenomenology.

Importantly, the workspace model exposes a misleading aspect of talk of cognitive accessibility. What it is for representations to be in the workspace (I_3) involves both actuality (sent to the workspace) and potential (can be accessed by consuming mechanisms without further processing). The representations that are actually in the workspace are in active contact with the consuming systems, and the consuming systems can (potentially do) make use of those representations. We might speak of the representations in I_3 (in the workspace) as cognitively accessible in the narrow sense (in which consuming mechanisms make use of what is already there), and representations in the union of I_3 and I_2 as cognitively accessible in the broad sense. It is narrow cognitive accessibility that Dehaene et al. identify with phenomenology. When I speak of phenomenology overflowing cognitive accessibility, I mean that the capacity of phenomenology is greater than that of the workspace—so it is narrow accessibility that is at issue. In the rest of this paper, I will be using "cognitive accessibility" in the narrow sense. The thesis of this paper is that phenomenology does not include cognitive accessibility in the narrow sense. Here we see that as theory – driven by experiment – advances, important distinctions come to light among what appeared at first to be unified phenomena (Block & Dworkin 1974 on temperature; Churchland 1986, 1994, 2002 on life and fire).

But what is wrong with the broad sense? Answer: The broad sense encompasses *too much*, at least if a necessary and sufficient condition of phenomenology at stake. Representations in I_2 can be "amplified if…attended to", but of course uncontroversially *unconscious* representations can be amplified too, if one shifts attention to what they represent,(Carrasco 2007). So including everything in I_2 in consciousness would be a mistake, a point I made (Block 1997) in response to the claim that consciousness correlates with a certain functional role by Chalmers (1997). No doubt a functional notion that is intermediate between narrow and broad could be framed but the challenge for the framer would be to avoid ad hoc postulation.

An experimental demonstration that shifting attention affects phenomenology to a degree sufficient to change a sub-threshold stimulus into a supra-threshold stimulus is to be found in a series of papers by Marisa Carrasco (Carrasco, et.al. 2004) in which she asked subjects to report the orientation of the one of a pair of gratings that had the higher contrast. She presented an attention-attracting dot on one side of the screen or the other slightly before the pair of gratings. She showed that the attention made a grating that was lower in contrast than the comparison seem higher in contrast. In subsequent work (Carrasco 2007), Carrasco has been able to show precisely measurable effects of attentional shifts on contrast and color saturation, but not hue.

This alleged conflation of potential phenomenology with actual phenomenology could be called the Refrigerator Light Illusion[v] (Block 2001), the idea being that just as someone might think the refrigerator light is always on, confusing its potential to be on with its actually being on, so subjects in the Sperling and Landman experiments might think that all the items register phenomenally because they can see any one that they attend to. In the rest of this section, I will argue against this allegation.

Let us begin by mentioning some types of illusions. There are neurological syndromes in which cognition about one's own experience is systematically wrong, for example subjects with anosognosia can complain bitterly about one neural deficit while denying another. And cognitive illusions can be produced reliably in normals (Piattelli-Palmarini 1994). To take a famous example, doctors are more reluctant to recommend surgical intervention if they are told that a disease has a mortality rate of 7% than if they are told it has a survival rate of 93%. Moving to a cognitive illusion that has a more perceptual aspect, much of vision is serial but subjects take the serial processes to be simultaneous and parallel (Nakayama 1990). For example, G. W. McConkie and colleagues (McConkie & Rayner 1975; McConkie & Zola 1979) created an eye-tracking setup in which subjects are reading from a screen of text but only the small area of text surrounding the fixation point (a few letters to the left and 15 to the right) is normal – the rest is perturbed. Subjects have the mistaken impression that the whole page contains normal text. Subjects suppose that the impression of all the items on a page is a result of a single glance, not realizing that building up a representation of a whole page is a serial process. These illusions all have a strong cognitive element.

Are the results from experiments like those of Sperling and Landman the result of cognitive illusions? One reason to think not is that the phenomenon that the Sperling and Landman experiments depend on do not require that subjects be asked to access any of the items. It is a simple matter to show subjects arrays and ask them what they see without asking them to report any specific items (as was done first in Gill & Dallenbach, 1926). This suggests that the analysis of subjects' judgments in the partial report paradigms as based on cognition – of noticing the easy availability of the items – is wrong. A second point is that cognitive illusions are often, maybe always, curable. For example, the framing illusion mentioned above is certainly curable. However, I doubt that the Sperling and Landman phenomenology is any different for advocates of the Dehaene view. Third, the sense that in these experiments so much of the perceptual content slips away before one can grab hold of it cognitively does not seem any kind of a cognition but rather is percept-like.

Recall, that in the Sperling experiment, the results are the same whether the stimulus is on for 50 ms or 500 ms. Steve Schmidt has kindly made a 320 ms demo that is available on my Web site at
www.nyu.edu/gsas/dept/philo/faculty/block/demos/Sperling320msec.mov.
See for yourself.

The suggestion that the putative illusion has a perceptual or quasi-perceptual nature comports with the way Dan Dennett and Kevin O'Regan describe the sparse representations allegedly revealed by change "blindness" (Dennett 1991; O'Regan 1992).[vi] Their idea is that the way it *seems that it seems* is – supposedly – not the way it *actually seems*. They allege not a mismatch between appearance and external reality as in standard visual illusions but rather a mismatch between an appearance and an appearance of an appearance. We could call this alleged kind of illusion in which the introspective phenomenology does not reflect the phenomenology of the state being introspected a *hyper-illusion.*

But are there any clear cases of hyper-illusions? I don't know of any. One candidate is the claim, often made, that although the "self" is really a discontinuous stream of experiences, we have the illusion that it is a continuous existent (Strawson 2003). But this alleged hyper-illusion is suspect, being perhaps more a matter of failing to experience the gappiness rather than actually experiencing non-gappiness. Further, subjects' introspective judgments led to the prediction investigated by Sperling, Landman, and Sligte. One should have an empirical reason to judge that this experimentally confirmed introspective judgment is wrong.

Subjects in the Landman experiment are looking right at the rectangles for half a second, a long exposure, and it is not hard to see the orientations clearly. It does not appear to them as if something vaguely rectangularish is coming into view, as if from a distance. In (c) of Landman, they see all the rectangle orientations for up to 1.5 seconds in the Landman version and up to 4 seconds in the Sligte version. It is hard to believe that people are wrong about the appearances for such a long period.

Dehaene and colleagues revisit this issue. Here is the relevant passage:

> The philosopher Ned Block, however, has suggested that the reportability criterion underestimates conscious contents (Block 2005). When we view a complex visual scene, we experience a richness of content that seems to go beyond what we can report. This intuition led Block to propose a distinct state of "phenomenal consciousness" prior to global access. This proposal receives an apparent confirmation in Sperling's iconic memory paradigm. When an array of letters is flashed, viewers claim to see the whole array, although they can later report only one subsequently cued row or column. One might conclude that the initial processing of the array, prior to attentional selection of a row or column is already phenomenally conscious. (Block 2005, Lamme 2003)
>
> However, those intuitions are questionable, because viewers are known to be over-confident and to suffer from an "illusion of seeing". [(O'Regan & Noe 2001).] The change blindness paradigm demonstrates this "discrepancy between what we see and what we think we see" (Simons & Ambinder 2005). In this paradigm, viewers who claim to perceive an entire visual scene fail to notice when an important element of the scene changes. This suggests that, at any given time, very little of the scene is actually consciously processed. Interestingly, changes that attract attention or occur at an attended location are immediately detected. Thus, the illusion of seeing may arise because viewers know that they can, at will, orient attention to any location and obtain conscious information from it. (Dehaene et al. 2006)

Dehaene and his colleagues propose to use the change "blindness" results to back up their view of the Sperling result. But the issues in these two paradigms are pretty much the same – our view of one is conditioned by our view of the other. Further, as I mentioned earlier, the first form of the Landman et al. experiment (See Figure 2 part (a)) is itself an experiment in the same vein as the standard change "blindness" experiments. The subject sees 8 things clearly but has the capacity (in the sense of Cowan's K) to make comparisons for only 4 of them. And so the Landman et al. experiment – since it gives evidence that the subject really does see all or almost all the rectangles – argues against the interpretation of the change "blindness" experiments given by Dehaene and his colleagues.

