Overflow, Access and Attention

Abstract

In this reply to 32 critics I start by clarifying the overflow argument. I explain why the distinction between generic and specific phenomenology is important and why we are justified in acknowledging specific phenomenology in the overflow experiments. Other issues discussed are the relations among report, cognitive access and attention, panpsychic disaster, the mesh between psychology and neuroscience and whether consciousness exists.

R1. Introduction

I have learned a great deal from reading the commentaries and I am gratified that so many respondents are sympathetic to separating phenomenal consciousness from cognitive access to it, a stark contrast to the responses to an earlier BBS paper (Block, 1995) in which I argued for similar though slightly stronger views. I don't flatter myself with the supposition that I have convinced anyone—the main factor is that a wide range of accumulating evidence increasingly supports separating phenomenal consciousness and cognitive access. (Evidence beyond what was described in the target article is mentioned in **Gopnik**, **Izard, Quinn & Most, Koch & Tsuchiya**, **Lamme**, **Landman & Sligte, Malach and Snodgrass & Lepisto**.)

The empirical core of my argument concerned evidence to the effect that the capacity of the phenomenal system is higher than the capacity of the cognitive access system that underlies reportability of phenomenal states, i.e. what I called "overflow". Many of the respondents (**Burge**, **Byrne**, **Hilbert** & **Siegel, Grush**, **Jacob**, **Kouider, Gardelle & Dupoux, Landman & Sligte, O'Regan & Myin, Naccache** & **Dehaene, Papineau, Spener, Van Gulick)** commented on that argument, and my discussion of their responses is in R2. There are two main issues, one concerning the distinction between generic and specific phenomenology (R2.1) and the other concerning hyperillusions (R2.2). If the overflow argument is correct, the cognitive system underlying reporting phenomenology is distinct from the system underlying phenomenology, a matter that raises the issue of the role of reports in an empirical investigation of consciousness. I argued that reports are the starting point but that they can be rejected on the basis of the method of "inference to the best explanation". **Naccache** & **Dehaene, Lau & Persaud, Papineau, Prinz, Sergent & Rees** criticize this reasoning. My response—in R3—makes use of contributions by **Koch & Tsuchiya, Malach** and **Snodgrass & Lepisto**. The overflow argument presupposes a view of the cognitive access system that drew many comments on both sides (**Balog, Clark, Gopnik, Harman, Kentridge, Lau & Persaud, Izard, Quinn & Most, Levine, Malach, Rosenthal, Shanahan & Baars, Snodgrass & Lepisto** and **Tye**) and is discussed in R4. R5 concerns the relation between consciousness and attention in relation to comments by **Lycan**, **Koch & Tsuchiya and Prinz.** I gave another empirical argument in addition to the overflow argument, an argument based on the claim that a theory that explains the mesh between psychology and neuroscience is to be preferred other things equal to a theory that doesn't, and this argument was discussed by **Burge, Grush, Lamme, Landman & Sligte, Hulme & Whitely, Prinz, Snodgrass & Lepisto** and **van Gulick**. I reply to them in R6. Finally, **McDermott** questions whether consciousness even exists, and that issue is discussed in R7.

R2: THE OVERFLOW ARGUMENT

R2.0 Introduction

I appealed to the Sperling, Landman and Sligte experiments in the overflow argument. In the Sperling experiment, subjects exposed to a short initial stimulus have the impression that they are aware of up to 32 specific letter-shapes but can report only about 3 or 4 letters. However, if a specific row is cued, subjects can report any 3 or 4 letters in that row, confirming their initial impression. Landman and Sligte get similar results using a paradigm in which subjects give a much more minimal response—as to whether a specific cued item has changed orientation--thus minimizing interference between the phenomenal representation and the subjects' own response. (See **Landman & Sligte**.) The overflow argument says that the capacity of phenomenology is much greater than the working memory system underlying report, so they must to some extent be distinct systems.

R2.1 Generic/Specific Phenomenology

The distinction between generic and specific phenomenology was crucial in the overflow argument although I did not give it a name or draw enough attention to it. I will explain it by example. In the Landman experiment, the relevant generic phenomenology would be the phenomenal presentation *that there is* a circle of rectangles. The relevant specific phenomenology would be a phenomenal presentation that specifies for each of the rectangles (or anyway, most of them) whether they are horizontal or vertical. For the Sperling experiment, the relevant generic/specific difference would be that between a phenomenal presentation *that there is* an array of alphanumeric characters and a phenomenal presentation of specific shapes of all or most items in the array. I argued that in the Sperling, Landman and Sligte experiments, there was specific phenomenology involving all or almost all of the items as well as generic phenomenology. This distinction figures—in those terms—in the commentaries by **Burge**, **Grush** and **Levine**. **Kouider, Gardelle & Dupoux** express it as the lower/higher distinction, **Papineau** as the scene/item distinction, **Sergent & Rees** as the scene gist/detail distinction, and in the target article as well as in **Byrne**, **Hilbert** & **Siegel**, **Naccache & Dehaene**, **Jacob, Spener** and **van Gulick,** the distinction is deployed without special terminology.

My argument was that before the cue, there is specific phenomenology for all or almost all items (and also generic phenomenology, but that does not figure in the argument). This specific phenomenology is what justifies the claim that the capacity of the phenomenal system is more than 4, whereas the capacity of the access system is 4 or less and thus that the two systems cannot completely coincide. A number of the critiques (most pointedly, **Papineau** and **Byrne**, **Hilbert** & **Siegel)** challenge the premise that there are more than 4 items of specific phenomenology before the cue. It is important to recognize that the objectors have to agree that before the cue, there are *specific* (not just generic) visual representations of all or almost all of the 8-32 items in the Sperling, Landman and Sligte experiments. There have to be such specific representations given that any location can be cued with high accuracy of response. The locus of controversy is *whether those specific representations are phenomenal*.

Here is my evidence for claiming that there is specific phenomenology for all or almost all the items in the overflow experiment?

(1) As **Burge** notes, subjects in overflow experiments (including me) often testify that their responses are based on specific phenomenology that was there all along. (Rogier Landman tells me that the extent to which subjects evince specific phenomenology may be correlated with how well they do in the experiments.) **Shanahan & Baars** say "It is worth noting that subjects in such experiments believe they are simply reporting the contents of their recent visual consciousness, even after the visual stimulus has disappeared from view". In the Sligte experiment, subjects have up to 4 seconds before the cue comes on, so their judgments about their phenomenology are not rushed.

(2) Subjects are attending to arrays in full view in good viewing conditions for half a second in the Landman, Sligte and some versions of the Sperling experiments, more than enough time for specific phenomenology. (**Burge** also makes this point.)

(3) If there is only generic phenomenology before the cue, and if the cue causes the generic phenomenology to be replaced by specific phenomenology, then there is a shift from generic to specific phenomenology. The fact that subjects report no such phenomenological shift might not be strong evidence against this view, but it is some evidence. The vast literature on this topic (including two PhD theses I have read) contains no mention of such a thing as far as I know. I can testify myself that even looking for such a shift, one does not experience it. The point of the example in the target paper of the rectangle coming into view as if from a distance was to compare that phenomenal shift with the absence of such a shift in the overflow experiments. If there are some items of specific phenomenology before the cue, and different items of specific phenomenology after the cue, then one might expect that to be noticed as well.

(4) There is evidence mentioned in the target article that cortical persistence obtains at all levels of the visual system and thus at the phenomenal level. In particular, there is evidence (mentioned in the target paper) that the persistence exists at levels where depth and motion are represented. As **Lamme** and **Landman & Sligte** note, persisting representations obtain at a stage of visual processing past figure-ground segregation and feature binding, properties that "are more associated with conscious processes." Thus there is a neural case for phenomenal persistence.

