Towards the Neuronal Substrate of Visual Consciousness

by

Christof Koch

This article is adapted from the chapter by C. Koch, entitled *Towards the neuronal substrate of visual consciousness*. This chapter appears in *Towards a Science of Consciousness: The First Tucson Discussions and Debates*, the Proceedings of the 1994 Tuscon conference that are edited by S.R. Hameroff, A.W. Kaszniak and A.C. Scott and that will be published by MIT Press in early 1996.

Abstract

What is the relationship between a visual percept and the underlying neuronal activity in parts of the brain? This manifesto reviews the theoretical framework of Crick and Koch for answering these questions based on the neuroanatomy and physiology of mammalian cortex and associated subcortical structures. This evidence suggests that primates are not directly aware of neural activity in primary visual cortex, although they may be aware of such activity in extrastriate cortical areas. Psychophysical evidence in humans supporting this hypothesis is discussed.

Basic Assumptions

What is the relationship between "awareness," in particular visual awareness, and neuronal activity in the nervous system? Within a larger context, this question is sometimes known as the "Mind-Body" problem, and has been asked since antiquity. Over the last six years, Francis Crick and the author have attempted to understand this relationship by focusing on the neural correlate of consciousness, the NCC. This chapter provides an overview of the framework we advocate and provides some of our more recent speculations. We have previously (Crick and Koch 1990a,b, 1992; Koch and Crick, 1995a,b) described our general approach to the problem of visual awareness. In brief, we believe the next important step is to find experimentally the
neural correlates of various aspects of visual awareness; that is how best to explain our subjective mental experience in terms of the behavior of large groups of nerve cells. At this early stage in our investigation we will not worry too much about many fascinating but at the moment unrewarding aspects of the problem, such as the exact function of visual awareness, what species do and what species do not have awareness, different forms of awareness (such as dreams, visual imagination etc.) and the deep problem of qualia.. We here restrict our attention mainly to results on humans and on the macaque monkey, since their visual systems appear to be somewhat similar and, at the moment, we cannot obtain all the information we need from either of them separately.

Our main assumption is that, at any moment, the firing of some but not all the neurons in what we call the visual cortical system (which includes the neocortex and the hippocampus as well as a number of directly associated structures, such as the visual parts of the thalamus and possibly the claustrum) correlates with visual awareness. Yet, visual awareness is highly unlikely to be caused by the firing of all neurons in this system that happen to respond above their background rate at any particular moment. If at any given point in time only 1% of all the neurons in cortex fire significantly, about one billion cells in sensory, motor and association cortices would be active and we would never be able to distinguish any particular event within this vast sea of active nerve cells. We strongly expect that the majority of neurons will be involved in carrying out computations, while only a much smaller number will express the results of these computations. It is probably only the latter that we become aware of. There is already preliminary evidence from the study of the firing of neurons during binocular rivalry that in area MT of the macaque monkey only a fraction of neurons follow the monkey’s percept (Logothetis and Schall 1989). We can thus usefully ask the question: what are the essential differences between those neurons whose firing does correlate with the visual percept and those whose firing does not. Are these “awareness" neurons of any particular cell type? Exactly where are they located, how are they connected and is there anything special about their patterns of firing?

At this point, in may be useful to state our fundamental assumptions. They are:

- 1. To be aware of an object or an event, the brain has to construct an explicit, multi-level, symbolic interpretation of part of the visual scene.

By explicit we mean that one such neuron (or a few closely associated ones) must be firing above background at that particular time in response to the feature they symbolize. The pattern of color dots on a TV screen, for instance, contains an “implicit" representation of, say, a person’s face, but only the dots and their
locations are made explicit here; an explicit face representation would correspond to a light that is wired up in such a manner that it responds whenever a face appears somewhere on the TV screen. By multi-level we mean, in psychological terms, different levels such as those that correspond, for example, to lines or to eyes or to faces. In neurological terms we mean, loosely, the different levels in the visual hierarchy (see Felleman and Van Essen 1991).

By symbolic, as applied to a neuron, we mean that neuron’s firing is strongly correlated with some “feature” of the visual world and thus symbolizes it (this use of the word “symbol” should not be taken to imply the existence of a homunculus who is looking at the symbol). The meaning of such a symbol depends not only on the neuron’s receptive field (i.e., what visual features the neuron responds to) but also to what other neurons it projects to (its projective field). Whether a neural symbol is best thought of as a scalar (one neuron) or a vector (a group of closely associated neurons as in population coding in the superior colliculus; Lee, Rohrer and Sparks 1988) is a difficult question that we shall not discuss here.

