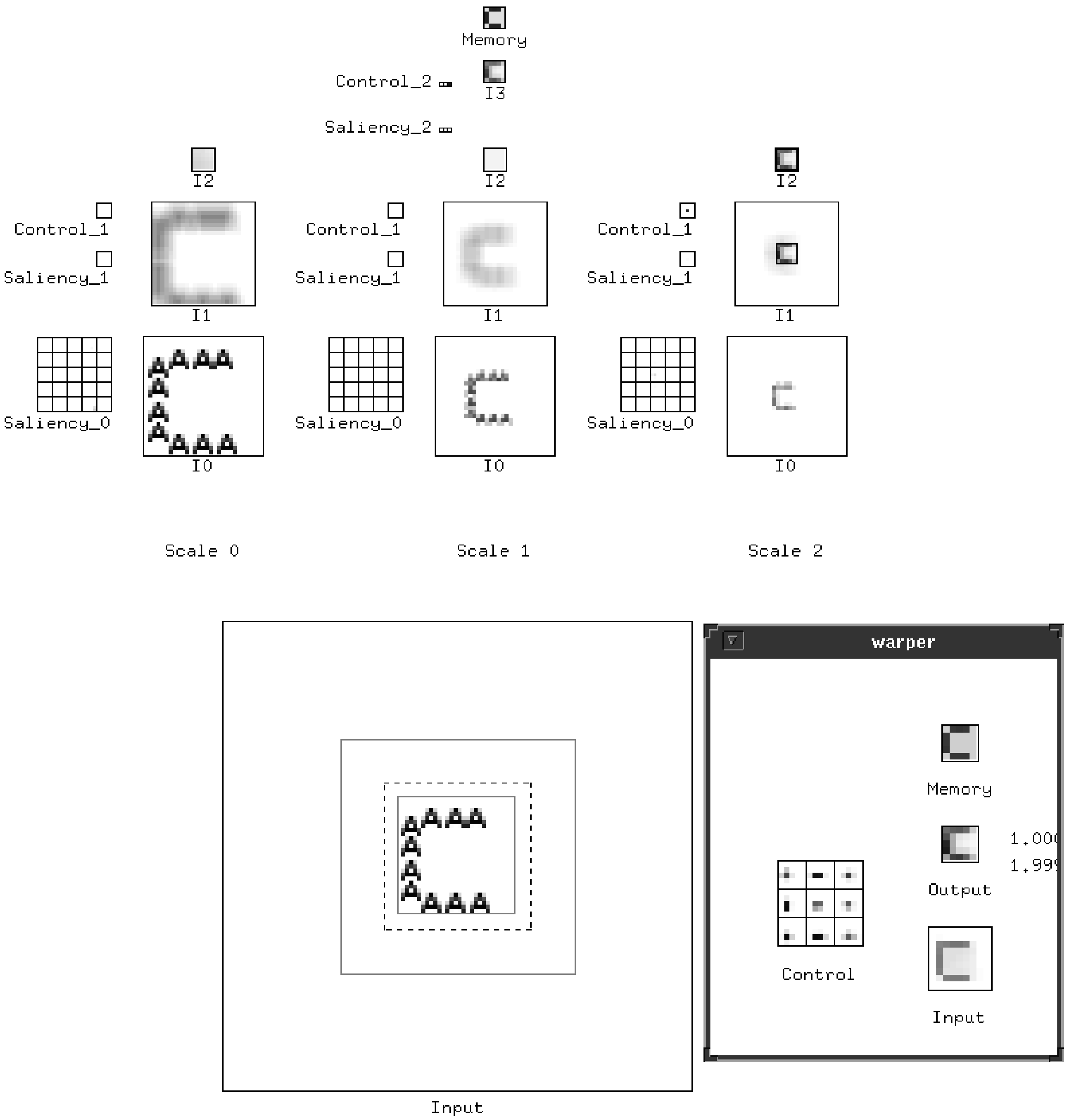

Figure 2.28: **Simulation of the stack circuit (local vs. global).** The circuit is shown attending to the global 'C' shape.

