

## 15 Paradox and Cross-Purposes in Recent Work on Consciousness

This is an expanded and revised version of a commentary on all the papers in a special issue of *Cognition* (April, 2001) on the state of the art in the neuroscience of consciousness. It came out in Block, N. (2007). *Functionalism, Consciousness and Representation*. Cambridge MA, MIT Press.

Several papers in Dehaene 2001 “see convergence coming from many different quarters on a version of the neuronal global workspace model” (Dennett 2001). On the contrary, even within that volume, there are commitments to very different perspectives on consciousness. And these differing perspectives are based on tacit differences in philosophical starting places that should be made explicit. Indeed, it is not clear that different uses of “consciousness” and “awareness” in Dehaene 2001 can be taken to refer to the same phenomenon. More specifically, I think there are three different concepts of consciousness in play there. The global-workspace model makes much more sense on one of these than on the others.

Part of the point of this comment is that “consciousness” and “awareness” are ambiguous terms. To avoid tedious use of “scare-quotes” and constant reminders that authors being discussed may mean something different from other authors, I will often just speak of consciousness in whatever sense is at issue in the work of the author being discussed.

### The Paradox of Recent Findings about Consciousness

The most exciting and puzzling results described in the special issue of *Cognition* and the volume edited by Dehaene appear in a linked set of experiments reported by Kanwisher 2001, Driver and Vuilleumier 2001, and Dehaene and Naccache 2001. Kanwisher notes that “neural correlates of perceptual experience, an exotic and elusive quarry just a few years ago, have suddenly become almost commonplace findings” (p. 98). And she backs this up with impressive correlations between neural activation on the one hand and indications of perceptual experiences of faces, houses, motion, letters, objects, words, and speech on the other. Conscious perception of faces whether rivalrous or not correlates with activity in the fusiform face area (FFA) but not the parahippocampal place area (PPA). And conversely for perception of places. This work is especially extensive in vision, where what I will refer to as the *ventral stream*, a set of occipital-temporal pathways, is strongly implicated in visual experience. Apparently,

the further into the temporal cortex, the more dominant the correlation with the percept. (The precise pathway depends on the subject matter, the different areas determining the different contents of consciousness.)

As Kanwisher notes, the FFA and PPA were selected for scrutiny in these experiments, not because of any association with consciousness but because it was known that they specialize in these sorts of stimuli. These areas are not as strongly activated by most other stimuli that are not places or faces. Thus the neural basis of consciousness is not localized in one set of cells, but rather in the very areas that do the perceptual analysis. Nonetheless, in a broader sense, this work does suggest a single neural basis for visual consciousness, because all visual stimuli affect areas of a single stream of processing, albeit different parts of that stream. Although finding the neural basis for visual consciousness would be exciting, it would be foolish to suppose it would immediately yield an understanding of *why* it is the neural basis. That understanding will no doubt require major ideas of which we now have no glimmer.

So we apparently have an amazing success: identification of the neural basis of visual consciousness in the ventral stream. Paradoxically, what has also become commonplace is activation of the *very same ventral stream pathways without awareness*. Damage to the inferior parietal lobes has long been known to cause visual extinction, in which, for example, subjects appear to lose subjective experience of stimuli on one side when there are stimuli on both sides, yet show signs of perception of the stimuli—for instance, the extinguished stimuli often facilitate responses to nonextinguished stimuli. (Extinction is associated with visual neglect in which subjects do not notice stimuli on one side; e.g., neglect patients often do not eat the food on one side of the plate.) Driver and Vuilleumier point out that the ventral stream is activated for extinguished stimuli (i.e., which the subject claims not to see). Rees et al. (2000) report studies of a left-sided neglect and extinction patient on face and house stimuli. Stimuli presented *only* on the left side are clearly seen by the patient, but when there are stimuli on both sides, the subject acknowledges just the stimulus on the right. However, the “unseen” stimuli show activation of the ventral pathway that is the same in location and temporal course as the seen stimuli. Further, studies in monkeys have shown that a classic “blindness” syndrome is caused by massive cortical ablation that spares most of the ventral stream but not the inferior parietal and frontal lobes (Nakamura and Mishkin 1980, 1986, as cited in Lumer and Rees 1999). Kanwisher notes that dynamic visual gratings alternating with a gray field—both very faint stimuli—showed greater activation for the gratings in V1, V2, V3A, V4v, and MT/MST despite the subjects saying they saw only a uniform gray field. Dehaene and Naccache note that processing of a masked number word proceeds all the way through the occipital-temporal pathway to a motor response even though subjects were at chance in discriminating presence from absence and in discriminating words from non-words: “An entire stream of per-

ceptual, semantic, and motor processes, specified by giving arbitrary verbal instructions to a normal subject, can occur outside of consciousness" (p. 9).

Is the difference between conscious and unconscious activation of the ventral pathway just a matter of the degree of activation? As Kanwisher notes, Rees et al. (2000) found activations for extinguished face stimuli that were as strong as for conscious stimuli. And evidence from ERP studies using the attentional blink paradigm show that neural activation of meaning is no less when the word is blinked and therefore not consciously perceived than when it is not, suggesting that it is not lower neural activation strength that accounts for lack of awareness. Further, in a study of neglect patients, McGlinchey-Berroth et al. (1993) showed that there is the same amount of semantic priming from both hemifields, despite the lack of awareness of stimuli in the left field, again suggesting that it is not activation strength that makes the difference. The upshot is that something in addition to activation strength must be playing a role.

Driver and Vuilleumier put the paradox as follows: "How then can the patient remain unaware of a contralesional stimulus, even when it can still activate the pathways that are most often considered to support conscious experience?" The paradox then is that our amazing success in identifying the neural correlate of visual experience in normal vision has led to the peculiar result that in masking and neglect, that very neural correlate occurs without, apparently, subjective experience.

What is the missing ingredient, X, which, added to ventral activation (of sufficient strength), constitutes conscious experience? Kanwisher and Driver and Vuilleumier, despite differences of emphasis, offer pretty much the same proposal as to the nature of X: (1) activation of the ventral stream supplies the contents of consciousness, (2) X is what makes those ventral contents conscious, (3) X includes binding perceptual attributes with a time and a place, that is, the representation of a token event, and (4) the neural basis of X is centered in the parietal cortex. If this is true, it is extremely significant, suggesting that the causal basis of all experience is spatiotemporal experience.

But I have a number of doubts about this proposal.

1. The proposal is wrong if there can be unbound but nonetheless conscious experiences—for example, experienced color not located in space or experienced shape and color not attached to one another. When a person has a visual experience of a *ganzfeld*, in which a color fills the subject's whole field of vision, that visual experience is apparently unbound and yet conscious. (I am indebted to correspondence with Ann Treisman on the *ganzfeld*.) But perhaps we need more clarity on what binding is before we can take this as evidence against the proposal. Friedman-Hill, Robertson, and Treisman (1995) and Wojciulik and Kanwisher (1998) discuss a patient (RM) with bilateral parietal damage who has binding problems. In many tasks, RM's level of illusory conjunctions (e.g., reporting a blue X and a red O when seeing a red X and a blue O) is high. Wojciulik and Kanwisher discuss a number of tasks in which RM is at chance, as

in the case of reporting which of two words is colored (rather than white). Perhaps RM has bound but illusory experiences—for instance, if the stimulus is a green “short” and white “ready,” he experiences a green “ready” and a white “short.” Or perhaps RM experiences green, white, “short,” and “ready” but without colors bound to words. (I have not been able to tell which is right from the published literature.) The binding hypothesis may withstand this putative disconfirmation, however, since as Wojciulik and Kanwisher report, he appears to be binding “implicitly” as indicated by his normal interference in a Strooplike task. (It takes him longer to name the colored word if he is presented with a green “brown” and a white (i.e., noncolored) “green” than a green “green” and a white “brown.”)

2. Weiskrantz and his colleagues (Kentridge, Heywood, and Weiskrantz 1999) have reported that attention can be guided to a flashed dot in the blind field by an arrow in the sighted field. Further, the patient, GY, learns when the contingencies are changed so that the arrow is misleading about where the dot will appear. A straightforward conjunction experiment requires a choice between four options:  $A \& B$ ,  $A \& \neg B$ ,  $\neg A \& B$ , and  $\neg A \& \neg B$ . If we want to know whether a blindsight patient visually perceives the conjunction of say red and moves vertically, we would have to give him choices among all four options. Unfortunately, the usual blindsight choice is between two, which may introduce skepticism about whether binding can be detected in blindsight. However, GY can choose among four options (DeGelder et al. 1999). Can a demonstration of binding in the blind field be far off?