- 2. Awareness results from the firing of a coordinated subset of cortical (and possible thalamic) neurons that fire in some special manner for a certain length of time, probably for at least one- or two-hundred milliseconds. This firing needs to activate some type of short-term memory by either strengthening certain synapses or maintaining an elevated firing rate or both. Experimental studies involving short-term memory tasks in the temporal lobe of the monkey (Fuster and Jervey 1981) have provided evidence of elevated firing rates for the duration of the interval during which an item needs to be remembered. It is at present not possible to assess empirically to what extent synapses undergo a short-term change during a memory task in the animal.

We are assuming that the semi-global activity that corresponds to awareness has to last for some minimum time (of the order of 100 msec) and that events within that time window are treated by the brain as approximately simultaneous. An example would be the flashing for 20 msec of a red light followed immediately by 20 msec of a green light in the same position. The observer sees a transient yellow light (corresponding to the mixture red and green) and not a red light changing into a green light (Efron 1973). Other psychophysical evidence shows that visual stimuli of less than 120--130 msec produce perceptions having a subjective duration identical to those produced by stimuli of 120--130 msec (Efron 1970a,b).

- 3. Unless a neuron has an elevated firing rate and unless it fires as a member of such an (usually temporary) assembly, its firing will not directly symbolize some feature of awareness at that moment.
These ideas, taken together, place restrictions on what sort of changes can reach awareness. An example would be the awareness of movement in the visual scene. Both physiological and psychophysical studies have shown that movement is extracted early in the visual system as a primitive (by the so-called short-range motion system; Braddick 1980). We can be aware that something has moved (but not what has moved) because there are neurons whose firing symbolizes movement as such, being activated by certain changes in luminance. To know what has moved (as opposed to a mere change of luminance) there must be active neurons somewhere in the brain that symbolize, by their firing, that there has been a change of that particular character.

3.1 As a corollary, we formulate our activity principle: Underlying every direct perception is a group of neurons strongly firing in response to that stimulus that come to symbolize it. An example is the “Kanizsa triangle” illusion, in which three Pacmen are situated at the corners of a triangle, with their open mouths facing each other. Human observers see a white triangle with illusory lines, even though the intensity is constant between the Pacmen. As reported by von der Heydt, Peterhans and Baumgartner (1984), cells in V2 of the awake monkey strongly respond to such illusory lines. Another case is the filling-in of the blind spot in the retina (Fiorani et al. 1992). Since we do not have neurons that explicitly represent the blind spot and events within it, we are not aware of small objects whose image projects onto them and can only infer such objects indirectly.

A semi-global activity that corresponds to awareness does not itself symbolize a change within that short period of awareness unless such a change is made explicitly by some neurons whose firing makes up the semi-global activity (for what else but another group of neurons can express the notion that a change has occurred?). These ideas are very counter-intuitive and are not easy to grasp on first reading, since the “fallacy of the homunculus” slips in all too easily if one doesn’t watch out for it.

3.2 It follows that active neurons in the cortical system that do not take part in the semi-global activity at the moment can still lead to behavioral changes but without being associated with awareness. These neurons are responsible for the large class of phenomena that bypass awareness in normal subjects, such as automatic processes, priming, subliminal perception, learning without awareness and others (Tulving and Schacter 1990; Kihlstrom 1987) or take part in the computations leading up to awareness. In fact, we suspect that the majority of neurons in the cortical system at any given time are not directly associated with awareness!

The elevated firing activity of these neurons also, of course, explain blindsight and similar clinical phenomena where patients with cortical blindness can point fairly
accurately to the position of objects in their blind visual field (or detect motion or
color) while strenuously denying that they see anything (Weiskrantz 1986; Stoerig

We have argued (from the experiments on binocular rivalry) that the firing of some
cortical neurons does not correlate with the percept. It is conceivable that all cortical
neurons may be capable of participating in the representation of one percept or
another, though not necessarily doing so for all percepts. The secret of visual
awareness would then be the type of activity of a temporary subset of them,
consisting of all those cortical neurons which represent that particular percept at that
moment. An alternative hypothesis is that there are special sets of “awareness"
neurons somewhere in cortex (for instance, layer 5 bursting cells; see below).
Awareness would then result from the activity of these special neurons.

