

14 How Not to Find the Neural Correlate of Consciousness

There are two concepts of consciousness that are easy to confuse with one another, access consciousness and phenomenal consciousness. However, just as the concepts of water and H₂O are different concepts of the same thing, so the two concepts of consciousness may come to the same thing in the brain. The focus of this paper is on the problems that arise when these two concepts of consciousness are conflated. I will argue that John Searle's reasoning about the function of consciousness goes wrong because he conflates the two senses. And Francis Crick and Christof Koch fall afoul of the ambiguity in arguing that visual area V1 is not part of the neural correlate of consciousness. Crick and Koch's work raises issues that suggest that these two concepts of consciousness may have different (though overlapping) neural correlates—despite Crick and Koch's implicit rejection of this idea.

I will start with two quotations from Searle. You will see what appears to be a contradiction, and I will later claim that the appearance of contradiction can be explained if one realizes that he is using two different concepts of consciousness. I am not going to explain yet what the two concepts of consciousness are. That will come later, after I have presented Searle's apparent contradiction and Crick and Koch's surprising argument.

Searle's Apparent Contradiction

Searle discusses my claim that there are two concepts of consciousness, arguing that I have confused modes of one kind with two different kinds:

There are lots of different degrees of consciousness, but door knobs, bits of chalk, and shingles are not conscious at all. . . . These points, it seems to me, are misunderstood by Block. He refers to what he calls an "access sense of consciousness." On my account there is no such sense. I believe that he . . . [confuses] what I would call peripheral consciousness or inattentiveness with total unconsciousness. It is true, for example, that when I am driving my car "on automatic pilot" I am not paying much attention to the details of the road and the traffic. *But it is simply not true that I am totally unconscious of these phenomena. If I were, there would be a car crash. We need*

therefore to make a distinction between the center of my attention, the focus of my consciousness on the one hand, and the periphery on the other. (Italics added)¹

Note that Searle claims that if I became unconscious of the road while driving, the car would crash. Now compare the next argument:

*The epileptic seizure rendered the patient totally unconscious, yet the patient continued to exhibit what would normally be called goal-directed behavior. . . . In all these cases, we have complex forms of apparently goal-directed behavior without any consciousness. Now why could all behavior not be like that? Notice that in the cases, the patients were performing types of actions that were habitual, routine and memorized. . . . normal, human, conscious behavior has a degree of flexibility and creativity that is absent from the Penfield cases of the unconscious driver and the unconscious pianist. Consciousness adds powers of discrimination and flexibility even to memorized routine activities. . . . one of the evolutionary advantages conferred on us by consciousness is the much greater flexibility, sensitivity, and creativity we derive from being conscious.*²

Note that according to the first quotation, if I were to become unconscious (and therefor unconscious of the road and traffic), my car would crash. But in the second quotation, he accepts Penfield's description "totally unconscious" as applying to the case of the petit mal patient who drives home while having a seizure. Thus we have what looks like a contradiction.

Crick and Koch's Argument

I will now shift to Crick and Koch's recent article in *Nature*³ arguing that V1 (the first major processing area for visual signals) is not part of the neural correlate of consciousness (what they call the NCC). Crick and Koch say that V1 is not part of the neural correlate of consciousness because V1 does not directly project to frontal cortex. (They extrapolate (tentatively) from the fact that no direct connections are known in macaques to no connections in humans.) Their reasoning makes use of the premise that part of the function of visual consciousness is to harness visual information in the service of the *direct* control of reasoning and decision-making that controls behavior. On the hypothesis that the frontal areas are involved in these mental functions, they argue that a necessary condition of inclusion in the NCC is direct projection to frontal areas. Though something seems right about their argument, it has nonetheless puzzled many readers. The puzzle is this: Why couldn't there be conscious activity based in V1 despite its lack of direct connection to frontal cortex? This is Pollen's⁴ worry: "I see no a priori necessity for neurons in perceptual space to communicate directly with those in decision space." The possibility of conscious activity based in V1 is especially salient in the light of Crick and Koch's suggestion that visual consciousness is reverberatory activity in pyramidal cells of the lower layers of the visual cortex involving connections to the thalamus.⁵ For one wonders how they have ruled out the possibility that such

activity *exists* in V1 despite the lack of direct connection between V1 and frontal cortex. They do not address this possibility at all. The overall air of paradox is deepened by their claim that that “Our hypothesis is thus rather subtle; if it [no direct connection] turns out to be true it [V1 is not part of the neural correlate of consciousness] will eventually come to be regarded as completely obvious” (p. 123). But the reader wonders why this is true at all, much less obviously true. When such accomplished researchers say such puzzling things, one has to wonder if one is understanding them properly.