Figure 4. Compare this with Figure 1 without looking at the two figures side by side. There is a difference that can be hard to see.

Dehaene et al. say, "The change blindness paradigm demonstrates this discrepancy between what we see and what we think we see." But this claim is hotly contested in the experimental community, including by one of the authors that they cite. As I mentioned earlier (footnote iii), many psychologists would agree that initial interpretations of change "blindness" went overboard and that rather than seeing the phenomenon as a form of inattentional blindness, one might see it as a form of inattentional inaccessibility (Block 2001). That is, the subject takes in the relevant detail of each of the presented items, but they are not conceptualized at a level that allows the subject to make a comparison. As Fred Dretske (2004) has noted, the difference between the two stimuli in a change blindness experiment can be one object that appears or disappears, and one can be aware of that object that constitutes the difference without noticing that there is a difference.

Compare Figure 1 with Figure 4. It can be hard for subjects to see what the difference between Figure 1 and Figure 4, even when they are looking right at the feature that changes. The idea that one cannot see the feature that changes strains credulity.

Two of the originators of the change "blindness" experiments, Dan Simons and Ron Rensink, (2005b) have since acknowledged that the "blindness" interpretations are not well supported by the "change blindness" experiments. In a discussion of a response by

Alva Noë (2005), they summarize (Simons and Rensink 2005a):

> We and others found the 'sparse representations' view appealing (and still do), and initially made the overly strong claim that change blindness supports the conclusion of sparse representations (Rensink et al. 1997; Simons 1997). We wrote our article because change blindness continues to be taken as evidence for sparse – or even absent – representations, and we used O'Regan and Noë's influential paper (O'Regan & Noe 2001) as an example. However, as has been noted for some time… this conclusion is logically flawed

I have been appealing to what the subjects say in Sperling-like experiments about seeing all or almost all the items. However, there is some experimental confirmation of what the subjects say in different paradigms. Geoffrey Loftus and his colleagues (Loftus & Irwin 1998) used a task devised by Vincent Di Lollo (1980) and his colleagues using a 5 by 5 grid in which all but one square is filled with a dot. They divided the dots into 2 groups of 12, showing subjects first one group of 12 briefly, then a pause, then the other group of 12 briefly. The subjects always were given partial grids, never whole grids. Subjects were asked to report the location of the missing dot – something that is easy to do if you have a visual impression of the whole grid. In a separate test with no missing dots, subjects were asked to judge on a scale of 1 to 4 how temporally integrated the matrix appeared to be. A 4 meant one complete matrix appeared to have been presented whereas a 1 meant that two separate displays had been presented. The numerical ratings are judgments that reflect phenomenology – how complete the grids looked. The length of the first exposure and the time between exposures was varied. This experiment probes persistence of phenomenology without using the partial report technique that leads Dehaene and his colleagues to suggest the Refrigerator Light Illusion. The result is that subjects' ability to judge which dot was missing correlated nearly perfectly with their phenomenological judgments of whether there appeared to be a whole matrix as opposed to two separate partial matrices. That is, the subjects reported the experience of seeing a whole matrix if and only if they could pick out the missing dot, thus confirming the subjects phenomenological reports.

To sum up: (1) the subjects' introspective judgments in the experiments mentioned are that they see all or almost all of the items. Dehaene and his colleagues seem to agree since that is entailed by the claim that the introspective judgments are illusory. (2) This introspective judgment is not contingent on subjects' being asked to report items as would be expected on the illusion hypothesis. (3) This introspective judgment leads to the prediction of partial report superiority, a prediction that is borne out. (4) The accuracy of the subjects' judgments is suggested by the fact that subjects are able to recover both size and orientation information with no loss. (5) These results cohere with a completely different paradigm – the Loftus paradigm just mentioned. (6) Dehaene and colleagues offer no other empirical support than the corresponding theory of the change "blindness" results that raise exactly the same issues.

The conclusion of this line of argument is, as mentioned before, that phenomenology overflows cognitive accessibility and so phenomenology and cognitive access are based at least partly in different systems with different properties. I will be moving to the promised argument that appeals to the mesh between psychology and neuroscience after I fill in some of the missing premises in the argument, the first of which is the evidence for a capacity of visual working memory of roughly four or less.

12. Visual Working Memory

At a neural level, we can distinguish between memory that is coded in the active firing of neurons – and ceases when that neuronal firing ceases – and structural memory that depends on changes in the neural hardware itself, for example change in strength of synapses. The active memory—which is active in the sense that it has to be actively maintained-- is sometimes described as "short term" – a misdescription since it lasts as long as active firing lasts, which need not be a short time if the subject is actively rehearsing. In this paper, the active memory buffer is called "working memory".

You may have heard of a famous paper by George Miller called "The magical number seven, plus or minus two: Some limits on our capacity for processing information" (Miller 1956). Although Miller was more circumspect, this paper has been widely cited as a manifesto for the view that there is a single active memory system in the brain that has a capacity of seven plus or minus two "items." What is an item? There are some experimental results that fill this notion in a bit. For example, Huntley-Fenner and colleagues (2002) showed that infants' visual object tracking system – which, there is some reason to believe, makes use of working memory representations – does not track piles of sand that are poured, but does track them if they are rigid. One constraint on what an item might be comes from some experiments that show that although we can remember only about four of them, we can also remember up to four features of each one. Luck and Vogel asked subjects to detect changes in a task somewhat similar to the Landman et. al. task already mentioned. They found that subjects could detect changes in four features (color, orientation, size, and the presence or absence of a gap in a figure) without being significantly less accurate than if they were asked to detect only one feature. (Luck & Vogel 1997; Vogel et al. 2001).

In the 50 years since Miller's paper, reasons have emerged to question whether there really is a single active memory system as opposed to a small number of linked systems connected to separate modalities and perhaps separate modules – for example, language. Some brain injuries damage verbal working memory but not spatial working memory (Basso et al. 1982), and others have the opposite effect (Hanley et al. 1991). And evidence has accumulated that the capacity of these working memories – especially visual working memory – is actually lower than seven items (Cowan 2001, Cowan et al. 2006).

The suggestion of seven items was originally made plausible by experiments demonstrating that people, read lists of digits, words or letters, can repeat back about seven of them. Of course, they can repeat more items if the items can be "chunked." Few

Americans will have trouble holding the nine letters 'FBICIAIRS' in mind, because the letters can be chunked into 3 acronyms.

More relevant to our discussion, visual working memory experiments also come up with capacities in the vicinity of four – or fewer than four items. (For work that suggests fewer than four see McElree 2006). Whether there is one working memory system that is used in all modalities or overlapping systems that differ to some extent between modalities, this result is what is relevant to the experiments discussed above. Indeed, you have seen three examples in this paper: the Sperling, Landman, and Sligte experiments themselves! I will briefly mention a few other quite different paradigms that have come up with the same number. One such paradigm involves the number of items that people – and monkeys – can effortlessly keep track of. For example, at a rhesus macaque monkey colony on a small island off of Puerto Rico, Marc Hauser and his colleagues did the following experiment: Two experimenters find a monkey relaxing on its own. Each experimenter has a small bucket and a pocket full of apple slices. The experimenters put down the buckets and one at a time, they conspicuously place a small number of slices in each bucket. Then they withdraw and check which bucket the monkey goes to to get the apple slices. The result is that for numbers of slices equal to or smaller than four, the monkeys overwhelmingly choose the bucket with more slices. But if either bucket has more than four, the monkeys choose at random. In particular, monkeys chose the greater number in comparison of one versus two, two versus three and three versus four, but chose at random in cases of four versus five, four versus six, four versus eight and, amazingly, three versus eight. The comparison of the three versus four case (where monkeys chose more) and the three versus eight case (where they chose at random) is especially telling (Hauser et al. 2000). The eight apple slices simply overflowed working memory storage. Infant humans show similar results, although typically with a limit more in the vicinity of three rather than four (Feigenson et al. 2002). Using graham crackers instead of apple slices, Feigenson and colleagues found that infants would crawl to the bucket with more crackers in the cases of one versus two and two versus three but were at chance in the case of one versus four. Again, four crackers overflows memory storage. In one interesting variant, infants are shown a closed container into which the experimenter – again conspicuously – inserts a small number of desirable objects (for example, M&Ms). If the number of M&Ms is one, two, or three, the infant continues to reach into the container until all are removed, but if the number is more than three, infants reach into the container just once (Feigenson & Carey 2003).

