Clearing the Ground

We assume that when people talk about 'consciousness', there is something to be explained. While most neuroscientists acknowledge that consciousness exists, and that at present it is something of a mystery, most of them do not attempt to study it, mainly for one of two reasons:

1. They consider it to be a philosophical problem, and so best left to philosophers.
2. They concede that it is a scientific problem, but think it is premature to study it now.

We have taken exactly the opposite point of view. We think that most of the philosophical aspects of the problem should, for the moment, be left on one side, and that the time to start the scientific attack is now.

We can state bluntly the major question that neuroscience must first answer. It is probable that at any moment some active neuronal processes in your head correlate with consciousness, while others do not: what is the difference between them? In particular, are the neurons involved of any particular neuronal type? What is special (if anything) about their connections? And what is special (if anything) about their way of firing? The neuronal correlate of consciousness is often referred to as the NCC. Whenever some information is represented in the NCC it is represented in consciousness.

In approaching the problem, we made the tentative assumption (Crick and Koch, 1990) that all the different aspects of consciousness (pain, visual awareness, self-consciousness, and so on) employ a basic common mechanism or perhaps a few such mechanisms. If one could understand the mechanism for one aspect, then, we hope, we will have gone most of the way towards understanding them all.

We made the personal decision (Crick and Koch, 1990) that several topics should be set aside or merely stated without further discussion, for experience had shown us that valuable time can be wasted arguing about them without coming to any conclusion.

1. Everyone has a rough idea of what is meant by being conscious. For now, it is better to avoid a precise definition of consciousness because of the dangers of premature definition. Until the problem is understood much better, any attempt at a formal definition is likely to be either misleading or overly restrictive, or both. If this seems evasive, try defining the word 'gene'. So much is now known about genes that any simple definition is likely to be inadequate. How much more difficult, then, to define a biological term when rather little is known about it.
2. It is plausible that some species of animals – in particular the higher mammals – possess some of the essential features of consciousness, but not necessarily all. For this reason, appropriate experiments on such animals may be relevant to finding the mechanisms underlying consciousness. It follows that a language system (of the type found in humans) is not essential for consciousness – that is, one can have the key features of consciousness without language. This is not to say that language does not enrich consciousness considerably.
3. It is not profitable at this stage to argue about whether simpler animals (such as octopus, fruit flies, nematodes) or even plants are conscious (Nagel, 1997). It is probable, however, that consciousness correlates to some extent with the degree of complexity of any nervous system. When one clearly understands, both in detail and in principle, what consciousness involves in humans, then will be the time to consider the problem of consciousness in much simpler animals. For the same reason, we will not ask whether some parts of our nervous system have a special, isolated, consciousness of their own. If you say, ‘Of course my spinal cord is conscious but it’s not telling me’, we are not, at this stage, going to spend time arguing with you about it. Nor will we spend time discussing whether a digital computer could be conscious.
4. There are many forms of consciousness, such as those associated with seeing, thinking, emotion, pain, and so on. Self-consciousness – that is, the self-referential aspect of consciousness – is probably a special case of consciousness. In our view, it is better left to one side for the moment, especially as it would be difficult to study self-consciousness in a monkey. Various rather unusual states, such as the hypnotic state, lucid dreaming, and sleep walking, will not be considered here, since they do not seem to us to have special features that would make them experimentally advantageous.

Visual Consciousness

How can one approach consciousness in a scientific manner? Consciousness takes many forms, but for an initial scientific attack it usually pays to concentrate on the form that appears easiest to study. We chose visual consciousness rather than other forms, because humans are very visual animals and our visual percepts are especially vivid and rich in information. In addition, the visual input is often highly structured yet easy to control.

The visual system has another advantage. There are many experiments that, for ethical reasons, cannot be done on humans but can be done on animals. Fortunately, the visual system of
primates appears fairly similar to our own (Tootell et al., 1996),
and many experiments on vision have already been done on
animals such as the macaque monkey.

This choice of the visual system is a personal one. Other
neuroscientists might prefer one of the other sensory systems. It
is, of course, important to work on alert animals. Very light
anaesthesia may not make much difference to the response of
neurons in macaque V1, but it certainly does to neurons in
cortical areas like V4 or IT (inferotemporal).

Why Are We Conscious?
We have suggested (Crick and Koch, 1995a) that the biological
usefulness of visual consciousness in humans 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 this interpretation directly available, for
a sufficient time, to the parts of the brain that contemplate and
plan voluntary motor output, of one sort or another, including
speech.

Philosophers, in their carefree way, have invented a creature
they call a ‘zombie’, who is supposed to act just as normal people
do but to be completely unconscious (Chalmers, 1995). This
seems to us to be an untenable scientific idea, but there is now
suggestive evidence that part of the brain does behave like a
zombie. That is, in some cases, a person uses current visual input
to produce a relevant motor output, without being able to say
what was seen. Milner and Goodale (1995) point out that a frog
has at least two independent systems for action, as shown by
Ingle (1975). These may well be unconscious. One is used by
the frog to snap at small, prey-like objects, and the other for jumping
away from large, looming discs. Why does not our brain consist
simply of a series of such specialized zombie systems?

