6 Binding and selective visual attention

6.1 The binding problem

There are an almost infinite number of possible, different objects that we are capable of seeing. There cannot be a single neuron, often referred to as a grandmother cell, for each one. The combinatorial possibilities for representing so many objects at all different values of depth, motion, color, orientation and spatial location are simply too staggering. This does not preclude the existence of sets of somewhat specialized neurons responding to very specific and ecological highly significant objects, such as the neurons responding to aspects of faces in inferotemporal cortex of primates[24].

It seems likely that at any moment any individual object in the visual field is represented by the firing of a *set* of neurons. This would not cause any special problem if the members of the set were in close proximity (implying that they probably interact somewhat), received somewhat similar inputs and projected to somewhat similar places. But because any object will have different characteristics (form, color, motion, orientation) that are processed in several different visual areas, it is highly reasonable to assume that seeing any one object often involves neurons in many different visual areas. The problem of how these neurons temporarily become active as a unit is often described as 'the binding problem'. As an object seen is often also heard, smelt, or felt, this binding must also occur across different sensory modalities.

Our experience of perceptual unity thus suggests that the brain in some way binds together, in a mutually coherent way, all those neurons actively responding to different aspects of a perceived object. In other words, if you are currently paying attention to a friend discussing some point with you, neurons in area MT that respond to the motion of his face, neurons in V4 that respond to its hue, neurons in auditory cortex that respond to the words coming from his face and possibly the memory traces associated with recognition of the face all have to be 'bound' together, to carry a common label identifying them as neurons that jointly generate the perception of that specific face.

6.2 Types of binding

Binding can be of several types. In a sense a neuron responding to an oriented line can be considered to be binding a set of points. The inputs to such a neuron are probably determined by genes and by developmental processes that have evolved due to the experience of our distant ancestors. Other forms of binding, such as that required for the recognition of familiar objects such as the letters of a well-known alphabet, may be acquired by frequently repeated experience; that is, by being overlearnt. This probably implies that many of the neurons involved have as a result become strongly connected together. (Recall that most cortical neurons have many thousands of connections and that initially many of these may be weak.) Both these types of binding are likely to have a large but limited capacity. These are the types of binding with which Damasio[7] is mainly concerned.

The binding we are especially concerned with is a third type, being neither epigenetically determined nor overlearnt. It applies particularly to objects whose exact combination of

responding to fairly simple features (such as an oriented edge) occupying a tiny part of the visual field.

The neurons in the higher cortical areas respond to more complex features, e.g., aspects of faces, and their receptive fields cover much larger areas. The different cortical areas respond, in general, to different features. For example, the neurons in area MT are mostly interested in motion and depth, those in area V4 in color and shape, and those in 7a in position in space relative to the head or the body. *So far no single area has been found whose neurons correspond to everything we see.* How is it, then, that we seem to have a single coherent visual picture of the scene before us?

The obvious suggestion is that at any one time the relevant neurons in many cortical areas cooperate together to form some sort of global activity and that this global activity corresponds to visual awareness. This is the hypothesis that we shall try to fill out with psychological and neurobiological details.

5.2 Convergence zones

Which particular cortical areas contribute directly to visual awareness? For example, is the first visual area, V1, closely involved? Damasio[7] has also suggested that many cortical areas take part at any moment at several different levels. It is probably significant that Damasio arrived at this idea largely from considering the detailed effects of various types of brain damage, evidence rarely used by other theorists. He suggests that the synchronous activity in the various cortical areas is coordinated by feedback projections from 'convergence zones'.

What exactly convergence zones means in neural terms is somewhat obscure. We suggest that convergence zones may mainly refer to the neurons (or a subset of them) that project 'backwards' (see Figure 1), such as those that project from the second visual area back to the first one. Thus the term zones may turn out to convey a somewhat misleading impression, since such neurons may be sprinkled fairly uniformly through certain layers of the neocortex. Nevertheless the basic idea is a valuable one. It points to the possibility that, in addition, convergence zones may also exist in other places, such as the thalamus or the claustrum. (The claustrum[21] is a thin sheet of cells located just below the neocortex of the insula. It receives axons from almost all cortical areas and, in return, projects back to them. Its caudal region is largely visual.)