(5) In the target article, I mentioned Di Lollo's paradigm using a 5 by 5 grid in

which all but one of the squares is filled with a dot. Subjects see a partial grid with 12 of the dots filled in, then after a delay, another partial grid with a different 12 dots filled in. The subjects' task is to report which square has a missing dot, something they can do easily if they have a visual impression as of the whole matrix of dots. Loftus & Irwin (1998) show that subjects' ability to do the task correlates nearly perfectly with their phenomenological judgments of whether there appears to be a whole matrix rather than two partial matrices. When writing the target paper, I did not know about the variant by Brockmole, Wang & Irwin (2002) in which the appearance of the 2nd partial grid was delayed by as long as 5 seconds and in which subjects were told that a good strategy was to "imagine the dots still being present after they disappeared" (p. 317). The subjects' memory capacity for the 12 dots in the 1st grid can be computed by the type of errors made. When the delay between the 1st and 2nd partial grids is 100 msecs, the subjects' retention capacity falls from 12 to 4.1 of the 12 dots in the first partial grid. The striking result was that with delays over 100 msecs, subjects' capacity *increased,* asymptoting at a delay of about 1.5 seconds at which time their capacity was 10 of 12 dots, and the capacity stayed that high for delays up to 4-5 seconds. Independent estimates of the time to generate a mental image by Kosslyn (Kosslyn, Thompson, & Ganis, 2006) are between 1 and 2 seconds, and the authors argue that the subjects were following instructions, generating a visual image of the first array, and integrating that visual image with the percept of the 2nd array. This result constitutes converging evidence for high capacity specific phenomenology: since the subjects could do the task well, the Loftus & Irwin result suggests they have a visual impression of the whole matrix, and in any case visual imagery is phenomenal. So the representations are phenomenal, and the capacity of 12 dots is substantially more than 4 items. The upshot is that there is a completely different paradigm in which the evidence favors high capacity specific phenomenology.

(6) The evidence mentioned by **Kouider, Gardelle & Dupoux** which will be discussed below.

Kouider, Gardelle & Dupoux suggest what they take to be an *alternative* to what I am suggesting but what I take to be a *version* of it. They hypothesize that Sperling-like paradigms result from "partial awareness: subjects have a transient and degraded access to fragments of all the letters in the grid." Kouider & Dupoux take comfort from a common observation that subjects in experiments involving masked stimuli often report seeing bits and pieces of stimuli. What are Kouider & Dupoux saying about specific phenomenology? Zero specific phenomenology is not compatible with their view, since they say subjects are to some degree conscious of and have access to "fragments of all the letters in the grid". And they clearly don't envision full specific phenomenology, so they must envision *partial specific phenomenology*. Their evidence involves cases (Kouider & Dupoux, 2004) of brief masked presentations of color words and pseudo-color words (e.g. 'green' vs. 'gener') in which both have the same effect on subsequent identification of colored stimuli, facilitation in the case of congruent stimuli (e.g. 'green' or 'gener' followed by green) and inhibition in the case of incongruent stimuli ('green' or 'gener' followed by red). They predict, plausibly, that in

Sperling experiments that include some letter-like symbols that are not letters, subjects would treat false letters as similar to real letters. More generally, Kouider & Dupoux (2004; 2007) give evidence that genuine semantic priming requires at least partial phenomenology and that totally unconscious stimuli can only have non-semantic effects. (But see (Abrams & Grinspan, 2007).) To the extent that I disagree with Kouider, et. al., it is on just *how* degraded the specific phenomenology is. One of the experiments in Sligte, et. al. used bars that differed in increments of 45^{o}, and subjects still showed high capacities—slightly under 8 for stimuli of 16 bars with the cues presented at 1000 msecs after the stimulus offset. The specific phenomenology is good enough to make those distinctions between say 135^{o} and 180^{o} pretty well. In the Sperling experiment, the specific phenomenology was good enough for subjects to distinguish among the 26 letters of the alphabet. No doubt if there had been pseudo-letters, subjects would have made more errors, but the 26-way distinction is still respectable. It is worth noting that in the experiment by Kouider & Dupoux, stimuli were presented at 29 msecs or 43 msecs and also masked, an intervention aimed at making them harder to see. Landman's and Sligte's stimuli were normally unmasked and presented for 500 msecs (And Sperling's stimuli yield the same results with a 500 msec presentation—cf. also **Burge**), so the fragmentariness of the phenomenology in Kouider & Dupoux, 2004 could be predicted to be greater than in the overflow experiments. The upshot is that Kouider, Gardelle & Dupoux have presented further evidence for high capacity specific phenomenology, just what my overflow argument relies on.

Byrne, **Hilbert** & **Siegel** are not impressed with the first of my reasons. They raise a plausible objection, that the generic/specific distinction is pretty abstruse, so how can the responses of subjects who don't know the distinction provide support for specific phenomenology? However, when subjects say (cf. **Shanahan & Baars**) that in reporting the letters in the cued row or in telling whether the cued rectangle changed orientation that they are simply reading their answers off of the visual impression that was in existence before the cue, they are evincing specific phenomenology whether or not they could state the distinction between specific and generic phenomenology.

Readers may be thinking that subjects may have seen the orientations/identities of a few of the items, but that subjects cannot be sure that they saw all or most. Recall that *specific representations of all or almost all the items before the cue (though perhaps fragmentary representations) have to be postulated* to explain the fact that subjects can report the items no matter which row is cued. So the options would appear to be that there was no specific phenomenology before the cue or that there was specific phenomenology involving all or most of the items, even if fragmentarily. As I just mentioned, subjects' testimony (and lack of surprise in what they can do) suggest the latter.

Papineau hypothesizes that the presence of generic phenomenology is my basic reason for postulating specific phenomenology. I should have drawn more attention to my reasons as I have now done.

Van Gulick notes that the "movie screen of the mind" view would say that you can't have generic phenomenology without specific phenomenology, implicitly suggesting that I am relying on the "movie screen of the mind" view plus the fact of generic phenomenology to argue for specific phenomenology. (**Grush** uses what would appear to be the movie screen view of pictorial representation but the other way around from what Van Gulick ascribes to me: he says generic phenomenology cannot be an image because if it were an image, clear contents would have to present the details.) I reject the principle--applied by both Van Gulick and Grush--that pictorial representation has to specify the relevant details. I call this principle the "photographic fallacy" (Block, 1983). More specifically, the photographic fallacy supposes that pictorial representations have to represent details of anything in view in the manner of a prototypical photograph. To see the fallacy, note that an impressionist painter might represent a hand in broad brush strokes that do not explicitly represent the number of fingers or whether one of them has a ring.

Van Gulick also argues that phenomenology in the overflow cases may be partial, weak or somewhat indeterminate, but that access may also be limited in the same way, so there is no evidence for a discrepancy. With regard to the issue of whether degree of phenomenology is matched by degree of access, there is experimental evidence to the contrary mentioned by **Sergent & Rees** (Sergent & Dehaene, 2004). Cognitive access appears to be more of a binary phenomenon than Van Gulick supposes.

Spener argues that the fine details of introspective grasp of specific phenomenology are thoroughly expectation-driven. In support, she mentions the disagreements in the philosophical literature about whether there is a determinate number of phenomenologically represented speckles on the speckled hen or not.

I think **Spener** exaggerates the problem. Phenomenological disagreement can be to some extent settled. The speckled hen case in particular has been illuminated by work on the different "grain" of vision as opposed to attention. Patrick Cavanagh and colleagues have shown that the grain of vision is about a 60^{th} of a degree at the fovea. For example, in order for a grating to be distinguishable from a gray field, the individual lines have to subtend more than a 60^{th} of a degree. However, in order to *attend* to visible lines and to be able to *move* attention from one to another, they must subtend at least 5-10 60ths of a degree. How is this relevant to representations of speckles? Plausibly the phenomenological disagreement about represented speckles stems from conflating seeing with attending. The speckles in the standard example in which they subtend more than 1 60^{th} of a degree but less than 5-10 60ths of a degree, are visible but not attendable, so that one can see them (and thus phenomenally represent them) but not count them or do anything else that requires moving one's attention from one to another. If one is looking right at the speckles for sufficient time, the phenomenology really does determinately represent many of the individual speckles--specific phenomenology as opposed to generic phenomenology—but the speckles have an "elusive" quality because one cannot attend to them.

Grush argues for an illusion in which generic phenomenology presents an object as affording answers to certain queries, so when the answers to the queries are filled in, the subject does not notice. One might call this putative illusion the affordance illusion. (Grush gives it a less mnemonic name.) The affordance illusion is supposed to explain change-"blindness" and inattentional "blindness" in terms of sparse generic representations. Of course the same idea, if it worked, could be used to undermine my reason 3 for specific phenomenology. The one argument for this illusion that Grush offers is that if there were an affordance illusion, it would serve to explain change "blindness" in terms of sparse representations. But as I noted in the target article, the Landman and Sligte experiments are themselves evidence against the sparse representations view of change "blindness".

Jacob raises the issue of whether what I say in the target article is compatible with what I have said against representationism. I say yes because the distinction between generic and specific phenomenology requires only that in the cases in question, phenomenology *has* representational content of the appropriate kinds, and that is compatible with the view (that I hold) that there is more to phenomenology than representational content.