Chapter 3

Neurobiological substrates and mechanisms

We now turn to the issue of how the routing circuit we built up in the previous chapter may be implemented in the brain of the macaque monkey. The major areas that will be of interest to us, along with their anatomical relationships, are shown in Figure 3.1. Information from the retino-geniculo-striate pathway enters the visual cortex through area V1 in the occipital lobe and proceeds through a hierarchy of cortical visual areas that can be subdivided into two major functional streams (Ungerleider and Mishkin, 1982). The so-called "form" pathway leads ventrally through V4 and inferotemporal cortex (IT) and is mainly concerned with object identification, regardless of position or size. The so-called "where" pathway leads dorsally into the posterior parietal complex (PP), and seems to be concerned with the locations and spatial relationships among objects, regardless of their identity. The pulvinar, a sub-cortical nucleus of the thalamus, makes reciprocal connections with all of these cortical areas (cf. Robinson and Petersen, 1992) and also receives a projection from the superior colliculus, a midbrain structure involved in directing overt visual attention (i.e., eye movements).

The following sections describe how the dynamic routing circuit may be mapped onto this collection of neural hardware. We begin by describing how the multiscale stack and the intermediate stages of the routing circuit may be represented in the cortex. We then turn to the proposed neural substrates for control, and possible gating mechanisms for modulating connection strengths.

3.1 Routing circuit substrates

Since we are primarily interested in how invariant representations are formed for object recognition, we will focus on the "form" pathway from V1 to IT. [1] Figure 3.2 shows a schematic of the main cortical areas in this pathway. Each area is drawn to scale according to its relative size in one dimension (square root of the area). The

[1] More generally, we can conceive of routing taking place in the other visual processing streams as well—for example, in the motion pathway for making fine discriminations of motion (Nowlan and Sejnowski, 1993; Van Essen and Anderson, 1990).