We suggest that such an arrangement is inefficient when very
many such systems are required. Better to produce a single but
complex representation and make it available for a sufficient
time to the parts of the brain that make a choice among many
different but possible plans for action. This, in our view, is what
seeing is about. As pointed out to us by Ramachandran and
Hirstein (1997), it is sensible to have a single conscious
interpretation of the visual scene, in order to eliminate
hesitation.

Milner and Goodale (1995) suggest that in primates there are
two systems, which we shall call the on-line system and the
seeing system. The latter is conscious, while the former, acting
more rapidly, is not. The general characteristics of these two
systems and some of the experimental evidence for them are
outlined below in the section on the on-line system. There is
anecdotal evidence from sports. It is often stated that a trained
tennis player reacting to a fast serve has no time to see the ball;
the seeing comes afterwards. In a similar way, a sprinter is
believed to start to run before he consciously hears the starting
pistol.

The Nature of the Visual Representation
We have argued elsewhere (Crick and Koch, 1995a) 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 eyes
or faces. In neurological terms, we mean, loosely, the different
levels in the visual hierarchy (Felleman and Van Essen, 1991).

The important idea is that the representation should be
explicit. We have had some difficulty getting this idea across
(Crick and Koch, 1995a). By an explicit representation, we mean
a smallish group of neurons which employ coarse coding, as it is
called (Ballard et al., 1983), to represent some aspect of the
visual scene. In the case of a particular face, all of these neurons
can fire to somewhat face-like objects (Young and Yamane,
1992). We postulate that one set of such neurons will be all of
type (say, one type of pyramidal cell in one particular layer
or sublayer of cortex), will probably be fairly close together, and
will all project to roughly the same place. If all such groups of
neurons (there may be several of them, stacked one above the
other) were destroyed, then the person would not see a face,
though he or she might be able to see the parts of a face, such as
the eyes, the nose, the mouth, etc. There may be other places in
the brain that explicitly represent other aspects of a face, such as
the emotion the face is expressing (Adolphs et al., 1994).

Notice that while the information needed to represent a face
is contained in the firing of the ganglion cells in the retina, there
is, in our terms, no explicit representation of the face there.

How many neurons are there likely to be in such a group? This
is not yet known, but we would guess that the number to
represent one aspect is likely to be closer to 10^2–10^3 than to
10^2–10^4.

A representation of an object or an event will usually consist of
representations of many of the relevant aspects of it, and these
are likely to be distributed, to some degree, over different parts
of the visual system. How these different representations are
bound together is known as the binding problem (von der
Malsburg, 1995).

Much neural activity is usually needed for the brain to
construct a representation. Most of this is probably unconscious.
It may prove useful to consider this unconscious activity as the
computations needed to find the best interpretation, while the
interpretation itself may be considered to be the results of these
computations, only some of which we are then conscious of. To
decide from our perception, the results probably have something
of a winner-take-all character.

As a working hypothesis we have assumed that only some
types of specific neurons will express the NCC. It is already
known (see the discussion under ‘Bistable Percepts’) that the
firing of many cortical cells does not closely correspond to what
the animal is currently seeing. An alternative possibility is that
the NCC is necessarily global (Greenfield, 1995). In one extreme
form this would mean that, at one time or another, any neuron in
cortex and associated structures could express the NCC. At this
point we feel it more fruitful to explore the simpler hypothesis –
that only particular types of neurons express the NCC – before
pursuing the more global hypothesis. It would be a pity to miss
the simpler one if it were true. As a rough analogy, consider a
typical mammalian cell. The way its complex behavior is
controlled and influenced by its genes could be considered to be
largely global, but its genetic instructions are localized, and
coded in a relatively straightforward manner.

Where is the Visual Representation?
The conscious visual representation is likely to be distributed
over more than one area of the cerebral cortex and possibly over
certain subcortical structures as well. We have argued (Crick
and Koch, 1995a) that in primates, contrary to most received
opinion, it is not located in cortical area V1 (also called the striate
cortex or area 17). Some of the experimental evidence in
support of this hypothesis is outlined below. This is not to say
that what goes on in V1 is not important, and indeed may be
crucial, for most forms of vivid visual awareness. What we
suggest is that the neural activity there is not directly correlated with what is seen.

We have also wondered (Crick, 1994) whether the visual representation is largely confined to certain neurons in the lower cortical layers (layers 5 and 6). This hypothesis is still very speculative.

What is Essential for Visual Consciousness?
The term ‘visual consciousness’ almost certainly covers a variety of processes. When one is actually looking at a visual scene, the experience is very vivid. This should be contrasted with the much less vivid and less detailed visual images produced by trying to remember the same scene. (A vivid recollection is usually called a hallucination.) We are concerned here mainly with the normal vivid experience. (It is possible that our dimmer visual recollections are mainly due to the back pathways in the visual hierarchy acting on the random activity in the earlier stages of the system.)

Some form of very short-term memory seems almost essential for consciousness, but this memory may be very transient, lasting for only a fraction of a second. Edelman (1989) has used the striking phrase ‘the remembered present’ to make this point. The existence of iconic memory, as it is called, is well-established experimentally (Coltheart, 1983; Gegenfurtner and Sterling, 1993).

