

How can we find the neural correlate of consciousness?

Ned Block

There are two concepts of consciousness, access consciousness and phenomenal consciousness. But just as the concepts of water and H₂O are different concepts of the same thing, so the two concepts of consciousness might come to the same thing in the brain. Some recent papers by Crick and Koch raise issues that suggest that these two concepts of consciousness might have different (though overlapping) neural correlates, despite Crick and Koch's implicit rejection of this idea.

Trends Neurosci. (1996) 19, 456–459

FOR MANY YEARS, consciousness was an embarrassment to neuroscience. It was not a fit topic for research and only appeared in journals in the often wildly dualistic remarks of eminent neuroscientists who had developed a speculative bent later in life. An atmosphere developed in which it seemed that neuroscience would have nothing to say about what is perhaps the most salient aspect of our mental life. But in recent years, this neglect of consciousness has ended, even reversed. It is hard to keep up with the new books and articles on consciousness by neuroscientists, psychologists and philosophers, and one of the best features of the revival of interest in this fascinating and important subject is that the thinkers in different fields are actually paying attention to one another's contributions.

Some of the most interesting, level-headed and influential contributions to the new work have been by Francis Crick and Christof Koch. But there has been a persistent problem in their thinking about consciousness – one that is common in neuroscience – that comes out especially clearly in their recent article in *Nature*¹ arguing that visual area V1 is not part of the neural correlate of consciousness (what they call the NCC). This paper will argue that the explicit conclusion of that article differs importantly from its implicit upshot.

The puzzle

Crick and Koch maintain that area 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 have been identified in macaques to the lack of similar connections in humans.] Their reasoning makes use of the entirely reasonable premise that part of the function of visual consciousness is to harness visual information in the service of the 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 in V1 despite its lack of direct connection to frontal cortex? This is Pollen's² 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 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 thalamus³. 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.

In this note, I argue that these pieces all fit together once one realizes that Crick and Koch's argument trades on a conflation of two concepts of consciousness. Once the two concepts are separated out, the argument turns out to be trivial on one reading and not clearly compelling on the other reading. That is the negative point, 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 suggest implicitly and that deserves to be opened up to public scrutiny. Furthermore, 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 consciousness⁴. Phenomenal consciousness is just experience; access consciousness is a kind of direct control. More exactly, a representation is access-conscious if it is 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

Ned Block is at the Dept of Philosophy, New York University, New York, NY 10003-6688, USA.

elicited). Hartmann *et al.*⁵ 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 which 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 Goodale⁶ 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 might even exist in a limited form. What about the converse, phenomenal consciousness without access consciousness? For an illustration at the conceptual level, consider the familiar phenomenon in which one notices that the refrigerator has just switched off. Sometimes one has the feeling that one has been hearing the noise all along, but without noticing it until it turned off. One of the many possible explanations of what happens in such a case illustrates phenomenal consciousness without access consciousness: before the refrigerator turned off, you had the experience (phenomenal consciousness) of the noise (let us suppose) but there was insufficient attention directed towards 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 reported freely. (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 turns off and one notices the change. The degree of phenomenal consciousness and access consciousness might 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 uncon-

scious in the Freudian sense is that it comes out in dreams, slips, fleeing from red rooms and the like rather than controlling behavior directly. Thus, in principle, an image can be unconscious in one sense (not poised for access), yet experienced and therefore conscious in another sense (phenomenally).

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 guiding behavior 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, it is trivial. Let me explain.

Access consciousness

Suppose they mean access consciousness. Then the claim that V1 is not part of the neural correlate of consciousness amounts to little more than the evidential base that they appeal to, namely that events in V1 do not control behavior directly. Crick and Koch provide a number of considerations that seem to be meant to persuade us that if it is true that events in V1 do not control behavior directly, then those events are not conscious. If consciousness equals access consciousness, these considerations are not necessary because the argument is trivial. Once Crick and Koch tell us that V1 is not connected directly 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 is one that is poised for the direct control of (among other things) behavior. To add that V1 is not part of the neural correlate of access consciousness is like saying of a known sleeping pill that it is dormitive. On this reading, we can understand Crick and Koch’s remark about their thesis that ‘if it [V1 is not connected directly 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’. On 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 connected directly 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 connected directly to frontal areas, 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.

The triviality of the conditional – if neurons in V1 are not connected directly to frontal areas then they are not part of the neural correlate of access consciousness – can be seen very starkly in Crick and Koch’s assumption that visual consciousness is used in harnessing visual information for guiding behavior. The thesis that the neural machinery of phenomenal consciousness has this function is substantive and plausible, but the thesis that the neural machinery of

access consciousness has this function is trivial⁴. To say that the neural machinery of access consciousness is used in guiding behavior is saying that the neural machinery of access consciousness is used in a kind of access.