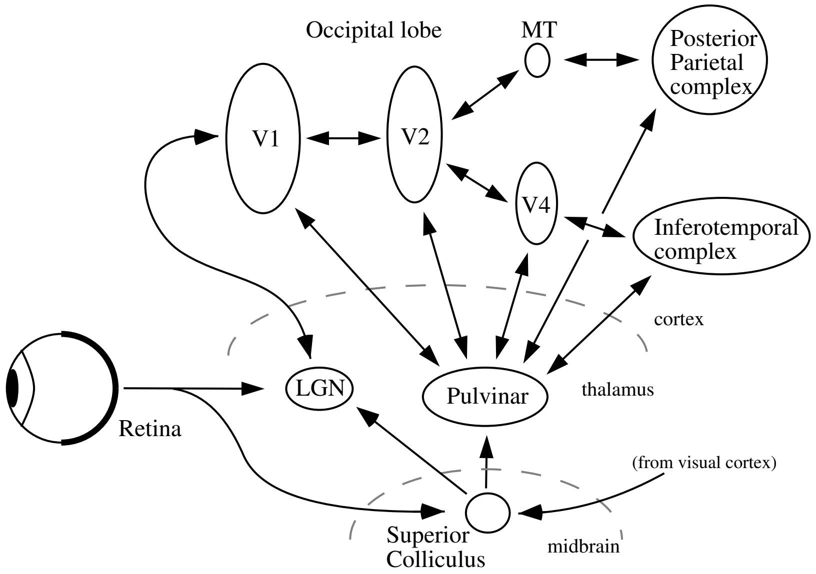

Figure 3.1: **Neurobiological substrates.**

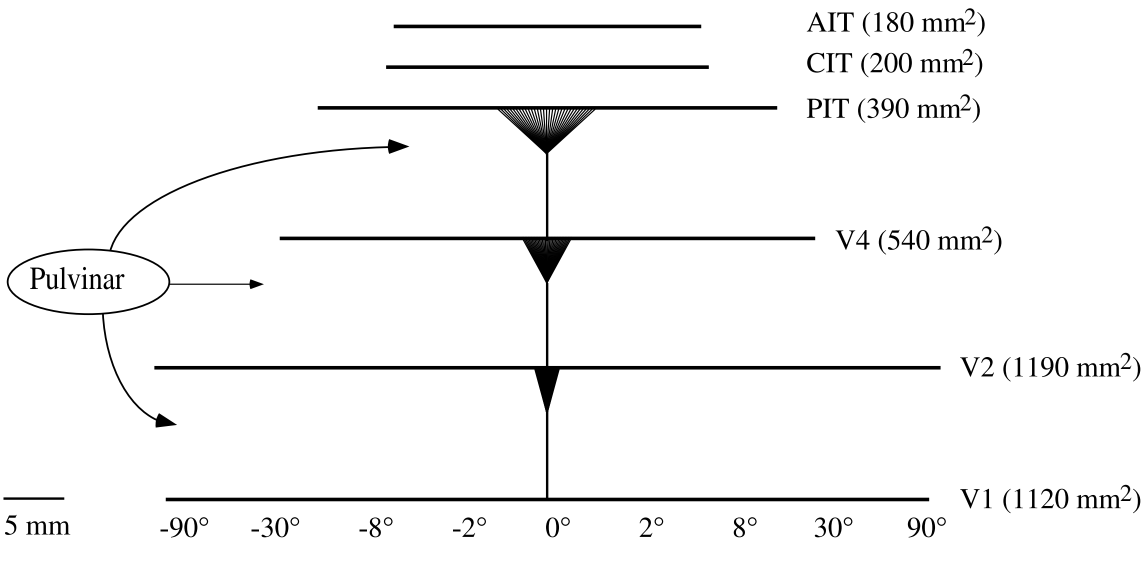

Figure 3.2: **Schematic of the main cortical areas in the "form" pathway.** The size of each area is drawn to scale ($\sqrt{\text{area}}$). The fan-in at each stage is drawn to comprise 30 outputs ($\sim$ 1000 in 2D), and the divergence increases by a factor of two at each stage. The pulvinar makes connections with all these visual areas, and is proposed to be a source of the control signals for modulating connection strengths (see Section 3.2 below).

fan-out shown at each stage is derived from the fact that receptive field size and visual field overlap between hemispheres approximately doubles at each stage in this pathway (Gattass et al., 1985), in addition to the anatomical observation that areal inter-connections increase their divergence and become more patchy in higher stages of the visual cortical hierarchy (Rockland, 1992; Van Essen et al., 1986; Van Essen et al., 1990; Van Essen and DeYoe, 1993; DeYoe and Sisola, 1991). This is only a rough depiction, however, and more data are needed to construct a firmer quantitative picture.

In previous work (Olshausen, Anderson, and Van Essen, 1993), it was described how the multistage circuit of Figure 2.4 could be mapped onto these stages of cortical visual processing. Here, we shall propose that area V1 forms a multiscale "stack" representation of the retinal image, and that the major intermediate visual areas in the form pathway (i.e., V2, V4) serve as intermediate stages for routing visual information from the stack in V1 into a progressively more position and scale invariant representation at higher stages. This process culminates with information being represented within an object-centered reference frame at the initial stages of IT (i.e., PIT). The size of each area would be expected to correspond roughly to the number of "sample nodes" in each area. Each node would correspond to a vector of features extracted in that area for a particular position and size of visual space (e.g., a full set of orientations in V1). Cells at higher stages of IT (i.e., CIT, AIT, or in the STS) would then perform their analyses on the contents of the window of attention, with variations in position and scale removed (e.g., face cells). Although we do not yet fully understand the nature of form processing occurring in this pathway—especially in the intermediate stages—we will assume we can leave this as an unknown and deal with issues of routing independently.