3. Why take X to be binding rather than just attention or rather sufficient attention? (Perhaps the attention in the blind field is too low for X.) Tipper and Behrman (1996) show a neglect patient a “barbell” consisting of two circles joined by a line, with target words flashed in both circles. The patient does not recognize the target words on the left. But if the barbell is rotated so that the circle that was on the left is now on the right, the subject does not recognize the words on the right. (Caramazza and Hillis (1990) obtained similar results.) The usual explanation is that the subject’s attention travels with the object that was initially on the left. So it seems attention is crucial to whether a stimulus is extinguished. Perhaps attention determines binding, and binding determines consciousness (in the presence of the right kind of activation). But anyone who pursues this hypothesis should investigate whether we need the intermediary. (Milner and Goodale (1995) propose that consciousness is ventral stream activity plus attention, and a similar view is advocated by Prinz (2000).)

Rees et al. (2000) make two suggestions as to (in my terms) what X is. One is that the difference between conscious and unconscious activation is a matter of neural synchrony at fine timescales. The finding that ERP components P1 and N1 revealed differences between left-sided “unseen” stimuli and left-sided seen stimuli supports this idea. Driver and Vuilleumier mention preliminary data to the same effect. As they

note, ERP is probably more dependent on synchrony than fMRI. Their second suggestion is that the difference between seen and “unseen” stimuli might be a matter of interaction between the classic visual stream and the areas of the parietal and frontal cortex that control attention. Since both proposals concern hypothetical mechanisms of attention, there may be no difference between them and the attention hypothesis.

Whether or not any of these proposals are right, the search for X seems to me the most exciting current direction for consciousness research. The search for X is a diagnostic for the main difference of opinion in Dehaene 2001. Kanwisher and Driver and Vuilleumier and I give it prominence. Dehaene and Naccache, Dennett, and Jack and Shallice do not. (Parvisi and Damasio are engaging different issues.) More on what the sides represent below.

Surprisingly, given her proposal that X is binding, Kanwisher also gives a second answer: “that awareness of a particular element of perceptual information must entail not just a strong enough neural representation of that information, but also access to that information by most of the rest of the mind/brain” (p. 16). What’s going on here? Why two solutions to one problem? Are these meant as exclusive alternatives? Or are they both supposed to be true?

The answer is found in the rationale given by Kanwisher for the access condition. She appeals to a

common intuition about perceptual awareness (e.g., Baars 1988), if you perceive something, then you can report on it through any output system [in my terms, the information is globally available—NB]. . . . Perceptual information that could be reported through only one output system and not through another just would not fit with most people’s concept of a true conscious percept. . . . It seems that a core part of the idea of awareness is that not only effector systems, but indeed most parts of the mind have access to the information in question. (p. 16)

Common intuition gives us access to the meanings of our words and our concepts but not necessarily to what they are concepts of. The rationale for saying that the concept of consciousness does not apply in the absence of global availability is like the rationale for calling a darkening of the skin “sunburn” only if the sun causes it. The identical skin change—spelled out in molecular terms—could fail to fit the concept of sunburn if it had a different cause. The suggestion is that the concept of consciousness only applies to states that are globally accessible.

But that leads to a question: Could there be ventral stream activation plus X (whatever X turns out to be) that is not widely broadcast and therefore does not deserve to be called “consciousness” in this “access” sense that Kanwisher is invoking? Kanwisher mentions that the neural synchrony involved in binding might also play a role in broadcasting. But the hypothesis serves to make salient the *opposite* idea. Whatever role synchrony plays in making a representation phenomenal is unlikely to be exactly the same as the role it plays in subserving broadcasting. And even if it is the same,

what would prevent the causal path to broadcasting from being blocked? Even if we make it a condition on X that X cause the reliable broadcast of the contents of the activated area, any reliable mechanism can fail or be damaged, in which case we would have activation plus X without broadcasting. If such a thing happened, no doubt one concept of “awareness” (e.g., global accessibility) would not apply to it. But maybe another concept—phenomenality—would.

What is phenomenality? What it is like to have an experience. When you enjoy the taste of wine, you are enjoying gustatory phenomenality. Sensations are the paradigms of phenomenality, but other experiences such as thinking to oneself also are phenomenal.

Any appeal to evidence to back a theory of consciousness depends on a pretheoretical concept of consciousness to supply a starting point. We have now seen two such concepts, phenomenality and global accessibility.<sup>1</sup> The import of this distinction for the current discussion is that the X that makes contents phenomenal might not be the same as the X that makes them accessible. Without being clear about the distinction between phenomenality and accessibility, we will have no hope of ever discovering any difference in the neural bases of these two properties.

Dehaene and Naccache state the global-accessibility view as follows: “An information becomes conscious... if the neural population that represents it is mobilized by top-down attentional amplification into a brain-scale state of coherent activity that involves many neurons distributed throughout the brain. The long-distance connectivity of these “workspace neurons” can, when they are active for a minimal duration, make the information available to a variety of processes including perceptual categorization, long-term memorization, evaluation and intentional action” (p. 2). Or for short, consciousness is being broadcast in a global neuronal workspace. Dennett, advocating a similar view, takes consciousness to be cerebral celebrity, fame in the brain.

The proposal that consciousness is ventral activation plus X (e.g., neural synchrony) is based on a different starting point, a different concept of consciousness than the proposal that consciousness is cerebral celebrity or global neuronal broadcasting. (I will ignore one difference, namely, that the first is a theory of *visual* consciousness and the second is a theory of consciousness *simpliciter*.) We could see the two types of proposals as responses to different questions. The question that motivates the ventral activation plus X type of proposal is: What is the neural basis of phenomenality? The question that motivates the global neuronal broadcasting type of proposal is: What makes neuronal representations available for thought, decision, reporting, and control of action, the main types of access? We can try to force a unity by *postulating* that it is a condition on X that it promote access, but that is a verbal maneuver that only throws smoke over the difference between the concepts and questions. Alternatively, we could, *hypothesize* rather than postulate that X plus ventral stream activation as a matter of fact is the

neural basis of global neuronal broadcasting. Note, however, that the neural basis of global neuronal broadcasting might exist but the normal channels of broadcasting nonetheless be blocked or cut, again opening daylight between phenomenality and global accessibility, and showing that we cannot think of the two as one. (An analogy: rest mass and relativistic mass are importantly different from a theoretical point of view despite coinciding for all practical purposes at terrestrial velocities. Failure of coincidence even if rare is theoretical dynamite if what you are after is the scientific nature of consciousness.)

Driver and Vuilleumier suggest that we should see X in part in terms of winner-takes-all functions. But this hypothesis is more of a different way of putting the question than an answer to it if winner-takes-all means “winner gets broadcast.”

Many of us have had the experience of suddenly noticing a sound (say, a jackhammer during an intense conversation), at the same time realizing that the sound has been going on for some time even though one was not attending to it. If the subject did have a phenomenal state before the sound was noticed, that state was not broadcast in the global neuronal workspace *until it was noticed*. If this is right, there was a period of *phenomenality without broadcasting*. Of course, this is merely anecdotal evidence. And the appearance of having heard the sound all along may be a false memory. But the starting point for work on consciousness is introspection, and we would be foolish to ignore it.

If we take seriously the idea of phenomenality without access, there is a theoretical option that should be on the table, one that I think is worth investigating—that the X that makes ventral contents accessible is not the same as the X that makes ventral contents phenomenal. Perhaps visual phenomenality is ventral stream activation alone or ventral stream activation plus one of the trio of synchrony, binding, and attention. If visual phenomenality is, say, ventral stream activation plus neural synchrony, then there is a distinct issue of what makes visual phenomenality accessible, which might be another of the items mentioned, say, binding. The idea would be that the claims of extinction patients not to see extinguished stimuli are in a sense wrong—they really do have phenomenal experience of these stimuli without knowing it. A similar issue will arise in the section to follow, in which I will focus on the relation between phenomenality and a *special case* of global accessibility, reflexive or introspective consciousness, in which the subject not only has a phenomenal state but also has another state that is about the phenomenal state, say, a thought to the effect that he has a phenomenal state.

