

Some Reflections on Visual Awareness

F. CRICK* AND C. KOCH†

*The Salk Institute, La Jolla, California 92037; †Computation and Neural Systems Program, 216-76, California Institute of Technology, Pasadena, California 91125

We have recently published a paper entitled *Towards a Neurobiological Theory of Consciousness* (Crick and Koch 1990) that outlined a sketch of such a theory. Our aim was not to produce as complete a theory of consciousness as possible but to indicate promising lines of experimental work, mainly neurobiological, that might lead eventually toward a solution of the problem. We made the plausible assumption that all forms of consciousness (e.g., seeing, thinking, and pain) employ, at bottom, rather similar mechanisms and that if one form were understood, it would be much easier to tackle the others. We then made the personal choice of the mammalian visual system as the most promising one for an experimental attack.

This choice means that fascinating aspects of the subject, such as volition, intentionality, and self-consciousness, to say nothing of the problem of qualia, have had to be left on one side. We have also not dwelt on unusual psychological states, such as hypnosis, dreaming, lucid dreaming, and sleepwalking, to say nothing of meditative states, since we do not see any special advantage in studying them experimentally at this stage.

Our method is to combine what is known experimentally from both psychological and neurobiological experiments with plausible theoretical arguments of a general nature. We have not so far attempted detailed computer simulations, since it is unlikely that we understand the details of the system well enough for such modeling to provide decisive answers. In the long run, however, modeling is essential, and even at this stage, well-conceived computer simulations should suggest plausible general principles and thus guide the experimental attack on what is obviously a system of enormous complexity and subtlety.

In our previous paper (Crick and Koch 1990), we listed a number of topics and assumptions that we did not intend to discuss. This list will not be repeated here in toto except to say that we have assumed the higher mammals have a form of visual awareness fairly similar to ours, so that we can, with discretion, combine experimental results on humans with those from monkeys, cats, and so on.

THE THEORY IN OUTLINE

Our knowledge of the visual system of the macaque monkey, which is probably fairly similar to that of

humans, suggests that there is no single place in the brain that corresponds to what we see. (For a detailed description of the macaque visual system, see Van Essen et al., this volume, and for the human visual system, see Zeki, this volume.) Thus, some mechanism must unite some of the neural activity in the many different visual areas to provide our global percept of the scene before us.

The problem of how to do this is sometimes referred to as the "binding problem." We distinguish three sorts of binding. The first is built in by genes acting on the developmental processes. The second is built up by experience, probably by strengthening existing but weak connections. An example might be neurons that respond to a familiar word. The third—the ad hoc binding problem—is what concerns us here. It is needed because the number of conceivable visual objects is so vast that there cannot be a special neuron (a "grandmother cell") for each one of them.

It is suggested that this binding is done mainly by the synchronized firing of the relevant neurons (as proposed by von der Malsburg and Schneider in 1986) and that such synchronized firing usually takes the form of semisynchronous oscillations in the 40–70-Hz range, sometimes called γ oscillations. (For a review of much of the recent experimental work on the "40 Hz" oscillations, see Singer et al., this volume.)

Psychological experiments show that we do not see everything in the visual scene at once. There appears to be a rapid, parallel process of handling the incoming visual information, followed by a further process that takes time and thus can be roughly described as "serial." This latter process is often referred to as "focal attention." It can happen to some extent without eye movements, as shown by Posner (1986), and is probably more rapid than eye movements, themselves another, slower, form of visual attention.

Whereas the initial, parallel, processing steps deal with the many "features" already encoded in the brain, we surmise that the steps involving focal attention serve temporarily to bind features that are not already coded there by fairly strong connections.

It has been suggested for many years that an essential feature of visual awareness is the temporary storage of what we have just seen in short-term memory (often called working memory) that decays within a few seconds. It is thought to have a somewhat limited capacity at any one moment, probably in the range of four to seven objects.

The deciphering by the brain of the visual scene before us is not an easy task, since in mathematical terms the problem is "ill-posed" (Poggio et al. 1985); that is, the information coming into our eyes is often ambiguous and can only be interpreted by using "constraints" built into the system by epigenetic processes and by previous experience. What needs to enter awareness is the result of these "computations," not the details of the computations themselves. We therefore suggest:

1. It is the result of these neural computations on an attended object that are expressed by phase-locked 40-Hz oscillations.
2. These phase-locked oscillations are the neural correlate of vivid visual awareness.
3. These oscillations then activate working memory.

These are our basic hypotheses. We also suggest very tentatively that there might be an even more transient form of awareness, linked to iconic memory, which we called "fleeting awareness." This is probably not associated with oscillatory firing patterns. Since this appears very difficult to study, we shall say no more about it here.

The reader is referred to our earlier paper (Crick and Koch 1990) both for a much fuller discussion of the rationale for these suggestions and for more extensive references. Our main object here is to build on that previous discussion in order to fill out some of these theoretical ideas. Our suggestions mostly concern ideas and phenomena that need to be studied, rather than being detailed proposals. We shall say little here about possible experimental tests, some of which are sketched in the earlier paper. We shall assume, without further discussion, that the 40-Hz oscillations are not an artifact and do indeed play the role we hypothesize for them. This assumption is far from being firmly established experimentally.

OSCILLATIONS

The Nature of Oscillations

The neurons involved do not strictly oscillate at 40 Hz. By "oscillate" we mean that when a neuron fires at 40 Hz, it does so at approximately 25-msec intervals. "Firing" here implies that the neuron produces an axonal spike, or sometimes two or three in very rapid succession. Sometimes a neuron may miss the opportunity to fire, but when it does fire it does so approximately "on the beat." The frequency at any one time is not especially regular. The mean frequency may vary somewhat with time and may vary also from neuron to neuron unless their phases have become locked together. At other times, such neurons may not oscillate but fire in a desynchronized manner. (For further details, see Singer et al., this volume.)