R2.2 Unconscious representation before the cue.

So far, I have been talking about the objection that before the cue, there is only generic phenomenology, and no specific phenomenology. However, there is a more radical view that says that until the cue appears, there is no phenomenological representation of the array at all, either generic or specific. Dehaene was arguably committed to this view by the refrigerator light illusion hypothesis as emphasized by **Burge,** but **Naccache** & **Dehaene** advocate the less radical thesis. The only advocate of the more radical thesis among the commentators is **O'Regan** & **Myin**. My main response is that the arguments I have given for specific phenomenology before the cue constitute a case for *some* phenomenology before the cue.

O'Regan & **Myin** note, correctly, that I am assuming that consciousness is a natural kind and has some kind of neural signature. They claim this hypothesis is speculative and that even if true, it would not solve the "hard problem". One of the points of my paper is that we can discover a lot about the neural basis of consciousness short of solving the hard problem. On the issue of speculativeness: I am not assuming that there is something physical in common to all possible cases of consciousness—e.g. including mammals, birds, octopi, conscious machines and conscious extra-terrestrials if there are any. My assumption is that there is a neural signature of consciousness in humans that is shared at least by other mammals with similar sensory systems. This is an assumption that I believe is shared by the field and looks promising so far. O'Regan & Myin note that I claim that phenomenal consciousness has effects on the basis of which we can find evidence about its nature. They interpret my (1995) as arguing for the opposite view, epiphenomenalism. Although Block (1995) took epiphenomenalism seriously, the upshot was that we have reason to reject it.

R 2.3 Hyperillusions

Naccache & Dehaene appear to agree with me (and disagree with

Byrne, **Hilbert** & **Siegel**) that subjects think they have specific phenomenology of all or almost all the items in the overflow experiments, but Naccache & Dehaene think subjects are suffering from what I called a hyper-illusion. In ordinary illusions, appearance misleads about reality, but in hyper-illusions, appearance of appearance misleads about appearance. (This may sound glib and meaningless and perhaps it is, but I think it is fair to describe a putative phenomenon in a way that brings out its peculiarity.) Naccache & Dehaene argue that "we all have the illusion of seeing a world in full color although color-sensitive cones are absent in the periphery of our retina." And **van Gulick** makes the same claim, citing Dennett's (1991) color marker demonstration: we hold a colored marker to the side of our visual field, bringing it gradually into the fovea, where we supposedly begin to see its color. However, the claim of illusion on the basis of retinal distribution of cones is analogous to (though not quite as obviously wrong as, for reasons I will get to) the idea that there is an illusion involved in seeing the world right side up given that our retina contains upside down images. What is on the retina is relevant to phenomenal experience *only to the extent that it affects the parts of the brain that determine phenomenology*. Assuming for the sake of the example that activity in V4 is the neural basis of color phenomenology, then the issue of whether the world genuinely appears colored in the periphery would be illuminated by the question of whether V4 has receptive fields in the periphery (that is, V4 contains groups of neurons that respond to and represent areas in the periphery) and not what is on the retina.

Where the two putative illusions differ is that there is a genuine issue of whether V4 does have receptive fields in the periphery and if so, why, given low color sensitivity in the retinal periphery? An obvious answer is: because visual phenomenology depends on integrating information over time. Our visual representations comprising the whole visual field are not built up in an instant! So color representation of something now in the periphery can derive from its representation recently in more central vision. Striking evidence for integration over time in some aspects of perception is provided by evidence that multicolor representations are built up serially—the system processes one color at a time—although locations and shapes are processed in parallel (Huang & Pashler, 2007; Huang, Treisman, & Pashler, 2007). There is another response to the question why cortical color representations would represent color in the periphery: it turns out that hue discrimination at 50^o eccentricity is *as good as in the fovea* (which subtends only 2^o—about twice the width of the thumbnail at arms length) if the size of the stimulus is magnified, and there is even some color sensitivity out to 80^o-90^o (Mullen, 1992).

Naccache & Dehaene mention another alleged hyper-illusion: the "moving window" experiment by McConkie in which subjects suppose there is a full page of normal text even though the text outside the small moving window is degraded or changed. The point just made about representations integrated over time applies here too.

R 2.4 Phenomenal Memory

Jacob notes that my talk of phenomenal memory is misleading. What I meant to be saying was that there is neural persistence at all levels of the visual system including the level that determines phenomenology. So instead of talking about phenomenal memory I should have just talked about phenomenal persistence. **Byrne**, **Hilbert** and **Siegel** say I don't need high capacity phenomenal persistence but only high capacity phenomenology. The role of phenomenal persistence in my paper was in connection with the "looking again" dialectic that I presented. The argument for the high capacity of the phenomenal system can be undermined by the idea that the high capacity derives from the information in the environment or the high capacity of low level vision, e.g. the retina. It would be hard to rule those alternatives out if not for the phenomenon of phenomenal persistence, which allows one to track where the phenomenal persistence is coming from. Recall that I argued that phenomenal persistence is at a level that involves binocularity (Engel), motion (Triesman), is not disturbed by a light mask that disturbs positive afterimages (Sligte) and is disturbed by a pattern mask that does not disturb retinal afterimages (Sligte). All these data provide reason for thinking that the phenomenal persistence is due to persistence in the underlying basis of phenomenology that is not being driven by earlier persistence. For these reasons, iconic memory is, as **Landman & Sligte** note, a window into phenomenal consciousness.

R3 REPORTS

The overflow argument entails that the cognitive system underlying reporting differs from the phenomenal system, raising the issue of the extent to which reports are pipelines to phenomenology.

R3.1 Are Reports Privileged?

Naccache & **Dehaene** pose a dilemma for me: "If one wants to define phenomenal consciousness differently from conscious reportability, then one should resist the temptation to make use of subjects' reports" as evidence for phenomenal consciousness." However, no one would think there is such a dilemma for, say acidity or heat. An acid is as a proton donor and heat as molecular kinetic energy. These are good scientific definitions but no one thinks that these definitions preclude any kind of evidence. Naccache & Dehaene see inconsistency looming: how can I rely on reports in the Sperling experiment while at the same time claiming that subjects' reports that they don't see something can be wrong? **Prinz** maintains that reports are "authoritative". My view is that reports are the *starting point* for building a theory of phenomenal consciousness, but can be rejected if the best explanatory theory requires it. The arguments for this view include the methodological points about inference to the best explanation, the sketch of an actual explanatory account on which reports can be wrong, and the points made forcefully by **Snodgrass & Lepisto** (and in Block, 2005) involving signal detection theory.

It is obvious that reports fail to be authoritative in that we can have conclusive evidence against the truth of introspective reports. As **Koch & Tsuchiya** note, in Anton's syndrome, subjects are blind but think and report that they see. More generally, anosognosics deny their perceptual and motor disabilities, making all sorts of false reports about their own experience. Introspective reports do have a certain priority: we have no choice but to start with reports in investigating consciousness. I am in complete agreement with Koch & Tsuchiya when they say "in the absence of compelling empirical evidence to the contrary…if the subject denies any phenomenal experience, this should be accepted…" One very notable form of empirical evidence that can conflict with report is evidence about subjects' decision process evaluated according to signal detection theory. As **Snodgrass & Lepisto** note, "contrary to many researchers' implicit assumptions, there is no such thing as an unmediated "subjective report"—ever."

In practice, **Naccache & Dehaene** use the methodology that I advocate, not the one that they themselves advocate. They concede that reports can be partial and biased and argue that "reports of a rich phenomenality cannot be taken at face value", citing the moving window paradigm and the claim that we have the illusion of seeing a world in full color. Further, both Naccache & Dehaene and **Sergent** & **Rees** suppose that reports have to be measured using high tech machinery, leading to a concern about what they could mean by 'report'. Is a report just (1) *any* evidence of consciousness? Or is it (2) evidence of consciousness from a subject's performing one voluntary action rather than another? Or is it (3) evidence of global broadcasting? If (1), their focus on report as the gold standard for evidence of consciousness is trivial since it just means that evidence is the gold standard of evidence. If (3), their claim that the evidence for consciousness is always evidence for global broadcasting is trivial, amounting to the claim that evidence for global broadcasting is evidence for global broadcasting. If (2), the claim that reports are the gold standard is substantive but probably false. As **Malach** (Hasson, Nir, Levy, Fuhrmann, & Malach, 2004) showed, you can get evidence about consciousness from people just watching a movie in a scanner, and not making any voluntary response. Another such case mentioned by **Sergent & Rees**, Lumer & Rees (1999) put subjects in a binocular rivalry experiment without requiring any voluntary response and found alternation between neural activations, providing evidence of the neural bases of different perceptual experiences without a voluntary response. The advantage of sticking to normal uses of words, is that we don't have to *guess* what people mean.