Psychophysical evidence for short-term memory (Potter, 1976; S. Subramaniam, I. Biederman and S.A. Madigan, submitted) suggests that if we do not pay attention to some part or aspect of the visual scene, our memory of it is very transient and can be overwritten (masked) by the following visual stimulus. This probably explains many of our fleeting memories when we drive a car over a familiar route. If we do pay attention (e.g. a child running in front of the car) our recollection of this can be longer lasting.

Our impression that at any moment we see all of a visual scene very clearly and in great detail is illusory, partly due to ever-present eye movements and partly due to our ability to use the scene itself as a readily available form of memory, since in most circumstances the scene usually changes rather little over a short span of time (O’Regan, 1992).

Although working memory (Baddeley, 1992; Goldman-Rakic, 1995) expands the time frame of consciousness, it is not obvious that it is essential for consciousness. It seems to us that working memory is a mechanism for bringing an item, or a small sequence of items, into vivid consciousness, by speech, or silent speech, for example. In a similar way, the episodic memory enabled by the hippocampal system (Zola-Morgan and Squire, 1993) is not essential for consciousness, though a person without it is severely handicapped.

Consciousness, then, is enriched by visual attention, though attention is not essential for visual consciousness to occur (Rock et al., 1992; Braun and Julesz, 1997). Attention is broadly of two types: bottom-up, caused by the sensory input; and top-down, produced by the planning parts of the brain. This is a complicated subject, and we will not try to summarize here all the experimental and theoretical work that has been done on it.

Visual attention can be directed to either a location in the visual field or to one or more (moving) objects (Kanwisher and Driver, 1992). The exact neural mechanisms that achieve this are still being debated. In order to interpret the visual input, the brain must arrive at a coalition of neurons whose firing represents the best interpretation of the visual scene, often in competition with other possible but less likely interpretations; and there is evidence that attentional mechanisms appear to bias this competition (Luck et al., 1997).

Recent Experimental Results
We shall not attempt to describe all the various experimental results of direct relevance to the search for the neuronal correlates of visual consciousness in detail but rather outline a few of them and point the reader to fuller accounts.

Action without Seeing

Classical blindsight
This will already be familiar to most neuroscientists. It is discussed, along with other relevant topics, in an excellent book by Weiskrantz (1997). It occurs in humans (where it is rare) when there is extensive damage to cortical area V1 and has also been reproduced in monkeys (Cowey and Stoerig, 1995). In a typical case, the patient can indicate, well above chance level, the direction of movement of a spot of light over a certain range of speed, while denying that he sees anything at all. If the movement is less salient, his performance falls to chance; if more salient (that is, brighter or faster), he may report that he had some ill-defined visual percept, considerably different from the normal one. Other patients can distinguish large, simple shapes or colors. (For Weiskrantz’s comments on Gazzaniga’s criticisms, see pp. 152–153; and on Zeki’s criticisms, see pp. 247–248.)

The pathways involved have not yet been established. The most likely one is from the superior colliculus to the pulvinar and from there to parts of visual cortex; several other known weak anatomical pathways from the retina and bypassing V1 are also possible. Recent functional magnetic resonance imaging of the blindsight patient G.Y. directly implicates the superior colliculus as being active specifically when G.Y. correctly discriminates the direction of motion of some stimulus without being aware of it at all (Sahraie et al., 1997 – this paper should be consulted for further details of the areas involved).

The On-line System
The broad properties of the two hypothetical systems — the on-line system and the seeing system — are shown in Table 1, following the account by Milner and Goodale in their book, The Visual Brain in Action (1995), to which the reader is referred for a more extended account. For a recent review, see Boussaoud et al. (1996). The on-line system may have multiple subsystems (for eye movements, for arm movements, for body posture adjustment, and so on). Normally, the two systems work in parallel, and indeed there is evidence that in some circumstances the seeing system can interfere with the on-line system (Rossetti, 1998).

One striking piece of evidence for an on-line system comes from studies on patient D.F. by Milner, Perrett and their colleagues (1991). Her brain has diffuse damage produced by carbon monoxide poisoning. She is able to see color and texture very well but is very deficient in seeing orientation and form. In spite of this, she is very good at catching a ball. She can ‘post’ her hand or a card into a slot without difficulty, though she cannot report the slot’s orientation.

It is obviously important to discover the difference between the on-line system, which is unconscious, from the seeing system, which is conscious. Milner and Goodale (1995) suggest that the on-line system mainly uses the dorsal visual stream. They propose that rather than being the ‘where’ stream, as suggested by Ungerleider and Mishkin (1982), it is really the ‘how’ stream.
This might imply that all activity in the dorsal stream is unconscious. The ventral stream, on the other hand, they consider to be largely conscious. An alternative suggestion, due to Steven Wise (personal communication and Boussaoud et al., 1996), is that direct projections from parietal cortex into premotor areas are unconscious, whereas projections to them via prefrontal cortex are related to consciousness.