I mentioned above that some studies have shown that people can recall about four items including a number of features of each one. However, other studies (Xu 2002) have suggested smaller working memory capacities for more complex items. Xu & Chun (2006) have perhaps resolved this controversy by showing that there are two different systems with somewhat different brain bases. One of these systems has a capacity of about four spatial locations or objects at four different spatial locations, independent of complexity, the other a smaller capacity depending on the complexity of the representation. The upshot for our purposes is that neither visual working memory system has a capacity higher than four.

This section is intended to back up the claim made earlier about the capacity of working memory – at least visual working memory. I move now to a quick rundown on working memory and phenomenology in the brain with an eye to giving more evidence that we are dealing with at least partially distinct systems with different properties.

13. Working memory and phenomenology in the brain

Correlationism in its metaphysical form (which, you may recall, regards cognitive accessibility as part of phenomenology) would have predicted that the machinery underlying cognitive access and underlying phenomenal character would be inextricably entwined in the brain. But the facts so far can be seen to point in the opposite direction, or so I will argue.

In many of the experiments mentioned so far, a brief stimulus is presented, then there is a delay before a response is required. What happens in the brain during the delay period? In experiments on monkeys using this paradigm, it has been found that neurons in the upper sides of the prefrontal cortex (dorsolateral prefrontal cortex) fire during the delay period. And errors are correlated with decreased firing in this period. (Fuster 1973; Goldman-Rakic 1987). Further, damage to neurons in this area has been found to impair delayed performance, but not simultaneous performance, and damage to other memory systems does not interfere with delayed performance (except possibly damage to parahippocampal regions in the case of novel stimuli (Hasselmo & Stern 2006)). Infant monkeys (1.5 months old) are as impaired as adult monkeys with this area ablated, and if the infant area is ablated the infants do not develop working memory capacity. It appears that this prefrontal area does not itself store sensory signals but rather is the main factor in maintaining representations in sensory, sensorimotor and spatial centers in the back of the head. As Curtis and D'Esposito (2003) note, the evidence suggests that this frontal 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." That is, the frontal area is coupled to and maintains sensory representations in the back of the head that represent, for example, color, shape, and motion. (See Supèr et al. 2001a for an exploration of the effect of this control on the posterior regions.) The main point is that as the main control area for working memory, this prefrontal area is the main bottleneck in working memory, the limited capacity system that makes the capacity of working memory what it is.

So the first half of my brain-oriented point is that the control of working memory is in the front of the head. The second half is that, arguably, the core neural basis of visual phenomenology is in the back of the head. I will illustrate this point with the example of one kind of visual experience of motion (typified by optic flow). But first a caution: No doubt the neural details presented here are wrong or at least highly incomplete. We are still in early days. My point is that the evidence does point in a certain direction, and more important, we can see how the issues I have been talking about could be resolved empirically.

Here is a brief summary of *some* of the vast array of evidence that the core neural basis of one kind of visual experience of motion is activation of a certain sort in an region in the back of the head centered on the area known as V5[vii]. The evidence includes:

> Activation of V5 occurs during motion perception (Heeger et al. 1999).
>
> Microstimulation to monkey V5 while the monkey viewed moving dots influenced the monkey's motion judgments, depending on the directionality of the cortical column stimulated (Britten et al. 1992).
>
> Bilateral (both sides of the brain) damage to a region that is likely to include V5 in humans causes akinetopsia, the inability to perceive – and to have visual experiences as of motion. (Akinetopsic subjects see motion as a series of stills.) (Rees et al. 2002a, Zihl et al. 1983)
>
> The motion aftereffect – a moving afterimage – occurs when subjects adapt to a moving pattern and then look at a stationary pattern. (This occurs, for example, in the famous "waterfall illusion.") These moving afterimages also activate V5 (Huk et al. 2001).
>
> Transcranial magnetic stimulation (TMS[viii]) applied to V5 disrupts these moving afterimages (Theoret et al. 2002).
>
> V5 is activated even when subjects view "implied motion" in still photographs, for example, of a discus thrower in mid-throw. (Kourtzi & Kanwisher 2000)
>
> TMS applied to visual cortex in the right circumstances causes stationary phosphenes[ix] – brief flashes of light and color. (Kammer 1999) When TMS is applied to V5, it causes subjects to experience moving phosphenes (Cowey & Walsh 2000).

However, mere activation over a certain threshold in V5 is not enough for the experience as of motion; the activation probably has to be part of a recurrent feedback loop to lower areas (Kamitani & Tong 2005; Lamme 2003; Pollen 2003; Supèr et al. 2001a). Pascual-Leone and Walsh (2001) applied TMS to both V5 and V1 in human subjects with the TMS coils placed so that the stationary phosphenes determined by the pulses to V1 and the moving phosphenes from pulses to V5 overlapped in visual space. When the pulse to V1 was applied roughly 50 ms later than to V5, subjects said that their phosphenes were mostly stationary instead of moving. The delays are consonant with the time for feedback between V5 and V1, which suggests that experiencing moving phosphenes depends not only on activation of V5 but also on a recurrent feedback loop in which signals go back to V1 and then forward to V5. Silvanto and colleagues (2005a, 2005b) showed subjects a brief presentation of an array of moving dots. The experimenters pinpointed the precise time – call it t – at which zapping V5 with TMS would disrupt the perception of movement. Then they determined that zapping V1 either 50 ms before t or 50 ms after t would also interfere with the perception of the moving dots. But zapping V5 a further 50 ms after that (100 ms after t) had no effect. They argue that in zapping V1 50 ms before t, they are intercepting the visual signal on its way to V5 and in zapping V1 50 ms after t, they are interfering with the recurrent loop. These results suggest that one V1-V5-V1 loop is the core neural basis for at least one kind of visual experience as of motion (and also necessary for that kind of experience in humans).

Recurrent loops also seem to be core neural bases for other types of contents of experience (Supèr et al. 2001a). The overall conclusion is that there are different core neural bases for different phenomenal characters. (Zeki and his colleagues have argued for a similar conclusion, using Zeki's notion of micro-consciousness [Pins & Ffytche 2003, Zeki 2001]).[x]

14. Neural coalitions

But there is a problem in the reasoning of the last section. Whenever a subject reports such phenomenology, that can only be via the activation of the frontal neural basis of global access. And how do we know whether those frontal activations are required for – indeed are part of – the neural basis of the phenomenology? Metaphysical correlationists say they are; epistemic correlationists say we can't know. This section will draw together strands that have been presented to argue that both kinds of correlationism are wrong because we have empirical reason to suppose that activation of working memory circuits are not part of the neural basis of phenomenology (not part of either the core or total neural basis).

A preliminary point: (Pollen forthcoming) summarizes evidence that prefrontal lobotomies on both sides and other frontal lesions do not appear to decrease basic perceptual content such as luminance or color. Frontal damage impairs access but it doesn't dim the bulb (Heath et al. 1949). Still, it could be said that *some* degree of frontal activation, even if minimal, is part of the background required for phenomenal consciousness, and, epistemic correlationists would allege, once there is so much frontal damage that the subject cannot report anything at all, there is no saying whether the person has any phenomenal consciousness at all.