The theory that consciousness is ventral stream activation plus, for example, neural synchrony, and the theory that consciousness is broadcasting in the global neuronal workspace, are instances of the two major rival approaches to consciousness in the philosophical literature, *physicalism* and *functionalism*. The key to the difference is that

functionalism identifies consciousness with a role, whereas physicalism identifies consciousness with a physical or biological property that fills or implements or realizes that role in humans. Global availability could be implemented in many ways, but the human biological implementation involves specific electrical and chemical quantities, which, according to the physicalist, are necessary for consciousness. By contrast, functionalism in its pure form is implementation independent. As Dennett says, “The proposed consensual thesis is . . . that this global availability . . . is, all by itself, a conscious state” (p. 2). Consciousness is defined as global accessibility, and although its human implementation depends on biochemical properties specific to us, the functionalist says that artificial creatures without our biochemistry could implement the same computational relations. Thus functionalism and physicalism are incompatible doctrines since silicon implementations of the functional organization of consciousness would not share our biological nature. The rationale is expressed in Dennett’s statement that “handsome is as handsome does, that matter matters only because of what matter can do” (p. 233). He says that “functionalism in this broad sense is so ubiquitous in science that it is tantamount to a reigning presumption of all science” (p. 8). I disagree. The big question for functionalists is this: “How do you know that it is broadcasting in the global workspace that makes a representation conscious as opposed to something about the *human biological realization* of that broadcasting that makes it conscious?” There is a real issue here with two legitimate sides. The biological point of view is represented here by the hypothesis of ventral stream activation plus, for instance, neural synchrony, which in one natural way of filling in the details requires a specific biological realization.<sup>2</sup>

Thus the search for X is not well defined. It could be the search for what makes ventral contents phenomenal or it could be the search for what makes ventral contents accessible. Presumably, phenomenality itself is *part* of what makes ventral contents accessible, but if there is a real possibility of phenomenality without accessibility, we should expect to find something more that jointly, together with phenomenality, makes for accessibility. When I have given talks on this topic, I am sometimes told that X is global broadcasting itself. But this proposal has to be evaluated in light of the ambiguity just mentioned. If we take the search for X to be the search for what makes ventral contents accessible, to take X to be global broadcasting itself is to take global broadcasting to make ventral contents accessible—a thesis on a par with trumpeting dormitivity as what makes sleeping pills work. On the other hand, if the search for X is the search for what makes ventral contents phenomenal, then saying X is global broadcasting is a substantive claim, but not one anyone can claim to know to be true. My own view is that the other way around is more plausible—that phenomenality greases the wheels of accessibility.

This section has concerned two concepts of consciousness, phenomenality and global accessibility. (I have not brought up the issue of whether global accessibility



should really be considered a concept of consciousness, given that it does not require phenomenality.) In the next section, I add a third.

### What Are Experiments “about Consciousness” Really about?

Merikle, Smilek, and Eastwood describe Debner and Jacoby’s (1994) “exclusion” paradigm, in which subjects follow instructions not to complete a word stem with the end of a masked word just presented to them—only if the word is presented consciously (lightly masked). If the word is presented unconsciously (heavily masked), the subjects are more likely than baseline to disobey the instructions, completing the stem with the very word that was presented.

But what is the “conscious/unconscious” difference in this experiment? Perhaps in the case of the conscious presentation, the subject says to himself something on the order of (though maybe not this explicitly) “I just saw ‘reason,’ so I’d better complete the stem ‘rea’ with something else, say ‘reader.’” (I am not saying the monologue has to be experienced by the subject on every trial. Perhaps it could be automatized if there are enough trials.) And in the case of the unconscious presentation, there is no internal monologue of this sort. If so, the sense of the “conscious/unconscious” difference that is relevant to this experiment has something to do with the presence or absence of whatever is required for an internal monologue, perhaps something to do with introspection. Tony Jack tells me that many of his subjects in this paradigm complained about how much effort was required to follow the exclusion instructions, further motivating the hypothesis of an internal monologue.

We get some illumination by attention to another experimental paradigm described by Merikle and Joordens (1997), the “false recognition” paradigm of Jacoby and Whitehouse (1989). Subjects are given a study list of 126 words presented for half a second each. They are then presented with a masked word, word<sub>1</sub>, and an unmasked word, word<sub>2</sub>. Their task is to report whether word<sub>2</sub> was old (i.e., on the study list) or new (not on the study list). The variable was whether word<sub>1</sub> was lightly or heavily masked, the former presentations being thought of as “conscious” and the latter as “unconscious.” The result, confining our attention just to cases in which word<sub>1</sub> = word<sub>2</sub>, is that subjects were much more likely to mistakenly report word<sub>2</sub> as old when word<sub>1</sub> was unconsciously presented than when word<sub>1</sub> was consciously presented. (When word<sub>1</sub> was consciously presented, they are less likely than baseline to mistakenly report word<sub>2</sub> as old; when word<sub>1</sub> was unconsciously presented, they were more likely than baseline to err in this way.) As before, the explanation would appear to be that when word<sub>1</sub> was consciously presented, the subjects were able to use an internal monologue of the following sort (though perhaps not as explicit): “Here’s why ‘reason’ (word<sub>2</sub>) looks familiar—because I just saw it (as word<sub>1</sub>),” thereby explaining away the familiarity of word<sub>2</sub>. But when word<sub>1</sub> was *unconsciously* presented, the subjects were not able to

engage in this monologue and consequently mistakenly blamed the familiarity of word<sub>2</sub> on its appearance in the original study list.

Any reasoning that can reasonably be attributed to the subject in this paradigm concerns the subject thinking about why a word (word<sub>2</sub>) *looks familiar* to the subject. For it is only by *explaining away* the familiarity of word<sub>2</sub> that the subject is able to decide that word<sub>2</sub> was not on the study list. (If you have a hypothesis about what is going on in this experiment that does not appeal to the subject's explaining away the familiarity, I would like to hear it. Of course, I would allow that the monologue could be automatized. I suppose a skeptic might think it has already been automatized even before the experiment starts by natural versions of the experiment.) Thus in the "conscious" case, the subject must have a state that is *about the subject's own perceptual experience* (looking familiar), and so the sense of "conscious" that is relevant here is what might be termed a "reflexive" sense. An experience is conscious in this sense just in case it is the object of another of the subject's states—for example, one has a thought to the effect that one has that experience.<sup>3</sup> The reflexive sense of "consciousness" contrasts with phenomenality, which perhaps attaches to some states, which are not the objects of other mental states. Reflexivity is phenomenality plus something else (reflection), and that opens up the possibility in principle for phenomenality without reflexivity. For example, it is at least conceptually possible for there to be two people in pain, one of whom is introspecting the pain, the other not. (Perhaps infants or animals can have pain but do not introspect it.) The first is reflexively conscious of the pain, but both have phenomenally conscious states, since pain is by its very nature a phenomenally conscious state. Reflexivity (of the sort we are considering) involves phenomenality plus another state, one that is about the phenomenal state. Note that reflexivity might occur but be *unconscious* (in the phenomenal sense). (Perhaps this could occur in blindsight or in a Freudian process.) There is clearly a *conceptual* distinction between phenomenality and reflexivity, even if one or the other does not deserve to be called "consciousness." (Oddly, I find some interlocutors want to bar one, some the other; we would be better off not using words like "consciousness" and "awareness" since there is so little terminological agreement about them. I would rather use "phenomenality" and "reflexivity.") In any case, the salient empirical question is whether phenomenality and reflexivity come to the same thing in the brain.

What is the relation between reflexivity and the notion of global accessibility discussed in the last section? Global accessibility does not logically require reflexivity, since global accessibility only requires access to the response modes that the organism actually has. Perhaps a dog or a cat does not have the capacity for reflection.<sup>4</sup> Reflexivity is a special kind of access, one that requires intellectual resources that may not be available to every being that can have conscious experience.

There is another aspect to both experimental paradigms just discussed, which motivates taking seriously the hypothesis that the reflexively *unconscious* case might possi-

bly be phenomenally *conscious*. In another variant of the exclusion paradigm reported by Debner and Jacoby (1994), heavy masking was replaced by divided attention. Subjects were presented with pairs of words flanked by digits, such as “4reason5,” and then given stems consisting of the first three letters of the word (“rea\_\_\_”) to complete. There were two conditions. In the “conscious” condition, they were told to ignore the digits. In the “unconscious” condition, they were told to report the sum of the digits before completing the stem. The results were the same as before, both qualitatively and quantitatively: in the “conscious” condition, the subjects were much more likely than baseline to follow the instructions and complete the stem with a word other than “reason,” whereas with “unconscious” presentations, subjects were much more likely than baseline to violate the exclusion instructions, completing the stem with “reason.” Merikle and Joordens (2001) report corresponding results for the false-recognition paradigm with divided attention substituted for heavy masking. The added significance of this variant is that it makes one wonder whether there was a fleeting phenomenal consciousness of “reason” as the subject’s eyes moved from the 4 to the 5 in “4reason5.”