Our suspicion is that while these suggestions about two systems are on the right lines, they are probably oversimple. The little that is known of the neuroanatomy would suggest that there are likely to be multiple cortical streams, with numerous anatomical connections between them (Distler et al., 1993). This is implied in Figure 1, a diagram often used by Fuster (Fuster, 1997: see his Fig. 8.4). In short, the neuroanatomy does not suggest that the sole pathway goes up to the highest levels of the visual system, and from there to the highest levels of the prefrontal system and then down to the motor output. There are numerous pathways from most intermediate levels of the visual system to intermediate frontal regions.

We would therefore like to suggest a general hypothesis: that the brain always tries to use the quickest appropriate pathway for the situation at hand. Exactly how this idea works out in detail remains to be discovered. Perhaps there is competition, and the fastest stream wins. The postulated on-line system would be the quickest of these hypothetical cortical streams. This would be the zombie part of you.

**Bistable Percepts**

Perhaps the present most important experimental approach to finding the NCC is to study the behavior of single neurons in the monkey’s brain when it is looking at something that produces a bistable percept. The visual input, apart from minor eye movements, is constant; but the subject’s percept can take one of two alternative forms. This happens, for example, when one looks a drawing of the well-known Necker cube.

It is not obvious where to look in the brain for the two alternative views of the Necker cube. Allman suggested a more practical alternative: to study the responses in the visual system during binocular rivalry (Myerson et al., 1981). If the visual input into each eye is different, but perceptually overlapping, one usually sees the visual input as received by one eye alone, then by the other one, then by the first one, and so on. The input is constant, but the percept changes. Which neurons in the brain mainly follow the input, and which the percept?

This approach has been pioneered by Logothetis and his colleagues working on the macaque visual system. They trained the monkey to report which of two rival inputs it saw. The experiments are difficult, and elaborate precautions had to be taken to make sure the monkey was not cheating. The fairly similar distribution of switching times strongly suggests that monkeys and humans perceive these bistable visual inputs in the same way.

The first set of experiments (Logothetis and Schall, 1989) studied neurons in cortical area MT (medial temporal, also called V5), since they preferentially respond to movement. The stimuli were vertically drifting horizontal gratings. Only the first response was recorded. Of the relevant neurons, only ~35% were modulated according to the monkey’s reported percept. Surprisingly, half of these responded in the opposite direction to the one expected.

The second set of experiments (Leopold and Logothetis, 1996) used stationary gratings. The orientation was chosen in each case to be optimal for the neuron studied, and orthogonal to it in the other eye. They recorded how the neuron fired during several alterations of the reported percept. The neurons were in foveal V1/V2 and in V4. The fraction following the percept in V4 was similar to that in MT, but a rather smaller fraction of V1/V2 neurons followed the percept. Also, here, but not in V4, none of the cells were anticorrelated with the stimulus.

The results of the third set of experiments (Sheinberg and Logothetis, 1997) were especially striking. In this case the visual inputs tried included images of humans, monkeys, apes, wild animals, butterflies, reptiles and various man-made objects. The rival image was usually a sunburst-like pattern (see Fig. 2). If a new image was flashed into one eye while the second eye was fixating another pattern, the new stimulus was the one that was always perceived (‘flash suppression’). Recordings were made in the upper and lower banks of the superior temporal sulcus (STS) and inferior temporal cortex (IT). Overall, ~90% of the recorded neurons in STS and IT were found to reliably predict the perceptual state of the animal. Moreover, many of these neurons responded in an almost all-or-none fashion, firing strongly for one percept, yet only at noise level for the alternative one.

More recently, Bradley et al. (1998) have studied a different...
bistable percept in macaque MT, produced by showing the monkey, on a television screen, the two-dimensional projection of a transparent, rotating cylinder with random dots on it, without providing any stereoscopic disparity information. Human subjects exploit structure-from-motion and see a three-dimensional cylinder rotating around its axis. Without further clues, the direction of rotation is ambiguous and observers first report rotation in one direction, a few seconds later, rotation in the other direction, and so on. The trained monkey responds as if it saw the same alternation. In their studies on the monkey, about half the relevant MT neurons Bradley et al. recorded from followed the percept (rather than the ‘constant’ retinal stimulus).

These are all exciting experiments, but they are still in the early stages. Just because a particular neuron follows the percept, it does not automatically imply that its firing is part of the NCC. The NCC neurons may be mainly elsewhere, such as higher up in the visual hierarchy. It is obviously important to discover, for each cortical area, which neurons are following the percept (Crick, 1996). That is, what type of neurons are they, in which cortical layer or sublayer do they lie, in what way do they fire, and, most important of all, where do they project? It is, at the moment, technically difficult to do this, but it is essential to have this knowledge, or it will be almost impossible to understand the neural nature of consciousness.