The multiscale "stack"

In order to propose a quantitative model for a multiscale stack representation in V1, we need to specify 1) the highest resolution available as a function of eccentricity, and 2) the resolution ratio between adjacent levels of the stack. For the primate visual system, the highest resolution available at eccentricity E is given by

$$\delta(E) = .01(E + 1.3) \ \text{deg}\,, \tag{3.1}$$

where δ gives the one-dimensional spacing between samples nodes in the retina (Van Essen and Anderson, 1990). In two dimensions, each sample node would cover an area of approximately δ^2. The resolution ratio between adjacent levels can be inferred from the spatial-frequency bandwidths of V1 cells, since an efficient coverage of the spatial frequency domain would require that the spacing in spatial frequency be related to bandwidth. Since the bandwidths of V1 cells tend to hover in the range of 1 to 1.5 octaves (De Valois et al., 1982), we will assume that resolution approximately doubles for each successive level of the stack.

Given these constraints, a stack comprising approximately 6 levels would suffice to cover the visual field up to $\pm 50°$ eccentricity (a total of $100°$), as shown in Figure 3.3. Beyond this eccentricity, retinal ganglion cell sample spacing is no longer

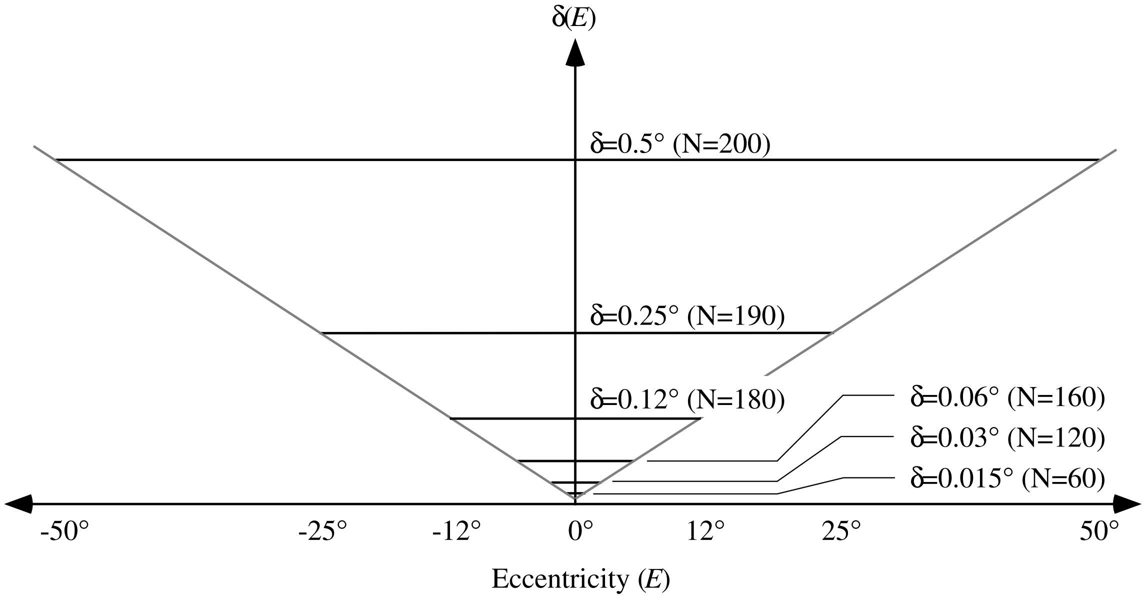

Figure 3.3: **A six-level "stack" model for V1.**
δ denotes the sample spacing and N denotes the number of nodes in each level of the stack.