What is the status of the “unconscious” percepts in these experiments? Two theoretical options come to the fore:

1. The “unconscious perceptions” are *both* phenomenally and reflexively unconscious. (In this case, the exclusion and false-recognition paradigms are about consciousness in both senses.)
2. The “unconscious perceptions” are (fleeting) phenomenally conscious but reflexively unconscious.

A third option, that they are phenomenally unconscious but “reflexively conscious,” seems less likely because the reflexive consciousness would be “false”—that is, subjects would have a state “about” a phenomenal state without the phenomenal state itself. That hypothesis would require some extra causal factor that produced the false recognition and would thus be less simple. One argument in favor of the second hypothesis is that subjects in experiments with near-threshold stimuli often report a mess of partial perceptions that they cannot hang onto. Some critics have disparaged the idea of fleeting phenomenal consciousness in this paradigm. But what they owe us is evidence for the first option or else a reason to think it is the default view. A fourth option, that there is both phenomenal and reflexive consciousness, seems doubtful given that it is the very absence of reflexive consciousness that explains the results.

What about the fact, detailed in the first half of Dehaene and Naccache, that reportable phenomenal experience of a stimulus is systematically correlated with the ability to perform a vast variety of operations with the stimulus, while nonreportable stimulus presentation is associated with a limited, encapsulated set of processing options? This certainly is evidence for a correlation between reflexivity and accessibility. But what

does it tell us about phenomenality? First, consider whether it provides evidence that phenomenality and reflexivity go together. It would be question begging to take the evidence provided by Dehaene and Naccache as evidence of a correlation of phenomenality itself (as opposed to reports of phenomenality) with reflexivity. For the very issue we are considering is whether some of those cases of limited encapsulated processing might involve a flicker of phenomenality. Of course, the cases of phenomenality that subjects *report* are reflexively conscious. The issue is whether there are *unreported* cases of phenomenality. Broadening our focus, the same point applies to the supposition that this evidence supports a correlation between phenomenality and accessibility. (In addition, though the considerations presented by Dehaene and Naccache do show a correlation between reflexivity and accessibility in alert adult humans, we cannot generalize to infants or dazed adults or nonhumans.)

It may be said that although there is no evidence for preferring option 1 to 2, 1 is preferable on methodological grounds. Here is a way of putting the point: “How are we going to do experiments on consciousness without taking at face value what people say about whether or not they saw something? For example, if we gave up this methodology, we would have to reject blindsight work.” But I am not suggesting abandoning that methodology. We can hold onto the methodology because it is the best we have while at the same time figuring out ways to test it. No one promised us that work on consciousness was going to be easy! In the next section, I will suggest a methodological principle that will help in thinking about how to get evidence on this issue.

Let me tie the issue of this section in with that of the last—the issue stemming from the fact that the classic ventral stream can be activated without reports of awareness. There are three options about the ventral stream in, say, extinction that deserve further consideration:

1. The ventral stream is not activated enough for either phenomenality or reflexivity. (As I mentioned, this one seems disconfirmed.)
2. The ventral stream is activated enough for phenomenality, but that is not sufficient for reflexivity (nor, more generally, for accessibility). Something else (one of the Xs mentioned earlier) is required (possibly not exactly the same extra ingredient for both reflexivity and accessibility).
3. There is no phenomenality or reflexive consciousness of the extinguished stimuli, but what is missing is not activation level but something else.

Again, what reason do we have for regarding option 2 (phenomenality without reflexivity) as less likely than option 1 or 3? I suggest none.

Dehaene and Naccache argue that durable and explicit information maintenance is one of the functions of consciousness. One of their items of evidence is Sperling’s (1960) experiment on iconic memory. Sperling flashed arrays of letters (e.g., 3 by 3) to subjects for brief periods (e.g., 50 milliseconds). Subjects typically said that they could

see all or most of the letters, but they could report only about half of them. Were the subjects right in saying that they could see all the letters? Sperling tried signaling the subjects with a tone. A high tone meant the subject was to report the top row, a medium tone indicated the middle row, and so on. If the tone was given immediately after the stimulus, the subjects could usually get all the letters in the row, whatever row was indicated. But once they had named those letters, they usually could name no others. Why did the information decay? One possibility is that the subjects had *phenomenal* images of all (or almost all) of the letters, and what they lacked was access consciousness and reflexive consciousness of their identities. For subjects report that they see all the letters (Sperling 1960; Baars 1988, 15), suggesting phenomenal experience of all of them. If so, durable and explicit information maintenance may be *a function of reflexive consciousness or of access consciousness without being a function of phenomenality*.

Dehaene and Naccache suggest that the introspective judgments that fuel my phenomenal/access distinction can be accounted for by postulating three levels of accessibility. The two extremes are  $I_1$ , total inaccessibility, and  $I_3$ , global accessibility. Level  $I_2$  consists of representations that are connected to the global workspace and that can be ushered into it by the application of attention. They suggest that the letters in the Sperling phenomenon are in  $I_2$  until attention is applied to only some of them, at which point those representations enter  $I_3$ .

But where does phenomenality come into this system? One option is that both  $I_2$  and  $I_3$  are phenomenal, in which case  $I_2$  representations are phenomenal without being globally accessible, as I suggested. Another option—the one favored by Dehaene and Naccache—is that only representations in the global workspace ( $I_3$ ) are phenomenal. Their proposal is geared toward explaining away the *appearance* that the subjects saw each letter, claiming that the source of the subjects' judgment is that they *could potentially* see each letter by focusing on its location. In other words, their proposal is that the subjects mistake potential phenomenality for actual phenomenality, and this yields the appearance of phenomenality without access. Let us call this the Refrigerator Light illusion, the allusion being to the possibility that technologically naive people might have the illusion that the refrigerator light is always on because it is always on when they look.

Note, however, that phenomenally active location is not enough to capture the experience of Sperling's subjects. Subjects do not report seeing an array of blobs at locations that turn into letters when they attend to them. Subjects report seeing an array of letters. Subjects in a related masking experiment (to be discussed below) were able to give judgments of brightness, sharpness, and contrast for letters that they could not report, and they also seemed aware that the stimuli were letters. Speaking as a subject in the Sperling experiment, I am entirely confident that subjects could give such judgments.

The natural way for functionalists such as Dehaene and Naccache to respond would be to say that both the phenomenal and reflexive contents of the subjects in the Sperling experiment include features such as letterlike and features of degrees of sharpness, brightness, and contrast. Thus, they would say, early vision gives subjects experience of these features that are both phenomenally and reflexively conscious, so there is no discrepancy. I share the functionalist view that the subjects have reflexive consciousness of the letterlikeness, sharpness, brightness, and contrast of the letters. My disagreement with them is that I also allow phenomenal consciousness of the shapes themselves without any reflexive consciousness of them. I say the subjects have phenomenal experience of the shapes; the functionalists say the appearance that the subjects have phenomenal experience of shapes is a case of the Refrigerator Light illusion fostered by the fact that the subjects could potentially access the shapes.

At this point, the reader may feel that there is little to choose between the two points of view. But there are two considerations that I believe tip the balance in favor of phenomenality without access. The first is that the functionalist position does not accommodate what it is like for the subjects as well as does phenomenality without access. Speaking as a subject, what it is like for the subjects is experiencing all or most of the letter shapes. An analogy: suppose you are one of a group of subjects who report definitely seeing red. The hypothesis that you and the other subjects have an experience as of red accommodates what it is like for the subjects better than the hypothesis that all of you are under the illusion that you have an experience as of red but are really experiencing green. Postulating an illusion is an extreme measure.

Second and more impressive, there is another hypothesis that applies in this case that also applies in the case of some other phenomena (to be discussed). The functionalist appeal to the Refrigerator Light illusion by contrast applies less well or not at all in these other cases. Thus the phenomenality-without-access hypothesis has the advantage of more generality, whereas the functionalist has the disadvantage of ad hoc postulation.

Let me fill in the phenomenality-without-access idea a bit. One picture is that the subjects in the Sperling experiment are phenomenally conscious of the letter shapes, but do not have the attentional resources to apply letter concepts or even shape concepts of the sort one applies to unfamiliar shapes when one has plenty of time. Phenomenal experience of shapes does not require shape concepts but reflexive consciousness being an intentional state does require shape concepts, concepts that the subjects seem unable to access in these meager attentional circumstances. Another option is that shape concepts are applied but the subjects do not have the attentional resources to harness those concepts in reflexive consciousness, and the neglected representations decay.