In the rest of this section, I will give my argument against this view, the one that the second half of the paper has been leading up to: If we suppose that the neural basis of the phenomenology does *not* include the workspace activations, we can appreciate a neural mechanism by which phenomenology can overflow cognitive accessibility.

There is convincing evidence that the neural processes underlying perceptual experience can be thought of in terms of neural network models (see Koch 2004, ch. 2, pp. 19, 20). In visual perception, coalitions of activation arise from sensory stimulation and compete for dominance in the back of the head, one factor being feedback from the front of the head that differentially advantages some coalitions in the back. Dominant coalitions in the back of the head trigger coalitions in the front of the head that themselves compete for dominance, the result being linked front and back winning coalitions. Support for this sort of model comes from, among other sources, computerized network models that have confirmed predictive consequences. (See Dehaene et al. 1998, 2006; Dehaene & Nacchache 2001)) Furthermore, some recent experiments (Sergent & Dehaene 2004) provide another line of evidence for this conclusion that is particularly relevant to the themes of this paper.

This line of evidence depends on a phenomenon known as the "attentional blink." The subject is asked to focus on a fixation point on a screen and told that there will be a rapid sequence of stimuli. Most of the stimuli are "distractors," which in the case of the Sergent & Dehaene version are black nonsense letter strings. The subject is asked to look for two "targets," a white string of letters, either XOOX or OXXO, and a black name of a number, for example, "five." One or both targets may be present or absent in any given trial. At the end of the series of stimuli, the subjects have to indicate what targets they saw. In the *standard* attentional blink, subjects simply indicate identity of the target or targets. The standard finding is that if the subject saw the first target (for example, XOOX), and if the timing of the second target is right, the second target (for example, five) is unlikely to be reported at certain delays (so long as it is followed by distractors that overwrite the phenomenal persisting representation, as in Figure 5). In this setup, a delay of about 300 ms makes for maximum likelihood for the second target to be "blinked." Sergent and Dehaene used a slight modification of this procedure in which subjects were asked to manipulate a joystick to indicate just how visible the number name was. One end of the continuum was labeled maximum visibility and the other was total invisibility. (See Figure 5.)

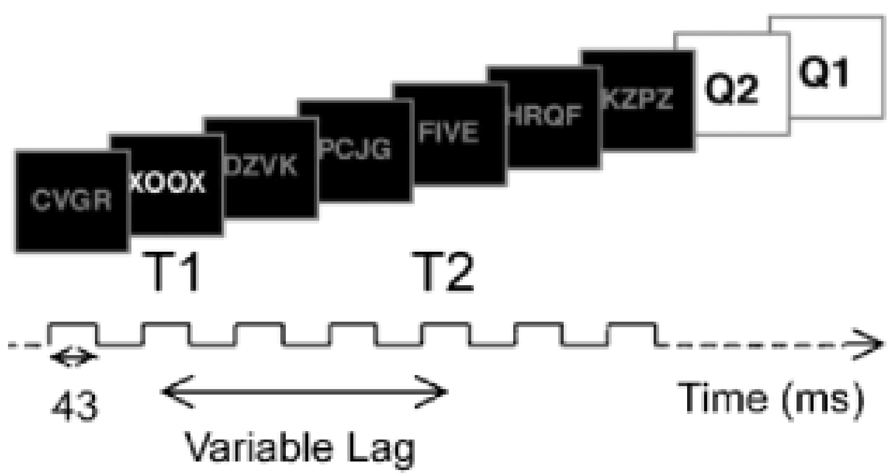

Figure 5. The Attentional Blink. A sequence of visual stimuli in which the first target is a white string of letters, either XOOX or OXXO and the second target is the name of a number. At the end of the series the subject is asked to indicate how visible target 2 was and whether target 1 was present and if so in which form.

The interesting result was that subjects tended to indicate that target 2 was either maximally visible or maximally invisible; intermediate values were rarely chosen. This fact suggests a competition among coalitions of neurons in the front of the head with winners and losers and little in between. Usually, the coalition representing the number word either wins, in which case the subject reports maximum visibility, or it loses, in which case subjects report no second target and there is no cognitive access to it at all.

I have guessed (Block 2005) that there can be coalitions in the back of the head that lose by a small amount and thus do not trigger winning coalitions in the front, but that are nonetheless almost as strong as the back of the head coalitions that do trigger global broadcasting in the front. The subject sees many things, but only some of those things are attended to the extent that they trigger global broadcasting. A recent study (Kouider et al. 2006) suggests that indeed there are strong representations in the back of the head that do not benefit from attention and so do not trigger frontal activations. (See also Tse et al. 2005 for convergent results.)

Kouider et al. contrasted a subliminal and supraliminal presentation of a stimulus, a lower-case word. In the subliminal case, the stimulus was preceded and succeeded by masks, which have the effect of decreasing the visibility of the stimulus (and, not incidentally, decreasing recurrent neural activation – see Supèr et al. 2001b). In the supraliminal case, the masks closest to the stimulus were omitted. The supraliminal but not the subliminal stimulus could be identified by the subjects when given a forced choice. In the key manipulation, the subject was told to look for an upper case word, ignoring everything else. In those conditions of distraction, subjects claimed that they were aware of the lower case stimuli in the supraliminal case but that they could hardly identify them because they were busy performing the distracting task on the upper case stimulus (which came later). The difference between the supraliminal and subliminal stimuli in conditions of distraction was almost entirely in the back of the head (in occipitotemporal areas). Supraliminal stimuli activated visual areas in the back of the head strongly but did not activate frontal coalitions.[xi] The strong activations in the back of the head did, however, *modulate* frontal activity.

Kouider et al. (2006) and Dehaene et al. (2006) acknowledge that there are highly activated losing coalitions in the back of the head. They argue that such losing coalitions are the neural basis of "preconscious" states – because they cannot be reported. But the claim that they are not conscious *on the sole ground of unreportability* simply assumes metaphysical 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.

Here is the argument that the second half of this paper has been building up to. If we assume that the strong but still losing coalitions in the back of the head are the neural basis of phenomenal states (so long as they involve recurrent activity), then we have a neural mechanism which explains why phenomenology has a higher capacity than the global workspace. If, on the contrary, we assume that the neural basis of phenomenology includes workspace activation, then we do not have such a mechanism. That gives us reason to make the former assumption. If we make the former assumption – that workspace activation is not part of the neural basis of phenomenology – we have a mesh between the psychological result that phenomenology overflows cognitive accessibility and the neurological result that perceptual representations that do not benefit from attention can nonetheless be almost as strong (and probably recurrent) as perceptual representations that do benefit from attention. The psychological argument from overflow showed that the machinery of phenomenology is at least to some extent different from that of cognitive accessibility, since something not in cognitive accessibility has to account for the greater capacity of phenomenology. What the mesh argument adds is that the machinery of phenomenology does not include the machinery of cognitive accessibility. Indeed, one could go further, arguing that they do not even overlap.

Of course my conclusion that the neural machinery of cognitive access is not partially constitutive of phenomenology leaves room for *causal* influence in both directions. And it may be that top-down causal influence is almost always involved in making the phenomenal activations strong enough. But that is compatible with the possibility of the relevant amplification happening another way, for example, by recurrent loops confined to the back of the head or even by stimulation by electrodes in the brain, and that is enough to show that top-down amplification is not *constitutively* necessary.

My first conclusion then is that the overlap of the neural machinery of cognitive access and the neural machinery of phenomenology can be empirically investigated. Second, there is evidence that the latter does not include the former. These points are sufficient to refute the correlationism of the sort advocated by Dehaene and his colleagues and to answer the question posed at the beginning of the paper.