scale invariant (i.e., no longer adheres to the linear relationship of Equation 3.1), and it also begins to be somewhat unreasonable to suppose that objects beyond this size [2] would be recognizable as a whole. The number of sample nodes in 1D for each level of the stack is given by

$$N = \frac{2\,E}{\delta(E)} = \frac{2\,E}{.01(E+1.3)} \tag{3.2}$$

which will equal approximately 200 for $E \gg 1.3°$. At eccentricities near or below $1.3°$ the number of nodes within a level will be fewer. The total number of sample nodes for the entire stack will thus be on the order of $6 \times 200^2 = 240,000$, which is about equal to the total number of sample nodes delivered by the optic nerve for the central 100° (80% of the total) when one takes into account the fact that information is divided into on- and off-channels, magno and parvo streams, and different spectral bands (Van Essen and Anderson, 1990). The highest resolution level of the stack will have a mean sample spacing of about .015°, which implies a peak spatial-frequency tuning for each sample node in the range of 15 cy/deg, assuming that the optimum spatial-frequency for a given level is somewhat less than half the sampling frequency. The lowest frequency nodes will have a spacing of about 0.5° with an expected peak spatial-frequency tuning of about 1 cy/deg or lower.

Figure 3.4 illustrates the stack in cortical dimensions, where space has been loga-

[2]This would be the extent of a fully stretched hand when held 3 inches from the eye.

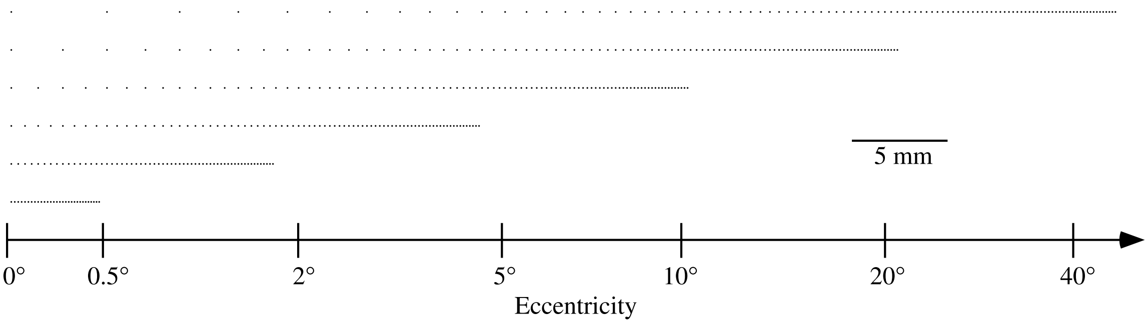

Figure 3.4: **The stack in cortical dimensions.**
Visual space has been logarithmically compressed according to Equation 3.3. Sample nodes for the lower resolution levels are distributed sparsely in the foveal cortex, but densely in the periphery.

rithmically compressed according to

$$X_c(E) = 10 \log(\frac{E + 0.8}{0.8}) \tag{3.3}$$

where X_c gives the cortical distance in millimeters from the origin (i.e., fovea) of V1. (Equation 3.3 was obtained by integrating the formula for cortical magnification factor, $10(E+.8)^{-1.1}$ mm/deg, as given by Van Essen et al. (1984), and rounding the exponent down to -1.0 to make the integration simple.) In foveal V1, the highest resolution nodes would be spaced by about 200μ, and the lowest frequency nodes would be spaced by about 6 mm. With increasing eccentricity, the spacing between low resolution nodes will decrease, and the total number of levels will decrease as well until only the lowest resolution level is left. The spacing between the lowest resolution nodes at the largest eccentricity (50°) will be the same as the spacing between the highest resolution nodes in the fovea ($\approx 200\mu$). Note that since the density of sampling nodes decreases by a factor of four (in 2D) for each octave decrease in resolution, the total density does not vary appreciably with eccentricity, even though there are many more levels of the stack represented in the fovea than in the periphery.