R 3.2 Panpsychic Disaster

A number of commentators argued that once you give up the special authority of reports, you will have no way of avoiding attributing consciousness to lampposts. **Papineau** notes that I regard some states as uncontroversially unconscious and wonders "what makes a state 'uncontroversially unconscious' if it is not that subjects tell us so." He argues that once we allow that a state can be conscious even though normal subjects systematically deny it, there may be no uncontroversially unconscious states. **Prinz** says "Block must either concede that reports are authoritative or deny that we can rule out the possibility of conscious states in V1, the LGN, and the retinae." **Lau** & **Persaud** say that the methodology I rely on to support recurrent loops as necessary assumes that subjects are to be believed when they report that they don't see something. But given that I allow for phenomenology without access, they wonder, how can I rule out that these subjects have phenomenology despite reporting that they don't? **Sergent & Rees** and **Naccache & Dehaene** would no doubt agree. But these critiques ignore the obvious failure of authoritativeness of reports just mentioned and the power of inference to the best explanation to sort the good from the bad attributions of consciousness.

Perhaps what **Naccache & Dehaene** and **Sergent & Rees** are supposing is that if biomarkers of consciousness conflict with actual verbal report, verbal report wins. But there is no evidence that *always wins*. If items of evidence conflict, the right scientific procedure is to find an explanation of the conflict, if need be doing further experiments to see which item of evidence is misleading.

R4 ACCESS

Thus far, I have talked about the overflow argument (R2) and the methodology of reports (R3). I now turn to issues concerning the cognitive accessibility system.

R 4.1 Awareness, Access And Agency

Levine notes that the idea of a phenomenal consciousness that has nothing to do with any kind of access-relation to the subject doesn't really seem like any kind of consciousness at all. A major difference between my position in the target article and Block (1995) is that in the target article I acknowledge this point and accommodate it within a framework that avoids any constitutive connection between that access-relation (which I called Awareness in the target article and Levine calls awareness-access) and *cognitive* access, which I identify in the target article with broadcasting in the global workspace. **Jacob** wonders why I have changed my mind, noting that the kind of awareness I now say is a necessary part of conscious experience is not full-fledged self awareness of the kind a rational thinking creature might sometimes have but that a mouse will presumably not have. In 1995, the only option I saw for explaining awareness-access in non-cognitive terms was as a kind of phenomenal property I called "me-ishness". But now I see that awareness-access can be adequately understood in terms of "same-order" and deflationary theories, so there is no need for cognitive or other "higher-order" accounts.

Levine makes a plausible case that the resistance that many feel to the idea that the machinery of phenomenal consciousness is separable from the machinery underlying report (i.e. broadcasting in the global workspace) stems from conflating broadcasting-access with awareness-access, and he mentions two interesting suggestions for ways that the relation between awareness-access and the self could be further elucidated. One of them is that there is no reason why subjectivity cannot involve a fragmented subject. Yes, but the subject cannot be *so* fragmented that the experience is not for-the-subject. If GK has the face experience on the left that he denies, what keeps him from acknowledging it is his inability to attend to the left side of space. But that is compatible with the face experience on the left being part of the *same visual field* as whatever he is seeing on the right. (I use this term as is standard in philosophy to mean the entirety of visual space, including left and right.) That is, the relations in the visual field would be normal and have normal awareness-access, but the subject would have poor broadcast-access to them. So we can make sense of the idea that experience of the face on the left is an experience *for-him* despite some fragmentation of the self.

Levine's hypothesis illuminates **Rosenthal's** critique, in which he asks: if phenomenology necessarily involves awareness as I say it does, and given that awareness normally results in cognitive access, why would cognitive access fail in the GK case? Rosenthal is assimilating or at least supposing too tight a connection between awareness-access and broadcasting access. GK has awareness-access on the left if he has phenomenology on the left, but it is his lack of broadcasting-access that explains why he can't report what is on the left. Rosenthal claims that higher-order theories of consciousness explain why ordinary phenomenology always involves some awareness of it, whereas the global-workspace theories I favor as theories of access do not. But the global workspace theories are theories of *cognitive* access, not theories of *awareness*-access. Again, **Levine's** hypothesis is confirmed. The competition to higher-order theories as theories of awareness-access is the same-order theory and deflationary theory not the global workspace theory.

Rosenthal suggests that infant pain might have "weak" phenomenology and that weak phenomenology may be what I am supposing GK's face experience has too. I am not sure what "weak" phenomenology is supposed to be, but I guess it is supposed to be the phenomenology you get when there is no global broadcasting and no higher order thought. There is no reason to suppose that infant pain or GK's face experience is any less *intense* than adult pain (cf **Gopnik** and **Izard, Quinn & Most**). **Gopnik** adds another layer of evidence to what I presented in the target article about infant phenomenology and **Malach** adds another layer of evidence relevant both to infant and adult phenomenology. Endogenous attention matures late, making it even less likely than I said that infants are capable of higher order thought. Inhibitory neurotransmitters, a major contributor to unconscious states mature late, making it less likely that infants' perceptual states are unconscious. Younger infants are slower to habituate, another source of unconscious perceptual states (although this may be because younger infants are slower to encode stimuli rather than anything to do with the machinery of habituation). Evidence against *both* higher order thought *and* unconscious states puts Rosenthal in a vise, since those exhaust his options. Izard, Quinn & Most also give suggestive evidence that emotion and action systems reveal phenomenal states that may not involve either global broadcasting or higher order thought. The upshot is to increase the empirical squeeze on higher order thought theories of consciousness, thereby increasing the attractiveness of same-order theories.

Rosenthal says standard same order views are just as cognitive as HOT. Rather than argue about the texts he cites, let me just say that Caston's influential paper on Aristotle's same order theory (Caston, 2002) emphasizes some of the advantages of the same-order view over higher order theories that I mentioned in the target article, namely that there is no need to postulate that whenever there is a token conscious pain, there is also a token thought about it, and that there is no puzzle about why my thought about a teacup or your pain does not make the teacup or the pain conscious, whereas a higher order thought about my own pain does. (See also **Malach's** Figure B which depicts a way of thinking about the same-order account.)

Astonishingly, **Rosenthal** claims in regard to infant pain that "because nonconscious pains have largely the same causal connections to behavior and to a strong desire that they cease, they are just about as bad." It is well known that pain in infants has bad effects, but I have never heard of any experimental evidence to the effect that those bad effects do not depend on whether the pain is conscious or not, nor does Rosenthal cite any. Newborns who are circumcised without anesthesia or analgesia are more stressed by vaccination even 6 months later (Taddio, Goldbach, Ipp, Stevens, & Koren, 1995). But this evidence does not approach the issue of whether the infant pain is conscious. I imagine that Rosenthal is concluding from the evidence that higher order thought is unlikely in infants and the evidence that pain in infants has bad effects to the conclusion that unconscious pain in infants has bad effects. But this response suggests that Rosenthal is retreating from a substantive empirical claim to a trivial linguistic claim. Rosenthal can if he likes simply use 'unconscious' to mean something on the order of 'unaccompanied by higher order thought'. But pains that are unconscious in that trivial sense cannot be supposed to be pains there is nothing it is like to have. The danger of promoting such a triviality into a substantive thesis can be seen in Peter Carruthers' infamous claim (Carruthers, 1989, 1992) that given that pains in dogs, cats, sheep, cattle, pigs, and chickens are not available to be thought about, these pains are not felt and hence are not appropriate objects of sympathy or concern and are of no moral significance. (Carruthers (1999) backpedals, not on the ground that animal pains are appropriate objects of concern or sympathy, but on the ground that the frustration of animal desires are of moral significance.) Drug testers and slaughterhouses take notice! Given the evidence presented in the target paper and in **Gopnik** and in **Izard, Quinn & Most**, I really don't see how advocates of this unattractive view can avoid applying it to human infants.