I will argue that once the two concepts of consciousness are separated out, the argument turns out to be trivial on one reading and not clearly compelling on the other reading. That’s the critical part of my comment on Crick and Koch, but I have two positive points as well. I argue that nonetheless their conclusion about V1 should be accepted, but for a different reason, one that they implicitly suggest and that deserves to be opened up to public scrutiny. Further, I argue that the considerations that they raise suggest that the two concepts of consciousness correspond to different neural correlates despite Crick and Koch’s implicit rejection of this idea.

The Two Concepts

The two concepts of consciousness are *phenomenal* consciousness and *access* consciousness.⁶ Phenomenal consciousness is just *experience*; access consciousness is a kind of direct control. More exactly, a representation is access conscious if it is actively poised for direct control of reasoning, reporting and action.

One way to see the distinction between the two concepts is to consider the possibility of one without the other. Here is an illustration of access without phenomenal consciousness. In Anton’s Syndrome, blind patients do not realize that they are blind (though implicit knowledge of blindness can often be elicited). Hartmann et al.⁷ report a case of “Reverse Anton’s Syndrome” in which the patient does not realize that he is *not* really blind. The patient regards himself as blind, and he is at chance at telling whether a room is illuminated or dark. But he has a small preserved island of V1 that allows him to read single words and recognize faces and facial expressions if they are presented to the upper right part of the visual field. When asked how he knows the word or the face, he says “it clicks” and denies that he sees the stimuli. There is no obvious factor in his social situation that would favor lying or self-deception. In addition to the damage in V1, he has bilateral parietal damage, including damage to the left inferior parietal lobe. Milner and Goodale⁸ have proposed that phenomenal consciousness requires ventral stream activity plus attention, and that the requisite attention can be blocked by parietal lesions. So perhaps this is a case of visual access without visual phenomenal consciousness. (Note that Milner and Goodale’s account is not in conflict with Crick and Koch’s claim that V1 is not part of the NCC if activity in V1 is not the object of attentional processes.)

So we see that access consciousness without phenomenal consciousness makes sense and may even exist in a limited form. What about the converse, phenomenal consciousness without access? For an illustration at the conceptual level, consider the familiar phenomenon in which one notices that the refrigerator has just gone off. Sometimes one has the feeling that one has been hearing the noise all along, but without noticing it until it went off. One of the many possible explanations of what happens in such a case illustrates phenomenal consciousness without access consciousness: Before the refrigerator went off, you had the experience (phenomenal consciousness) of the noise (let us suppose) but there was insufficient attention directed toward it to allow direct control of speech, reasoning or action. There might have been *indirect* control (the volume of your voice increased to compensate for the noise) but not direct control of the sort that happens when a representation is poised for free use as a premise in reasoning and can be freely reported. (It is this free use that characterizes access consciousness.) On this hypothesis, there is a period in which one has phenomenal consciousness of the noise without access consciousness of it. Of course, there are alternative hypotheses, including more subtle ones in which there are degrees of access and degrees of phenomenality. One might have a moderate degree of both phenomenal consciousness of and access to the noise at first, then filters might reset the threshold for access, putting the stimulus below the threshold for direct control, until the refrigerator goes off and one notices the change. The degree of phenomenal consciousness and access consciousness may always match. Although phenomenal consciousness and access consciousness differ conceptually (as do the concepts of water and H₂O), we do not know yet whether or not they really come to the same thing in the brain.