Further, this theoretical picture leads to predictions. One prediction is that in the Sperling, Landman, and Sligte experiments the representations of the unaccessed items will prove to involve recurrent loops. Another upshot is that if the activations of the fusiform face area mentioned earlier in the patient GK turn out to be recurrent activations, we would have evidence for phenomenal experience that the subject not only does not know about but in these circumstances cannot know about. The fact that the fusiform face activations produced in GK by the faces he says he doesn't see are almost as strong as the activations corresponding to faces he does see suggests that top-down amplification is not necessary to achieve strong activations.

The mesh argument argues that workspace activation is not a constitutive part of phenomenology. And given that actual workspace activation is not a constitutive part of phenomenology, it is hard to see how anyone could argue that potential workspace activation is a constitutive part. Further, as noted a few paragraphs back, it is doubtful that potential workspace activation is even causally necessary to phenomenology.

15. Conclusion

If we want to find out about the phenomenological status of representations inside a Fodorian module, we should find the neural basis of phenomenology in clear cases and apply it to neural realizers inside Fodorian modules. But that requires already having decided whether the machinery of access should be included in the neural kinds in clear cases, so it seems that the inquiry leads in a circle. This paper has been about breaking out of that circle. The abstract idea of the solution is that all the questions have to be answered simultaneously, tentative answers to some informing answers to others. The key empirical move in this paper has been to give meshing answers to psychological and neural considerations about overflow.[xii]

References

Adelson, E. H. (1978) Iconic storage: The role of rods. *Science* 201:-544–46.

Aimola Davies, A. M. (2004) Disorders of spatial orientation and awareness: Cognitive and behavioral rehabilitation: From neurobiology to clinical practice. The Guilford Press.

Alkire, M. T. & Miller, J. (2005) General anesthesia and the neural correlates of consciousness. *Progress in Brain Research* 150:-229–44.

Aristotle (1955) On dreams. In: *Aristotle. Parva naturalia*, ed. W. D. Ross, pp.175–223 . Clarendon Press.

Armstrong, D. M. (1977) The causal theory of the mind. *Neue Heft für Philosophie* 11: 82–95.

(1978) What is consciousness? *Proceedings of the Russellian Society* 3:-65–76.

Baars, B. (1988) *A cognitive theory of consciousness*, Cambridge University Press.

(1997) *In the theater of consciousness: The workspace of the mind*,. Oxford.

Bartels, A. & Zeki, S. (2004) Functional brain mapping during free viewing of natural scenes. *Human Brain Mapping 21*.

Basso, A., Speinnler, H., Vallar, G. & Zanobio, M. E. (1982) Left hemisphere damage and selective impairment of auditory verbal short term memory: A case study. *Neuropsychologia* 20:-263–74.

Block, N. (1978) Troubles with functionalism. *Minnesota Studies in the Philosophy of Science* 9:-261–325.

(1980) What is functionalism? In: *Readings in the Philosophy of Psychology*, ed. N. Block, pp.171-84 . Harvard University Press.

(1990) Consciousness and accessibility. *Behavioral and Brain Sciences* 13(4):596–98.

(1992) Begging the question against phenomenal consciousness. *Behavioral and Brain Sciences* 15(2) .

(1995a) How many concepts of consciousness? *Behavioral and Brain Sciences* 18(2):272–84.

(1995b) On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18(2):227–47.

(1997) Biology versus computation in the study of consciousness. *Behavioral and Brain Sciences* 20(1):159–66.

(2001) Paradox and cross purposes in recent work on consciousness. *Cognition* 79(1–2):197–220.

(2005) Two neural correlates of consciousness. *Trends in Cognitive Sciences* 9(2):46–52.

Block, N. & Dworkin, G. (1974) IQ, Hheritability, and Iinequality I. *Philosophy and Public Affairs 3*(4):331–409.

Boghossian, P. A. (2006) *Fear of knowledge: Against relativism and constructivism*,. Oxford University Press.

Brentano, F. (1874/1924) *Psychologie vom Empirischen Standpunkt*, Felix Meiner.

Britten, K. H., Shadlen, M. N., Newsome, W. T. & Movshon, A. (1992) The analysis of visual motion: A comparison of neuronal and psychophysical performance. *Journal of Neuroscience* 12(4)745–65.

Burge, T. (2006) Reflections on two kinds of consciousness. In: *Philosophical Essays, Volume II: Foundations of Mind*, ed T. Burge, pp. 392-419. Oxford University Press.

Byrne, A. (2004) What phenomenal consciousness is like. In: *Higher-order theories of consciousness*, eds. R. Gennaro & John Benjamins.

Carrasco, M. (2007) Visual attention alters appearance: Psychophysical studies of subjective experience. In: Oxford Companion to Consciousness, ed. P. Wilken, T. Bayne & A. Cleeremans, Oxford University Press

Carrasco, M., Ling, S. & Read, S. (2004) Attention alters appearance. Nature Neuroscience 7(3), 308-313.

Carruthers, P. (2000) *Phenomenal consciousness: A naturalistic theory*,. Cambridge University Press.

Caston, V. (2002) Aristotle on consciousness. *Mind* 111(444):751–815.

Cattell, J. M. (1885) The inertia of the eye and brain. *Brain* 8(3):295–312.

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

(1997) Availability: The cognitive basis of experience. *Behavioral and Brain Sciences* 20(1):148–49 .

(1998) On the search for the neural correlate of consciousness. In: *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates*, ed- S. Hameroff, A. Kaszniak, & A. Scott, pp. 219–30. MIT Press.

(2000) What is a neural correlate of consciousness? In: *Neural correlates of consciousness: Empirical and conceptual issues*, ed. T. Metzinger, pp. 17–40 . MIT Press.

Churchland, P. S. (1986) *Neurophilosophy: Toward a unified science of the mind-brain*., MIT Press.

(1994) Can neurobioloby teach us anything about consciousness? *American Philosophical Association Proceedings* 67(4):23–40.

(2002) *Brain-wise: Studies in neurophilosophy*, MIT Press.

(2005) A neurophilosophical slant on consciousness research. *Progress in Brain Research* 149:285–92.

Cohen, J. (2002) The grand illusion illusion. *Journal Of Consciousness Studies* 9(5–6): 141–57.

Coltheart, M. (1980) Iconic memory and visible persistence. *Perception and Psychophysics* 27(3):183–228.

Cowan, N. (2001) The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 24:87–185.

Cowan, N., Morey, C. & Chen, Z. (2007) The legend of the magical number seven. In: *Tall tales about the brain: Things we think we know about the mind, but ain't so*, ed. S. Della Sala.. Oxford University Press.

Cowey, A. & Walsh, V. (2000) Magnetically induced phosphenes in sighted, blind and blindsighted subjects. *NeuroReport* 11(3):269–73.

Crick, F. & Koch, C. (1995) Are we aware of neural activity in primary visual cortex? *Nature* 375:121–23.

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

Damasio, A. (1999) *The feeling of what happens: Body and emotion in the making of consciousness*, Harcourt Brace and Co.

de Fockert, J. W., Rees, G., Frith, C. D. & Lavie, N. (2001) The role of working memory in visual selective attention. *Science* 291(1):803–06.

Dehaene, S. & Changeux, J.P. (2004) Neural mechanisms for access to consciousness. In: *The Cognitive Neurosciences III*, ed. M. Gazzaniga, pp. 1145–58. MIT Press.

Dehaene, S., Changeux, J.-P., Nacchache, L., Sackur, J. & Sergent, C. (2006) Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences* 10:204–11.

Dehaene, S., Kerszberg, M. & Changeux, J.P. (1998) A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences* 95:14,529–34.

Dehaene, S. & Nacchache, L. (2001) Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition* 79:1–37.

Demkiw, P. & Michaels, C. (1976) Motion information in iconic memory. *Acta Psychologica* 40:257–64.

Dennett, D. (1988) Quining Qualia. In: *Consciousness in contemporary science*, ed. A. Marcel & E. Bisiach, pp. 381–414. Oxford University Press.

(1991) *Consciousness explained.* Little Brown.

(1993) The message is: There is no *me*dium. *Philosophy and Phenomenological Research* 53(4):889–931.