Figure 2. The activity of a single neuron in the superior temporal sulcus (STS) of a macaque monkey in response to different stimuli presented to the two eyes (taken from Sheinberg and Logothetis, 1997). In the upper left panel a sunburst pattern is presented to the right eye without evoking any firing response (ineffective stimulus). The same cell will fire vigorously in response to its ‘effective’ stimulus, here the image of a monkey’s face (upper right panel). When the monkey is shown the face in one eye for a while, and the sunburst pattern is flashed onto the monitor for the other eye, the monkey signals that it is ‘seeing’ this new pattern and that the stimulus associated with the rival eye is perceptually suppressed (‘flash suppression’; lower left panel). At the neuronal level, the cell shuts down in response to the ineffective yet perceptual dominant stimulus following stimulus onset (at the dotted line). Conversely, if the monkey fixates the sunburst pattern for a while, and the image of the face is flashed on, it reports that it perceives the face, and the cell will now fire strongly (lower right panel). Neurons in V4, earlier in the cortical hierarchy, are largely unaffected by perceptual changes during flash suppression.

Electrical Brain Stimulation
An alternative approach, with roots going back to Penfield (1958), involves directly stimulating cortex or related structures in order to evoke a percept or behavioral act. Libet and his colleagues (Libet, 1993) have used this technique to great advantage on the somatosensory system of patients. They established that a stimulus, at or near threshold, delivered through an electrode placed onto the surface of somatosensory cortex or into the ventrobasal thalamus required a minimal stimulus duration (between 0.2 and 0.5 s) in order to be consciously perceived. Shorter stimuli were not perceived, even though they could be detected with above-chance probability, using a two-alternative, forced-choice procedure. In contrast, a skin or peripheral sensory-nerve stimulus of very short duration could be perceived. The difference appears to reside in the amount and type of neurons recruited during peripheral stimulation versus direct central stimulation. Using sensory events as a marker, Libet also established (1993) that events caused by direct cortical stimulation were backdated to the beginning of the stimulation period.

In a series of classical experiments, Newsome and colleagues (Britten et al., 1992) studied the macaque monkey’s performance in a demanding task involving visual motion discrimination. They established a quantitative relationship between the performance of the monkey and the neuronal
discharge of neurons in its medial temporal cortex (MT). In 50% of all the recorded cells, the psychometric curve — based on the behavior of the entire animal — was statistically indistinguishable from the neurometric curve — based on the averaged firing rate of a single MT cell. In a second series of experiments, cells in MT were directly stimulated via an extracellular electrode (Salzman et al., 1990) (MT cells are arranged in columnar structure for direction of motion). Under these conditions, the performance of the animal shifted in a predictable manner, compatible with the idea that the small brain stimulation caused the firing of enough MT neurons, encoding for motion in a specific direction, to influence the final decision of the animal. It is not clear, however, to what extent visual consciousness for this particular task is present in these highly overtrained monkeys.

**The V1 Hypothesis**

We have argued (Crick and Koch, 1995a) that one is not directly conscious of the features represented by the neural activity in primary visual cortex. Activity in V1 may be necessary for vivid and veridical visual consciousness (as is activity in the retinas), but we suggest that the firing of none of the neurons in V1 directly correlates with what we consciously see. [For a critique of our hypothesis, see Pollen (1995), and our reply, Crick and Koch (1995b).]

Our reasons are that at each stage in the visual hierarchy the explicit aspects of the representation we have postulated are always recoded. We have also assumed that any neurons expressing an aspect of the NCC must project directly, without recoding, to at least some of the parts of the brain that plan voluntary action — that is what we have argued seeing is for. We think that these plans are made in some parts of frontal cortex (see below).

The neuroanatomy of the macaque monkey shows that V1 cells do not project directly to any part of frontal cortex (Crick and Koch, 1995a). Nor do they project to the caudate nucleus of the basal ganglia (Saint-Cyr et al., 1990), the intralaminar nuclei of the thalamus (L.G. Ungerleider, personal communication), the claustrum (Sherk, 1986) nor to the brainstem, with the exception of a small projection from peripheral V1 to the pons (Fries, 1990). It is plausible, but not yet established, that this lack of connectivity is also true for humans.

The strategy to verify or falsify this and related hypotheses is to relate the receptive field properties of individual neurons in V1 or elsewhere to perception in a quantitative manner. If the structure of perception does not map to the receptive field properties of V1 cells, it is unlikely that these neurons directly give rise to consciousness. In the presence of a correlation between perceptual experience and the receptive field properties of one or more groups of V1 cells, it is unclear whether these cells just correlate with consciousness or directly give rise to it. In that case, further experiments need to be carried out to untangle the exact relationship between neurons and perception.

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 Bushbaum, 1988). This form of (partial) color constancy is often called the Land effect. It has been shown in the anesthetized monkey (Zeki, 1980, 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. If the same results could be obtained in a behaving monkey, it would follow that it would not be directly aware of the ‘color’ neurons in V1.

**Some Experimental Support**

In the last two years, a number of psychophysical, physiological and imaging studies have provided some support for our hypothesis, although this evidence falls short of proving it (He et al., 1995; Kolb and Braun, 1995; Cumming and Parker, 1997; summarized in Koch and Braun, 1996; but see Morgan et al., 1997). Let us briefly discuss two other cases.

When two isoluminant colors are alternated at frequencies beyond 10 Hz, humans perceive only a single fused color with a minimal sensation of brightness flicker. In spite of the perception of color fusion, color opponent cells in primary visual cortex of two alert macaque monkeys follow high-frequency flicker well above heterochromatic fusion frequencies (Gur and Snodderly, 1997). In other words, neuronal activity in V1 can clearly represent certain retinal stimulation yet is not perceived. This is supported by recent fMRI studies on humans by Engel et al. (1997).