Phenomenal consciousness

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 Koch⁷ 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.

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)?’ (Ref. 8, p. 92).

Crick and Koch often use ‘aware’ and ‘conscious’ as synonyms, as does Crick in his book³ *The Astonishing Hypothesis*. For example, the thesis of the paper in *Nature*¹ 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’ (Ref. 1, p. 122). What could ‘consciousness’ mean here? ‘Consciousness’ cannot mean access consciousness, because 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 a distinction between access consciousness and phenomenal consciousness, 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 much of what Crick and Koch say and of much of the empirical work on consciousness. 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 cortex³. 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 access consciousness does involve the frontal cortex. Even if they are wrong about this, 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 17th 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 source B, but when they measured degree of heat by how much ice a heat source could melt in a given time, source B came out hotter than source A (Ref. 9). The concept of degree of heat was a mongrel concept, one that lumps together things that are very different⁴.

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). 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. 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 non-phenomenal 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. We should accept the methodological principle: at this early stage of inquiry, accept what people say 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 accepted implicitly 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 'broadcasted' in the brain, but perhaps there are other mechanisms of broadcasting. (Blindsight would be a weak example.) 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 both of the reasons given here are independent of

whether or not there is a direct connection between V1 and frontal areas. Even if there is a direct connection between V1 and frontal areas, given the point just made about color constancy, we would have reason for regarding the representations in the primary visual cortex that feed directly to frontal cortex as access conscious without being phenomenally conscious. Thus evidence for a direct connection would provide empirical support for the proposition that the neural correlates of the two concepts of consciousness are distinct.

The assumption that phenomenal consciousness is part of access consciousness is very risky empirically. 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 by the fact that normal people can discriminate 1400 different frequencies from one another¹⁰. There are many more phenomenal experiences than there are concepts of them.

Despite these disagreements, I admire Crick's and Koch's work on consciousness greatly and have written a very positive review of Crick's book¹¹. Crick has written¹² 'No longer need one spend time ...[enduring] the tedium of philosophers perpetually disagreeing with each other. Consciousness is now largely a scientific problem'. I think this conceptual issue shows that even if consciousness is largely a scientific issue, it is not entirely so. There is still some value in a collaboration between philosophers and scientists on this topic.

Selected references

- 1 Crick, F. and Koch, C. (1995) *Nature* 375, 121–123
- 2 Pollen, D. (1995) *Nature* 377, 293–294
- 3 Crick, F. (1994) *The Astonishing Hypothesis*, Scribners
- 4 Block, N. (1995) *Behav. Brain Sci.* 18, 227–247
- 5 Hartmann, J.A. et al. (1991) *Brain Cognit.* 16, 29–40
- 6 Milner, A.D. and Goodale, M.A. (1995) *The Visual Brain in Action*, Oxford University Press
- 7 Crick, F. and Koch, C. (1995) *Nature* 377, 294–295
- 8 Crick, F. and Koch, C. (1995) *Sci. Am.* 12, 92
- 9 Wiser, M. and Carey, S. (1983) *Mental Models: When Heat and Temperature Were One* (Gentner, D. and Stevens, A., eds), Erlbaum
- 10 Raffman, D. (1995) *Conscious Experience: On the Persistence of Phenomenology* (Metzinger, T., ed.), Schoningh
- 11 Block, N. (1996) *Contemp. Psychol.* 41, 427–429
- 12 Crick, F. (1996) *Nature* 379, 485–486

Acknowledgements

I am grateful to audiences at the 1996 Consciousness conference in Tucson, at the 1996 Cognitive Science conference at the University of Sienna, and at the Dept of Experimental Psychology, University of Oxford. I am also grateful to Francis Crick, Martin Davies, Christof Koch, David Milner and to the editor for comments on a previous draft.

TINS Editorial Policy

Trends in Neurosciences is the leading neuroscience review journal (ranked No. 1 in the Neurosciences category of *SCI Journals Citation Reports*® 1994, with an Impact Factor of 20.194), publishing timely and wide-ranging feature articles that enable neuroscientists to keep up to date in and around their own research fields.

Reviews form the foundations of each issue and offer concise and authoritative summaries of important areas of neuroscientific research; *Meeting Reports* and *Research News* articles describe recent developments within a particular field; *Perspectives* discuss areas of historical, social and more oblique interest – offering interesting insights into current research as it applies to neuroscience; *Viewpoints* include more controversial or synthetic ideas. The *Debate* section provides a series of alternative viewpoints on a currently hotly debated topic. *Perspectives on Disease* discuss how basic neurobiological studies can be applied to neurological disease; *Techniques* offers readers a guide to new methods and their use.

TINS articles are specially commissioned by the Editor, in consultation with the Advisory Editorial Board. All submissions are subject to peer and editorial review – commissioning does not guarantee publication.