

# Are we aware of neural activity in primary visual cortex?

Francis Crick\* & Christof Koch†

\* The Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA

† Computation and Neural Systems Program, 139-74, California Institute of Technology, Pasadena, California 91125, USA

**It is usually assumed that people are visually aware of at least some of the neuronal activity in the primary visual area, V1, of the neocortex. But the neuroanatomy of the macaque monkey suggests that, although primates may be aware of neural activity in other visual cortical areas, they are not directly aware of that in area V1. There is some psychophysical evidence in humans that supports this hypothesis.**

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. In recent years, we have been trying to understand this relationship from a neuronal, reductionist point of view by making the following two basic assumptions.

(1) For an animal to be aware of some aspect of a visual object, there has to be a group of similar neurons (to allow coarse coding) located somewhere in the cortical system (which includes the neocortex and the hippocampus, as well as directly associated parts, such as the thalamus), the firing of which is correlated with that feature in the visual scene. We shall not discuss here the so-called binding problem (that is, how different aspects of a single visual object are bound together), nor the idea that this binding may require correlated firings (such as the  $\gamma$  oscillations) of sets of neurons in different cortical areas or the detailed nature of the neuronal correlate of consciousness (NCC)<sup>1-5</sup>.

(2) 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). In general, this large expanse of cortex is called 'frontal' cortex and includes motor cortex, premotor and prefrontal areas.

We have argued elsewhere<sup>5-7</sup> that to be aware of an object or event the brain has to construct a multilevel, explicit, symbolic interpretation of part of the visual scene. By multilevel 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<sup>8</sup>.

What exactly do we mean by 'explicit'? This is not an easy question to answer. The pattern of coloured dots on a television screen, for instance, may contain an implicit representation of a person's face, but only the dots and their location are explicitly represented on the screen. A simple neural network (such as a one-layer perceptron) could be constructed to respond whenever a red dot appeared at one particular place on the screen, or even to a particular fixed set of dots. Such a simple network cannot be tuned up to respond to a particular face that might appear anywhere on the TV screen, and varied in size, orientation or facial expression.

It is tempting to say that an explicit representation of a feature means that in the brain there is only a single neuron (otherwise called a 'grandmother' cell) that responds to that particular feature, and to that feature alone. We think this very improbable, if only because the firing of a single neuron in the cortical system is, by itself, unlikely to activate strongly any of its postsynaptic targets. Furthermore, 'normal' visual input is unlikely to activate such a cell by itself without also activating its neighbours.

We believe that any particular visual feature will be coarse coded (that is, spread out over a set of similar neurons), and indeed there is some evidence for this in the case of faces<sup>9</sup>. These 'face neurons' do not respond to a single spot of light (as retinal ganglion cells do) but are activated mainly by face-like objects, each neuron of the set being activated in a somewhat different way. This type of firing of a set of neurons is considered to be the neural symbol for a face, and the collective firing of this set makes the face-like aspect of the face explicit. (This use of the term 'symbol' should not be taken to imply the existence of a disembodied homunculus who is looking at it.) The meaning of such a symbol depends not only on the receptive fields of these neurons but also to which other neurons they project (their projective fields), and exactly how they are connected to them.

The firing of a single retinal ganglion cell might represent explicitly a particular spot of light in the visual field, but could not represent a face. An explicit face representation can only occur at a higher level in the visual system, after much further processing, especially as the 'face-neurons' are found experimentally to be relatively indifferent to the location and orientation of the face in the visual field<sup>10</sup>; if they were not there would have to be too many of them. Note that a person is not necessarily aware of all such representations. Something more is required for awareness: see (2) above.

We have not previously discussed in detail the need for projections from the visual system to the front of the brain, although we have mentioned it in passing<sup>7</sup>. 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<sup>8,11,12</sup>), the information is recoded at each step. This is 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<sup>13,14</sup>.

A neuron that is firing, but whose axon has been inactivated by local anaesthetics, can contribute little, if anything, to immediate visual awareness, unless one favours some dualistic position<sup>15</sup>. Thus the main destination of the axon—loosely, its projective field—is clearly important.

## Prefrontal brain areas and planning

Our second 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 that contemplate, plan and execute voluntary motor outputs (of one sort or another).