Liss (1968) contrasted subjects' responses to brief unmasked stimuli (one to four letters) with their responses to longer lightly masked stimuli. He asked for judgments of

brightness, sharpness, and contrast as well as what letters they saw. He found that lightly masked 40-msec stimuli were judged brighter and sharper than unmasked 9-msec stimuli, even though the subjects could report three of four of the letters in the unmasked stimuli and only one of four in the masked cases. He says: "The Ss commented spontaneously that, despite the high contrast of the letters presented under backward masking, they seemed to appear for such brief duration that there was very little time to identify them before the mask appeared. Although letters presented for only 7 msec with no masking appeared weak and fuzzy, their duration seemed longer than letters presented for 70 msec followed by a mask" (p. 329).

What is especially intriguing about the Liss experiment is the suggestion of a double dissociation between phenomenal clarity and the ability to form a conceptual representation. The masked stimuli were relatively high in phenomenal clarity but low in conceptualization, whereas the unmasked stimuli were higher in conceptualization but lower in phenomenal clarity. (There is a third level, perceptual representation, intermediate between phenomenal and conceptual, which this experiment casts little light on.)

As in the Sperling phenomenon, a natural hypothesis is that the subjects were phenomenally conscious of all the masked letter shapes, but could not apply the letter concepts (and perhaps could not apply perceptual representations) required for reflexive consciousness of all of them. Or, as before, perhaps they did briefly apply the letter concepts, but with insufficient attention those conceptual representations dissolved. And as before, there is an alternative functionalist hypothesis—that the contents of both the subjects' phenomenal states and their reflexive states are the same and include the features sharp, high contrast, bright, and letterlike without any specific shape representation. A major difference between Sperling and Liss is that in the Liss experiment, there is no evidence that the subjects were able to access any letter they chose. Sperling asked them to report an indicated row; Liss did not. In the Liss experiment, subjects were trying to grab all the letters they could, and they could get only about one of four when masked. Thus the Refrigerator Light illusion hypothesis applied by Dehaene and Naccache to the Sperling phenomenon gets no foothold in the Liss phenomenon. The subjects' conviction that they saw all four of the masked letters would have to be explained in some other way, and that makes the functionalist position ad hoc compared with the hypothesis of phenomenality without reflexivity. The third and final stage of this argument will be presented in the next section in the discussion of the grain of vision, where I will mention a third experimental paradigm—one that is completely different from that of either Sperling or Liss and that also does not fit the Refrigerator Light illusion hypothesis but does suggest phenomenality without access.

Dennett takes a stand similar to that of Dehaene and Naccache, arguing that potential and actual fame in the brain are all that are needed to handle such phenomena.

The Liss experiment just described suggests phenomenality without fame or even potential fame. In addition, potential fame without any hint of phenomenality is often reported. Many people have representations of direction (which way is north) and time without (apparently) phenomenality or the illusion of phenomenality. As soon as they ask themselves the question of what time it is or which way is north, they “just know.” Before the knowledge popped into mind, it was potentially famous but with no phenomenality or illusion of phenomenality.

Moving back to the main subject of this section, we have seen three concepts of consciousness: phenomenality, reflexive consciousness, and access consciousness. Can we blame the disagreements among our authors on different concepts of consciousness?

This ecumenical stance is especially helpful in reading Parvisi and Damasio and Jack and Shallice. Parvisi and Damasio characterize consciousness as follows: “Core consciousness occurs when the brain’s representation devices generate an imaged, nonverbal account of how the organism’s own state is affected by the organism’s interaction with an object, and when this process leads to the enhancement of the image of the causative object, thus placing the object saliently in a spatial and temporal context” (p. 137). This would be a mysterious account of phenomenality, since the images mentioned in it presumably *already* have phenomenality, making the nonverbal account unnecessary. And the account would make little sense as an account of access consciousness, since a *thought* can be access conscious without involving such images, much less images of a causative object or the enhancement of them. The account is best construed as a characterization of reflexive consciousness, since it emphasizes the knowledge of the subject of how that subject has been affected by an interaction, and thus involves reflection.

Jack and Shallice propose that a conscious process is one in which a supervisory system directly selects one of a number of competing schemata plus its arguments. But they do not give us any evidence against the possibility of either phenomenality or global access without supervisory selection, or supervisory selection without phenomenality or global access. Phenomenality might be a matter of activation plus binding, which, as far as we know, could occur in an organism that does not have a supervisory system, or even in an organism that has a supervisory system, without its activity, or even with its activity, without its selecting one of a number of competing schemata. Access might be a matter of broadcasting in a system that contains no supervisor. Conversely, it would appear at first glance that there could be supervisory selection of the sort they suggest without phenomenality or global access. They do give evidence that supervisory selection among schemata does lead to encoding of specific episodes, but they do not argue that this encoding requires either phenomenality or global accessibility. If Jack and Shallice were advancing a theory of phenomenality or of access consciousness, there would be a heavy burden on them to justify it, a burden that they give no hint of acknowledging. But as a theory of reflexive consciousness it makes



much more sense. Reflexive consciousness involves one aspect of the mind monitoring another aspect, like a sensory state, so in one sense of “supervisory,” reflexive consciousness necessarily involves a supervisory system. (They make a similar point.) Jack and Shallice would still owe us an account of why there cannot be reflexive consciousness where the supervisory system focuses on a sensory state without choosing among competing schemata. But at least with reflexive consciousness they are in the right ballpark.

Jack and Shallice may be skeptical about the global workplace account. Shallice (1975) argued that there is reason to think that there is more than one “workplace” for different functions, and no global one. He was criticizing Atkinson and Shiffrin’s (1971) “idea that in some sense consciousness can be ‘equated’ with the short-term store” (Shallice 1975, 270). And Jack and Shallice note that it is unlikely that representational codes in different modules match. But the version of the global-workspace model advocated by Dehaene and Naccache does not extend to broadcasting within modules. They are not committed to the idea that conscious experiences of, say, color are available to the phonology module, nor do Jack and Shallice suggest any such thing.

Perhaps Jack and Shallice think that representations that have been selected by a supervisory system of a certain sort are as a matter of fact globally accessible in an appropriately qualified sense, but that does not address the issue of why their definition characterizes a *necessary* condition for global accessibility (of an appropriately limited sort). Could a machine be made that has globally accessible representations that are not the result of selection of competing schemata by a supervisory system? They do not say why not.

Though Jack and Shallice give an account that makes sense as a theory of reflexive consciousness (and maybe as an account of access consciousness restricted to humans), they have ambitions for its application to any process that is *phenomenally* the same. They say that “tasks involving Type-C processes should either actually require the subject to make an introspective judgment, or be phenomenologically similar to tasks that do” (p. 15). Also, Dehaene and Naccache make it clear that they see their stance as applying to (as I would put it) phenomenality. Just after the words quoted earlier, they say that “we postulate that this global availability of information through the workspace is what we subjectively experience as a conscious state” (p. 1). Someone (like myself) who believes in phenomenality as distinct from its function would naturally think that phenomenality *causes* the global availability of information, not that phenomenality *is* the global availability of information (although, given our ignorance about the fundamental nature of phenomenality, I am not prepared to rule that option out a priori). In sum, these theories are best seen as theories of reflexivity or global accessibility rather than as theories of phenomenality, but their advocates claim phenomenality nevertheless.

### Defining Reflexivity

Rosenthal (1997) defines reflexive consciousness as follows:  $S$  is a reflexively conscious state of mine  $\leftrightarrow$   $S$  is accompanied by a thought—arrived at noninferentially and nonobservationally—to the effect that I am in  $S$ . He offers this “higher order thought” (HOT) theory as a theory of phenomenal consciousness. It is obvious that phenomenal consciousness without HOT and HOT without phenomenal consciousness are both *conceptually* possible. For example, perhaps dogs and infants have phenomenally conscious pains without higher-order thoughts about them. For the converse case, imagine that by biofeedback and neural imaging techniques of the distant future, I learn to detect the state in myself of having the Freudian unconscious thought that it would be nice to kill my father and marry my mother. People can learn to lower blood pressure via biofeedback. Perhaps they could learn to detect brain states. I could come to know—noninferentially and nonobservationally—that I have this Freudian thought even though the thought is not phenomenally conscious. Since there are conceptually possible counterexamples in both directions, the issue is the one discussed above of whether reflexivity and phenomenality come to the same thing in the brain.

