Our basic hypothesis at the neural level is that it is useful to think of consciousness as being correlated with a special type of activity of perhaps a subset of neurons in the cortical system. Consciousness can undoubtedly take different forms, depending on which parts of the cortex are involved, but we hypothesize that there is one basic mechanism (or a few such) underlying them all. At any moment consciousness corresponds to a particular type of activity in a transient set of neurons that are a subset of a much larger set of potential candidates. The problem at the neural level then becomes:

1. Where are these neurons in the brain?
2. Are they of any particular neuronal type?
3. What is special (if anything) about their connections?
4. What is special (if anything) about the way they are firing?

5 Visual awareness

At this point we propose to make a somewhat arbitrary personal choice. Since we hypothesize that there is a basic mechanism for consciousness that is rather similar in different parts of the brain (and, in particular, in different parts of the neocortex), we propose that the visual system is the most favorable for an initial experimental approach. The visual system has several well-known advantages as an experimental system for investigating the neuronal basis of consciousness. Unlike language, it is fairly similar in man and the higher primates. There is already much experimental work on it, both by psychophysicists and by neuroscientists. Moreover we believe that it will be easier to study than the more complex aspects of consciousness associated with self-awareness. From now on, then, we shall discuss not consciousness in all its aspects but visual awareness and only the awareness that does not involve the laying down of long-term episodic memory. Johnson-Laird[1] has called this 'bare awareness', which does indeed convey the idea that it is simpler than self-awareness. Although here we mainly consider visual perception, many of the processes we discuss are also likely to be used in visual imagery.

5.1 The mammalian visual system

The visual system of mammals is complex and we only give an outline description. In the macaque monkey there are many distinct visual areas in the neocortex, perhaps as many as two dozen[20]. One reason for these multiple areas is that to handle all activity in one single very large neural net, with everything connected to everything else, would make the brain both cumbersome and prohibitively large. Loosely speaking these areas are connected in several hierarchies, with the processing within any one area being largely parallel (see Figure 1). There are also many back projections (and some cross connections) between cortical areas, and also from the cortex back to the thalamus, the functions of all of which are unknown.

The dominant output of the retina projects—through the lateral geniculate nucleus of the thalamus—to the primary visual cortex at the back of the brain. This area has neurons

Structures in the midbrain or hindbrain, such as the cerebellum, are probably not essential for consciousness. It remains to be seen whether certain other structures, such as the thalamus, the basal ganglia and the claustrum, all intimately associated with the neocortex, are closely involved in consciousness. We shall include these structures together with the cerebral cortex as 'the cortical system'.

4.3 Split brains

The study of persons with 'split brains' gives us information about some of the pathways that are (or are not) involved in concsciousness. It is a well-established fact[14] that for persons whose corpus callosum—the massive fiber bundle connecting the two cortical hemispheres—has been cut, the left-hand side of the brain (for right-handed people) is not aware of the activity in the visual system taking place on the right side, whereas in a normal person it is. (We leave on one side the somewhat controversial matter as to whether the right side, on its own, is as conscious as the left side.) This strongly suggests that some of the information associated with consciousness can traverse the normal corpus callosum. It also suggests that such information, with the exception of some emotional states, cannot be transmitted from one side of the cortex to the other by the subcortical pathways that remain intact in this operation.

4.4 Blindsight

It now seems generally agreed that blindsight is a genuine phenomenon[15]. Certain people with cortical blindness can point fairly accurately to the position of objects in their blind visual field, while denying that they see anything. It was previously suspected that the neural pathway was subcortical, through the superior colliculus, but this has recently been brought into question[16]. The motor output appears to be voluntary and the subject is certainly aware what movements he is making. It is now an urgent matter to decide experimentally, by comparative work on humans and monkeys, exactly which neural pathways are used in blindsight, since this information may suggest which neural pathways are used for consciousness and which not.

We shall not discuss here other cases in which people respond to a stimulus such as an auditory or visual cue, but claim to be unaware of the stimulus (for example, subliminal perception and priming,[17,18] and also galvanic skin responses[19]). Such stimuli may affect ongoing mental processes, including higher order processes, without being registered in working or long-term memory.

4.5 Neuronal firing

What is the general character of neural behaviors associated, or not, with consciousness? We think it plausible that consciousness in some sense requires neuronal activity, in which we include not only neurons that fire action potentials but also non-spiking neurons such as amacrine cells.

cannot lay down new episodic memories but appear in all other respects to be fully conscious. Procedural memory,[11] i.e. that part of memory responsible for highly automated procedures, such as typing or swimming, is probably largely unconscious.

Consciousness appears to be a process that is fairly immediate. In other words, it does not take too long to become conscious of, at least, straightforward things. Exactly how long is not clear, but figures somewhere in the 50–250 ms time range might be reasonable, rather than a few seconds.