We shall assume, without further discussion, that certain neurons in the cerebral cortex (and perhaps elsewhere) have an inherent tendency to fire at about

40 Hz in the sense described above. This may be due to their intrinsic properties (as suggested by Llinás, this volume) or to circuit properties or, more likely, to both together. For example, a time delay of about 7–13 msec in the production of a short burst of inhibition might help to promote oscillations in the 40–70-Hz range.

The Number of Oscillations

How many distinct oscillations can exist in the brain at any one moment? This is a difficult question. It is unlikely to be only one, since there would then be little point in having oscillations. For instance, the activity leading to awareness need only be the "best" set of neurons, all firing very strongly together, say at 400 Hz, and no oscillations would be required. One of the advantages of having oscillations is that more than one coherent set of neurons can be active at any one moment without interfering too much with each other.

Moreover, it is an obvious advantage to maintain the representations of several objects simultaneously, since then further processing on this group of objects can proceed more rapidly and more efficaciously. This may be especially important for language, and for handling ideas in general.

Two different oscillations may have the same frequency but different phases. If these oscillations spread so that they overlap, they are likely either to cancel each other out in the overlap region or to coalesce to form one single set of oscillations of the same phase over the whole region. Oscillations of different frequencies may be able to interpenetrate more easily, but how much they will interfere with each other remains to be seen, since the system is probably highly nonlinear. It is by no means clear, on theoretical grounds, how many oscillations can exist together without significant interference. It may depend somewhat on whether they largely overlap in certain brain areas or whether, on the other hand, they dominate rather separate brain regions. An educated guess (meaning a guess with no sound theoretical foundations) might be four in the first case (heavy overlap) and up to seven or so in the second (little overlap).

It seems highly plausible that these restrictions on the maximum number of simultaneous oscillations are the main causes of the well-known limitations on the capacity, at any one moment, of the attentional or memory mechanism (for a somewhat different viewpoint, see Allport 1989).

It will thus be important, as more facts become known, to describe for a variety of cases the time of buildup of one set of oscillating neurons, the time between establishing one oscillatory set and establishing the next one, the length of time for which any one of them persists, and thus the number active at any particular moment. In interpreting the psychological data, it will be important to know whether an oscillatory set persists without interruption or whether it has lapsed but been recalled by a rapid readdressing system.

The Buildup of Oscillations

How easy is it for oscillations to build up and phase-lock? It is plausible to assume that for a simple, straightforward object, the time is likely to be shorter than for a more complex, ambiguous one. If there are several possible interpretations of the incoming information, it may take some time for the one particular interpretation to dominate its rivals and establish itself. In the case of "rivalry," when the percepts alternate, as in the well-known case of the Necker cube, we assume that the oscillations that first became established eventually habituate somewhat so that the other interpretation gets the upper hand by establishing the oscillations relevant to it and, in doing so, pushes down its rival. After a delay, it is then itself pushed down, and so on.

For complex objects, there is likely to be a constant struggle between rival sets of oscillations (strictly, between rival sets of neurons each trying to establish a coherent set of oscillations), except in very straightforward cases when one set of neurons is so dominant that it captures the oscillations without any significant challenge from other sets. The set of neurons that establishes strongly developed oscillations corresponds to those that embody the best interpretation of relevant parts of the visual scene, taking account of all the various mutual excitatory and inhibitory contacts already established by previous experience. We do not discuss at all any of the neural computations that underlie finding this best interpretation, that is, the visual algorithms computing the optical flow field, structure-from-motion, etc.

In all this, we have assumed that there will be competition between different sets of neurons, all attempting to set up coalitions to support each other's oscillations. Exactly how this is done is a very complicated matter that we shall not discuss here.

Oscillations and Awareness

It should not be assumed that all neurons showing some degree of 40-Hz oscillations in their autocorrelograms, or even sets of neurons showing some degree of phase-locking (see Singer et al., this volume), are the neural correlates of visual awareness. They may represent phases in the struggle to establish a coherent set of oscillations. This may be especially true in lightly anesthetized cats whose exact state of awareness we can only guess at.

What kind of oscillation corresponds, then, to full perceptual awareness? Two extreme views are possible. In the first view, visual awareness would require phase-locked sets of only very strong oscillations. (Strong here means that the peaks of the autocorrelation function are high and the troughs low.) Individual neurons of this type are occasionally seen among the oscillating neurons studied experimentally (Engel et al. 1990; Gray et al. 1990). In the second view, it could be

argued that even for a fully established percept, the strengths of the oscillations needed are more distributed so that oscillations of various strengths contribute, since in this way more information might be conveyed. In both these cases, we lack any secure data on the fraction of cortical neurons, in one cortical region, that need to oscillate to produce full visual awareness. A different experimental approach suggests that the fraction of active neurons to lead to visual awareness may be quite small (see Newsome et al., this volume). Interpretation of the experimental results would be simpler if the first alternative were true, but the second sounds almost equally plausible.

The Importance of Short Chains

At what levels in the visual system do neurons interact to build up these phase-locked oscillations? We shall argue that a general principle may be involved. Rather simple simulations (Kammen et al. 1989) suggest that it is difficult to phase-lock oscillations quickly between two neurons or groups of neurons if they are not fairly directly connected but instead have many distinct neurons in the shortest chain connecting them. We believe that there has been intense evolutionary pressure for mechanisms that can build up coherent sets of oscillations as quickly as possible. Thus, as a general principle, we suggest that all pathways that help to synchronize oscillations will turn out to have as few intervening neurons as possible.

Since the coalitions that form are likely to involve interactions both locally and globally, we suggest that these synchronizing pathways will occur at many levels in the system. In particular, we expect such connections to occur within a cortical area, both at short distances (under 0.5 mm) and to some extent up to longer distances (a few mm) and also between different cortical areas, and especially between areas connected at adjacent levels in the hierarchy. This may well involve some of the so-called "back" pathways (see Van Essen et al., this volume). In addition, we think there will likely be pathways that help to synchronize whole sets of cortical areas. Obvious places to look for such pathways are the thalamus (Jones 1985) and the claustrum (Sherk 1986).