4. Such neurons must project directly to some part of the front of the cortex; in
particular, to those areas in front of the primary motor area (M1, also called area 4).
Such areas are called "frontal" or "prefrontal" but, unfortunately, the terminology is
usually not very precise. We have only recently (Crick and Koch, 1995a) discussed
the need for projections from the visual system to the front of the brain. Our basic
argument assumes that in going from one visual area to another further up in the
visual hierarchy--- that is, further away from the retina (Zeki and Shipp, 1988;
Felleman and Van Essen, 1991; Young, 1992), the information is recoded at each
step. This is certainly broadly compatible with the known fact that the "features" to
which a neuron responds become more complex in going from the primary visual
cortex, V1 (also called striate cortex, or area 17), to the higher levels in the visual
hierarchy, such as the inferotemporal cortical areas (Ungerleider and Mishkin, 1982;
Maunsell and Newsome, 1987).

Prefrontal Brain Areas and Planning

This last assumption is based on the broad idea of the biological usefulness of visual
awareness (or, strictly, of its neural correlate). This is: to produce the best current
interpretation of the visual scene, in the light of past experience either of ourselves or
of our ancestors (embodied in our genes), and to make it available, for a sufficient
time, to the parts of the brain which contemplate, plan and execute voluntary motor
outputs (of one sort or another).

Exactly how these prefrontal and premotor cortical areas operate is currently
unknown, though there is now fragmentary evidence about the behavior of some of
them. Even in the macaque, the details of neuroanatomical connections between all
these areas have not yet been worked out in as much detail as they have for most of
the visual areas of the macaque (Fuster, 1989; 1993, Barbas, 1992; Gerfen, 1994).
It is probably a general rule that the further---connectionwise---a prefrontal area is from the primary motor area M1, the longer the time scale of the planning it is engaged in (Fuster, 1989; Birbaumer et al., 1990). Moreover, these cortical areas are all heavily involved with the basal ganglia (which include the neostriatum, the globus pallidus and the substantia nigra), whose main function, we speculate, is to provide a bias back to these areas (as well as to the superior colliculus in the midbrain) to influence the next step in their processing; that is, to assist some behaviors that involve a sequence of activities. The subject is additionally complicated for humans because of our highly developed language system and its usefulness for expressing our "thoughts" (in silent speech, for example).

Fortunately, at this stage, the details of the behavior of these "frontal" areas need not concern us. All we need to postulate is that unless a visual area has a direct projection to at least one of them, the activities in that particular visual area will not enter visual awareness directly.

**Primary Visual Cortex and its Connections**

This lack of any such projection appears to be true for area V1 of the macaque monkey, the almost exclusive recipient of the output of the lateral geniculate nucleus (LGN). V1 has no direct projections to the frontal eye fields (part of area 8), nor to the broad prefrontal region surrounding and including the principal sulcus (see Table 3 in Felleman and Van Essen, 1991); nor, as far as we know, to any other "frontal" area. Nor, for that matter, does primary visual cortex in the monkey project to the caudate nucleus of the basal ganglia (Saint-Cyr et al., 1990) nor to the intralaminar nuclei of the thalamus (L.G. Ungerleider, personal communication), the claustrum (Sherk, 1986) or the brainstem (with the exception of a small projection from peripheral V1 to the pons; Fries, 1990). V1 does, of course, provide the dominant visual input to most of the posterior visual cortical areas, including V2, V3, V4 and area MT. Among subcortical targets, the lower layers of V1 strongly project to the superficial layers of the superior colliculus (Sparks, 1986), the lateral geniculate nucleus and to the inferior and lateral pulvinar nuclei of the thalamus (Ungerleider et al., 1984; Robinson and Petersen, 1992).

We think it unlikely that information sent along the pathway from V1 to the superior colliculus, responsible for controlling and initiating eye movements, can produce conscious visual awareness. There is a multistage pathway from V1 to the colliculus, from there to the (inferior) pulvinar and thence to higher visual areas. This pathway may be involved in visual attention (Robinson and Petersen, 1992); but according to our arguments, it is not sufficiently direct or strong to produce, by itself, vivid visual awareness of the neural activities in V1.
The pathway from V1 to the colliculus might possibly be used to produce involuntary eye movements, so that psychophysical tests, using eye movements as the response, might show a form of blindsight in which subjects respond above chance while denying that they see anything. It is also possible that this or other pathways can produce vague feelings of some sort of awareness.

**Primary Visual Cortex and Awareness**

Our hypothesis (Crick and Koch, 1995a,b) is too speculative to be convincing as it stands, since we are not yet confident as to how to think correctly about most of the operations of the brain, and especially about the detailed function of the so-called "back pathways." Many readers will find these suggestion counter-intuitive. We would ask them: Do you believe that you are directly aware of the activity in your retina? Of course, without your retinas, you cannot see anything. If you do not believe this, what is the argument that you are directly aware of the neural activity in V1?