Once one sees the distinction, one sees many pure uses of both concepts. For example, the Freudian unconscious is *access* unconscious. A repressed memory of torture in a red room could in principle be a phenomenally vivid image; what makes it unconscious in the Freudian sense is that it comes out in dreams, slips, fleeing from red rooms and the like rather than directly controlling behavior. Thus in principle an image can be unconscious in one sense (not poised for access), yet experienced and therefore conscious in another sense (phenomenally).

Searle's Contradiction

Let's go back to Searle's (apparent) contradiction. You will recall that he says that if he were to become unconscious of the details of the road and traffic, the car would crash: "When I am driving my car 'on automatic pilot' I am not paying much attention to the details of the road and the traffic. But it is simply not true that I am totally unconscious of these phenomena. If I were, there would be a car crash." But he also says that Penfield's famous unconscious driver is "totally unconscious" yet manages to drive home. (Searle has never suggested that he changed his mind on this issue between 1990 and 1992.) Note that there is no room for resolving the contradiction via appeal

to the difference between “conscious” and “conscious of.” If Penfield’s driver is “totally unconscious,” then he is not conscious *of* anything. And thus we have a conflict with the idea that if one were to become unconscious of the road and traffic, the car would crash. Can we resolve the contradiction by supposing that what Searle thinks is that *normally* if one were to become unconscious of the road the car would crash, but the Penfield case is an abnormal exception? Not likely, since Searle’s explicit conclusion is that consciousness adds flexibility, creativity, and sensitivity to action—suggesting that he thinks that consciousness is simply not necessary to routine activities like driving home.

I think that appeal to the access/phenomenal distinction does serve to resolve the contradiction. The resolution is that Searle is presupposing that the Penfield petit mal seizure case loses phenomenal consciousness but still has sufficient access consciousness to drive. But when he says that if he were unconscious of the road the car would crash, he is thinking of loss of both phenomenal and access consciousness—and it is the loss of the latter that would make the car crash.

I find that audiences I have talked to about this issue tend to divide roughly evenly. Some use “conscious” to mean phenomenal consciousness—to the extent that they control their uses. Others use “conscious” to mean either access consciousness or some kind of self-consciousness. But Searle’s error shows how easy it is for people to mix the two concepts together, whatever their official stance is.

How Crick and Koch’s Argument Depends on a Conflation

Crick and Koch argue that V1 is not part of the neural correlate of consciousness because V1 does not project to frontal cortex. Visual consciousness is used in harnessing visual information for directly guiding reasoning and decision making and direct projection to frontal cortex is required for such a use. But what concept of consciousness are Crick and Koch deploying? They face a dilemma. If they mean phenomenal consciousness, then their argument is extremely interesting but unsound: their conclusion is *unjustified*. If they mean access-consciousness, their argument is *trivial*. Let me explain.

Let us look at their argument more closely. Here it is:

1. Neural machinery of visual consciousness harnesses visual information for direct control of reasoning and decision making
2. Frontal areas subserve these functions
3. V1 does not project directly to frontal cortex
4. SO V1 is not part of neural correlate of consciousness

Note that the “direct” in premise 1 is necessary to generate the conclusion. But what reason is there to suppose that there cannot be *some* neural machinery of visual consciousness—V1, for example—that is part of the machinery of control of reasoning

and decision making, but only indirectly so? If by “consciousness” we mean phenomenal consciousness, there is no such reason, and so premise 1 is unjustified. But suppose we take “consciousness” to mean *access consciousness*. Then premise 1 is trivially true. Of course the neural machinery of access consciousness harnesses visual information for *direct* control since access consciousness just *is* direct control. But the trivial interpretation of premise 1 trivializes the argument. For to say that *if* V1 does not project directly to areas that control action, *then* V1 is not part of the neural correlate of access consciousness is to say something that is very like the claim that *if* something is a sleeping pill, then it is dormitive. Once Crick and Koch tell us that V1 is not directly connected to centers of control, nothing is added by saying that V1 is not part of the neural correlate of consciousness in the *access* sense. For an access-conscious representation just *is* one that is poised for the direct control of reasoning and decision making.