Because of our general principle, we suspect that more indirect pathways, for example, from cortical areas high in the Van Essen hierarchy to those much lower down, will be less useful for this purpose, unless there are fairly direct connections between them.

A general theoretical problem of considerable interest is the design of efficient systems that can quickly set up phase-locked oscillations. We strongly suspect that the design would be hierarchical, with feedback pathways that spread somewhat, together with some semi-global and global pathways. Although the detailed design would obviously depend on the problems any particular system has to solve, there may be theoretical arguments that favor certain styles of design. Whether this is true remains to be seen.

The Problem of Time Delays

In considering mechanisms for phase-locking, it is of the first importance to allow for the inevitable time delays in the pathways. Already simple simulations (Kammen et al. 1989; Sporns et al. 1989) have shown that delays in excitation of more than a quarter of a cycle make phase-locking difficult, although presumably delays of a full cycle will work. Of course, if a synchronizing pathway produces inhibition, time delays of about half a cycle between onset of excitation and onset of inhibition would be expected.

An interesting possibility is that the neurons with an inherent tendency to oscillate at about 40 Hz are mainly some of the inhibitory neurons, as indeed Llinás's results suggest (see Llinás, this volume). Inhibitory neurons in the cortex only project locally (i.e., within a single cortical area). Thus, if they sometimes oscillate in error, these errors will at first be fairly local. It may be that what mainly matters for perception is oscillations in sets of pyramidal cells (all of which are excitatory), since pyramidal cells usually project over long distances, both to other cortical areas and subcortically.

Delays in transmission down dendrites might in some cases produce unacceptably long delays for synchronization purposes. Most inhibitory neurons have many excitatory synapses on their somas, whereas excitatory neurons (such as pyramidal cells) have none, the synapses on their somas being almost entirely inhibitory (for review, see Crick and Asanuma 1986). Thus the circuit:

excitation → activation of → inhibition of a
inhibitory cell pyramidal cell

can be a fast one, since it may not be subject to the delays of dendritic transmission.

For this reason, we expect that such "somatic" circuits will be used for fast processes, such as those needed to phase-lock 40-Hz oscillations. The main function of certain types of inhibitory cells (e.g., basket cells) may be to generate and phase-lock 40-Hz oscillations, even though the expression of these oscillations is mainly required in appropriate sets of pyramidal cells. This is, of course, quite distinct from the more traditional role of subtraction or vetoing assigned to them.

That inhibitory cells may be able to help the phase-locking of oscillations is suggested by the quite independent work of W.W. Lytton and T.J. Sejnowski (in prep.). By detailed modeling of the various ion currents, etc., they have shown how inhibitory cells, such as basket cells and chandelier cells, could help entrain an existing 40-Hz oscillation in a pyramidal cell, partly via a mechanism related to "anodal break excitation."

A set of coherent oscillations may not always build up in the same way, i.e., always starting in the same visual area. It seems more likely that in many cases, one cortical region will set up a strong coalition very rapidly and that the effects of this powerful alliance will then spread to other cortical areas where they can assist

further coherent additions to the coalition. Sudden transitions from nonoscillating to oscillating behavior may sometimes be due to such effects.

Oscillations and Gestalt

The roles we are suggesting for the 40-Hz oscillations bring to mind some of the ideas of the Gestalt psychologists. What we have referred to as an object would be better termed a "gestalt." A transient coalition of neurons, all expressing the phase-locked oscillations, is probably built up by neuronal interactions that correspond somewhat to the "laws" of gestalt psychology, such as proximity, continuity of direction, common fate, good continuation, and even symmetry. What to a psychologist is a particular gestalt would, to the neuroscientist, be expressed by a particular set of phase-locked oscillating neurons. Today, these laws are interpreted within the current computational vision framework as constraints needed to arrive at a unique perceptual solution (Poggio et al. 1985).

A psychologist might complain that this reformulation does nothing more than define the problems confronting him in other terms, and moreover, in terms he is not certain how to manipulate. This is true, but this disadvantage is quite outweighed by the advantage of being able to tie cognitive ideas to neuronal behavior, so that several distinct types of experimental evidence can be brought to bear on the same problem. If the brain does in fact behave in the way we are suggesting, then explanations at the psychological level will have to be recast in terms that more closely correspond to the underlying neurobiological reality.

ATTENTION

The idea that attention and consciousness are closely connected dates back to the last century. It is impossible here to give a detailed review of all the current theories of visual attention, if only for limitations of space. Instead, we shall try to organize the various ideas into broad classes. Because of the complexity of the subject, both psychologically and neurobiologically, our discussion must be regarded as preliminary.

Almost all the psychologists involved have accepted that in vision there is an early, rapid, highly parallel process, followed by a slower one that usually takes longer and is often considered to be serial.

We shall deal only briefly with the early phase, often associated with iconic memory (Coltheart 1983). This memory is thought to have a large capacity but to be transient, with decay times in the range of a fraction of a second. All workers have considered it to contain visual "features," but some have emphasized simple features, such as orientation or color, whereas others have argued that it also contains more elaborate features, such as letters, numerals, and familiar words (see the discussion in the earlier parts of Allport 1989). In our terms, anything that has been already "built in"

(and thus does not need the 40-Hz oscillations to relate its features) is likely to be processed in parallel, often rather rapidly, and to a high level in the system unless this early processing produces conflicts in interpretation. By a high level, we mean a level at which the neurons respond to such things as familiar words or faces.

This early phase is often called "preattentive." In some psychological models (Treisman 1988), an object can "pop out" into awareness by the parallel process alone. In others (Koch and Ullman 1985; Duncan and Humphreys 1989; Cave and Wolfe 1990), it needs the serial stage for this. Pop-out in these latter models is then the first, rapid step of the serial process.