Snodgrass & Lepisto give a very plausible argument concerning Jacoby style "exclusion" experiments that they involve phenomenally conscious stimuli or parts of stimuli which the subject has little confidence in having seen and hence no higher order belief in having seen it. (There was also an argument to this effect in Block (2001).) For example, the subject is instructed to complete the stem 'rea__' with an ending that is not a word that might have just been flashed briefly. If the word is 'reason' and the subject saw all or part of it, but has low confidence, the subject is more likely to complete 'rea___' with 'son' than if no word at all was flashed. As noted in (Snodgrass, 2002), subjects who are penalized for errors do better at excluding, suggesting on a signal detection analysis that they really did see the word (or parts of it as suggested in **Kouider, Gardelle & Dupoux**) but had low confidence and no higher order state that would lead to a report. This provides another piece of empirical evidence against the higher order point of view. Perhaps the advocates of the higher order point of view will take the degree of confidence to be an index of degree of consciousness, but the signal detection models show that consciously seeing the stimulus and confidence that one has seen it can vary independently, each being influenced by a variety of different factors. (For example, changing the catch rate can influence confidence without influencing perception. See (Supèr, Spekreijse, & Lamme, 2001) and (Block, 2005).)

Clark argues for a view opposed to **Rosenthal** and to **Levine** and me, that for a representation to be phenomenal, it must be "available for use in the planning and selection of deliberate, stored knowledge-exploiting, goal-reflecting and goal-responsive actions." He makes a cogent case, but it doesn't involve any empirical evidence. Rather, it is an a priori conceptual appeal. "To count as a conscious experience, an informational state must" be available for "rational choices and considered actions". One and the same condition of the skin, intrinsically described, will *count* as sunburn if it is caused by the sun but not if it is caused by a fire. Similarly, according to Clark's view, one and the same neural-informational state, intrinsically described, can count as a conscious experience if it is properly connected to rational choices and considered actions and not if it isn't. But our ordinary concept of consciousness is too vague to support this kind of claim. Further, one reasonable precisification of our ordinary concept of consciousness will make it the concept of a natural kind. The key feature of that natural kind-property is the way it feels to have it. My own view is that there is an ordinary concept of *phenomenal* consciousness. **Gopnik** gives evidence that infant consciousness may "not be accessible for goal-directed planning nor be the subject of the endogenous attention that accompanies such planning." I don't give this as *evidence* against Clark's view since conceptual claims are not subject to evidence. Instead, if Gopnik is right, Clark's conceptual claim would dictate that our concept of consciousness does not apply to infants. But the absurdity of this step strongly indicates that our concept of consciousness—or at least one of our concepts of consciousness--is the concept of a phenomenal feel that has no conceptually necessary relation to agency of the sort **Clark** describes.

Malach presents exciting evidence that intense experience such as watching an engrossing movie (a spaghetti western) activates an "extrinsic" sensory system in the same ways in different people; but activates an "intrinsic" system based in frontal areas (Hasson et al., 2004) less strongly and not in the same ways in different people, Malach's intrinsic system overlaps considerably with a "default' system that is active when the subject is "doing nothing" and that is inactive when performing intense goal-directed tasks (Raichle, MacLeod, Snyder, Powers, & Gusnard, 2001). In one study, Malach and colleagues (Goldberg, Harel, & Malach, 2006) used the same stimuli (pictures and audio clips) in an introspective task and a difficult rapid categorization task but with different instructions. In the introspective task, subjects were asked to categorize their emotional reaction as positive, negative or neutral. In the categorization task, the stimuli were presented at triple the rate and the subjects were asked to categorize the stimuli, e.g. as animals or not. Subjects also rated their own self-awareness and their self-awareness was high in the introspective task and low in the rapid categorization task. Part of the intrinsic system—see Malach's Figure A-- was activated both in the introspective task and other "self" relevant tasks, including judging the application of emotionally neutral words to oneself. This area is suppressed in the rapid categorization task. The upshot is that the intrinsic system is less active when the subject is doing a demanding perceptual task, or watching Clint Eastwood, and that a part of it that is especially tuned to self-relevant tasks is especially suppressed in difficult perceptual tasks. This study presents a challenge to views such as **Lycan's**, **Harman's**, **Rosenthal's** and **Van Gulick's** that emphasize the importance of the self in phenomenal experience because it provides evidence that the self appears to be lost in intense perceptual activity.

R4.2 Fractionation Of Access

As I have already mentioned in commenting on **Clark** and **Rosenthal**, the evidence presented in **Gopnik** and **Izard, Quinn & Most** is highly relevant to the issues of higher order thought and agency that connect with consciousness. However, I want to make a few more remarks about developmental evidence.

1. The main methodological point made by Gopnik and Izard, Quinn & Most is well taken. Adult capacities fractionate in infants and by studying that fractionation and the process by which the distinct capacities and abilities are integrated, we can learn a lot about adult consciousness and cognition.
2. Much of the behavior that seems at first glance to be a reflection of cognitive access in infants may bypass global broadcasting altogether. One example from adults mentioned by Izard, Quinn & Most serves to illustrate the point: dorsal visual system guidance of action as exemplified in Goodale and Milner's famous patient DF who can post a letter in a variable angled slot about as well as normal persons and avoid obstacles, but does not have the globally broadcast visual contents that are created by the ventral visual system. DF only knows about the angle of the slot by noticing her own actions (Milner & Goodale, 2007). The data presented by Gopnik and Izard, Quinn & Most might make one wonder about the extent to which infants even have global broadcasting. But young infants—at least by 10 months--do show working memory systems that function in much the way those of adults do, albeit with a slightly lower capacity (Feigenson & Carey, 2003, 2005; Feigenson, Carey, & Hauser, 2002).
3. I agree with Gopnik's cluster of points having to do with the preponderance of exogenous (stimulated from outside) over endogenous (internally directed) attention in infants. However, I am not sure that this fact suggests that infants have a lantern of consciousness instead of a spotlight. Why can't exogenous attention focus the spotlight?
4. Although I agree in general with Gopnik's point that infant cognition is not geared towards goal directed planning in the early months, infants are

certainly capable of goal-directed action by 5 months, and there is evidence of some capacity for goal directed action (and perception of it) at 3 months (Sommerville, Woodward, & Needham, 2005).

R 4.3 Kinds Of Access

Representationism is the view that the phenomenology of an experience is the experience's representational (intentional) content. **Harman**, **Jacob**, **Lycan** and **Tye** are advocates of this view, I am an opponent. However, I am willing to allow that every phenomenological state has representational content, even that the phenomenology consists—in part—in its having that representational content. I think that phenomenology goes beyond representational content, but this opposition played no role in the target article and will play no role in this reply. With this bone of contention cleared away, Harman's argument can be streamlined.

Harman argues that clearing up some unclarities renders the claim that phenomenology does not require cognitive access non-scientifically true or non-scientifically false. In my weight-lifting this morning, I experienced "flow". Among my visual experiences was seeing the barbell. Is the cognitive access in question to the experience of seeing the barbell or to the barbell itself? If the cognitive access in question is to the experience of seeing the barbell, then we know from non-scientific observations of flow (Csikszentmihalyi, 1975) that cognitive access to the experience itself destroys the experience, or at least to its flow aspect. However, cognitive access to the barbell itself is guaranteed by the experience being *of* the barbell. In neither case, Harman concludes, is there any need for science.

Recall that my paper accepts the **Shanahan & Baars**/**Naccache & Dehaene** idea of access as global broadcasting. What is globally broadcast is a representation with both representational and phenomenal content. What makes the access cognitive is that the consuming systems in terms of which global broadcasting is defined include the cognitive machinery of reasoning, planning, problem solving, categorization, memory and the like. **Harman** assumes that what makes access cognitive is just that the accessed state is about something. But a frog might have a perceptual state that is about a fly without engaging any cognitive machinery. Let's go back to the first step of Harman's argument, that access has to be to the experience or to the barbell. In the framework that I was using, access is neither to the experience nor the barbell but to the content of the state since that is what is broadcast. So Harman's dilemma leaves out the crucial case. It may seem as if the issue between me and Harman is verbal—what is meant by 'cognitive' and 'access'. However, I chose the global broadcasting framework because it seems the best empirical framework for thinking about access and accessibility. Within that framework, the issue of phenomenology without access becomes an empirical issue. So, contrary to Harman, the issue is doubly scientific. Scientific considerations go into choosing the terms of the debate, and then once they are chosen, scientific considerations determine the answer.

Balog, like **Harman**, is concerned with the question "Access to what?"