On this reading, we can understand Crick and Koch’s remark about their thesis that “if it [V1 is not directly connected to centers of control] turns out to be true it [V1 is not part of the neural correlate of consciousness] will eventually come to be regarded as completely obvious.” In the access-consciousness interpretation, this remark is like saying that if it turns out to be true that barbiturates cause sleep, their dormitivity will eventually come to be regarded as completely obvious.

To avoid misunderstanding, I must emphasize that I am not saying that it is a triviality that neurons in V1 are not directly connected to frontal areas. That is an empirical claim, just as it is an empirical claim that barbiturates cause sleep. What is trivial is that if neurons in V1 are not directly connected to centers of control, then neurons in V1 are not part of the neural correlate of access consciousness. Similarly, it is trivial that if barbiturates cause sleep, then they are dormitive.

That was the “access-consciousness” interpretation. Now let us turn to the phenomenal interpretation. On this interpretation, their claim is very significant, but not obviously true? How do we know whether activity in V1 is phenomenally conscious without being access conscious? As mentioned earlier, Crick and Koch’s own hypothesis that phenomenal consciousness is reverberatory activity in the lower cortical layers makes this a real possibility. They can hardly rule out this consequence of their own view by fiat. Crick and Koch⁹ say, “We know of no case in which a person has lost the whole prefrontal and premotor cortex, on both sides (including Broca’s area), and can still see.” But there are two concepts of seeing, just as there are two concepts of consciousness. If it is the phenomenal aspect of seeing that they are talking about, they are ignoring the real possibility that patients who have lost these frontal areas *can* see.

Crick and Koch attempt to justify the “directly” by appeal to representations on the retina. These representations control but not directly; and they are not conscious either. Apparently, the idea is that if representations do not control directly, then they are not conscious. But this example cuts no ice. Retinal representations have *neither*

phenomenal *nor* access consciousness. So they do not address the issue of whether V1 representations might have phenomenal but not access consciousness.

So Crick and Koch face a dilemma: their argument is either not substantive or not compelling.

Is the Point Verbal?

Crick and Koch often seem to have phenomenal consciousness in mind. For example, they orient themselves towards the problem of “a full accounting of the manner in which subjective experience arises from these cerebral processes. . . . Why do we experience anything at all? What leads to a particular conscious experience (such as the blueness of blue)? Why are some aspects of subjective experience impossible to convey to other people (in other words, why are they private)?”¹⁰

Crick and Koch often use “aware” and “conscious” as synonyms, as does Crick in *The Astonishing Hypothesis*. For example, the thesis of the paper in *Nature*³ is that V1 is not part of the neural correlate of consciousness and also that V1 is not part of the neural correlate of visual awareness. But sometimes they appear to use “awareness” to mean access-consciousness. For example, “All we need to postulate is that, unless a visual area has a direct projection to at least one of [the frontal areas], the activities in that particular visual area will not enter visual *awareness* directly, because the activity of frontal areas is needed to allow a person to report *consciousness*” (p. 122; emphases added). What could “consciousness” mean here? “Consciousness” cannot mean *access* consciousness, since reporting is a kind of accessing, and there is no issue of *accessing* access consciousness. Consciousness in the sense in which they mean it here is something that might conceivably exist even if it cannot be reported or otherwise accessed. And consciousness in this sense might exist in V1. Thus when they implicitly acknowledge an access/phenomenal consciousness distinction, the possibility of phenomenal without access consciousness looms large.

My point is not a verbal one. Whether we use “consciousness” or “phenomenal consciousness,” “awareness,” or “access consciousness,” the point is that there are two different concepts of the phenomenon or phenomena of interest. We have to acknowledge the possibility in principle that these two concepts pick out different phenomena. Two versus one: that is not a verbal issue.

Are the Neural Correlates of the Two Kinds of Consciousness Different?