4 General neurobiological aspects

4.1 When

When is an animal conscious? This turns out not to be a straightforward problem. It seems likely that the essential features of consciousness are probably not usually present in slow-wave sleep, nor under a deep anaesthetic. Rapid Eye Movement (REM) or dreaming sleep is another matter. It seems to us that a limited form of consciousness often occurs in REM sleep. Although cognition is not completely normal and memories cannot be put into the long-term store, nevertheless dreams seem to have some of the attributes of consciousness. Whether there are any experimental advantages in studying REM sleep to help us understand consciousness remains to be seen. At the moment we do not see any.

The really difficult case is that of rather light anaesthesia or states induced by modern receptor-specific anaesthetic agents[12]. It seems likely that in some such states the brain is at least partly conscious. This would not matter were it not for the large amount of experimentation on mammals in this state. We return to this problem later.

4.2 Where

Where in the brain are the neural correlates of consciousness? One of the traditional answers—that consciousness depends on the reticular activating system in the midbrain—is misleading. Certainly the relevant parts of the brain need to be activated in various ways, but this is a little like believing that the characteristic part of a television set is its electrical supply. It is more likely that the operations corresponding to consciousness occur mainly (though not exclusively) in the neocortex and probably also in the paleocortex, associated with the olfactory system, since local damage to the cerebral cortex often removes particular aspects of consciousness (as, for example, in face agnosia).

The hippocampal system (the allocortex) is a little more complicated. A person with complete bilateral damage to the hippocampal system and to all the higher order association cortices that are connected with it, appears to have most aspects of consciousness[13] (including short-term conjunctions over at least three modalities) so the hippocampal system is unlikely to be essential. But it could be argued that what goes on in the normal hippocampus does indeed reach consciousness, so it may be important to consider in detail the inputs and outputs of the hippocampal system.

we believe that any adequate theory of consciousness should explain how we see color at all.

This outlines the framework within which we will address the problem of consciousness. We next ask what psychology can tell us about the phenomenon.

3 The cognitive approach

The most effective way to approach the problem of consciousness would be to use the descriptions of psychologists and cognitive scientists and attempt to map different aspects of their models onto what is known about the neuroanatomy and neurophysiology of the brain. Naturally we have attempted to do this but have not found it as useful as one might hope, although such models do point to the importance of attention and short-term memory and suggest that consciousness should have easy access to the higher, planning levels of the motor system.

A major handicap is the pernicious influence of the paradigm of the von Neumann digital computer. It is a painful business to try to translate the various boxes labeled 'files', 'CPU', 'character buffer', and so on occurring in psychological models, each with its special methods of processing, into the language of neuronal activity and interaction. This is mainly because present-day computers make extensive use of precisely-detailed pulse-coded messages. There is no convincing evidence that the brain uses such a system and much to suggest that it does not. For these reasons we will not attempt to comment in detail on these psychological models but only make rather general remarks.

Johnson-Laird[1] proposes that the brain is a complex hierarchy of largely parallel processors—an idea that is almost certainly along the right lines—and that there is an operating system at the top of the hierarchy. He considers the conscious mind to correspond to this operating system. Thus the mechanism of consciousness expresses the results of some of the computations the brain makes but not the details of how they were done. If there is an operating system of this type it is not easy at the moment to see any particular brain area in which it is located[7]. Johnson-Laird also lays emphasis on self-reflection and self-awareness, topics that we have decided to leave on one side.

An alternative view, due to Jackendoff[2] is that consciousness is not associated with the highest levels in the hierarchy but with the intermediate levels. He arrives at this point of view by considering in detail the language system, the visual system and the music system. His arguments are certainly suggestive, though not completely compelling.

Both authors emphasize the intimate relationship between consciousness and working memory. They also both envisage a serial process (that we may loosely identify as attention) on top of the parallel processes. Thus clearly we shall have to consider the mechanisms of attention and of working memory.

The type of memory that will be of most interest to us is short-term memory, whether it be very short (a fraction of a second; sometimes called 'iconic' memory[8] if it is visual) or short-term memory proper (meaning a few seconds) sometimes called working memory[9,10]. These are so important that we shall take up each of these subjects later. It does not seem essential for consciousness that the brain should be able to put anything into the long-term episodic memory system,[11] since people with certain brain damage (to be discussed below)

odd and, at first sight, so difficult to understand that only a rather special explanation is likely to work. The general nature of consciousness may be easier to discover than more mundane operations, such as shape-from-shading, that could, in principle, be explained in many different ways. Whether this is really true remains to be seen.

The following topics will be left on one side or our attitude to them stated without further discussion, for experience has shown that otherwise much time can be frittered away in fruitless argument about them.

1. Everyone has a rough idea of what is meant by consciousness. We feel that it is better to avoid a precise definition of consciousness because of the dangers of premature definition. Until we understand the problem much better, any attempt at a formal definition is likely to be either misleading or overly restrictive, or both.

2. Arguments about what consciousness is for are probably premature, although such an approach may give a few hints about its nature. It is, after all, a bit surprising that one should worry too much about the function of something when we are rather vague about what it is.

3. We shall assume that some species of animals, and 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.

3.1. From this 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 may not enrich consciousness considerably.

3.2. We consider that it is not profitable at this stage to argue about whether 'lower' animals, such as octopus, *Drosophila*, or nematodes, are conscious. It is probable, though, that consciousness correlates to some extent with the degree of complexity of any nervous system.