She mentions, in effect, three different options: (1) global broadcasting of the nonconceptual contents of perception or (2) of conceptual contents of perception or (3) Global broadcasting of introspective conceptual contents that are concerned with the phenomenal character of the experience itself and that involves phenomenal concepts, where a phenomenal concept is a concept whose tokening itself involves phenomenology. (See papers in (Alter & Walter, 2006)) Balog argues that my interpretation of what subjects in the overflow experiments say "crucially relies on the subjects' introspective report of the phenomenality of their entire visual experience, including those aspects of the experience whose content is not access conscious" (in sense 2), and this motivates her to suggest a sense of 'access'--sense 3–in which there is no gap between phenomenality and access.

Balog is responding to a problem that comes up repeatedly in the commentaries (especially **Naccache & Dehaene** and **Sergent & Rees**): how can I use reports to demonstrate inaccessible (including non-reportable) phenomenology? My reply is that in some cases, one infers unreportable or unreported phenomenal states via theoretical considerations. In the overflow experiments, the situation is this: generic contents are reportable, and one of those generic contents is that one has had specific phenomenology. That generic content need not deploy any phenomenal concepts in the sense of a concept whose tokening itself involves phenomenology. So there is no need to appeal to phenomenal concepts.

Balog says my view of access is given by 2—that I define it in terms of conceptual contents. There is some justification for that in my (1995, see especially footnote 11), which was written before I had adopted the global broadcasting model of access. My model of access in 1995 was dominated by inference and other reasoning, and it is often taken as part of the definition of nonconceptual content that it can play no role in inference. However, once I had adopted the global broadcasting model (Block, 2002) the issue of whether the broadcast contents were conceptual or non-conceptual no longer looms large. I don't see why broadcast contents cannot be a mix.

Tye notes that in one sense of 'accessible', phenomenology does not overflow accessibility. The sense he has in mind is that for each letter that the subject sees, the subject can demonstrate it with, e.g. "that letter". He thinks that this point deprives my claim that phenomenology overflows accessibility from its "bite". Recall that I distinguished between wide and narrow accessibility. In my terms, what **Tye** has done is to provide a another wide sense of 'accessibility'—demonstration as opposed to identification as in Sperling or comparison as in Landman and Sligte. Demonstration, identification and comparison are all different and have different cognitive demands. Demonstration is in a way more demanding since (see **Koch & Tsuchiya**), identification can take place without attention or with limited attention, whereas demonstration requires attention. As I noted in the discussion of **Spener**, the grain of vision is finer than the grain of attention, so you can see more than you can attend to. But whether or not Tye's version of wide accessibility is wide enough for his purposes, my point is about something different. The overflow point in the target article was this: the

Sperling, Landman and Sligte experiments show that the phenomenal system is at least in part distinct from the global workspace because the capacity of the phenomenal system is greater than that of the global workspace. This point can be defended without being concerned with different senses of 'accessibility', so I do not agree that providing another wide sense of accessibility affects the bite.

R4.4 Phenomenology And Working Memory

Lau & **Persaud** suppose that the apparent difference in capacity between phenomenology and working memory is just a consequence of deterioration of information as in the "Telephone" or "Chinese whisper game" in which information is whispered from one person to another, becoming ever more garbled. However, what decreases in the Chinese whisper game is *quality* or *amount* of information, not *capacity* of an informational system. Lau & Persaud note that there are cases in which forced-choice responses overestimate phenomenology (blindsight) or yield inconsistent measures of phenomenology. They conclude that forced-choice reports are not ideal for measuring phenomenology. My replies:

1. Forced choice measures in the Landman and Sligte experiments are measures of working memory, not phenomenology
2. The Sperling version of the partial-report superiority effect used free-recall, not forced choice as in Landman and Sligte and got similar results. It is the use of different methodologies with the same results that avoid difficulties of the sort that Lau & Persaud are mentioning.

Shanahan & Baars say their aim is to defend GWT (Global Workspace Theory) from my challenge. More specifically, they say I use an outdated and imprecise concept of access and perpetuate a common misunderstanding of GWT that conflates the workspace with working memory. My main disagreement with them is simple: they say what is broadcast in the global workspace is all there is to consciousness, whereas I argue that it is an empirical issue whether there is more to consciousness and that the tentative answer is yes there is. Their remarks do not seem to speak to this main issue. On the "misunderstanding" of 'working memory': the issue here is purely verbal. I use the term 'working memory' to mean what Baars & Franklin call "active working memory" (Baars & Franklin, 2003). One of the first things that strikes a reader of the "working memory" literature is that the term "working memory" is used differently by different theorists. Cowan (2007) notes "Working memory has been conceived and defined in three different, slightly discrepant ways". It would take too long to explain all three ways, but what I can say briefly is that a pretty standard model (Cowan, 2005) includes three ingredients: an activated part of long-term memory (often called "short-term memory"), attentional processes that operate to broadcast perceptual and short-term memory contents and what is broadcast in the global workspace. I tend to use 'short-term memory' for the activated part of long term memory and 'working-memory' for what is broadcast in the global workspace, but judging from their commentary, Shanahan & Baars prefer to use both 'short-term memory' and 'working memory' for what I call 'short-term memory'. The bottleneck of working memory on any definition I have seen including that of Shanahan & Baars is the capacity of the global workspace, so if that bottleneck is what is of interest, my use of 'working memory' is apt. Whether or not I am attacking global workspace theory depends on whether Shanahan & Baars' controversial ideas about consciousness are part of it. Baars gets credit for many of the original ideas about the global workspace, but now that the ideas have been widely adopted, he no longer gets to dictate the terminology. My overflow argument could be put entirely in terms of global workspace terminology, excluding all talk of memory. Here it is: The capacity of phenomenology is greater than the capacity of the global workspace, so phenomenology must contain machinery not part of the machinery of the global workspace.

R 4.5 Report and Color

Kentridge describes an intriguing case: cerebral achromatopsics can recognize borders without being able to distinguish the colors on either side of the borders. Of course guitar strings that are less than a JND (just-noticeable-difference) apart in pitch create interference "beats" that allow one to tell that they are different. And color patches that are less than a JND apart can be discriminated by normal subjects if they share a border. As I understand it, (Kentridge, Heywood, & Cowey, 2004) rule out such explanations, so phenomenal presentation of colors that the subject cannot report is a live option in the case that Kentridge describes. This case is as much a challenge for my views as an opportunity, since there is an issue of how one can tell whether Kentridge's patient MS can see colors that he can't report or whether he has some kind of unconscious appreciation of color differences without phenomenal presentations of the colors themselves.

R5 ATTENTION

R 5.1 Phenomenology And Attention

Lycan asks me (3 times in 871 words) to explain how there could be a form of awareness that is intrinsic to consciousness but does not necessarily involve attention or cognitive access. He notes (and I agree) that any form of awareness (worthy of the name) is representational and that there can be representation without attention. However, he claims that cases of representation without attention have to be *sub*-personal. According to Lycan, representing by the whole person has to involve attending. This claim is in part empirical and I believe the evidence goes against it.

One kind of evidence derives from unconscious perception, more specifically perception that is unconscious because the person is not paying attention. To pick one of many examples: In visuospatial extinction, as mentioned in the target article, subjects often claim not to be able to see something on the left if there is a stimulus that competes for attention on the right. But as first noted by Volpe, LeDoux, & Gazzaniga (1979), (and nailed down by Verfaellie, Milberg, McGlinchey-Berroth, Grande, & D'Esposito (1995)), the subject's claim not to see the object on the left is often combined with the ability to make comparisons between the thing on the left and the thing on the right. The thing on the left is certainly *seen* and, it is usually supposed, unconsciously so. Unconscious seeing is still seeing, and seeing by the person, not by a part of the person. This is seeing and therefore representing without, as far as we can tell, attention (whether or not the seeing on the left is unconscious). Thus Lycan's principle: no personal-level representation without attention is empirically suspect.

In the discussion of **Spener**, I mentioned another type of evidence against the claim that there is no personal level representation without attention, namely that one can see more than one can attend to. Seeing the speckles on the speckled hen is personal and unconscious, but there is no attention to the individual speckles.

Lycan (1996; 2004) follows Locke and Armstrong in holding that what makes a state conscious is that it is the object of an "inner sense", which Lycan cashes out in terms of attention. The points I just made is not directly relevant to that view, but rather to Lycan's way of supporting it in terms of personal level representation requiring attention. However, as noted by **Koch & Tsuchiya** (and Koch & Tsuchiya, 2007) *conscious* gist perception of peripheral non-targets persists for certain kinds of stimuli even when attention is maximally drained off by a demanding central task. So there is direct evidence against Lycan's version of the inner sense thesis as well.