The study by He et al. (1996) is based on a common visual aftereffect (see Fig. 5a). If a subject stares for a fraction of a minute at a horizontal grating, and is then tested with a faint grating at the same location to decide whether it is oriented vertically or horizontally, the subject’s sensitivity for detecting a horizontal grating will be reduced. This adaptation is orientation specific — the sensitivity for vertical gratings is almost unchanged — and disappears quickly. He and colleagues projected a single patch of gratings onto a computer screen some 25° from the fixation point. It was clearly visible and their subjects showed the predictable orientation-selective adaptation effect. Adding one or more similar patches of gratings to either side of the original grating — which remained exactly as before — removed the lines of the grating from visibility; it was now ‘masked’. Subjectively, one still sees ‘something’ at the location of the original grating, but one is unable to make out its orientation, even when given unlimited viewing time. Yet despite this inability to see the adapting stimulus, the aftereffect was as strong and as specific to the orientation of the ‘invisible’ grating as when the grating was visible (see Fig. 36). What this shows, foreshadowed by earlier experiments (Blake and Fox, 1974), is that visual awareness in such cases must occur at a higher stage in the visual hierarchy than orientation-specific adaptation. This aftereffect is thought to be mediated by oriented neurons in V1 and beyond, implying that at least in this case the neurons which mediate visual awareness must be located past this stage.

Our ideas regarding the absence of the NCC from V1 are not disproven by positron emission tomography experiments showing that in at least some people V1 is activated during visual imagery tasks (Kosslyn et al., 1995), though severe damage to V1 is compatible with visual imagery in patients (Goldenberg et al., 1995). 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. We hope that further neuroanatomical work will make our hypothesis plausible for humans, and that further neurophysiological studies will show it to be true for most primates. If correct, it would narrow the search to areas of the brain farther removed from the sensory periphery.

**The Frontal Lobe Hypothesis**

As mentioned several times, we hypothesize that the NCC must
have access to explicitly encoded visual information and directly project into the planning stages of the brain, associated with the frontal lobes in general and with prefrontal cortex in particular (Fuster, 1997). We would therefore predict that patients unfortunate enough to have lost their entire prefrontal cortex on both sides (including Broca's area) would not be visually conscious, although they might still have well-preserved, but unconscious, visual-motor abilities. No such patient is known to us (not even Brickner's famous patient; for an extensive discussion of this, see Damasio and Anderson, 1993). The visual abilities of any such 'frontal lobe' patient would need to be carefully evaluated using a battery of appropriate psychophysical tests.

The fMRI study of the blindsight patient G.Y. (Sahraie et al., 1997) provides direct evidence for our view by revealing that prefrontal areas 46 and 47 are active when G.Y. is visually aware of a moving stimulus.

Large-scale lesion experiments carried out in the monkey suggest that the absence of frontal lobes leads to complete blindness (Nakamura and Mishkin, 1980, 1986). One would hope that future monkey experiments reversibly inactivate specific prefrontal areas and demonstrate the specific loss of abilities linked to visual perception while visual-motor behaviors – mediated by the on-line system – remain intact.

It will be important to study the pattern of connections between the highest levels of the visual hierarchy – such as inferotemporal cortex – and premotor and prefrontal cortex. In particular, does the anatomy reveal any feedback loops that might sustain activity between IT and prefrontal neurons (Crick and Koch, 1998)? There is suggestive evidence (Webster et al., 1994) that projections from prefrontal cortex back into IT might terminate in layer 4, but these need to be studied directly.

**Gamma Oscillations**

Much has been made of the presence of oscillations in the gamma range (30–70 Hz) in the local-field potential and in multi-unit recordings in the visual and sensory-motor system of cats and primates (Singer and Gray, 1995). The existence of such oscillations remains in doubt in higher visual cortical areas (Young et al., 1992). We remain agnostic with respect to the relevance of these oscillations to conscious perception. It is possible that they subserve attention or figure-ground in early visual processing.

**Philosophical Matters**

There is, at the moment, no agreed philosophical answer to the problem of consciousness, except that most living philosophers are not Cartesian dualists – they do not believe in an immaterial soul which is distinct from the body. We suspect that the majority of neuroscientists do not believe in dualism, the most notable exception being the late Sir John Eccles (1994).

We shall not describe here the various opinions of philosophers, except to say that while philosophers have, in the past, raised interesting questions and pointed to possible conceptual confusions, they have had a very poor record, historically, at arriving at valid scientific answers. For this reason, neuroscientists should listen to the questions philosophers raise but should not be intimidated by their discussions. In recent years the amount of discussion about consciousness has reached absurd proportions compared to the amount of relevant experimentation.