Iconic memory may involve rapid synaptic modification (see Crick and Koch 1990); but whether it does or not, we believe that it will always involve the continued firing of the relevant neurons. How this firing is maintained is not yet understood. It may be due to processes inherent in the neurons themselves, but it seems likely that it will also involve reverberatory circuits of neurons of some sort or another.

Is There a Saliency Map?

The later, more serial stage has been explained by psychologists and theoreticians in several ways (Koch and Ullman 1985; Treisman 1988; Duncan and Humphreys 1989; Cave and Wolfe 1990; Van Essen and Anderson 1990). One of the main differences between these theories is whether a topographical saliency map exists or not. The basic idea of a saliency map (Koch and Ullman 1985) is that various visual areas send retinotopically mapped signals to a distinct area of the brain. This area does not code information on *what* it is that is salient but only *where* it is. Saliency is here meant to be understood in terms of simple operations, implemented by center-surround types of operations, e.g., a green object among many red ones or a moving stimulus among a stationary scene would both be very salient objects. In some versions, this area then selects, by some kind of winner-take-all mechanism (with or without noise in it), the most salient region in the visual field and directs attention to it. This attention can take the form either of boosting the firing of the neurons in that region of the visual field (for all retinotopically mapped areas) or of damping down the firing of regions outside the attentional spotlight, or both. There is neurophysiological evidence for both these effects in the region V4 of macaque monkeys (Moran and Desimone 1985; Spitzer et al. 1988).

From our point of view, we want the attentional mechanism to boost the appropriate 40-Hz oscillations in that region of the visual field. On the other hand, others (Duncan and Humphreys 1989) have argued that no saliency map is needed and that the struggle for saliency takes place in the cortical areas themselves. In our terminology, the struggle to establish the best set of

40-Hz oscillations would take place all over the visual cortex.

The Spotlight and the Oscillations

Let us consider three broad alternatives. In each case, there is first a highly parallel feature stage, either in the broad or the narrow sense of feature. We shall assume that in all three models, sets of 40-Hz oscillations compete with each other to some extent. In the models A and B, the spotlight promotes activity in the region of its beam and/or inhibits activity in the surroundings of the beam.

- A. The 40-Hz oscillations only start in the neurons related to an object when some form of spotlight lights them up. When the oscillations are set up, they can continue (for a time) while the spotlight lights up the neurons representing the next most salient object. There is a saliency mechanism to control the spotlight.
- B. The 40-Hz oscillations start up everywhere, as best they can. A spotlight (controlled by a crude saliency map) picks out the representations of the most salient item in several cortical areas and boosts them. When those oscillations are established they can continue, but the spotlight moves on to help the neurons corresponding to the next most salient item.
- C. There is no saliency map, with its associated spotlight. The 40-Hz oscillations all start together, but the competition is such that the best set of oscillations wins, although it does not suppress the others entirely (they simply grow more slowly). After a time, this first one "fatigues" while others press forward to become the most favored one(s).

Thus, in model C, there is no separate attentional mechanism, but rather competition between sets of 40-Hz oscillations (P. Braam, pers. comm.).

It may not be easy to decide definitely, by psychophysical experiments alone, whether a separate saliency map exists or not. We are inclined to favor the view that it does, for the reason that it is difficult enough to establish quickly the oscillation of the best set of neurons for even a single object within the visual field (including solving the so-called figure-ground problem) without requiring this complex mechanism to cope by itself with competition between objects. Since the exact order of processing objects may only matter in extreme cases, it seems better to let the order in which objects are processed be decided by a much cruder and faster mechanism. At the moment, model B above seems the most attractive.

Where Is the Saliency Map?

A saliency map for eye movements is known to exist in the superior colliculus of mammals. It would not be surprising if it were used by the brain to help decide the saliency needed for the faster attentional system. Since

the colliculus projects to the pulvinar, and since experimental evidence points to the thalamus, and the pulvinar in particular (see Desimone et al., this volume), as involved in attention, the pulvinar seems a promising place to look for saliency maps. The thalamic reticular nucleus may also be involved (Rafal and Posner 1987), although probably not exactly in the way previously suggested by one of us (Crick 1984), since that particular mechanism is now believed to occur only in slow-wave sleep (Steriade et al. 1990). There is also experimental evidence that the posterior parietal region is involved (Posner et al. 1987). Whatever its location, the attentional system must have access—in one way or another—to information from both cerebral hemispheres.

Since the thalamus has many distinct areas, including several visual ones, it would not be surprising if many of them were associated with one or another form of attention (see Allport 1989), so that even for vision, the single-map mechanism sketched above may well be too simple.

The Preattentive Stage

It is well established that the brain can process external stimuli, even to the point of initiating behavior, without consciously perceiving them. Such nonconscious, unconscious, or subconscious processes include blindsight, priming, implicit memory, subliminal perception, and automatic processes (Holender 1986; Kihlstrom 1987). For instance, patients with prosopagnosia, an inability to recognize familiar faces, show significant electrodermal skin conductance responses to pictures of persons they had previously known but were now unable to recognize, at the same time failing to show such an autonomic response to unfamiliar faces (Tranel and Damasio 1985).

We postulate that in these cases, the relevant stimuli give rise to strong firing responses in the associated cortical areas, without leading to phase-locked 40-Hz oscillations. Since it lacks the 40-Hz oscillations, this neuronal activity will not stimulate short-term memory and produce awareness. Nevertheless, this nonoscillatory (or weakly oscillatory) activity may be sufficient to leave a forward-reaching trace of the sort seen in priming and implicit memory (Tulving and Schacter 1990), perhaps partly by a non-Hebbian mechanism of synaptic modification. This possibly non-Hebbian mechanism is described more fully in our earlier paper (Crick and Koch 1990).

We can now suggest why some workers emphasize primitive features for the early visual stage, whereas others emphasize more complex features, such as familiar letters. The neurons that respond to the simpler features, being mapped retinotopically, may be the main ones that feed into the saliency map. Neuronal activity, as we have suggested, may very quickly reach the neurons in higher cortical areas that respond to more complex features, but because these features are

probably not mapped retinotopically, their activity will not feed into the retinotopic saliency map. The activity of these complex feature-detectors will not reach awareness unless there is time and sufficient saliency for them to build up phase-locked 40-Hz oscillations.