Are these characteristics consistent with the physiological and anatomical data on V1? Determining the number and range of spatial-frequency tuned cells in V1 is difficult because of the disparate and conflicting data, so a more in depth discussion of this issue is provided in Appendix B. To summarize, though, it appears that within foveal V1, the necessary range of spatial frequencies exists: Most cells have very small receptive fields with central excitatory zones on the order of 2-4 minutes in diameter, which puts their peak spatial-frequency in the range of 8-15 cy/deg (Parker and Hawken, 1988). Cells with larger receptive fields become fewer in number with increasing diameter, bottoming out with a very small number on the order of 1 cy/deg (DeValios et al., 1982; Tootell et al., 1988). There is also evidence that the range of peak spatial-frequencies is largest in the fovea, and that this range progressively decreases with eccentricity, as would be expected of a multiscale stack representation. It is not possible to determine whether the relative spacing of sample nodes at different resolutions is consistent with a stack model, but one can infer from the relative numbers of cells at each spatial-frequency that lower spatial frequency tuned cells would be spaced farther apart, assuming that cells are spread uniformly.

Intermediate stages and wiring constraints

The number and size of intermediate stages of routing required depends on the total input-output convergence and the maximum allowable fan-in (number of inputs per neuron). The input to the routing circuit for each scale will be a 2D array comprising approximately 200x200 nodes (as described above), and the output of the routing circuit is hypothesized to be a relatively small array, comprising on the order of 30x30 sample nodes. (This estimate is based largely on spatial acuity and recognition studies that provide hints about the resolution of the window of attention—see Discussion, Section 4.1). Thus, the total convergence for the routing circuit for each scale will be about 40,000:1. Since the maximum allowable fan-in is on the order of 1000 inputs per neuron (Cherniak, 1990; Douglas and Martin, 1990a), the routing circuit for each scale must be broken into several stages.

A nominal configuration would be for each routing circuit to be broken into two stages, as shown in Figure 3.5*a*. This circuit is simply a scaled-up version of the circuit described in the previous chapter (Fig. 2.9), where the middle layer is now composed of modules of size 30. On the right, the circuit is pictured in terms of its fan-out, which is more neurobiologically relevant. Although this architecture does not appear to pose any anatomical problems as drawn, it must be kept in mind that the routing circuits for each each scale will be superimposed in register in the cortex. Thus, the low resolution nodes will need to have a very great divergence in terms of cortical distance in order to span 30 nodes in the next cortical area (cf. Fig. 3.4), which is rather implausible. We can reduce the fan-in/fan-out at each stage by breaking the routing circuit into more stages, as shown in Figure 3.5*b,c*.

Note that since the divergence is not as great a problem for the smallest scales, fewer stages of routing would be required for these levels of the stack, which is consistent with the direct projections observed from V1 to V4 for the central visual field only (Yukie and Iwai, 1985). For example, a high resolution level of the stack may pass merely through V1-V4-IT, while a low resolution level may be routed through V1-V2-V4-IT. Also, since the spacing between inputs doubles at each stage, the circuit is consistent with the observed doubling in receptive field size and progressively greater divergence in connectivity patterns at each stage, as mentioned earlier. (There is a big jump at the top, though, but this is not a necessity, just a minimal configuration.)

The particular scheme adopted here of dedicating a distinct routing stream for each scale and switching between scales at the top stage is not a strict requirement. It is quite possible that switching between scales could be done at earlier stages of the circuit, thus allowing the resources at the top stages to be shared between scales. The scheme proposed here has the advantage that the routing circuits for each scale can work independently, which allows the possibility that the window of attention could be set in advance for the upcoming attentional fixation at that scale while attending