**The Problem of Qualia**

What is it that puzzles philosophers? Broadly speaking, it is qualia – the blueness of blue, the painfulness of pain, and so on. This is also the layman's major puzzle. How can you possibly explain the vivid visual scene you see before you in terms of the firing of neurons? The argument that you cannot explain consciousness by the action of the parts of the brain goes back...
subjectively the same as the blue color I see? If it turns out that ones.

conscious experience, only, in some cases, its relation to other

is that we can never explain to other people the nature of any

associated with a difference at the motor stage. The implication

difference

of a subjective experience. It is possible, however, to convey a

representation expressed by the firing of the neurons associated

the motor neurons is related, but not identical, to the explicit

communicate the exact nature of what we are conscious of.) To

(By 'private,' we mean that it is inherently impossible to

consider, an inevitable consequence of the way the brain works.
The reason that visual consciousness is largely private is, we

'Why Are We Conscious', what we think consciousness is for.

and Koch, 1995c). We have already explained, in the section

The Problem of Qualia

This problem of qualia is what Chalmers (1995) calls 'The

Hard Problem': a full account of the manner in which subjective

experience arises from cerebral processes. As we see it, the hard

problem can be broken down into several questions, of which

the first is the major problem: How do we experience anything

at all? What leads to a particular conscious experience (such as

the blueness of blue)? What is the function of conscious experience? Why are some aspects of subjective experience impossible to convey to other people (in other words, why are they private)?

We believe we have answers to the last two questions (Crick and Koch, 1995c). We have already explained, in the section 'Why Are We Conscious', what we think consciousness is for. The reason that visual consciousness is largely private is, we consider, an inevitable consequence of the way the brain works. (By 'private,' we mean that it is inherently impossible to communicate the exact nature of what we are conscious of.) To be conscious, we have argued, there must be an explicit representation of each aspect of visual consciousness. At each successive stage in the visual cortex, what is made explicit is recoded. To produce a motor output, such as speech, the information must be recoded again, so that what is expressed by the motor neurons is related, but not identical, to the explicit representation expressed by the firing of the neurons associated with, for example, the color experience at some level in the visual hierarchy.

It is thus not possible to convey with words the exact nature of a subjective experience. It is possible, however, to convey a difference between subjective experiences — to distinguish between red and orange, for example. This is possible because a difference in a high-level visual cortical area can still be associated with a difference at the motor stage. The implication is that we can never explain to other people the nature of any conscious experience, only, in some cases, its relation to other ones.

Is there any sense in asking whether the blue color you see is subjectively the same as the blue color I see? If it turns out that the neural correlate of blue is exactly the same in your brain as in mine, it would be scientifically plausible to infer that you see blue as I do. The problem lies in the word 'exactly'. How precise one has to be will depend on a detailed knowledge of the processes involved. If the neural correlate of blue depends, in an important way, on my past experience, and if my past experience is significantly different from yours, then it may not be possible to deduce that we both see blue in exactly the same way (Crick, 1994).

Could this problem be solved by connecting two brains together in some elaborate way? It is impossible to do this at the moment, or in the easily foreseeable future. One is therefore tempted to use the philosopher's favorite tool, the thought experiment. Unfortunately, this enterprise is fraught with hazards, since it inevitably makes assumptions about how brains behave, and most of these assumptions have so little experimental support that conclusions based on them are valueless. For example, how much is a person's percept of the blue of the sky due to early visual experiences?

The Problem of Meaning

An important problem neglected by neuroscientists is the problem of meaning. Neuroscientists are apt to assume that if they can see that a neuron's firing is roughly correlated with some aspect of the visual scene, such as an oriented line, then that firing must be part of the neural correlate of the seen line. They assume that because they, as outside observers, are conscious of the correlation, the firing must be part of the NCC. This by no means follows, as we have argued for neurons in V1.

But this is not the major problem, which is: How do other parts of the brain know that the firing of a neuron (or of a set of similar neurons) produces the conscious percept of, say, a face? How does the brain know what the firing of those neurons represents? Put in other words, how is meaning generated by the brain?

This problem has two aspects. How is meaning expressed in neural terms? And how does this expression of meaning arise? We suspect (Crick and Koch, 1995c) that meaning derives both from the correlated firing described above and from the linkages to related representations. For example, neurons related to a certain face might be connected to ones expressing the name of the person whose face it is, and to others for her voice, memories involving her and so on, in a vast associational network, similar to a dictionary or a relational database. Exactly how this works in detail is unclear.

But how are these useful associations derived? The obvious idea is that they depend very largely on the consistency of the interactions with the environment, especially during early development. Meaning can also be acquired later in life. The usual example is a blind man with a stick. He comes to feel what the stick is touching, not merely the stick itself. For an ingenious recent demonstration along similar lines, see Ramachandran and Hirstein (1997).

Future Experiments

Although experiments on attention, short-term and working memory, the correlated firing of neurons and related topics may make finding the NCC easier, at present the most promising experiments are those on bistable percepts. These experiments should be continued in numerous cortical and thalamic areas and need extending to cover other such percepts. It is also important to discover which neurons express the NCC in each case (which neuronal subtype, in what layer, and so on), how they fire (e.g. do they fire in bursts) and, especially, to where they project. To
assist this, more detailed neuroanatomy of the connectivity will be needed. This is relatively easy to do in the macaque but difficult in humans (Crick and Jones, 1993). It is also important to discover how the various on-line systems work, so that one can contrast their (unconscious) neuronal activity with the NCC.