(2001) Are we explaining consciousness yet? *Cognition* 79:221–37.

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(1–3):1–18.

Driver, J. & Vuilleumier, P. (2001) Perceptual awareness and its loss in unilateral neglect and extinction. *Cognition* 79(1–2):39–88.

Edelman, G. M. (2004) *Wider than the sky: The phenomenal gift of consciousness.* Yale University Press.

Editorial, Nature. (2006) Flickers of consciousness. *Nature* 443(7,108): 121–22.

Engel, G. R. (1970) An investigation of visual responses to brief stereoscopic stimuli. *Quarterly Journal of Experimental Psychology* 22:148–60.

Feigenson, L. & Carey, S. (2003) Tracking individuals via object files: Evidence from infants'' manual search. *Developmental Science* 6:568–84.

Feigenson, L., Carey, S. & Hauser, M. (2002) The representations underlying infants'' choice of more: Object files vs. analog magnitudes. *Psychological Science* 13(150–56).

Fodor, J. A. (1983) *The Modularity of Mind.* MIT Press.

Fuster, J. M. (1973) Unit activity in prefrontal cortex during delayed-response performance: Neuronal correlates of transient memory. *Journal of Neurophysiology* 36:61–78.

Gazzaniga, M., Ivry, R. & Mangun, G. (2002) *Cognitive neuroscience: The biology of the mind.* W. W. Norton.

Gill, N. F. & Dallenbach, K. M. (1926) A preliminary study of the range of attention. *The American Journal of Psychology* 37(2):247–56.

Goldman-Rakic, P. (1987) Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In: *Handbook of physiology, the nervous system, higher functions of the brain*, ed. F. Plum, pp. 373-417. American Physiological Society.

Grill-Spector, K., Sayres, R. & Ress, D. (2006) High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. *Nature Neuroscience* 9(1):177–85.

(2007) Corrigendum. *Nature Neuroscience* 10:133.

Hanley, J. R., Young, H. W. & Pearson, N. A. (1991) Impairment of visuo-spatial sketch pad. *Quarterly Journal of Experimental Psychology* 43a(101–25).

Harman, G. (1965) The inference to the best explanation. *The Philosophical Review* 74(1): 88–95.

Hasselmo, M. E. & Stern, C. E. (2006) Mechanisms underlying working memory for novel information. *Trends in cognitive sciences* 10(11):487–93.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G. & Malach, R. (2004) Intersubject synchronization of cortical activity during natural vision. *Science* 303:(1)634–40.

Hauser, M., Carey, S. & Hauser, L. (2000) Spontaneous number representation in semi-free-ranging rhesus monkeys. *Proceedings of the Royal Society of London: Biological Sciences* 267:829–33.

Haynes, J.-D. & Rees, G. (2006) Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience* 7:523–34.

Heath, R. G., Carpenter, M. B., Mettler, F. A. & Kline, N. S. (1949) Visual apparatus: Visual fields and acuity, color vision, autokinesis. In: *Selective partial ablation of the frontal cortex, a correlative study of its effects on human psychotic subjects; problems of the human brain*, ed. F. A. Mettler, pp. 489–91 . P. B. Hoeber, Inc.

Heeger, D. J., Boynton, G. M., Demb, J. B., Seideman, E. & Newsome, W. T. (1999) Motion opponency in visual cortex. *Journal of Neuroscience* 19(7):162–74.

Hopkin, M. (2006) 'Vegetative' patient shows signs of conscious thought. *Nature* 443(7,108):132–33.

Huk, A. C., Ress, D. & Heeger, D. J. (2001) Neuronal basis of the motion aftereffect reconsidered. *Neuron* 32:161–72.

Huntley-Fenner, G., Carey, S. & Solimando, A. (2002) Objects are individuals but stuff doesn't count: Perceived rigidity and cohesiveness influence infants' representations of small numbers of discreet entities. *Cognition* 85:203–21.

James, W. (1890) *Principles of Psychology.* Henry Holt.

Kamitani, Y. & Tong, F. (2005) Decoding the visual and subjective contents of the human brain. *Nature Neuroscience* 8:679–85.

Kammer, T. (1999) Phosphenes and transient scotomas induced by magnetic stimulation of the occipital lobe: Their topographic relationship. *Neuropsychologia* 37:191–98.

Kanwisher, N. (2001) Neural events and perceptual awareness. *Cognition* 79:89–113.

(2006) What's in a face? *Science* 311(5,761):617–18.

(2007) Does the fusiform face area contain subregions highly selective for nonfaces? *Nature Neuroscience* 10(1):3–4.

Kobes, B. (1995) Access and what it is like. *Behavioral and Brain Sciences* 18(2):260.

Koch, C. (2004) *The quest for consciousness: A neurobiological approach.* Roberts and Company.

Kouider, S., Dehaene, S., Jobert, A. & Le Bihan, D. (2006) Cerebral bases of subliminal and supraliminal priming during reading. *Cerebral Cortex*.

Kourtzi, Z. & Kanwisher, N. (2000) Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience* 12:48–55.

Kriegel, U. (2005) Naturalizing subjective character. *Philosophy and Phenomenological Research* 71:23–57.

Kriegel, U. & Williford, K. (2006) *Self-representational approaches to consciousness.* MIT Press.

Kriegeskorte, N., Sorger, B., Naumer, M., Scwarzbach, J., van den Boogert, E., Hussy, W., & Goebel, R. (2003) Human cortical object recognition from a visual motion flowfield. *Journal of Neuroscience* 23(4):1,451.

Lamme, V. (2003) Why visual attention and awareness are different. *Trends in Cognitive Sciences* 7:12–18.

(2006) Towards a true neural stance on consciousness. *Trends in Cognitive Sciences* 10(11):494–501.

Landman, R., Spekreijse, H. & Lamme, V. A. F. (2003) Large capacity storage of integrated objects before change blindness. *Vision Research* 43:149–64.

Laureys, S. (2005) The neural correlate of (un)awareness: Lessons from the vegetative state. *Trends in Cognitive Sciences* 9(2):556–59.

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

(2001) *Purple haze: The puzzle of consciousness.* Oxford University Press.

(2006) Conscious awareness and (self)representation. In: *Self-representational approaches to consciousness*, ed. U. Kriegel & K. Williford, pp. 173–97. MIT Press.

Llinás, R. R. (2001) *I of the vortex.* MIT Press.

Llinás, R. R., Ribary, U., Contreras, D. & Pedroarena, C. (1998) The neuronal basis for consciousness. *Philosophical Transactions of the Royal Society B* 353(1):841–49.

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

Long, G. (1980) Iconic memory: A review and critique of the study of short term visual storage. *Psychological Bulletin* 88:785–820.

(1985) The variety of visual persistence: Comments on Yeomas and Irwin. *Perception and Psychophysics* 38:381–85.

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

Lycan, W. (1996) *Consciousness and experience.* MIT Press.

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

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

McElree, B. (2006) Accessing recent events. In: *The psychology of learning and motivation, volume 46*, ed. B. H. Ross, pp. 155–201. Academic Press.

McGinn, C. (1991) *The problem of consciousness.* Oxford University Press.

McLaughlin, B. P. (1992) The rise and fall of British emergentism. In: *Emergence or reduction?: Prospects for nonreductive physicalism*, ed. A. Beckermann, H. Flohr & J. Kim, pp. 49–93. De Gruyter.

Merker, B. (2007) Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences* 30(1).

Metzinger, T. (2003) *Being no one.* MIT Press.

Miller, G. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63:81–97.

Minsky, M. (1986) *The society of mind.* Simon & Schuster.

Naccache, L. (2006) Is she conscious? *Science* 313(1)395–96.

Nagel, T. (1974) What is it like to be a bat? *The Philosophical Review* LXXXIII(4):435–50.

Nakayama, K. (1990) The iconic bottleneck and the tenuous link between early visual processing and perception. In: *Vision: coding and efficiency*, ed. C. Blakemore, pp. 411–22 . Cambridge University Press.