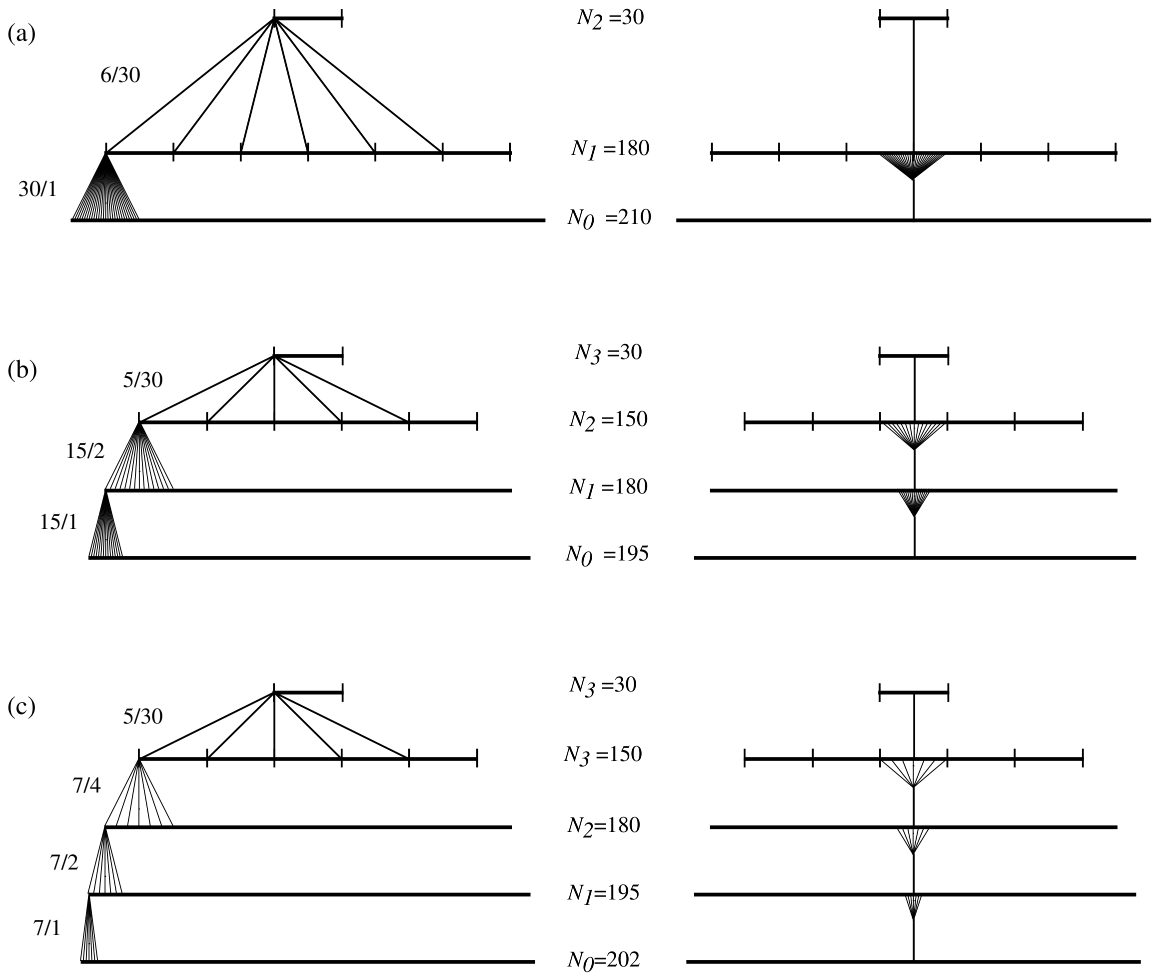

Figure 3.5: **Some possible multistage routing circuits.**
The numbers to the left of each stage (x/y) denote the fan-in (x) and the spacing between inputs (y) for the stage. N_l denotes the number of sample nodes in 1D for each layer. The vertical tick marks in the next to last layer denote the modules, of size 30 (equal to the size of the window of attention). *a*, a nominal configuration, corresponding to the circuit of Figure 2.9. The fan-in/fan-out in the first stage can be reduced by adding more intervening stages, as in *b,c*.

somewhere else at a different scale. More importantly, it allows for a wide variety of features to be computed in parallel at different scales in the intermediate stages (V2,V4), which would provide a richer set of "primitives" for object matching within the window of attention.

3.2 Control substrates

Postulating the substrates for control requires that we specify 1) the anatomical location of the control neurons, 2) bottom-up sources for driving the control neurons during blob search—i.e., the "saliency map," and 3) top-down sources for driving the control neurons during recognition. We consider here each of these issues in turn.