Perhaps there is evidence that the neural correlate of phenomenal consciousness is exactly the same as the neural correlate of access consciousness? The idea that this is a conceptual difference without a real difference would make sense both of Crick and Koch as well as other work in the field. But paradoxically, the idea that the neural

correlates of the two concepts of consciousness coincide is one which Crick and Koch themselves actually give us reason to *reject*. Their hypothesis about the neural correlate of visual *phenomenal* consciousness is that it is localized in reverberatory circuits involving the thalamus and the lower layers of the visual cortex.³ This is a daring and controversial hypothesis. But it entails a much less daring and controversial conclusion: that the localization of visual phenomenal consciousness *does not involve the frontal cortex*. However, Crick and Koch think that the neural correlate of consciousness—and here we would be charitable in interpreting them as meaning access consciousness—*does* involve the frontal cortex. It would not be surprising if the brain areas involved in visual control of reasoning and reporting are not exactly the same as those involved in visual phenomenality.

One way for Crick and Koch to respond would be to include the neural correlates of *both* access and phenomenal consciousness in the “NCC.” To see what is wrong with this, consider an analogy. The first sustained empirical investigation of heat phenomena was conducted by the Florentine Experimenters in the seventeenth century. They did not distinguish between temperature and heat, using a single word, roughly translatable as “degree of heat,” for both. This failure to make the distinction generated paradoxes. For example, when they measured degree of heat by the test “Will it melt paraffin?” heat source A came out hotter than B, but when they measured degree of heat by how much ice a heat source could melt in a given time, B came out hotter than A.¹¹ The concept of degree of heat was a *mongrel* concept, one that lumps together things that are very different.⁶

The suggestion that the neural correlate of visual consciousness includes both the frontal lobes *and* the circuits involving the thalamus and the lower layers of the visual cortex would be like an advocate of the Florentine experimenters’ concept of degree of heat saying that the molecular correlate of degree of heat includes both *mean* molecular kinetic energy (temperature) and *total* molecular kinetic energy (heat), an idea that makes little sense. The right way to react to the discovery that a concept is a *mongrel*, is to distinguish distinct tracks of scientific investigation corresponding to the distinct concepts, not to lump them together.

Another way for Crick and Koch to react would be to include both the frontal lobes and the circuits involving the thalamus and the lower layers of the visual cortex in the neural correlate of *phenomenal* consciousness. (Koch seems inclined in this direction in correspondence.) But this would be like saying that the molecular correlate of *heat* includes both mean and total molecular kinetic energy. Again, this makes little sense. The criteria that Crick and Koch apply in localizing visual phenomenal consciousness are very fine grained, allowing them to emphasize cortical layers 4, 5, and 6 in the visual areas. For example, they appeal to a difference in those layers between cats that are awake and cats that are in slow-wave sleep, both exposed to the same visual stimuli. No doubt there are many differences between the sleeping and the waking cats in areas

outside the visual cortex. But we would need a very good reason to include any of those other differences in the neural correlate of visual phenomenology as opposed, say, to the nonphenomenal cognitive processing of visual information.

A Better Reason for not Including V1 in the NCC

Though I find fault with one strand of Crick and Koch's reasoning about V1, I think there is another strand in the paper that does justify the conclusion, but for a reason that it would be good to have out in the open and to distinguish from the reasoning just discussed. (Koch tells me that what I say in this paragraph is close to what they had in mind.) They note that it is thought that representations in V1 do not exhibit the Land effect (color constancy). But our experience, our phenomenal consciousness, does exhibit the Land effect, or so we would all judge. Similarly, it appears that neurons in V1 are sensitive to gratings that are finer than people judge they can make out. We should accept the methodological principle: *at this early stage of inquiry*, don't suppose that people are wildly wrong about their own experience. Following this principle and assuming that the claim that cells in V1 do not exhibit color constancy is confirmed, then we should accept for the moment that representations in V1 are not on the whole phenomenally conscious. This methodological principle is implicitly accepted throughout Crick's and Koch's work.

An alternative route to the same conclusion would be the assumption that the neural correlate of phenomenal consciousness is "part of" the neural correlate of access consciousness (and so there can be no phenomenal without access consciousness). Phenomenal consciousness is automatically "broadcast" in the brain, but perhaps there are other mechanisms of broadcasting. So even if the "reverse Anton's syndrome" case turns out to be access without phenomenal consciousness, Crick and Koch's conclusion might still stand.

Note that neither of the reasons given here make any use of the finding that V1 is not directly connected to frontal areas.