Nakayama, K., He, Z. J. & Shimojo, S. (1995) Visual surface representation: A critical link between lower-level and higher-level vision. In: *Visual cognition: An invitation to cognitive science, volume 2*, ed. S. M. Kosslyn & D. Osherson, pp. 1–70. MIT Press.

Noë, A. (2002) Is the visual world a grand illusion? *Journal Of Consciousness Studies* 9(5–6):1–12.

(2004) *Action in perception.* MIT Press.

(2005) What does change blindness teach us about consciousness? *Trends in Cognitive Sciences* 9(5):218.

Noë, A. & Thompson, E. (2004) Are there neural correlates of consciousness? *Journal Of Consciousness Studies* 11(1):3–28.

O'Craven, K. & Kanwisher, N. (2000) Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience* 12:1,013–23.

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

O'Regan, K. (1992) Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology* 46(3):461–88.

Owen, A. M., Coleman, M. R., Boly, M. & Davis, M. H. (2006) Detecting awareness in the vegetative state. *Science* 313:313.

In E. Craig (Ed.), Routledge Encyclopedia of Philosophy. London: Routledge. Retrieved April 22, 2007, from http://www.rep.routledge.com/article/V015SECT4

Papineau, D. (2002) *Thinking about consciousness.* Oxford University Press.

Pascual-Leone, A. & Walsh, V. (2001) Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science* 292:510–12.

Peirce, C. S. (1903) *Collected papers of Charles Sanders Peirce.* Harvard University Press.

Phillips, W. A. (1974) On the distinction between sensory storage and short-term visual memory. *Perception and Psychophysics* 16:283–90.

Piattelli-Palmarini, M. (1994) *Inevitable illusions: How mistakes of reason rule our minds.* Wiley.

Pins, D. & Ffytche, D. H. (2003) The neural correlates of conscious vision. *Cerebral Cortex* 13:461–74.

Pollen, D. A. (1999) On the neural correlates of visual perception. *Cerebral Cortex* 9:4–19.

(2003) Explicit neural representations, recursive neural networks and conscious visual perception. *Cerebral Cortex* 13:807–14.

(forthcoming) The fundamental requirement for primary visual Perception. Submitted to *Nature Reviews Neuroscience*

Potter, M. C. (1993) Very short-term conceptual memory. *Memory & Cognition* 21(2):151–61.

Prinz, J. (2000) A neurofunctional theory of consciousness. *Consciousness and Cognition* 9(2):243–59.

Putnam, H. (1975) The meaning of ‘meaning.’ *Minnesota Studies in the Philosophy of Science* 7:131–93.

(1981) *Reason, truth and history.* Cambridge University Press.

Pylylshyn, Z. (2003) *Seeing and visualizing.* MIT Press.

Qiu, J. (2006) Does it hurt? *Nature* 444(9):143–45.

Quine, W. V. (1969) Natural kinds. In: *Ontological reality and other essays*, ed. W. V. Quine, pp.114–39 . Columbia University Press.

Rees, G., Kreiman, G. & Koch, C. (2002a) Neural correlates of consciousness in humans. *Nature Reviews Neuroscience* 3:261–70.

Rees, G., Wojciulik., E., Clarke, K., Husain, M., Frith, C. & Driver, J. (2000) Unconscious activation of visual cortex in the damaged right hemisphere of a parietal patient with extinction. *Brain* 123(8):1,624–33.

Rees, G., Wojciulik., E., Clarke, K., Husain, M., Frith, C. & Driver, J. (2002b) Neural correlates of conscious and unconscious vision in parietal extinction. *Neurocase* 8:387–93.

Rensink, R., O'Regan, J. Kevin & Clark, James J. (1997) To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science* 8:368–73.

Rosenthal, D. (2005) *Consciousness and mind.* Oxford University Press.

Ross, W. D., ed. (1961) *Aristotle's De anima*, Clarendon Press.

Rousselet, G., Fabre-Thorpe, M. & Thorpe, S. (2002) Parallel processing in high-level visual scene categorization. *Nature Neuroscience* 5:629–30.

Searle, J. (1992) *The rediscovery of the mind.* MIT Press.

(2005) Consciousness: What we still don't know. *The New York Review of Books* 52(1) 36–39.

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

Shoemaker, S. (1981) Some varieties of functionalism. *Philosophical Topics* 12:93–119.

Siegel, S. (2006a) The contents of perception. In: The Stanford encyclopedia of philosophy, ed. E. N. Zalta,. Stanford Encyclopedia of Philosophy. http://plato.stanford.edu/archives/spr2006/entries/perception-contents/

(2006b) Which properties are represented in perception? In: Perceptual experience, ed. T. Gendler & J. Hawthorne, pp.481–503 . Oxford University Press.

Silvanto, J., Cowey, A., Lavie, N. & Walsh, V. (2005a) Striate cortex (V1) activity gates awareness of motion. *Nature Neuroscience* 8(2):143–44.

Silvanto, J., Lavie, N. & Walsh, V. (2005b) Double dissociation of V1 and V5/MT activity in visual awareness. *Cerebral Cortex*. 15: 1736–41

Simons, D. J. (1997) Change blindness. *Trends in Cognitive Sciences* 1:261–67.

Simons, D. J. & Ambinder, M. (2005) Change blindness: Theory and consequences. *Current Directions in Psychological Science* 14:44–48.

Simons, D. J. & Rensink, R. (2005a) Change blindness, representations and consciousness: Reply to Noe. *Trends in Cognitive Sciences* 9(5):219.

(2005b) Change blindness: Past, present and future. *Trends in Cognitive Sciences* 9(1):16–20.

Slater, R., Cantarella, A., Gallella, S., Worley, A., Boyd, S., Meek, J. & Firzgerald, M. (2006) Cortical pain responses in human infants. *The Journal of Neuroscience* 26(14):3,662–66.

Sligte, I. G., Lamme, V. A. F. & Scholte, H. S. (2006). *Capacity limits to awareness.* Paper presented at the Association for the Scientific Study of Consciousness, Oxford, U. K.

Smith, D. W. (1986) The structure of (self-)consciousness. *Topoi* 5:149–56.

Snodgrass, M. & Shevrin, H. (2006) Unconscious inhibition and facilitation at the objective detection threshold. *Cognition* 101(1):43–79.

Sosa, E. (2002) Privileged access. In: *Consciousness: New philosophical perspectives*, ed. Q. Smith & A. Jokic, pp. 273–94 . Oxford University Press.

Sperber, D. (2001) In defense of massive modularity. In: *Language, brain and cognitive development: Essays in honor of Jacques Mehler*, ed. E. Dupoux, pp. 47–57 . MIT Press.

Sperling, G. (1960) The information available in brief visual presentations. *Psychological Monographs* 74(498): whole issue.

Stoerig, P. (2001) The neuroanatomy of phenomenal vision: A psychological perspective. In: *Cajal and consciousness: Scientific approaches to consciousness on the centennial of Ramon y Cajal's Textura*, ed. P. C. Marijuan, pp. 176–94. Annals of the New York Academy of Sciences.

Strawson, G. (2003) What is the relation between an experience, the subject of the experience, and the content of the experience? *Philosophical Issues* 13(1): 279–315.

Supèr, H., Spekreijse, H. & Lamme, V. A. F. (2001a) A neural correlate of working memory in the monkey primary visual cortex. *Science* 293:120–24.

(2001b) Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). *Nature Neuroscience* 4(3):304–10.

Theoret, H., Kobayashi, M., Ganis, G., Di Capua, P. & Pascual-Leone, A. (2002) Repetitive transcranial magnetic stimulation of human area MT/V5 disrupts perception and storage of the motion aftereffect. *Neuropsychologia* 40(13):2,280–87.

Thomas, N. (1999) Are theories of imagery theories of imagination? *Cognitive Science* 23(2):207–45.