To avoid misunderstanding, let us underline what our hypothesis does not say. We are not suggesting that the neural activity in V1 is unimportant. On the contrary, we believe the detailed processing in V1 is crucial for normal vision, though recent work (Barbur et al., 1993) has shown that V1 in at least one patient is not essential for some limited form of visual awareness related to motion perception. All we are hypothesizing is that the activity in V1 does not directly enter awareness. What does enter awareness, we believe, is some form of the neural activity in certain higher visual areas, since they do project directly to prefrontal areas. This seems well established for cortical areas in the fifth tier of the visual hierarchy, such as MT and V4. For areas in the intervening tiers, such as V2, V3, V3A, VP and PIP, we prefer to leave the matter open for the moment (see Table 3 in Felleman and Van Essen, 1991).

Our hypothesis was suggested by neuroanatomical data from the macaque monkey. For humans, we are less certain, due to the present miserable state of human neuroanatomy (Crick and Jones, 1993), but we surmise that our hypothesis, if true for the macaque monkey, is also likely to be true for apes and humans. To be established as correct, it also needs to fit with all the neurophysiological and psychological data. What kind of evidence would support it?

**Physiological and Psychophysical Evidence**

A possible example may make this clearer. It is well known that the color we perceive at one particular visual location is influenced by the wavelengths of the
light entering the eye from surrounding regions in the visual field (Land and McCann, 1971; Blackwell and Buchsbaum, 1988). This mechanism acts to partially compensate for the effects of differently colored illumination. A white patch surrounded by patches of many colors still looks fairly white even when illuminated by pink light. This form of (partial) color constancy is often called the Land effect.

It has been shown in the anesthetized monkey (Zeki, 1983; Schein and Desimone, 1990) that neurons in V4, but not in V1, exhibit the Land effect. As far as we know, the corresponding information is lacking for alert monkeys. Since we cannot voluntarily turn off the Land effect, it would follow---if the same results could be obtained in a behaving monkey---that it would not be directly aware of the "color" neurons in V1. Notice that if neurons in both V1 and V4 in the alert monkey did turn out to show the full Land effect, this would not, by itself, disprove our hypothesis, as we do believe that we are visually aware of certain neural activity in V4 that could be triggered by activity in V1.

Psychophysical experiments would support our hypothesis if they demonstrate that we are not aware of neuronal activity that is highly likely to occur in V1. Such experiments have been done recently by D. I. MacLeod and Sheng He (personal communication). In brief, they have shown that exposure to high-contrast gratings that are so finely spaced that they cannot be seen (that is, cannot be distinguished from a uniformly gray surface) can produce an orientation-selective loss in sensitivity of human subjects to slightly less finely spaced gratings that can be visually perceived. Due to neuronal convergence in higher cortical areas and the associated reduction in receptive field size, neurons sensitive to the very fine conditioning grating appear to be restricted to V1 (DeValois et al., 1982; Levitt et al., 1994). These psychophysical experiments are therefore compatible with the idea that certain neurons in V1 respond to very high spatial frequencies that we are not visually aware of. The support for our ideas would be greater if it were shown (by imaging methods such as PET or functional MRI) that these invisible gratings produced significant activity in V1 in humans. For an alert macaque, it might be possible to show experimentally that very finely spaced gratings, that activated certain neurons in V1, could not be reported by the monkey, although the animal could report less finely spaced ones.

There are neurons in V1 whose firing depends upon which eye the visual signal is coming through. Neurons higher in the visual hierarchy do not make this distinction; that is, they are typically binocular. We are certainly not vividly and directly aware of which eye we are seeing with (unless we close or obstruct one eye), though whether we have some very weak awareness of the eye of origin is more controversial (Pickersgill, 1961; Blake and Cormack, 1979). These well-known facts suggest that we are not vividly aware of much of what goes on in V1.
These ideas would not be disproved if it were shown convincingly that (for some people) V1 is activated during visual imagery tasks (see the debate among P.E. Roland and others in Roland et al., 1994). There is no obvious reason why such top-down effects should not reach V1. Such V1 activity would not, by itself, prove that we are directly aware of it, any more than the V1 activity produced there when our eyes are open proves this. This hypothesis (Crick and Koch, 1994), then is a somewhat subtle one, though we believe that if it turns out to be true, it will eventually come to be regarded as completely obvious. We hope that further neuroanatomical work will make it plausible for humans, and further neurophysiological studies will show it to be true for most primates. We have yet to track down the location and nature of the neural correlates of visual awareness. Our hypothesis, if correct, would narrow the search to areas of the brain farther removed from the sensory periphery.

Acknowledgments: All of the ideas discussed here were jointly developed with Dr. Francis Crick.

References