The assumption that phenomenal consciousness is part of access consciousness is very empirically risky. One empirical phenomenon that favors taking phenomenal without access consciousness seriously is the fact that phenomenal consciousness has a finer grain than access consciousness based on memory representations. For example, normal people can recognize no more than 80 distinct pitches, but it appears that the number of distinct pitch experiences is much greater. This is indicated (but not proven) by the fact that normal people can *discriminate* 1,400 different frequencies from one another.¹² There are many more phenomenal experiences than there are concepts of them.

Despite these disagreements, I greatly admire Crick's and Koch's work on consciousness and have written a very positive review of Crick's book.¹³ Crick has written "No

longer need one spend time . . . [enduring] the tedium of philosophers perpetually disagreeing with each other. Consciousness is now largely a scientific problem.”¹⁴ I think this conceptual issue shows that even if largely a scientific issue, it is not entirely one. There is still some value in a collaboration between philosophers and scientists on this topic.

Notes

This is a substantially revised version of a paper that appeared in *Trends in Neurosciences* 19(2): 456–459, 1996. I am grateful to audiences at the 1996 consciousness conference in Tucson, at the 1996 cognitive science conference at the University of Sienna, at the University of Oxford, Department of Experimental Psychology, at Union College Department of Philosophy, and at the Royal Institute of Philosophy. I am also grateful to Susan Carey, Francis Crick, Martin Davies, Christof Koch, and David Milner, as well as to the editor of *Trends in Neurosciences* for comments on a previous draft.

Reprinted in João Branquinho, ed., *The Foundations of Cognitive Science* (New York: Oxford University Press, 2001). Reprinted in *Intellectica* 31 (2000): 1–12. Reprinted in Anthony O’Hear, ed., *Current Issues in Philosophy of Mind*, Royal Institute of Philosophy Supplement (Cambridge: Cambridge University Press, 1998). German translation (“Das neurale Korrelat des Bewusstseins”) in Frank Esken and Heinz-Dieter Heckmann, eds., *Bewusstsein und Repraesentation* (Paderborn: Verlag Ferdinand Schoeningh, 1997). Italian translation, “Come trovare il correlato neurale della coscienza?”, in *Sistemi Intelligenti*, with comments by Italian philosophers, 1998. Reprinted in the proceedings of the 1996 conference on consciousness at Tucson S. Hameroff, A Kaszniak, and A. Scott, eds., *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates* (Cambridge: MIT Press, 1998). Spanish translation forthcoming in a book on consciousness edited by Jose Luis Blanco.

1. Searle, John. 1990. Who is computing with the brain? *Behavioral and Brain Sciences* 13(4): 632–634.
2. Searle, John. 1992. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
3. Crick, F., and Koch, C. 1995, May 11. Are we aware of neural activity in primary visual cortex? *Nature* 375: 121–123.
4. Pollen, D. 1995, September 28. Cortical areas in visual awareness. *Nature* 377: 293–294.
5. Crick, F. 1994. *The Astonishing Hypothesis*. New York: Scribner.
6. Block, N. 1995. On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18(2): 227–247.
7. Hartmann, J. A., et al. 1991. Denial of visual perception. *Brain and Cognition* 16: 29–40.
8. Milner, A. D., and Goodale, M. A. 1995. *The Visual Brain in Action*. Oxford: Oxford University Press.

9. Crick, F., and Koch, C. 1995. Untitled response to Pollen. *Nature* 377: 294–295.
10. Crick, F., and Koch, C. 1995, December. Why neuroscience may be able to explain consciousness. Sidebar in *Scientific American*, p. 92.
11. Wiser, M., and Carey, S. 1983. When heat and temperature were one. In D. Gentner and A. Stevens, eds., *Mental Models*. Mahwah, NJ: Erlbaum.
12. Raffman, D. 1995. On the persistence of phenomenology. In T. Metzinger, ed., *Conscious Experience*. Thorverton: Imprint Academic.
13. Block, N. 1996. Review of Francis Crick, *The Astonishing Hypothesis*. *Contemporary Psychology*, May.
14. Crick, F. 1996, August 2. Visual perception: Rivalry and consciousness. *Nature* 379: 485–486.