4. There are many forms of consciousness, such as those associated with seeing, thinking, emotion, pain and so on. We shall assume that self-consciousness, that is the self-referential aspect of consciousness, is merely a special case of consciousness and is better left on one side for the moment. Volition and intentionality will also be disregarded and also various rather unusual states, such as the hypnotic state, lucid dreaming and sleep walking, unless they turn out to have special features that make them experimentally advantageous.

5. No neural theory of consciousness will explain everything about consciousness, at least not initially. We will first attempt to construct a rough scaffold, explaining some of the dominant features and hope that such an attempt will lead to more inclusive and refined models.

6. There is also the problem of qualia[6]. Some argue that certain aspects of consciousness (such as whether the red I see is the same as the red you see), being essentially private, cannot in principle be addressed by any objective, scientific study. We feel that this difficult issue is, for the moment, best left on one side. Our present belief is that it *may* prove possible, in the fullness of time, to make it plausible that you see red as I do (assuming that psychophysical tests suggest you do). To decide whether one *can* make a plausible case or not we shall need to know the exact neural correlate in a human brain of seeing red. Whatever the outcome,

1 Introduction

It is remarkable that most of the work in both cognitive science and the neurosciences makes no reference to consciousness (or 'awareness'), especially as many would regard consciousness as the major puzzle confronting the neural view of the mind and indeed at the present time it appears deeply mysterious to many people. This attitude is partly a legacy of behaviorism and partly because most workers in these areas cannot see any useful way of approaching the problem. In the last few years several books have appeared[1–4] that address the question directly but most of these[1–3] have been written largely from a functional standpoint and so have said rather little about neurons and other machinery of the brain.

We suggest that the time is now ripe for an attack on the neural basis of consciousness. Moreover, we believe that the problem of consciousness can, in the long run, be solved only by explanations at the neural level. Arguments at the cognitive level are undoubtedly important but we doubt whether they will, by themselves, ever be sufficiently compelling to explain consciousness in a convincing manner. Attempting to infer the internal structure of a very complex system using a 'black-box' approach (i.e., manipulating the input variables while observing the output of the system) will never lead to unique answers. In short, such methods are not by themselves powerful enough ever to solve a problem, though they are good enough to suggest tentative solutions.

Our basic idea is that consciousness depends crucially on some form of rather short-term memory and also on some form of serial attentional mechanism. This attentional mechanism helps sets of the relevant neurons to fire in a coherent semi-oscillatory way, probably at a frequency in the 40–70 Hz range, so that a temporary global unity is imposed on neurons in many different parts of the brain. These oscillations then activate short-term (working) memory. In the later parts of the paper we shall be mainly concerned with visual awareness.

Before approaching the problem in detail, it seems sensible to describe our general approach to consciousness and to decide what aspects of it are best left on one side.

2 Prolegomenon to the study of consciousness

We make two basic assumptions. The first is that there is something that requires a scientific explanation. There is general agreement that we are not conscious of all the processes going on in our heads, though exactly which might be a matter of dispute. While we are aware of many of the results of perceptual and memory processes, we have only limited access to the processes that produce this awareness (e.g., "How did I come up with the first name of my grandfather?"). In fact, some psychologists[5] have argued that we have only very limited introspective access to the origins of even higher order cognitive processes. It seems probable, however, that at any one moment some active neuronal processes correlate with consciousness, while others do not. What are the differences between them?

The second assumption is tentative: that all the different aspects of consciousness, for example pain and visual awareness, employ a basic common mechanism or perhaps a few such mechanisms. If we understand the mechanisms for one aspect, we will have gone most of the way to understanding them all. Paradoxically, consciousness appears to be so

Abstract

Visual awareness is a favorable form of consciousness to study neurobiologically. We propose that it takes two forms: a very fast form, linked to iconic memory, that may be difficult to study; and a somewhat slower one involving visual attention and short-term memory. In the slower form an attentional mechanism transiently binds together all those neurons whose activity relates to the relevant features of a single visual object. We suggest this is done by generating coherent semi-synchronous oscillations, probably in the 40-70 Hz range. These oscillations then activate a transient short-term (working) memory. We outline several lines of experimental work that might advance the understanding of the neural mechanisms involved. The neural basis of very short-term memory especially needs more experimental study.

Key words: consciousness / attention / binding / 40 Hz oscillations / short-term memory

CALIFORNIA INSTITUTE OF TECHNOLOGY
COMPUTATION AND NEURAL SYSTEMS PROGRAM

January 28, 1991 CNS Memo 9

As published in *Seminars in The Neurosciences* **2:** 263–275 (1990).

TOWARDS A NEUROBIOLOGICAL THEORY OF CONSCIOUSNESS

Francis Crick† and Christof Koch*

† The Salk Institute
10010 N. Torrey Pines Road, La Jolla, CA 92037
* Computation and Neural Systems 216-76
California Institute of Technology, Pasadena, CA 91125

PASADENA, CALIFORNIA 91125