NEUROPHYSIOLOGICAL EXPERIMENTS

Since the pioneer studies of Mountcastle et al. (1981), there have been a number of demonstrations that for an alert monkey the response of certain neurons depends on whether the monkey is attending or not. Some of these experiments are mentioned in the review by Wise and Desimone (1988). Here, we concentrate on a set of experiments by Moran and Desimone (1985), since these show that the idea of a simple spotlight of attention is inadequate. In their review Wise and Desimone state: (our italics)

Neurons in area V4 ... have receptive fields so large that many stimuli typically fall within them. One might expect that the responses of such cells would reflect the properties of *all* stimuli inside their receptive fields. However it has been found that when a monkey restricts its attention to *one* location within a V4 ... cell's receptive field, the response of the cell is determined primarily by the stimulus at the attended location, *almost as if the receptive field "shrinks" around the attended stimulus*. For example, consider a cell that responds strongly to red stimuli and not to green when only a single stimulus appears inside its receptive field [see Fig. 1]. If red and green stimuli appear simultaneously at different locations within the field, and the animal focuses its attention on only the red one, the cell will respond strongly [Fig. 1A]. If, however, the animal attends to only the green stimulus, the cell will respond weakly or not at all to the red stimulus [Fig. 1B], even though the red stimulus is still inside the receptive field and the retinal stimulation is identical to the previous condition.

So far, these results could be explained by a small spotlight of attention, surrounded by a larger suppressive penumbra at a lower level in the visual hierarchy, since this suppression could account for the V4 neuron's poor response in Figure 1B (other neurons in V4 are presumed to respond to the green stimulus).

The next experiment shows that this explanation is too simple. In the case illustrated in Figure 1C, Moran and Desimone (1985) kept the two stimuli the same distance apart, but placed the attended green one outside the receptive field, while the unattended red one remained in its place within the receptive field. The neuron's response was not reduced, as it was in Figure 1B, but remained strong as shown in Figure 1C (for a further description of these experiments, together with those on IT [inferotemporal] neurons, see Moran and Desimone [1985]).

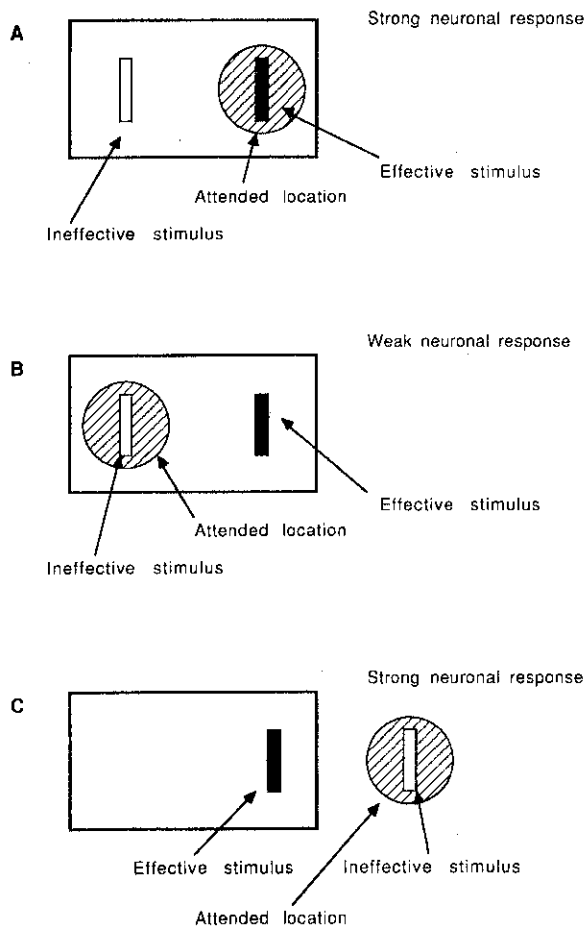


Figure 1. Simple diagrams to illustrate Moran and Desimone's (1985) results for one particular neuron in V4. The rectangle shows the receptive field of this neuron. The alert trained monkey fixated a point outside this field. This particular neuron responded only to red bars (shown black here) and not to green bars (shown open here). Thus, a red bar was an effective stimulus; a green bar an ineffective stimulus. For each trial the monkey was trained to pay attention (while maintaining fixation) to one particular place in the visual field, symbolized here by the circle, representing the hypothetical spotlight of attention. Note that the distance between the effective and the ineffective stimulus in C was the same as in A or B. The response of the neuron in each case is described to the right of each sketch. See text for a fuller discussion.

A simple spotlight, with a suppressive penumbra, will not explain this result, since the penumbra in the case of Figure 1C would have been expected to suppress the neuron and produce a poor response, whereas the observed response was strong. How is this very unexpected result to be explained?

A Possible Model

We suggest that there is indeed a spotlight of attention that, as implied in Figure 1, is smaller in area than a typical receptive field in V4. This spotlight provides some kind of boost or tag to certain of the neurons

whose receptive fields are in that location in the visual field. These neurons are probably at a lower level in the visual hierarchy (e.g., in V2), where receptive fields are smaller than those in V4. In addition, we suggest a more elaborate mechanism. This will only be outlined here in an oversimplified form.

Suppose that in V4, there are many different stacks of neurons. Within each stack, the neurons have receptive fields all the same size and all in the same place in the visual field, some responding to blue, others to green, others to red, and so on. We consider two states of a single stack: without retinotopic attention (within that receptive field) and with attention. We postulate that the members of the stack do not compete when attention there is lacking but do compete strongly when attention is present. In other words, attention facilitates competition between the different neurons in the same stack but not between those in different stacks.