If there are no actual counterexamples, the question arises of why. Is it supposed to be a basic law of nature that phenomenality and reflexivity co-occur? That would be a very adventurous claim. But if it is only a fact about us, then there must be a mechanism that explains the correlation, as the fact that both heat and electricity are carried by free electrons explains the correlation of electrical and thermal conductivity. But any mechanism breaks down under extreme conditions, as does the correlation of electrical and thermal conductivity at extremely high temperatures. So the correlation between phenomenality and reflexivity would break down too, showing that reflexivity does not yield the basic scientific nature of phenomenality.

Rosenthal’s definition of reflexivity has a number of ad hoc features. “Nonobservationally” is required to rule out (for example) a case in which I know about a thought I have repressed by observing my own behavior. “Noninferentially” is needed to avoid a somewhat different case in which I appreciate (nonobservationally) my own pain and infer a repressed thought from it. Further, Rosenthal’s definition involves a stipulation that the possessor of the reflexively conscious state is the same as the thinker of the thought—otherwise my thinking about your pain would make it a conscious pain. All these ad hoc features can be eliminated by moving to the following definition of reflexivity:  $S$  is a reflexively conscious state  $\leftrightarrow$   $S$  is phenomenally presented in a thought about  $S$ . This definition uses the notion of phenomenality, but this is no disadvantage unless there is no such thing apart from reflexivity itself. The new definition of reflexivity, requiring phenomenality as it does, has the additional advantage of making it clear that reflexivity is a kind of *consciousness*. (See Burge’s 1997 critique of my definition of access consciousness as constituting a kind of consciousness.)

### Is It Impossible in Principle to Empirically Distinguish Phenomenality from Reflexivity?

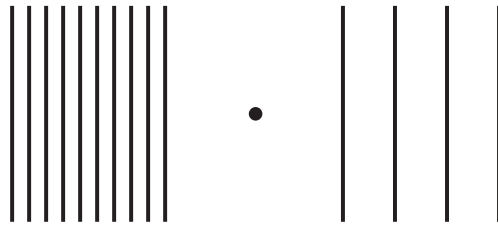
Some objectors think that the distinction between phenomenality and reflexivity has no real empirical significance. Here is a version of that view: “In order to ascertain empirically whether a phenomenal state is present or absent or what its content is, we require the subject’s testimony. But when a subject says that he did or did not see something, or that his state did or did not have a certain content, he is exhibiting presence or absence of the relevant reflexive consciousness too. So how can there ever be an empirical wedge between phenomenality and reflexivity or between phenomenal content and reflexive content?” Further, if the contents of phenomenal states are non-conceptual, how can we ever find out what they are by attention to what a subject says? (A similar but more difficult issue arises about the relation between phenomenality and global accessibility that I will not have the space to discuss.)

Here are some considerations that should loosen the grip of this pessimistic point of view. First, consider the common experience, mentioned earlier, of suddenly noticing that one has been hearing a noise for some time. Testimony at time  $t_2$  can be evidence for phenomenality at time  $t_1$  even though the subject did not notice the phenomenal experience at time  $t_1$ . That is, a phenomenal state does not have to be accompanied *simultaneously* by a reflection on it for there to be testimony about it. How do we know there wasn’t also a brief flash of reflexivity about the phenomenality at  $t_1$ ? There is no reason to believe there is any principled problem of discovering such a thing, since reflexivity is a kind of thought. (For example, if we discover a language-of-thought hypothesis to characterize thought in other circumstances, we could apply it here.)

Second, note that reflexivity involves phenomenality plus more—reflection on the phenomenality.<sup>5</sup> If this is right, we can see that whatever processes produce the reflection will—like all physical processes—sometimes misfire and we will have phenomenality without reflexivity. The prior probability then, of phenomenality without reflexivity is considerable. Jack and Shallice may think otherwise—their theory certainly presupposes otherwise—but they do not present a single empirical result that points in this direction. To the extent that they supply a case against phenomenality without reflexivity, it is entirely *philosophical*.

We can guess that phenomenality without reflexivity will happen when the machinery of reflection is damped down—perhaps in infants whose reflection machinery is undeveloped, in adults where it is permanently or temporarily damaged, or in animals where it is minimal to begin with. When we know that something very likely occurs and we have an idea of what makes it occur, we should not be pessimistic about our ability to find a reliable way of experimentally exploring it.

The best way to silence the pessimistic point of view is to canvas some empirical approaches. One line of evidence emerges from work by Cavanagh and his colleagues



**Figure 15.1**

Fixate on the dot at the center and attend to the lines on the right, with the page held at arm's length. (Distance is not very important with this display.) Subjects are capable of "stepping" (described in the text) from one line to another on the right (though with trouble on line 3) but not on the left. From Cavanagh, He, and Intriligator 1998.

that shows that the resolution of visual attention is five to ten times coarser than the resolution of vision itself (Cavanagh, He, and Intriligator 1998; He, Cavanagh, and Intriligator 1996; Intriligator and Cavanagh, forthcoming). The grain of visual attention is about 5–10 arc min (1 arc min is a sixtieth of a degree) at the fovea (the densest area of the retina), whereas the grain of vision is about 1 arc min at the fovea. What is meant by "grain" and "resolution"? In the experiments by Cavanagh and his colleagues, the resolution of vision is measured by such procedures as whether a subject can verbally distinguish a set of lines from a uniform gray field, and whether the subject can report the orientation of the lines. The resolution of visual attention can be measured by whether the subject can count the items to be resolved, but a better measure is a "stepping" procedure that is illustrated in figure 15.1. First, fixate on the dot in the middle. (This is necessary to avoid eye movements, and consequent complication in interpretation; whether subjects succeed in fixating can be checked with eye-tracking devices.) One line lights up; the subject is asked to focus on that one, then move, for example, one to the right, another to the right, one to the left, one to the right. Success is determined by checking which line the subject is focused on at the end. In the set of four lines on the right in figure 15.1, most subjects can step through the first, second, and fourth lines from the left, but the third tends to cause trouble. Most subjects cannot step through the lines on the left even though the lines on the left are visually resolvable. Attentional resolution can also be measured by a tracking task developed by Pylyshyn and his colleagues (Pylyshyn and Storm 1988).

How are these findings relevant to the distinction between phenomenality and reflexivity? Landolt (1891)—who apparently was the first to publish an observation of the phenomenon—asked observers to count finely spaced dots or stripes. As Intriligator and Cavanagh (forthcoming) note, Landolt's observers could not count the stripes or dots if their spacing was less than about 5 arc min, even though they could still see

them. Landolt says, “You get to a point where you can no longer count them at all, even though they remain perfectly and distinctly visible.”<sup>6</sup> (Landolt’s subjects looked right at them instead of fixating to a single spot to the side, but it turns out that that the eye movements did not matter much for his stimuli.) The individual lines remain “purely and distinctly visible”—in my terms, one is phenomenally conscious of them. And one can say roughly how many there are. But, to the extent that one cannot attend to them, one cannot apply concepts to them individually—for example, shape concepts. True, one has an impression of lines of a certain length (as on the left of figure 15.1), but to the extent that one cannot attend to individual items, one cannot distinguish the shape of one from another. If the items are gratings rather than lines, one cannot say what the orientation is; if they are letters, one can see that they are letters but not which letters they are. My suggestion is the same as the one I made in the case of Sperling and Liss, namely, that the subjects may have phenomenal awareness of the individual shapes without the attentional resources to apply shape concepts to them and thus without reflexive awareness of them. (This may be because they cannot apply perceptual representations and perceptual representations are necessary for conceptual representations.) Or alternatively, they may apply shape concepts to them but lack the attentional resources to harness those concepts in reflexive states. In either case we would have phenomenality without reflexivity.

There is an alternative hypothesis—that the contents of both the subject’s phenomenal states and their reflexive states are the same and include the feature “letterlike” without any specific shape. Cavanagh speaks of seeing a “texture.” There is a reason for preferring my sort of hypothesis, namely, that subjects find the individual items “perfectly and distinctly visible” in Landolt’s phrase. (Look at the figure. Doesn’t it seem to you that you see each line, rather than just a texture of a sort one has learned is produced by lines?) But perhaps subjects are under an illusion of some sort? Maybe, but if so, it is not the same as the “Refrigerator Light” illusion postulated by Dehaene and Naccache in the case of the Sperling phenomenon. You will recall that they supposed that the sense of subjects in the Sperling experiment that they saw all the letters derived from the fact that they could attend to any small number of their locations and be aware of the identity of the letter. But there is no such thing here. No matter how hard subjects try, they cannot ascertain the identity of stimuli that are crowded to the point of being below the grain of attention. This hypothesis, then, has more generality than the Refrigerator Light illusion hypothesis of Dehaene and Naccache.