To discover the exact role (if any) of the frontal cortex in visual perception, it would be useful to inactivate it reversibly by cooling and/or the injection of GABA agonists, perhaps using the relatively smooth cortex of an owl monkey.

Inevitably, it will be necessary to compare the studies on monkeys with similar studies on humans, using both psycho-physical experiments as well as functional imaging methods such as PET or fMRI. Conversely, functional imaging experiments on normal subjects or patients, showing, for instance, the involvement of prefrontal areas in visual perception (Sahraie et al., 1997; Weiskrantz, 1997), can provide a rationale for appropriate electrophysiological studies in monkeys. It would help considerably if there were more detailed architectonic studies of cortex and thalamus, since these can be done post-mortem on monkeys, apes and humans. The extremely rapid pace of molecular biology should soon provide a wealth of new markers to help in this endeavor.

To understand a very complex nonlinear system, it is essential to be able to interfere with it both specifically and delicately. The major impact of molecular biology is likely to be the provisions of methods for the inactivation of all neurons of a particular type. Ideally, this should be done reversibly on the mature animal (see, for example, No et al., 1996; Nirenberg and Meister, 1997). At the moment this is only practical on mice, but in future one may hope for methods that can be used on mature monkeys (perhaps using a viral vector), as such methods are also needed for the medical treatment of humans.

As an example, consider the question of whether the cortical feedback pathways — originating in a higher visual area (in the sense of Felleman and Van Essen, 1991) and projecting into a lower area — are essential for normal visual consciousness. There are at least two distinct types of back pathways (Salin and Bullier, 1995): one, from the upper cortical layers, goes back only a few steps in the visual hierarchy; the other, from the lower cortical layers, can also go back over longer distances. We would like to be able to selectively inactivate these pathways, both singly and collectively, in the mature macaque. Present methods are not specific enough to do this, but new methods in molecular biology should, in time, make this possible.

It will not be enough to show that certain neurons embody the NCC in certain — limited — visual situations. Rather, we need to locate the NCC for all types of visual inputs, or at least for a sufficiently large and representative sample of them. For example, when one blinks, the eyelids briefly (30–50 ms) cover the eyes, yet the visual percept is scarcely interrupted (blink suppression; Volkmann et al., 1980). We would therefore expect the NCC to be also unaffected by eye blinks (e.g. the firing activity should not drop noticeably during the blink) but not to blanking out of the visual scene for a similar duration due to artificial means. Another example is the large number of visual illusions. For instance, humans clearly perceive, under appropriate circumstances, a transient motion aftereffect. On the basis of fMRI imaging it has been found that the human equivalent of cortical area MT is activated by the motion aftereffect (in the absence of any moving stimuli; Tootell et al., 1995). The time course of this illusion parallels the time course of activity as assayed using fMRI. In order to really pinpoint the NCC, one would need to identify individual cells expressing this, and similar, visual aftereffects. We have assumed that the visual NCC in humans is very similar to the NCC in the macaque, mainly because of the similarity of their visual systems. Ultimately, the link between neurons and perception will need to be made in humans.

The problem of meaning and how it arises is more difficult, since there is not, as yet, even an outline formulation of this problem in neural terms. For example, do multiple associations depend on transient priming effects? Whatever the explanation, it would be necessary to study the developing animal to show how meaning arises; in particular, how much is built in epigenetically and how much is due to experience.

In the long run, finding the NCC will not be enough. A complete theory of consciousness is required, including its functional role. With luck this might illuminate the hard problem of qualia. It is likely that scientists will then stop using the term consciousness except in a very loose way. After all, biologists no longer worry whether a seed or a virus is ‘alive’; they just want to know how it evolved, how it develops, and what it can do.

Finale

We hope we have convinced the reader that the problem of the neural correlate of consciousness (the NCC) is now ripe for direct experimental attack. We have suggested a possible framework for thinking about the problem, but others may prefer a different approach; and, of course, our own ideas are likely to change with time. We have outlined the few experiments that directly address the problem and mentioned briefly other types of experiments that might be done in the future. We hope that some of the younger neuroscientists will seriously consider working on this fascinating problem. After all, it is rather peculiar to work on the visual system and not worry about exactly what happens in our brains when we ‘see’ something. The explanation of consciousness is one of the major unsolved problems of modern science. After several thousand years of speculation, it would be very gratifying to find an answer to it.

Notes

For helpful comments we thank David Chalmers, Leslie Orgel, John Searle and Larry Weiskrantz. We thank the J.W. Kieckhefer Foundation, the National Institute of Mental Health, the Office of Naval Research and the National Science Foundation.

Address correspondence to Dr Francis Crick, The Salk Institute, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA.

References


Britten KH, Shadlen MN, Newsome WT, Movshon JA (1992) The analysis


**Note Added in Proof**

The recent discovery of neurons in the inferior prefrontal cortex (IPC) of the macaque that respond selectively to faces — and that receive direct input from regions around the superior temporal sulcus and the inferior temporal gyrus that are well known to contain face-selective neurons — is of considerable interest (Scalaidhe et al., 1997). It raises the questions of why would face cells be represented in both IT and IPC. Do they differ in some important aspect?