Exactly how these prefrontal and premotor cortical areas operate is currently unknown, although there is now fragmentary evidence about the behaviour 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<sup>16-19</sup>.

It is probably a general rule that the further, in terms of the number of stages, a prefrontal area is from the primary motor area M1, the longer is the timescale of the planning in which it is engaged<sup>17,20</sup>. Moreover, these cortical areas are all heavily involved with the basal ganglia (which include the neostriatum, the globus pallidus and the substantia nigra), the main function of which may be 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 behaviours 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 behaviour 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, because the activity of frontal areas is needed to allow a person to report consciousness.

### Primary visual cortex and its connections

This lack of any direct projection to the frontal cortex 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 ref. 8), nor, as far as we know, to any other 'frontal' area (see Fig. 1). Nor does it project to the caudate nucleus of the basal ganglia<sup>21</sup>, to the intralaminar nuclei of the thalamus (L. G. Ungerleider, personal communication), the claustrum<sup>22</sup> or the brainstem, with the exception of a small projection from peripheral V1 to the pons<sup>23</sup>. V1 does, of course, provide the dominant visual input to most of the posterior visual cortical areas, including V2, V3, V4 and MT (also known as V5<sup>11</sup>). Among subcortical targets, V1 projects to the superior colliculus<sup>24</sup>, the LGN and the pulvinar<sup>25,26</sup>.

It is unlikely that information sent along the pathway from V1 to the superior colliculus (partly responsible for initiating and controlling eye movements) can produce conscious visual awareness. There is a multistage pathway going from V1 to the colliculus, from there to the pulvinar (a large, mainly visual, part of the thalamus) and thence to higher visual areas. This pathway may be involved in visual attention<sup>26</sup>, 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 psychophysical tests, using eye movements as the response, might show a form of blindsight<sup>27,28</sup> 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 is too speculative to be convincing as it stands, as 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'. Some readers may find our suggestion counterintuitive, partly because for many years V1 was the only visual cortical area that was worked on extensively. We would ask them: do you believe that you are directly aware of the activity in your retina? (Of course, without your retinae, you cannot see anything.) If you do not believe

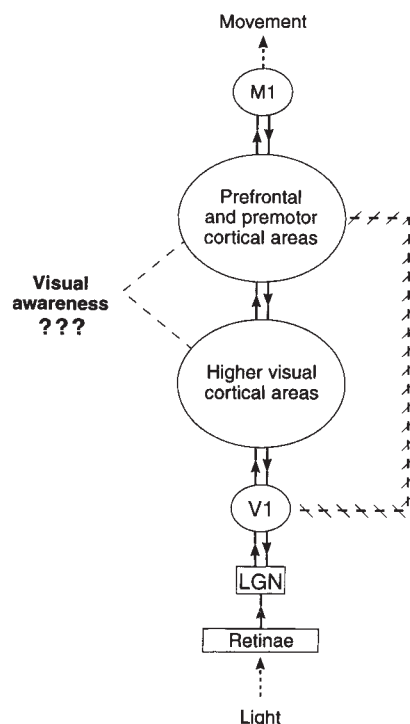


FIG. 1 An extremely schematic diagram of some of the brain areas involved. V1 is the primary visual cortex (also called the striate cortex or cortical area 17). M1 is the primary motor cortex (also called cortical area 4). LGN is the lateral geniculate nucleus of the thalamus. The main point of this paper is that the connection shown hatched on the right of the diagram is missing in the macaque monkey. Many relevant areas, such as the rest of the thalamus, the superior colliculus, the basal ganglia, the cerebellum, the brain stem and the claustrum, have been omitted for simplicity, as have the other sensory systems. The question marks indicate that, at the moment, we are unsure of the exact nature of visual awareness, although we argue here that it does not directly involve neurons in V1.

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 highly important for normal vision, although recent work<sup>29</sup> 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, because they 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 ref. 8).

Experiments suggest that only some of the active neurons in a cortical area are likely to produce direct visual awareness. During binocular rivalry<sup>30,31</sup>, the visual input is constant but the visual percept changes. Experiments on neurons in cortical area MT of the alert macaque monkey show that during binocular rivalry produced by two gratings moving in opposite directions, only a subset of the active neurons in cortical area MT follow the percept<sup>32</sup>. It will be of great importance to discover the layer in which these cells are located, their type and where they send their axons. The firing of most of the neurons in MT does not

depend on which of the two alternative percepts the monkey reports seeing.