In our theory, this boosting or tagging would take the form of synchronized oscillations. Thus, neurons in a lower level—whose receptive field is smaller than those in V4—oscillate in response to the attended red object, expressing the neuronal correlate of attention. These neurons, feeding into the V4 neuron Moran and Desimone (1985) recorded from, will induce similar phase-locked oscillations in this cell but not in the other neurons in this stack responding to, say, green or blue. Neurons in this stack will then compete against each other, such that the oscillatory and phase-locked neurons will win out over the neurons with less synchronized or desynchronized firing. The response in Figure 1B is weak because the spotlight has activated competition between all the neurons in that stack and has also tagged the attended ("green") neurons at a lower level, so that in the resulting competition in V4 the "green" neurons there suppress the "red" ones. In Figure 1C, the spotlight of attention, while promoting competition in other stacks, fails to do so for the particular stack illustrated. Thus, lacking competition among nonoverlapping stacks, all neurons in the stack respond strongly to any incoming stimulus of the appropriate characters.

As the results show, an unattended stimulus (as in Fig. 1C) can sometimes give a fairly strong response. However, Moran and Desimone (1985) also studied neurons in visual region IT (inferotemporal). Here, the receptive fields are so large that they often fill much of the visual field. Thus, in the case of Figure 1C, the strong response of the neuron in V4 would be attenuated when it reached IT. There the roles of the two stimuli would be described by a figure that corresponded conceptually to Figure 1B, since in IT both stimuli would lie in the same receptive field and thus in the same stack, and so would compete with each other. The unattended red stimulus would therefore be suppressed somewhat in IT, as indeed Moran and Desimone found.

We can express the essence of our idea in more familiar words by using the concept of columns suggested by Mountcastle (Edelman and Mountcastle 1978).

In oversimplified terms, when a macrocolumn is not attended to, it can have several outputs. When it is attended to, it will have only one. This particular formulation is not to be taken too literally.

The above model has been simplified for didactic purposes. The idea of sets of stacks that do not overlap is far too simplistic. A more realistic model would have neurons whose receptive fields overlapped to varying degrees. Broadly speaking, for any two neurons, the greater the overlap of their receptive fields, the greater the competition the model would allow between them. This competition would, of course, depend on local activation by the spotlight in one way or another.

How might this "facilitation of competition" be done? We suggest by involving the 40-Hz oscillations, assuming competition between oscillations is fiercer than that between nonoscillatory activity, an idea that is not implausible. Moreover, the boost referred to above might be due, at least in part, to the resonances produced by 40-Hz oscillations. In other words, whenever a neuron is oscillating, it shuts down the nonoscillating neurons in that stack, or, more precisely, shuts down those neurons whose receptive fields heavily overlap with its own.

Unfortunately, it is not known whether 40-Hz oscillations occurred in Desimone's experiments. We would hope that the neuron in cases like that of Figure 1A would show strong 40-Hz oscillations, whereas for cases like Figure 1C, we would expect 40-Hz oscillations to be weak or absent. Whether this is true remains to be seen. We think that the idea that attention facilitates local competition may shed an entirely new light on mechanisms of attention and deserves further detailed study.

Top-down Effects

It would take us too far afield to consider top-down effects (see Julesz, this volume), except to say that they are undoubtedly important. It would not be too surprising if some of them originated in frontal cortex. Such top-down effects might act directly to bias the saliency map, or more indirectly by biasing some of the neurons in the various visual areas of the neocortex that project to the saliency map. There may also be top-down effects that adjust the size of the attentional spotlight. In addition, there must be mechanisms for remembering the task in a psychophysical experiment and for recognizing a target object when it springs into awareness. We shall discuss none of these subjects here.

WORKING MEMORY

It is difficult to say much about the neurological basis of working memory because there has been so little experimental work on it. We have to distinguish two broad possibilities. In the first, neural activity is maintained for some seconds by one mechanism or another (see Crick and Koch 1990), with or without transient

synaptic modification. In the second, neuronal activity is not maintained, but only transient synaptic modifications take place. Both mechanisms may be used to some extent.

Let us first consider the activity taking place before working memory is activated. When a 40-Hz oscillation is both strongly active and phase-locked over many cortical areas, we expect this global activity to correspond to vivid visual awareness. There are likely to be certain regions that have easily formed this strong coalition (we call these the core of the activity) and have then spread their oscillations to other regions that we refer to as the periphery. These in their turn may activate further peripheral regions, but not necessarily to the extent of promoting phase-locked oscillations there. Some of these peripheral activities may correspond to what might be considered the "meaning" of the percept.

What will happen when, for one reason or another, this whole set of neurons ceases to oscillate in phase? In some cases, the core regions may maintain mainly nonoscillatory activity, perhaps with the assistance of transient synaptic modifications. We suggest that this is the neural correlate of working memory. Because of the lack of global oscillations, their activity is not correlated with vivid awareness; however, should the attentional spotlight return to it (or should competition from other oscillatory sets greatly decrease), the core will quickly set up all or part of the previous phase-locked oscillatory set and thus enter awareness again in a very short time.

Recall that in some cases, it may have taken an appreciable time for a set of coherent oscillations to be set up in the first place. If this activity has produced some transient synaptic modification, there may be no need to repeat this prolonged setup time when the memory is evoked again.

A second possibility is that, after all oscillations have died away, all the neurons concerned lapse into firing at background rates. Even in such a case the memory may still be retrievable if sufficiently strong (transient) synaptic modifications have taken place. In this case, however, a fairly strong and specific cue, directed to the right place, will be needed to start up the oscillatory activity again.

In the first case discussed above, strong cueing should not be needed because the self-sustained (nonoscillatory) activity should be able itself to act as a sufficient cue under the right circumstances. If something more is required to reactivate the memory, it need only be a rather unspecific nudge. One form of recalling a previous item in working memory is often called rehearsal. Rehearsal may explain the rather large differences in the decay time of working memory reported by different workers. For example, in the case of Damasio's patient Boswell (Damasio et al. 1985), the time quoted is about 40 seconds (A.R. Damasio, pers. comm.). In those studies on normal people (Baddeley 1986) in which any form of rehearsal was discouraged, the decay times tended to be only a few seconds.