One would certainly expect areas near to the hippocampus, such as the inferotemporal region, as well as those near to the higher levels of the motor system to be closely involved in visual awareness. How far awareness is directly correlated with every neocortical level of the visual system remains uncertain, as is the involvement of the thalamus, although certain parts of it, such as its reticular nucleus[22] and the pulvinar[23] may be involved with attention.

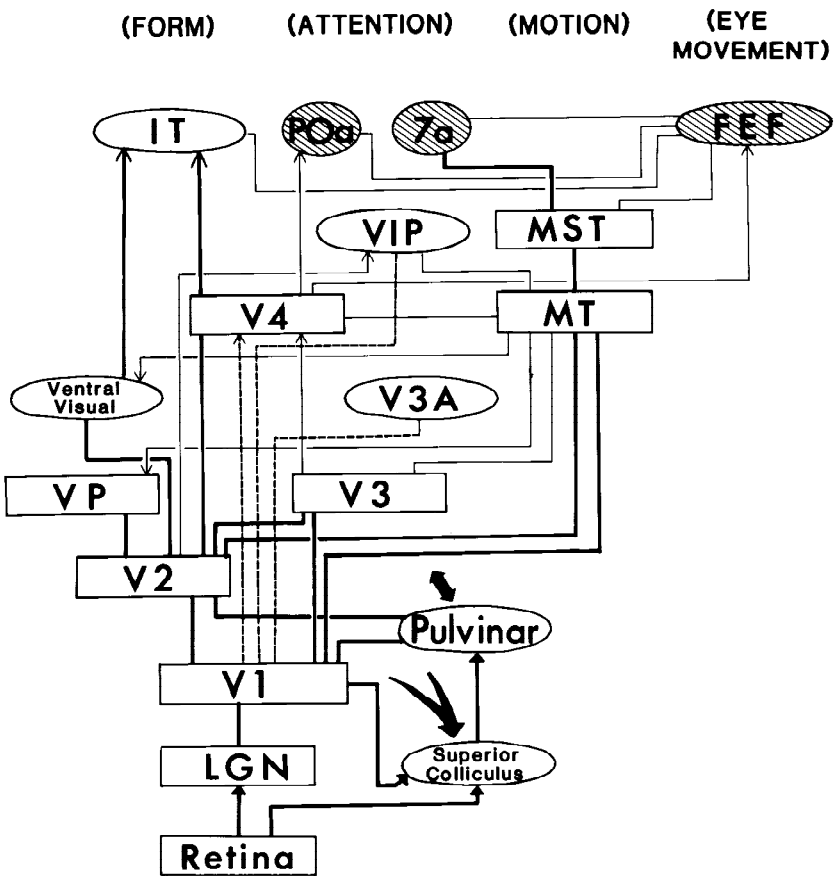

Figure 1:
Some of the major cortical and sub-cortical visual areas in the macaque monkey. Major pathways are indicated by bold lines, minor pathways by thin lines; dashed lines weak or uncertain. The shaded ellipsoids correspond to higher-level visual-motor areas. Most cortical areas project back to the superior colliculus as well as to the different maps in the pulvinar, which is part of the thalamus. Furthermore, every forward projection has usually an equally strong, if not stronger, back-projection associated with it (not shown). These feedback pathways usually terminate outside cortical layer IV, while the forward projection terminates most densely in this layer. Areas related to the control and expression of visual attention include the superior colliculus and the pulvinar, as well as areas POa and 7a, part of the posterior parietal cortex. Abbreviations: FEF—frontal eye fields; IT—inferior temporal lobe; LGN—lateral geniculate nucleus; MT, MST—motion-processing areas; V1—primary visual area; V2-4—higher-order visual processing areas; VIP—ventral interparietal area; VP—ventral posterior. The figure is from D. Van Essen and J. Maunsell, personal communication. For more information, see ref 20.