Our hypothesis was suggested by neuroanatomical data from the macaque monkey. For humans, we are less certain because of the present miserable state of human neuroanatomy<sup>33</sup>, 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 colour perceived at one particular visual location is influenced by the wavelengths of the light entering the eye from surrounding regions in the visual field<sup>34,35</sup>. This mechanism acts partially to compensate for the effects of differently coloured illumination. A white patch surrounded by patches of many colours still looks fairly white even when illuminated by pink light. This form of (partial) colour constancy is often called the Land effect<sup>34</sup>.

It has been shown in the anaesthetized monkey<sup>36-38</sup> 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. Because people 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 'colour' 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 people 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 people 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 grey surface) can produce an orientation-selective loss in sensitivity of human subjects to slightly less finely spaced gratings that can indeed be perceived. Because of neuronal convergence in higher cortical areas and the associated increase in receptive field size, neurons sensitive to the very fine conditioning grating appear to be restricted to V1 (refs 39, 40). (This is one of the reasons why it is frequently assumed that we are aware of neural activity in V1.) These psychophysical experiments are therefore compatible with the idea that certain neurons in V1 respond to very high

spatial frequencies of which we are not visually aware. (We did not know of these results when we first formulated our hypothesis.) The support for our ideas would be greater if it were shown (by imaging methods such as positron-emission tomography (PET) or functional magnetic resonance imaging (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.

The firing of some neurons in V1 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. Most people are certainly not vividly and directly aware of which eye they are seeing with (unless they close or obstruct one eye), although whether they have some very weak awareness of the eye of origin is more controversial<sup>41</sup>. These well-known facts suggest that people are not vividly aware of much of the activity in V1.

Our ideas would not be disproved if it were shown convincingly that (for some people) V1 is activated during visual imagery tasks (for debate, see ref. 42). There is no obvious reason why such top-down effects should not reach V1. Such V1 activity would not by itself prove that people are directly aware of it, any more than the V1 activity produced there when their eyes are open proves this.

Theories are rarely proved or disproved by a single experiment; it needs a coherent body of interlocking data to establish a theory as correct. It could be argued, in opposition to our view, that there is indeed a small subset of V1 neurons the firing of which directly expresses some unique aspect of visual awareness. These special neurons could easily be missed, making it difficult to disprove our idea. If it turns out that in many other cortical areas only a particular type of neuron expresses visual awareness (for example, some of those in the lower cortical layers), and that neurons of a similar type in V1 respond in ways of which we are not aware, this might be sufficient proof of our hypothesis.

Our hypothesis is thus rather subtle; although 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<sup>43</sup>. Our hypothesis, if correct, would narrow the search to areas of the brain further removed from the sensory periphery. □

Received 19 December 1994; accepted 25 February 1995.