Interestingly, adaptation to the orientation of gratings that the subject cannot report affects the detection of other gratings as much as uncrowded gratings whose orientation the subjects can report. He, Cavanagh, and Intriligator (1996) note that V1 is the first site of orientation processing, so the attentional effect of crowding must occur later in processing than V1. He and colleagues conclude that activation of neurons in

V1 is “insufficient for conscious perception” (p. 335), but although this result shows V1 is insufficient for reflexive consciousness, it is less effective in showing that activation in V1 is insufficient for phenomenality (Block 1996). Do not get me wrong. I am not saying activation in V1 is sufficient for phenomenality. I am making a logical point about comparative strength of evidence, the upshot of which is that an empirical wedge between phenomenality and reflexivity is possible. More generally, I allow that it could be discovered that, contrary to what I have been arguing, one is phenomenally aware of exactly the same features that one is reflexively aware of. I do not say that there is strong evidence for phenomenality without reflexivity. My point is that for stimuli that are below the level of attentional resolution but above the level of visual resolution, there is a better case for phenomenal awareness than for reflexive awareness.

If my picture of the attentional phenomena is accepted, it can also avoid the conclusion that many have drawn from the change-blindness and inattentional-blindness literature (Simons 2000), that there is an “illusion” of rich visual awareness of the world. As Cavanagh (1999) puts it, “But what of our feeling that we piece together our world in multiple glances, building up a reasonably complete model of a stable world around us? This concept of a rich model of the world does not hold up.” Vision, he says gives us a “*false* sense of ‘knowing what is out there.’” (This view, a version of the Refrigerator Light hypothesis, is strongly defended in O’Regan 1992.) We can avoid the idea that vision creates an illusion if our perception of the world is phenomenally rich but attentively sparse. That is, our phenomenal impression is accurate, but only the attended aspects of it are available for the tasks tapped in the change-blindness literature.<sup>7</sup>

Turning to something completely different. I will mention an old, somewhat anecdotal result, not because it is itself serious evidence for anything, but because it illustrates some methodological points.

I have in mind the strange phenomenon of aerodontalgia (Melzack and Wall 1988; Nathan 1985). Two American dentists in Britain in World War II noticed that pilots who went up in the unpressurized planes of the time often complained of pains that seemed to be recreations of pains of previous dental work, even though the dental work had been done under anesthesia. They hypothesized that the recreated pains all derived from dental work done under general anesthesia rather than local anesthesia and they put this hypothesis to the test by doing extractions under combinations of local and general anesthesia. The result was that they only got recreated pains for general anesthesia. For example, if they gave a pilot general anesthesia and also local anesthetic on the left side and then extracted teeth from both sides, they got recreated pains from the right side only. (They used a substitute for the unpressurized planes—stimulation of the nasal mucosa—since it turned out that the effect of the unpressurized cabins was mediated by sinus stimulation.)

My point is *not* that this is serious evidence for phenomenal states under general anesthesia. This is old work that was not done by behavioral scientists. I do not know of any replication. Further, even if replicated, there would be a problem since maybe what happened was that traces were laid down under general anesthesia without any phenomenal event, and then those traces later produced a phenomenal event. This would be representation of pain under general anesthesia rather than pain under general anesthesia. My points about this experiment are these:

1. Though the evidence is flawed, it is better evidence for phenomenality under general anesthesia than it is for reflexive consciousness under general anesthesia, since a reflexively conscious pain is a phenomenal event (pain is necessarily phenomenal) plus something else—reflection on it. The reflection is a species of thought, and there is reason to believe that anesthetic gets in the way of thought. So we are on better ground postulating a pain under general anesthesia than a pain plus reflection on it.
2. The methodological point is that *the reflexively conscious second pain can be evidence for the first pain even though the first pain isn't reflexively conscious—we don't need the subject's testimony about the first pain itself*. It is this feature of the aerodontalgia case that makes it methodologically interesting despite the obvious flaws and despite the fact that it is not in itself serious evidence for phenomenality without reflexivity. To see the interest of this methodological item, consider the objection “If you think of phenomenality as a purely subjective phenomenon—something only the subject can tell you about—how can you possibly get evidence for phenomenality that the subject can't report?” Answer: the sense of subjectivity in the objection is faulty. Subjectivity does not entail that only the subject can tell us about it.

Objection: “But you have admitted that this is far from conclusive evidence for phenomenality without reflexivity. Doesn't the principled problem arise again when you try to go from highly flawed evidence of the sort you are presenting to conclusive evidence of the sort the scientific community would be compelled to believe? How could we ever get more than a glimmer of evidence for phenomenality without reflexivity?” The answer is that if we can get many convergent though flawed sources of evidence—so long as the flaws are of different sorts and uncorrelated—we will have a convincing case. (Note that I am not saying that a lot of weak evidence adds up to strong evidence.) For example, there are different methods of dating rocks and artifacts based on isotopes. Those based on counting the products of radioactive decay, “daughter” isotopes (e.g., potassium-argon dating), have different flaws from those based on counting decay of the parent substance (e.g., carbon-14 dating), and other methods such as the fission-track method have still different flaws, but if a number of measures with different flaws agree, that is very convincing.

## Conclusion

The papers in Dehaene 2001 deploy three different concepts of consciousness:

1. *Phenomenality* experience. This is the concept of consciousness that is most directly the subject of the hypothesis discussed by Driver and Vuilleumier and Kanwisher that visual consciousness is ventral stream activation plus X.
2. *Access consciousness* global accessibility. This is the concept of consciousness most directly related to Dehaene and Naccache's account of consciousness as being broadcast in a global neuronal workspace and Dennett's account of consciousness as cerebral celebrity. Since this concept of consciousness does not require phenomenality, there is some doubt as to whether it is a full-fledged concept of consciousness. (See Burge 1997.)
3. *Reflexive consciousness* a special kind of access; a state is introspectively conscious just in case (roughly) it is the object of another state. Or alternatively (and better in my view) a state S is reflexively conscious just in case it is phenomenally presented in a thought about S. This is the concept of consciousness most clearly involved in reasoning about the false-recognition and exclusion experiments (Merikle, Smilek, and Eastwood) and is most appropriate to Jack and Shallice.

Some of the disagreements among the contributors to the volume can be explained by interpreting them as talking about different things.

Are the three kinds of consciousness aspects of a single thing? There are a number of ways of interpreting this question. One is the sense of aspects of a single thing in which the solid, liquid, and gaseous phase are aspects of a single substance. In this sense, being an aspect of a single thing requires that for any solid substance there be some conditions under which it would be gaseous. In this sense, I think it is a wide-open empirical question whether phenomenality and access consciousness are aspects of a single thing. But I suspect that it is less likely that reflexivity can be included with these two. If a lizard has phenomenality, must there be conditions under which it would have reflexive consciousness of that phenomenality? If you are doubtful, then you are doubtful whether all three kinds of consciousness are aspects of a single thing.<sup>8</sup>

## Notes

This is a somewhat expanded and revised version of a paper that appeared in *Cognition* 79: 1–2, April 2001. It is a commentary on all the papers in a special issue of that journal on the state of the art in the neuroscience of consciousness. (The special issue is available as a book edited by Stan Dehaene and published by MIT Press; see Dehaene 2001.) Two philosophers—Dan Dennett and I—were asked to comment on all the scientists' papers. (We both made some comments on each others' papers as well.) Dennett's paper is available at <http://ase.tufts.edu/cogstud/papers/>



cognition.fin.htm. The volume of *Cognition* is on the web and available on the NYU library site by going to [http://www.nyu.edu/library/bobst/research/sci/ej\\_c.htm](http://www.nyu.edu/library/bobst/research/sci/ej_c.htm), clicking on "Cognition," and then going to 79, 1–2. Other library sites will also have access to this journal.

1. Block 1997 says a representation is access conscious if it is poised for global control. Block 1995 adopts a more cumbersome formulation that lists various types of control. (The advantage of the cumbersome formulation is that it avoids ascribing consciousness to simple devices that nonetheless have global control in the sense of control over all response systems.) Since phenomenal consciousness is best thought of as an occurrence, *broadcast* for global control would be better. (See Burge 1997, which criticizes Block 1995.)

2. The problem for functionalists could be put like this: the specifically human realization of global availability may be necessary to consciousness—other realizations of global availability being "ersatz" realizations. Dennett responds to this point by arguing in effect that we can preserve functionalism by simply characterizing global availability in a more detailed way—at the level of biochemistry. But the utility of this technique runs out as one descends the hierarchy of sciences, because the lowest level of all, that of basic-level physics, is vulnerable to the same point. Putting the point for simplicity in terms of the physics of forty years ago, the causal role of electrons is the same as that of antielectrons. If you formulate a functional role for an electron, an antielectron will realize it. Thus an antielectron is an ersatz realizer of the functional definition of *electron*. Physics is characterized by symmetries that allow ersatz realizations.