Tong, F., Nakayama, K., Vaughan, J. T. & Kanwisher, N. (1998) Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* 21(4):753–59.

Tononi, G. & Edelman, G. M. (1998) Consciousness and complexity. *Science* 282:1,846–51.

Treisman, A., Russell, R. & Green, J. (1975) Brief visual storage of shape and movement. In: *Attention and performance*, ed. P. Rabbitt & S. Dornic, pp. 699–721. Academic Press.

Tsao, D. Y., Tootell, R. B. H. & Livingstone, M. S. (2006) A cortical region consisting entirely of face-selective cells. *Science* 311(5,761):670–74.

Tse, P. U., Martinez-Conde, S., Schlegel, A. A. & Macknik, S. L. (2005) Visibility, visual awareness and visual masking of simple unattended targets are confined to areas in the occipital cortex beyond human V1/V2. *Proceedings of the National Academy of Sciences* 102:17,178–83.

Uebel, T. (2006) Vienna Circle. *The Stanford Encyclopedia of Philosophy Fall 2006 Edition,* ed. E. N. Zalta. http://plato.stanford.edu/archives/fall2006/entries/vienna-circle

Vogel, E. K., Woodman, G. F. & Luck, S. J. (2001) Storage of features, conjunctions and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance* 27:92–114.

Weiskrantz, L. (1997) *Consciousness, lost and found*, Oxford University Press.

Weiskrantz, L., Cowey, A. & Hodinott-Hill, I. (2002) Prime-sight in a blindsight subject. *Nature Neuroscience* 5(101–102):101–102.

Wilken, P. (2001) *Capacity limits for the detection and identification of change: Implications for models of visual short-term memory.* University of Melbourne.

Wolfe, J. (1999) Inattentional amnesia. In: *Fleeting memories*, ed. V. Coltheart, pp. . MIT Press.

Xu, Y. (2002) Encoding colour and shape from different parts of an object in visual short-term memory. *Perception and Psychophysics* 64(1):260–80.

Xu, Y. & Chun, M. M. (2006) Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature* 440(March 2):91–95.

Yang, W. (1999) *Lifetime of human visual sensory memory: Properties and neural substrate.* New York University.

Zeki, S. (2001) Localization and globalization in conscious vision. *Annual Review of Neuroscience* 24:57–86.

Zeki, S. & Bartels, A. (1999) Toward a theory of visual consciousness. *Consciousness and Cognition* 8:225–59.

Zeki, S. & Ffytche, D. H. (1998) The Ridoch syndrome: Insights into the neurobiology of conscious vision. *Brain* 121:25–45.

Zihl, J., von Cramon, D. & Mai, N. (1983) Selective disturbance of movement vision after bilateral brain damage. *Brain 106*: 313–340.

END NOTES

[i] See (Siegel 2006a, 2006b) for discussion of what kind of thing the content of a phenomenal state is.

[ii] Phenomenal persistence and persistence of accessible information should be distinguished from what Koch, 2004, ch 9 calls gist perception. We have specialized detectors for certain kinds of scenes, and in learning to read, we develop similar detectors for words. These detections can take place in 100 msecs and seem to require no attention. See (Potter 1993, Rousselet, Fabre-Thorpe, & Thorpe 2002)

[iii] The inattentional blindness view can be found in (Rensink 1997); (Simons 1997); (Noë 2004, J. K. O'Regan & Noe 2001). Views more closely related to the inattentional inaccessiblity view can be found in articles by philosophers – (Block 2001); (Cohen 2002); (Dretske 2004) – and by psychologists – (Simons & Rensink 2005a, 2005b, Wolfe 1999)

[iv] The "cerebral celebrity" view of consciousness is not the view in Dennett's Consciousness Explained (Dennett 1991), but was introduced a few years after that, I think first in Dennett (1993). I argued for a distinct notion of "access-consciousness" in Block (1990, 1992).

[v] I used this term in Block (2001) but I discovered years later that Nigel Thomas published pretty much the same idea first, deriving it from Marvin Minsky's "Immanence Illusion". Minsky's (1986, section 15.5) Immanence Illusion is this: "Whenever you can answer a question without a noticeable delay, it seems as though that answer were already active in your mind." At least in what I have read, Minsky does not focus on the idea that potential phenomenology is supposed to be confused with actual phenomenology. Thomas does focus on phenomenology, arguing for a view similar to that of O'Regan and Noë mentioned earlier: "The seeming immediate presence of the visual world to consciousness does not arise because we have built a detailed internal representation of it, rather it is (like the ever shining fridge light) a product of the "immanence illusion" (Minsky 1986). For the most part, the visual perceptual instruments ask and answer their questions so quickly and effortlessly that it seems as though all the answers are already, and contemporaneously, in our minds." (219) (Thomas 1999).

[vi]Noë (2002, 2004) suggests an even more pervasive form of such an illusion – that all experience is a matter of potentiality, but precisely because it is so pervasive, he does not regard the view as one that postulates an illusion. See (Cohen 2002).

[vii] The first classical "visual" cortical area is V1; later classic "visual" areas include V2, V3, V4, V5. V5 has two names, 'MT' and 'V5' because it was identified and named by two groups. I put "visual" in scare quotes because there is some debate as to whether some of the classic "visual" areas are best thought of as multimodal and spatial rather than visual per se. The motion area I am talking about in the text is actually a complex including MT/V5 and surrounding areas and is often referred to as hMT+. See (Kriegeskorte et al. 2003)

[viii] TMS delivers an electromagnetic jolt to brain areas when placed appropriately on the scalp. The effect is to disrupt organized signals but also to create a signal in a quiescent area. Thus TMS can both disrupt moving afterimages and create phosphenes. A comparison is to hitting a radio: the static caused might interrupt good reception going on but also cause a noise when there is no reception. (I am indebted here to Nancy Kanwisher and Vincent Walsh.)

[ix] To experiences phosphenes for yourself, close your eyes and exert pressure on your eye from the side with your finger. Or if you prefer not to put your eyeball at risk, look at the following website for an artist's rendition: http://www.reflectingskin.net/phosphenes.html

[x] TMS stimulation directed to V1 may also stimulate V2 (Pollen 2003). Perhaps V2 or other lower visual areas can substitute for V1 as the lower site in a recurrent loop. Blindsight patients who have had blindsight for many years can acquire some kinds of vision in their blind fields despite lacking V1 for those areas. One subject describes his experience as like a black thing moving on a black background (Zeki & ffytche 1998). Afterimages in the blind field have been reported (Weiskrantz, Cowey, & Hodinott-Hill 2002). Stoerig (2001) notes that blindsight patients are subject to visual hallucinations in their blind fields even immediately after the surgery removing parts of V1, however, that may be due to a high level of excitation that spreads to other higher cortical areas that have their own feedback loops to other areas of V1 or to other areas of early vision such as V2. See also (Pollen 1999)

[xi] In a different paradigm (de Fockert, Rees, Frith, & Lavie 2001), working memory load can increase the processing of distractors.

[xii] Earlier versions of this paper were presented at the 2005 NYU Colloquium in Language and Mind; Georgia State University; MIT; Harvard Medical School; Dan Pollen's course at Harvard College; Dartmouth College; NYU Medical Center; Aarhus University, Denmark; Stockholm University, Sweden,; Göteborg University, Sweden; Umeå University, Sweden; Indiana University, Brown University; the Pacific APA 2006; University of Western Ontario; Cal State LA; UCLA; the Association for the Scientific Study of Consciousness; the Institut Jean Nicod; the University of Toronto and Katalin Balog's class at Yale. I am grateful to Thomas Nagel for his role as chief inquisitor when this material was defended at the NYU Mind and Language Colloquium and for some of his suggestions; and I am grateful to Katalin Balog, Alex Byrne, Tyler Burge, Susan Carey, Max Coltheart, Leonard Katz, Sean Kelly, Victor Lamme, Mohan Matthen, Alva Noë, Dan Pollen, David Rosenthal and Susanna Siegel for comments on a previous draft.