- Milner, P. M. *Psychol. Rev.* **81**, 521-535 (1974).
- von der Malsburg, C. & Schneider, W. *Biol. Cybern.* **54**, 29-40 (1986).
- Eckhorn, R. et al. *Biol. Cybern.* **60**, 121-130 (1988).
- Gray, C. M., König, P., Engel, A. K. & Singer, W. *Nature* **338**, 334-337 (1989).
- Crick, F. & Koch, C. *Semin. Neurosci.* **2**, 263-275 (1990).
- Crick, F. & Koch, C. *Scient. Am.* **267**, 152-159 (1992).
- Koch, C. & Crick, F. in *Large-scale Neuronal Theories of the Brain* (eds Koch, C. & Davis, J.) 93-109 (MIT Press, Cambridge, Massachusetts, 1994).
- Felleman, D. J. & Van Essen, D. *Cerebr. Cortex* **1**, 1-47 (1991).
- Young, M. P. & Yamane, S. *Science* **256**, 1327-1331 (1992).
- Desimone, R. *J. cogn. Neurosci.* **3**, 1-8 (1991).
- Zeki, S. & Shipp, S. *Nature* **335**, 311-317 (1988).
- Young, M. P. *Nature* **358**, 152-155 (1992).
- Ungerleider, L. G. & Mishkin, M. in *Analysis of Visual Behavior* (eds Ingle, D. J., Goodale, M. A. & Mansfield, R. J. W.) 549-586 (MIT Press, Cambridge, Massachusetts, 1982).
- Maunsell, J. H. R. & Newsome, W. T. A. *Rev. Neurosci.* **10**, 363-401 (1987).
- Eccles, J. C. *Proc. R. Soc. B* **227**, 411-428 (1986).
- Barbas, H. in *Advances in Neurology* Vol. 57 (eds Chauvel, P., Delgado-Escueta, A. V., Halgren, E. & Bancaud, J.) 91-115 (Raven, New York, 1992).
- Fuster, J. M. *The Prefrontal Cortex*, 2nd edn (Raven, New York, 1989).
- Fuster, J. M. *Curr. Opin. Neurobiol.* **3**, 160-165 (1993).
- Gerfen, C. R. in *Motor and cognitive functions of the prefrontal cortex* (eds Thierry, A.-M., Glowinski, J., Goldman-Rakic, P. S. & Christen, Y.) 78-92 (Springer, Berlin, 1994).
- Birbaumer, N., Elbert, T., Canavan, A. G. M. & Rockstroh, B. *Physiol. Rev.* **70**, 1-41 (1990).
- Saint-Cyr, J. A., Ungerleider, L. G. & Desimone, R. *J. comp. Neurol.* **298**, 129-156 (1990).
- Sherk, H. in *Cerebral Cortex* Vol. 5 (eds Jones, E. G. & Peters, A.) 467-499 (Plenum, New York, 1986).

- Fries, W. *Vis. Neurosci.* **4**, 205-216 (1990).
- Sparks, D. L. *Physiol. Rev.* **66**, 118-171 (1986).
- Ungerleider, L. G., Desimone, R., Galkin, T. W. & Mishkin, M. *J. comp. Neurol.* **223**, 368-386 (1984).
- Robinson, D. L. & Petersen, S. E. *Trends Neurosci.* **15**, 127-132 (1992).
- Cowey, A. & Stoerig, P. *Trends Neurosci.* **14**, 140-145 (1991).
- Cowey, A. & Stoerig, P. *Nature* **373**, 247-249 (1995).
- Barbur, J. L., Watson, J. D. G., Frackowiak, R. S. J. & Zeki, S. *Brain* **116**, 1293-1302 (1993).
- Blake, R. *Psychol. Rev.* **96**, 145-167 (1989).
- Myerson, J., Miezin, F. & Allman, J. *Behav. Analysis Lett.* **1**, 149-159 (1981).
- Logothetis, N. K. & Schall, J. D. *Science* **245**, 761-763 (1989).
- Crick, F. & Jones, E. *Nature* **361**, 109-110 (1993).
- Land, E. H. & McCann, J. J. *J. opt. Soc. Am.* **61**, 1-11 (1971).
- Blackwell, S. T. & Buchsbaum, G. *J. opt. Soc. Am.* **A5**, 1772-1780 (1988).
- Zeki, S. *Nature* **282**, 412-418 (1980).
- Zeki, S. *Proc. R. Soc. B* **217**, 449-470 (1983).
- Schein, S. J. & Desimone, R. *J. Neurosci.* **10**, 3369-3389 (1990).
- DeValois, R., Albrecht, D. G. & Thorell, L. G. *Vision Res.* **22**, 545-559 (1982).
- Levitt, J. B., Kiper, D. C. & Movshon, J. A. *J. Neurophysiol.* **71**, 2517-2542 (1994).
- Pickersgill, M. J. Q. *Jl exp. Psychol.* **11**, 168-172 (1961).
- Roland, P. E. & Gulyas, B., *inter alia Trends Neurosci.* **17**, 281-297 (1994).
- Crick, F. *The Astonishing Hypothesis* (Scribner's, New York, 1994).

ACKNOWLEDGEMENTS. F.C. is supported by the J. W. Kieckhefer Foundation. C.K. is supported by the Air Force Office of Scientific Research, the Office of Naval Research and the National Science Foundation. We thank L. G. Ungerleider, and D. I. MacLeod and S. He for allowing us to quote unpublished experimental work, and the Santa Fe Institute for their hospitality.