In cases of inattentional blindness, subjects say they don't see the stimulus—and this contrasts with the Sperling and Landman cases in which subjects clearly do see the stimulus. Relying on this contrast **Prinz** and **Grush** are convinced (against **Naccache** & **Dehaene**) that phenomenology outstrips *actual* broadcasting in the global workspace, but like **Snodgrass & Lepisto** they think I go too far in supposing that perhaps phenomenology outstrips even potential broadcasting, as in the GK case. Prinz, like Lycan, argues that the flaw in my position is to allow phenomenology without attention. Prinz's account is based on the idea that attention is necessary for phenomenology and for accessibility in the sense of potential for broadcast in the global workspace (Prinz, 2000, 2005). On both issues, I think the evidence is piling up against Prinz's view. First on the claim that attention is necessary for accessibility, i.e. potential broadcast: Dehaene and his colleagues (Dehaene, Changeux, Nacchache, Sackur, & Sergent, 2006; Kouider, Dehaene, Jobert, & Le Bihan, 2006) have convincingly shown that in conditions in which attention is maximally drained away by other tasks, representations that are as unattended as can be ensured by such conditions are nonetheless very strongly activated. They call this category "pre-conscious" as opposed to "unconscious" to indicate the strength of the activations, their strong influence and their potential to be globally broadcast. (Tse, Martinez-Conde, Schlegel, & Macknik, 2005) have obtained similar results. These representations are accessible in that with a shift of attention, they will be accessed. Whether or not attention is necessary for actual access, it is not necessary for potential access, contrary to Prinz.

On the more important issue of the relation between attention and phenomenology, as **Koch & Tsuchiya** note**,** the evidence points toward the conclusion that neither is necessary for the other ((Koch & Tsuchiya, 2007). On attention without phenomenology, subjects' attention is drawn by nudes of the opposite sex (for heterosexuals) in conditions in which the nudes are invisible according to experimental standards (Jiang, Costello, Fang, Huang, & He, 2006). (Whether the invisible pictures attract attention to one side of the visual field or the other is judged by performance on a subsequent task to which distribution of attention is crucial.) Kentridge, Heywood & Weiskrantz (1999) showed effects of attention in blindsight. On phenomenology without attention, again in a "dual task" paradigm, subjects are able to see and report certain kinds of scene gists (e.g. the presence of a face in the periphery) despite maximal siphoning off of attention by a very demanding task. Work by Alvarez & Oliva (2007) suggests that without attention (or much attention) conscious visual representations represent much more abstract properties of stimuli, for example center of mass rather than shape. As Koch & Tsuchiya point out (2007) it is difficult to make absolutely sure that there is no attention devoted to a certain stimulus, but given this limitation, this evidence points away from both of **Prinz's** claims. And the same point about the speckled hen case I made in discussion of **Spener** and **Lycan** also applies to **Prinz**.

R6 MESH BETWEEN PSYCHOLOGY AND NEUROSCIENCE

Burge argues that the conclusion of my paper can be derived from the psychological evidence alone without the argument from mesh with neuroscience. Here is a version of his argument. First, there is specific (not just generic) phenomenology prior to the cue in the Sperling-type experiments. Second, conscious events occur, they happen at a time. So we would have to have special reason to think that something that happens *after* a conscious event--or a disposition for something to happen after the conscious event--is necessary to the identity of the conscious event. If I see something red now, why should we think that the conditional occurrence of my later saying "I saw something red" if asked is partly constitutive of my sensation of red now? So the best explanation of the psychological evidence is that the machinery of the processes of access do not constitutively overlap with the machinery of consciousness itself.

Snodgrass & Lepisto say that to show that the machinery of consciousness and access are distinct, we do not need to show that there is phenomenality that cannot be accessed, but only to show that there is phenomenality that *is not in fact* accessed. And of course that is shown by the Sperling, Landman and Sligte experiments—IF as I have argued, those experiments reveal specific phenomenology without specific access. However, Snodgrass & Lepisto are neglecting the position that **Burge** refutes, the view that the machinery of access can be constitutively related to the machinery of phenomenology even via a connection of potential. Philosophers in the functionalist (and behaviorist) tradition have been impressed with the analogy between consciousness and dispositions like solubility. (For example: **Grush** and **Prinz** advocate potential access as a condition of phenomenology.) Burge's argument gives a powerful reason for dissatisfaction with this analogy. The mesh argument gives an experimentally based way of reaching Burge's conclusion about overlap of machinery. However, I did not try to directly rule out potential access as a constitutively necessary condition and so Burge's argument is a welcome addition.

Hulme & Whiteley note that inference to the best explanation requires a comparison (a point also noted by **van Gulick**). If there is only one horse in the race, it wins. The comparison I had in mind was between the model on which recurrent activation in the back of the head without global broadcasting is sufficient for phenomenology and a model on which global broadcasting is required. My point was that the former yields a mechanism of overflow whereas the latter does not. Hulme & Whiteley suggest another option: that recurrent activation is actually best thought of as a kind of binding. According to their hypothesis, activation of the relevant areas over threshold but without binding would produce amodal experience, and they give one example involving patient GY that could be interpreted that way. I would not want to base much on GY's descriptions since he has given so many somewhat different and perhaps incompatible descriptions of his experience. Also, I don't follow Hulme & Whiteley's comment on the disruption of moving phosphenes by zapping V1 with trans-cranial magnetic stimulation. The subjects say they have stationary phosphenes, which does not sound amodal. However, Hulme & Whiteley's model would be fine for my purposes since it shares a feature not shared by the competitor I was trying to refute, namely the view that a representation cannot be phenomenal unless it is globally broadcast.

Lamme says I have not integrated fully unconscious (neither accessible nor phenomenal) processing into the mesh argument and have not adequately considered arguments intrinsic to neuroscience. He and I agree almost completely about what is conscious and what is not and the reasoning why but not on the abstract methodological description. I say it is inference to the best explanation. He says the key is criteria that are intrinsic to neuroscience. Lamme says the issue is whether Dehaene's I_2 is more similar to I_1 or to I_3, using criteria intrinsic to neuroscience. He makes a convincing case that I_2 is indeed more similar to I_3 than to I_1, but I think he misdescribes his own reasoning. Similarity in neuroscience and everywhere is *always in a respect*—according to a certain metric. There is no *abstract* issue of whether I_2 is more like I_1 than I_3. This point was first rammed down the throats of philosophers by Nelson Goodman (1972) who pointed out that any two objects chosen at random are similar in an infinity of respects and dissimilar in another infinity of respects. E.g. you and I are similar in being more than 1 meter from the Eiffel Tower, more than 1.1 meter from the Eiffel Tower, etc. The same point can be made *within neuroscience*. *One can come up with neural similarity metrics that give opposite rankings*. Representations in I_2 are more similar to those in I_1 in than I_3 in that the category I_3 involves frontal activations whereas the representations in I_1 and I_2 do not. This respect of similarity is just as "*intrinsic to neuroscience*" as the one Lamme uses. Figure-ground organization, binding and perceptual organization are important dimensions of similarity not because they are neural but because they are associated with *known conscious processing* as Lamme's former and current students, **Landman & Sligte** note, and feature extraction is *excluded* because we have reason to think it happens unconsciously in the feed-forward sweep. This is classic inference to the best explanation reasoning of the sort that I am recommending.

R7 DOES CONSCIOUSNESS EVEN EXIST?

McDermott says that the ultimate theory of how the brain works "will of course not refer to anything like phenomenology, but only to neural structures," concluding that as science marches on, notions of phenomenal consciousness will give way to neurally specified cognitive access. I have two criticisms. First, why *replacement* rather than *reduction*? The distinction I am appealing to is described in every intro philosophy of science text, e.g. (Rosenberg, 2005)). To illustrate: The concept of 'phlogiston' has been *replaced* by the concept of oxygen. By contrast, we still have the concepts of heat and temperature: heat has been *reduced* to molecular kinetic energy: heat exists and is molecular kinetic energy. Reductionist physicalists (a category that includes people as diverse as me and the Churchlands) hold that phenomenal consciousness can be reduced in neuroscientific terms. McDermott speaks of the buzz saw that is cutting through the science of consciousness. But the buzz saw of the revolution in chemistry in the eighteenth century did not show that there was no such thing as heat, temperature, pressure or entropy, but rather that they could be understood in molecular terms, i.e. reduced rather than replaced. Of course, there are some cases to which the reduction/replacement distinction does not neatly apply. One much discussed example is the gene (Darden & Tabery, 2007) for which there is no straightforward answer to the question of whether there are genes and they are snippets of DNA or whether genes have been shown to not exist. Perhaps the most charitable interpretation of McDermott's remarks on life and the subject is that he predicts that the case of consciousness will end up resembling the case of the gene.