Although rehearsal may be obvious in some cases, it may sometimes escape introspection.

OTHER OSCILLATIONS

The main theoretical requirement for mechanisms that produce global activity among neurons is that they should promote correlated firing. Such in-phase firing of the incoming axons is likely to produce a larger effect at the soma than desynchronized firing at the same average rate, but this increase is likely, by itself, to be fairly small. A further advantage of oscillations as a form of coordinated firing is that they may be able to induce membrane-bound resonances if some neurons have a strong, but latent, tendency to fire at or near that particular frequency.

For the purpose of exposition, we have considered only γ oscillations (in the 40–80-Hz range), but no theoretical reason has been given so far why the brain should use these particular frequencies. On general grounds, one would expect the best oscillations to be as fast as possible, allowing for the unavoidable time delays in axonal, synaptic, and dendritic transmission. This argument is at the moment too imprecise to favor one frequency range over another.

The brain shows a whole range of oscillations, the most notable being the θ rhythm of the hippocampus and the α rhythm mainly associated with the neocortex. Exactly what these other rhythms are used for is unclear, but they should not be forgotten. The α rhythm tends to occur when the brain is awake but somewhat inactive. Duncan and Humphreys (1989) in their theory of attention propose a subsidiary mechanism that they call the "spreading suppression" of similar nontargets (the reader is referred to their paper for details). It is just conceivable that the α rhythm provides this hypothetical function. According to this view, the function of the α rhythm would be to hinder the establishment and spread of 40-Hz oscillations.

A major unanswered question concerns the interaction between different oscillations, either within a family or between families (e.g., between α rhythms and γ rhythms). Of course, even in a linear system, the interaction of two oscillations of different frequency will often produce "beats." The 40-Hz oscillations may perhaps be too irregular for such beats to be very prominent unless the oscillations become rather more regular when fully established. Nevertheless, it is conceivable that some mechanism based on beats tends to set a certain distance between the frequencies of two interacting sets of oscillations and, by this means, spaces simultaneous frequencies fairly uniformly over the allowed frequency range. The problem is complex and deserves more detailed theoretical study.

THE OVERALL PICTURE

Let us now try to describe the overall behavior of the neurons in the visual cortex of an alert animal. Whatever the details of the attentional mechanism, we ex-

pect that there will be several sets of phase-locked neurons oscillating at the same time and having somewhat different frequencies and/or phases. Some of these oscillations may be dying away while new ones may be struggling to grow. All this occurs against a large background of neurons firing in a more desynchronized or chaotic manner. Because we do not yet have a clear idea of the usual lifetime of one particular oscillation set, the time scale of all this transient activity is difficult to estimate precisely, but the changes are likely to be fairly fast, probably taking a small fraction of a second. For this reason, it may be difficult to study the oscillations in the brain of an alert monkey viewing a typical visual scene, since any particular set of neurons may oscillate in such a transient manner as to be difficult to pin down. An alert animal trained to fixate and looking at a straightforward rather uniform stimulus, such as a drifting grating, may be somewhat easier to study, but the results obtained may be so simple as to be misleading.

Eventually other devices may be needed. Perhaps the speed at the animal's attentional mechanism can be either slowed down or disabled, possibly by lesions to the parietal region (Posner et al. 1987). This slowing down may be one of the main effects of a state of light anesthesia. (It is also possible that an anesthetic reduces or prevents the effects of the oscillations on working memory, so that subjectively the subject is largely unconscious even though his brain may display oscillations to some extent.)

Radically new experimental methods may have to be invented if we are to understand what is going on. For example, it may be possible to follow many of the main features of the oscillations by recording multiunit activity (see Singer et al. or Bower; both this volume) in many places simultaneously. Perhaps an array of 10,000 "electrodes" could be placed on the exposed cortex of a monkey whose cortex had few convolutions, such as the owl monkey (or even on that of the humble, rather nonvisual rat; see Wilson and Bower, 1990). The results could be displayed on a television monitor, probably slowed down considerably, so that the dynamic features of the recording can be grasped more easily. Another possible technique that would allow monitoring of the 40-Hz oscillations simultaneously over large areas of cortex is that of voltage-sensitive dyes (Grinvald et al. 1988). It remains to be seen whether the time resolution and the signal-to-noise ratio can be brought to acceptable levels.

CONCLUSION

The object of this paper is not to put forward detailed theories but rather to outline a few of the ideas that have arisen rather naturally by assuming that the 40-Hz oscillations are doing what we are guessing they do. These considerations should alert neurobiologists both to ideas proposed by psychologists and to several rather complex dynamic activities in the brain that urgently need further study.

To psychologists, these ideas outline some neurobiological facts and possibilities that could provide new idioms with which to construct their models. For example, there may be two somewhat distinct forms of neural activity: (1) nonoscillatory or weakly oscillatory activity that does not produce awareness but can leave a forward-reaching trace of the sort seen in priming and (2) phase-locked oscillatory activity that does produce awareness and activates the working memory system.

To neural modelers, and to theorists in general, these ideas raise numerous theoretical problems, often very complex and difficult ones, although even simple models might give suggestive solutions of a general nature.

It is our personal view that none of these three approaches, by themselves, will provide definitive answers. On the other hand, we hope that a combined attack, with a constant interplay between theory and experiments of all types, may eventually lead to an understanding of what is clearly a highly complex nonlinear dynamic system whose behavior skips rapidly from one semistable state to another. Our suggestions should be regarded as a preliminary skirmish in such an attack.

ACKNOWLEDGMENTS

We thank Kyle Cave and Charles Gray for helpful comments on drafts of this manuscript. F.C. is supported by the J.W. Kieckhefer Foundation. C.K. is supported by the Air Force Office of Scientific Research, a Presidential Young Investigator Award from the National Science Foundation, and the James S. McDonnell Foundation.