For an introduction to issues about functionalism, the reader could consult the entries on consciousness or on functionalism in any of the truly excellent philosophy reference works that have been published in the last five years: *The Routledge Encyclopedia of Philosophy*, *The Oxford Companion to Philosophy*, *The Cambridge Companion to Philosophy*, *Blackwell's Companion to the Philosophy of Mind*, or the supplement to Macmillan's *The Encyclopedia of Philosophy*.

3. This definition is oversimple but complications will not matter here.

4. To avoid overattributing access consciousness, we have to specify the machinery instead of using the catchall "global." For a specific brain architecture that provides flexibility in the choices of devices the conscious information is passed to, see Dehaene, Kerszberg, and Changeux 1998.

5. I am ignoring the possibility that reflexivity might occur without the experience it is normally about.

6. "On arrive à un point où l'on ne peut plus les compter d'aucune façon, alors qu'ils demeurent encore parfaitement et distinctement visibles" (p. 385).

7. Wolfe (1999) advocates "inattentional amnesia" rather than inattentional blindness, which is close to the "inattentional inaccessibility" view I am advocating. Simons (2000) quotes Wolfe as suggesting that we might think of subjects' failure to notice in the change- and inattentional-blindness literature as "inattentional agnosia." (Agnosia is a centrally caused failure to achieve knowledge on the basis of sensation.) Assuming that agnosia involves failure of application of concepts to stimuli or failure to harness such concepts in the service of knowledge, there is another overlap between my view and Wolfe's.

8. I am grateful to Tyler Burge, Susan Carey, Nancy Kanwisher, Georges Rey, and Jesse Prinz for comments on an earlier version, and I am especially grateful to Stan Dehaene and Tony Jack for many rounds of debate on key issues on which we disagree. The paper has been much improved as a result of these controversies.

## References

- Atkinson, R. C., and Shiffrin, R. M. 1971. The control of short-term memory. *Scientific American* 225: 82–90.
- Baars, B. J. 1988. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
- Block, Ned. 1995. On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18(2): 1995. Reprinted in N. Block, O. Flanagan, and G. Güzelde, eds., *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: MIT Press.
- Block, Ned. 1996, October. How can we find the neural correlate of consciousness? *Trends in Neuroscience* 19: 456–459.
- Block, Ned. 1997. Biology vs. computation in the study of consciousness. *Behavioral and Brain Sciences* 20: 1.
- Burge, Tyler. 1997. Two kinds of consciousness. In N. Block, O. Flanagan, and G. Güzelde, eds., *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: MIT Press.
- Caramazza, A., and Hillis, A. E. 1990. Levels of representation, coordinate frames, and unilateral neglect. *Cognitive Neuropsychology* 7(5–6): 391–445.
- Cavanagh, P. 1999. Attention: Exporting vision to the mind. In C. Taddei-Ferretti and C. Musio, eds., *Neuronal Basis and Psychological Aspects of Consciousness*, 129–143. Singapore: World Scientific.
- Cavanagh, P., He, S., and Intriligator, J. 1999. Attentional resolution: The grain and locus of visual awareness. In C. Taddei-Ferretti and C. Musio, eds., *Neuronal Basis and Psychological Aspects of Consciousness*. Singapore: World Scientific.
- Debner, J. A., and Jacoby, L. L. 1994. Unconscious perception: Attention, awareness, and control. *Journal of Experimental Psychology: Learning, Memory and Cognition* 20: 304–317.
- DeGelder, B., Vroomen, J., Pourtois, G., and Weiskrantz, L. 1999. Non-conscious recognition of affect in the absence of striate cortex. *NeuroReport* 10: 3759–3763.
- Dehaene, S., ed. 2001. *The Cognitive Neuroscience of Consciousness*. Cambridge, MA: MIT Press.
- Dehaene, S. 2001. Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. In Dehaene 2001, 1–37.
- Dehaene, S., Kerszberg, M., and Changeux, J. P. 1998. A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences USA* 95: 14529–14534.

- Dennett, D. 2001. Are we explaining consciousness yet? In Dehaene 2001, 221–237.
- Driver, J., and Vuilleumier, 2001. Perceptual awareness and its loss in unilateral neglect and extinction. In Dehaene 2001, 39–88.
- Friedman-Hill, S., Robertson, L., and Treisman, A. 1995, August 11. Parietal contributions to visual feature binding: Evidence from a patient with bilateral lesions. *Science* 269: 853–855.
- He, Sheng, Cavanagh, Patrick, and Intriligator, James. 1996. Attentional resolution and the locus of visual awareness. *Nature* 383: 334–337.
- Intriligator, James, and Cavanagh, Patrick. Forthcoming. The spatial resolution of visual attention. *Cognitive Psychology* 43: 171–216.
- Jack, A., and Shallice, T. 2001. Introspective physicalism as an approach to the science of consciousness. In Dehaene 2001, 161–196.
- Jacoby, L. L., and Whitehouse, K. 1989. An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General* 118: 126–135.
- Kanwisher, N. 2001. Neural events and perceptual awareness. In Dehaene 2001, 89–113.
- Kentridge, R. W., Heywood, C. A., and Weiskrantz, L. 1999. Attention without awareness in blind-sight. *Proceedings of the Royal Society of London. Series B, Biological Sciences* 266(1430): 1805–1811.
- Landolt, E. 1891. Nouvelles recherches sur la physiologie des mouvements des yeux. *Archives d'ophthalmologie* 11: 385–395.
- Liss, P. 1968. Does backward masking by visual noise stop stimulus processing? *Perception and Psychophysics* 4: 328–330.
- Lumer, Erik, and Rees, Geraint. 1999. Covariation of activity in visual and prefrontal cortex associated with subjective visual perception. *Proceedings of the National Academy of Sciences* 96: 1669–1673.
- McGlinchey-Berroth, R., Milberg, W. P., Verfaellie, M., Alexander, M., and Kilduff, P. 1993. Semantic priming in the neglected field: Evidence from a lexical decision task. *Cognitive Neuropsychology* 10: 79–108.
- Melzack, R., and Wall, P. 1988. *The Challenge of Pain*. 2nd ed. New York: Penguin.
- Merikle, Philip, and Joordens, Steve. 1997. Parallels between perception without attention and perception without awareness. *Consciousness and Cognition* 6: 219–236.
- Merikle, Philip, Smilek, Daniel, and Eastwood, John. 2001. Perception without awareness: Perspectives from cognitive psychology. In Dehaene 2001, 115–134.
- Milner, A. D., and Goodale, M. A. 1995. *The Visual Brain in Action*. Oxford: Oxford University Press.
- Nathan, P. 1985. Pain and nociception in the clinical context. *Philosophical Transactions of the Royal Society of London, Series B* 308: 219–226.

- O'Regan, J. K. 1992. Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology* 46: 461–488.
- Parvisi, J., and Damasio, A. 2001. Consciousness and the brainstem. In Dehaene 2001, 135–160.
- Prinz, J. J. 2000. A neurofunctional theory of visual consciousness. *Consciousness and Cognition* 9(2): 243–259.
- Pylyshyn, Z., and Storm, R. 1988. Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision* 3: 179–197.
- Rees, Geraint, Wojciulik, Ewa, Clarke, Karen, Husain, Masud, Frith, Chris, and Driver, Jon. 2000. Unconscious activation of visual cortex in the damaged right hemisphere of a parietal patient with extinction. *Brain* 123: 1624–1633.
- Shallice, Tim. 1975. On the contents of primary memory. In P. M. A. Rabbit and S. Dornic, eds., *Attention and Performance V*. London: Academic Press.
- Simons, Daniel. 2000. Attentional capture and inattention blindness. *Trends in Cognitive Science* 4(4): 147–155.
- Sperling, George. 1960. The information available in brief visual presentations. *Psychological Monographs* 74(11): 1–29.
- Tipper, S. P., and Behrman, M. 1996. Object-centered not scene-based visual neglect. *Journal of Experimental Psychology: Human Perception and Performance* 22(5): 1261–1278.
- Wojciulik, E., and Kanwisher, N. 1998. Implicit visual attribute binding following bilateral parietal damage. *Visual Cognition* 5: 157–181.
- Wolfe, J. M. 1999. Inattentional amnesia. In V. Coltheart, ed., *Fleeting Memories*. Cambridge, MA: MIT Press.