My second criticism of McDermott is: why suppose that the reduction or replacement of the future will be in terms of access as opposed to lower level neuroscience, e.g. in terms of recurrent activation of neural connections? Computer scientists tend to assume—without argument—that anything a neuroscientist might discover about what consciousness is will be basically computational. They often assume it will be implementable in a silicon computer. The underlying disagreement here is between physicalist and functionalist reduction (or replacement). The difference is a form of a dispute about the mind/body problem that has been around in one form or another for ages and is discussed in detail in (Block, 2007, 2008).

REFERENCES

Abrams, R. L., & Grinspan, J. (2007). Unconscious semantic priming in teh absence of partial awareness. *Consciousness and Cognition, 16*, 942-953.

Alter, T., & Walter, S. (2006). *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford: Oxford University Press.

Alvarez, G. A., & Oliva, A. (2007). The representation of ensemble visual features outside the focus of attentiion. *Talk at 2007 Vision Sciences Society meeting, Sarasota Florida*.

Baars, B., & Franklin, S. (2003). How conscious experience and working memory interact. *Trends in Cognitive Sciences, 17*(4), 166-172.

Block, N. (1983). The Photographic Fallacy in the Debate about Mental Imagery. *Nous, 17*, 651-662.

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

Block, N. (2001). Paradox and cross purposes in recent work on consciousness. *Cognition, 79*(1-2), 197-220.

Block, N. (2002). Concepts of Consciousness. In D. Chalmers (Ed.), *Philosophy of Mind: Classical and Contemporary Readings* (pp. 206-218). Oxford: Oxford University Press.

Block, N. (2005). Two Neural Correlates of Consciousness. *Trends in Cognitive Sciences, 9*(2), 46-52.

Block, N. (2007). "Remarks on Chauvinism and the Mind-Body Problem". In N. Block (Ed.), *Consciousness, Function, and Representation* (Vol. 1). Cambridge: MIT Press.

Block, N. (2008). Functional Reduction. In T. Horgan & M. H. Sabates & D. Sosa (Eds.), *Supervenience in Mind*. Cambridge MA: MIT Press.

Brockmole, J. R., Wang, R. F., & Irwin, D. E. (2002). Temporal Integration Between Visual Images and Visual Percepts. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 315-334.

Carruthers, P. (1989). Brute Experience. *Journal of Philosophy, 86*, 258-269.

Carruthers, P. (1992). *The Animals Issue: Moral Theory in Practice*. Cambridge: Cambridge University Press.

Carruthers, P. (1999). Sympathy and Subjectivity. *Australasian Jouenal of Philosophy, 77*(1999).

Caston, V. (2002). Aristotle on Consciousness. *Mind, 111*(444), 751-815.

Cowan, N. (2005). Working-memory capacity limits in a theoretical context. In C. Izawa & N. Ohta (Eds.), *Human learning and memory: Advances in theory* *and applications*. Florence, Kentucky: Erlbaum.

Cowan, N. (2007). What are the differences between long-term, short-term and working memory? In W. Sossin & J.-C. Jacaille & V. F. Castellucci & S. Belleville (Eds.), *The essense of memory*. Amsterdam: Elsevier/Academic Press.

Csikszentmihalyi, M. (1975). *Beyond Boredom and Anxiety*. San Francisco: Jossey-Bass.

Darden, L., & Tabery, J. (2007). Molecular Biology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2007 Editin)*. Stanford CA: <http://plato.stanford.edu/archives/fall2007/entries/molecular-biology/>. .

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*, 204-211.

Feigenson, L., & Carey, S. (2003). Tracking individuals via object files: Evidence from infants' manual search. *Developmental Science, 6*, 568-584.

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition, 97*(3), 295-313.

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files vs. analog magnitudes. *Psychological Science, 13*, 150-156.

Goldberg, I. I., Harel, M., & Malach, R. (2006). When the Brain Loses Its Self: Prefrontal Inactivation during Sensorimotor Processing. *Neuron, 50*, 329-339.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science, 303*, 1634-1640.

Huang, L., & Pashler, H. (2007). A Boolean Map Theory of Visual Attention. *Psychological Review, 114*(3), 599-631.

Huang, L., Treisman, A., & Pashler, H. (2007). Characterizing the Limits of Human Visual Awareness. *Science, 317*, 823-825.

Jiang, Y., Costello, P., Fang, F., Huang, M., & He, S. (2006). A gener- and sexual orientation-dependent spatial attentional effect of invisible images. *Proceedings of the National Academy of Sciences, 103*(45), 17048-17052.

Kentridge, R. W., Heywood, C. A., & Weiskrantz, L. (1999). Attention without awareness in blindsight. *Proceedings of the Royal Society of London: Biological Sciences, 266*(1430), 1805-1811.

Kentridge, R. W., Heywood, S. C., & Cowey, A. (2004). Chromatic edges, surfaces and constancies in cerebral achromatopsia. *Neuropsychologia, 42*(6), 821-830.

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: two distinct brain processes. *Trends in Cognitive Sciences*.

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

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

Kouider, S., & Dupoux, E. (2004). Partial Awareness Creates the "Illusion" of Subliminal Semantic Priming. *Psychological Science, 15*(2), 75-81.

Kouider, S., & Dupoux, E. (2007). How "semantic" is response priming restricted to practiced items? A reply to Abrams & Grinspan (2007). *Consciousness and Cognition, 16*, 954-956.

Loftus, G., & Irwin, D. (1998). On the Relations among Different Measures of Visible and Informational Persistence. *Cognitive Psychology, 35*, 135-199.

Lumer, E. D., & Rees, G. (1999). Covatiation of activity in visual and prefrontal cortex associated with subjective visual perception. *Proceedings of the National Academy of Sciences, 96*(4), 1669-1673.

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

Lycan, W. (2004). The Superiority of HOP to HOT. In R. Gennaro (Ed.), *Higher-Order Theories of Consciousness* (pp. 93-114). Amsterdam: John Benjamins.

Milner, A. D., & Goodale, M. A. (2007). Two Visual Systems Re-Viewed. *Neuropsychologia*.

Mullen, K. T. (1992). Colour Vision as a Post-Receptoral Specialization of the Central Visual Field. *Vision Research, 31*(1), 119-130.

Prinz, J. (2000). A Neurofunctional Theory of Visual Consciousness. *Consciousness and Cognition, 9*(2), 243-259.

Prinz, J. (2005). A neurofunctional theory of consciousness. In A. Brook & K. Akins (Eds.), *Cognition and the brain: Phil.osophy and the neuroscience movement* (pp. 381-396). Cambridge, UK: Cambridge University Press.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., & Gusnard, D. A. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences, 98*(2), 676-682.

Rosenberg, A. (2005). *Philosophy of Science: A Contemporary Introduction*. London: Routledge.

Sergent, C., & Dehaene, S. (2004). Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychological Science, 15*, 720-728.

Snodgrass, M. (2002). Disambiguating conscious and unconscious inferences: Do exclusion paradigms demonstrate unconscious perception? *American Journal of Psychology, 115*(545-580).

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition, 96*, B1-B11.

Supèr, H., Spekreijse, H., & Lamme, V. A. F. (2001). A Neural Correlate of Working Memory in the Monkey Primary Visual Cortex. *Science, 293*, 120-124.

Taddio, A., Goldbach, M., Ipp, M., Stevens, B., & Koren, G. (1995). Effect of neonatal circumcision on pain responses during vaccination in boys. *The Lancet, 345*(8945), 291-292.

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*, 17178-17183.

Verfaellie, M., Milberg, W. P., McGlinchey-Berroth, R., Grande, L., & D'Esposito, M. (1995). Comparison of cross-field matching and forced-choice identification in hemispatial neglect. *Neuropsychology, 9*, 427-434.

Volpe, B. T., LeDoux, J. E., & Gazzaniga, M. S. (1979). Information processing of visual stimuli in an extinguished field. *Nature, 282*(722-724).