REFERENCES

- Allport, A. 1989. Visual attention. In *Foundations of cognitive science* (ed. M.I. Posner), p. 631. MIT Press, Cambridge.
- Baddeley, A. 1986. *Working memory*. Oxford University Press, United Kingdom.
- Cave, K.R. and J.M. Wolfe. 1990. Modeling the role of parallel processing in visual search. *Cognit. Psychol.* **22**: 225.
- Coltheart, M. 1983. Iconic memory. *Philos. Trans. R. Soc. Lond. B* **302**: 283.
- Crick, F.H.C. 1984. The function of the thalamic reticular complex: The searchlight hypothesis. *Proc. Natl. Acad. Sci.* **81**: 4586.
- Crick, F. and C. Koch. 1986. Certain aspects of the anatomy and physiology of the cerebral cortex. In *Parallel distributed processing: Explorations in the microstructure of cognition* (eds. D. Rumelhart and J.L. McClelland), vol. 2, p. 333. MIT Press, Cambridge.
- Crick, F. and C. Koch. 1990. Towards a neurobiological theory of consciousness. *Semin. Neurosci.* (in press).
- Damasio, A.R., P.J. Eslinger, H. Damasio, G.H. Van Hoesen, and S. Cornell. 1985. Multimodal amnesic syndrome following bilateral temporal and basal forebrain damage. *Arch. Neurol.* **42**: 252.
- Duncan, J. and G.W. Humphreys. 1989. Visual search and stimulus similarity. *Psychol. Rev.* **96**: 433.
- Edelman, G. and V.B. Mountcastle. 1978. *The mindful brain: Cortical organization and the group selective theory of higher brain function*. MIT Press, Cambridge.
- Engel, A.K., P. König, C.M. Gray, and W. Singer. 1990. Stimulus-dependent neuronal oscillations in cat visual cortex: II. Inter-columnar interaction as determined by cross correlation analysis. *Eur. J. Neurosci.* (in press).
- Gray, C.M., A.K. Engel, P. König, and W. Singer. 1990. Stimulus-dependent neuronal oscillations in cat visual cortex: I. Receptive field properties and feature dependence. *Eur. J. Neurosci.* (in press).
- Grinvald, A., R.D. Frostig, E. Lieke, and R. Hildesheim. 1988. Optical imaging of neuronal activity. *Physiol. Rev.* **68**: 1285.
- Holender, D. 1986. Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behav. Brain Sci.* **9**: 1.
- Jones, E.G. 1985. *The thalamus*. Plenum Press, New York.
- Kammen, D.M., P.J. Holmes, and C. Koch. 1989. Cortical architecture and oscillations in neuronal networks: Feedback versus local coupling. In *Models of brain function* (ed. R.M.J. Cotterill), p. 273. Cambridge University Press, United Kingdom.
- Kihlstrom, J.F. 1987. The cognitive unconscious. *Science* **237**: 1445.
- Koch, C. and S. Ullman. 1985. Shifts in selective visual attention: Towards the underlying neural circuitry. *Hum. Neurobiol.* **4**: 219.
- Moran, J. and R. Desimone. 1985. Selective attention gates visual processing in the extrastriate cortex. *Science* **229**: 782.
- Mountcastle, V.B., R.A. Andersen, and B.C. Motter. 1981. The influence of attentive fixation upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *J. Neurosci.* **1**: 1218.
- Poggio, T., V. Torre, and C. Koch. 1985. Computational vision and regularization theory. *Nature* **317**: 314.
- Posner, M.I. 1986. *Chronometric explorations of mind*. Oxford University Press, United Kingdom.
- Posner, M.I., J.A. Walker, F.A. Friedrich, and R.D. Rafal. 1987. How do the parietal lobes direct covert attention? *Neurophysiologia* **25**: 135.
- Rafal, R.D. and M.I. Posner. 1987. Deficits in human visual spatial attention following thalamic lesions. *Proc. Natl. Acad. Sci.* **84**: 7349.
- Sherk, H. 1986. The claustrum and the cerebral cortex. In *Cerebral cortex, Volume 5: Sensory-motor areas and aspects of cortical connectivity* (eds. E.G. Jones and A. Peters), p. 467. Plenum Press, New York.
- Spitzer, H., R. Desimone, and J. Moran. 1988. Increased attention enhances both behavioral and neuronal performance. *Science* **240**: 338.
- Sporns, O., J.A. Gally, G.N. Reeke, Jr., and G.M. Edelman. 1989. Reentrant signaling among simulated neuronal groups leads to coherency in their oscillatory activity. *Proc. Natl. Acad. Sci.* **86**: 7265.
- Steriade, M., E.G. Jones, and R.R. Llinas. 1990. *Thalamic oscillations and signaling*. Wiley, New York.
- Tranel, D. and A.R. Damasio. 1985. Knowledge without awareness: An autonomic index of facial recognition by prosopagnosics. *Science* **228**: 1453.
- Treisman, A. 1988. Features and objects: The fourteenth Bartlett Memorial Lecture. *Q. J. Exp. Psychol.* **40A**: 201.
- Tulving, E. and D.L. Schacter. 1990. Priming and human memory systems. *Science* **247**: 301.
- Van Essen, D. and C.H. Anderson. 1990. Information processing strategies and pathways in the primate retina and visual cortex. In *An introduction to neural and electronic networks* (ed. S.F. Zornetzer et al.), p. 43. Academic Press, San Diego.
- von der Malsburg, C. and W. Schneider. 1986. A neural cocktail-party processor. *Biol. Cybern.* **54**: 29.
- Wilson, M. and J. Bower. 1990. Cortical oscillations and temporal interactions in a computer simulation of piriform cortex. *J. Neurophysiol.* (in press).
- Wise, S.P. and R. Desimone. 1988. Behavioral neurophysiology: Insights into seeing and grasping. *Science* **242**: 736.