Control neurons

The pulvinar nucleus of the thalamus is proposed as a major source of the control signals for routing information through the cortex. Since the pulvinar is reciprocally connected to all areas in the form pathway, it is a good candidate for modulating information flow from V1 to IT. The pulvinar also receives a massive projection from the superior colliculus, which is known to encode the direction of saccade targets and may also be involved in setting up attentional targets (Gattass and Desimone, 1991, 1992; Posner and Petersen, 1990). In addition, neurophysiological studies (Petersen et al., 1985, 1987), lesion studies (Desimone et al., 1990; Bender, 1988; Rafal and Posner, 1987) and PET studies (LaBerge and Buchsbaum, 1990; Corbetta et al., 1991) of the pulvinar suggest that it plays a role in engaging visual attention, or filtering out unattended stimuli. A detailed analysis of these studies and their relevance to the model is provided in Appendix C.

A subcortical nucleus such as the pulvinar also has the important property of being spatially localized while at the same time being able to communicate with vast areas of the visual cortex. The relative proximity of pulvinar neurons to each other would facilitate the competitive and cooperative interactions among the control neurons which are necessary for choosing a single position and size of the attentional window and for maintaining spatial relationships within the window. Although it is not known whether such interactions exist among pulvinar neurons, Ogren and Hendrickson (1979) have reported the existence of interneurons with elaborate dendritic trees approaching 600μ in diameter, which could mediate communication among pulvinar neurons. In addition, neuropharmacological experiments by Petersen et al. (1987) have shown that enhancing or depressing inhibition within the pulvinar can respectively slow down or speed up attentional shifts, which is suggestive of lateral inhibitory connections within the pulvinar. An analogous function might also be served by the reticular nucleus of the thalamus, which is an inhibitory structure through which pulvinar neurons project on their way to the cortex. One study in *Galago* (Conley and Diamond, 1990) has shown that the pulvinar projects quite diffusely into the reticular nucleus, which would be desirable for a winner-take-all type circuit.

To first order, it would make sense for each stage of the routing circuit to have its own set of control neurons. The anatomical subdivisions of the pulvinar correspond roughly with this scheme, insofar as the inferior pulvinar projects mainly to lower areas (V1, V2) and the lateral and medial pulvinar to higher areas (V4, IT). (A small fraction of inferior pulvinar neurons (10%) have been shown to project to both V1 and V2 (Kennedy and Bullier, 1985).) The control neurons for the lower stages would need to compete only locally, since these stages would be more concerned with making local adjustments in the position and scale of the window of attention. Control neurons at the highest stage would need to compete globally, since these stages are setting the position and scale of the window of attention for the entire scene.

The number of control neurons that would be required for the routing circuit depends on how many cortical synapses are modified by each control neuron. Theoretically, the minimal number of control neurons is given by

$$\#\text{ of control neurons} = \frac{(\#\text{ of output nodes}) \times (\text{fan-in per node})}{(\#\text{ of synapses per control block})}.$$

Assuming that the control blocks comprise a maximum of 1000 synapses each, then the minimum number of control neurons required for each stage for each scale of the routing circuit would be on the order of the number of output nodes of each stage (since the maximum fan-in per node is about 1000). The total number of control neurons required for each stage across all scales would be obtained by multiplying this number by 6 (the number of levels in the stack). Thus, for the circuit of Figure 3.5*b,c*, approximately $6 \times 40,000 = 240,000$ control neurons would be required for the first stage, approximately $6 \times 30,000 = 180,000$ for the second stage, etc., which is well within the estimated number of neurons in the pulvinar.[3] However, each output node in the circuit actually corresponds to a multitude of neurons representing various features, such as local orientation, texture, etc. Thus, each pulvinar control neuron would require an additional fan-out for controlling the inputs to all the neurons corresponding to an output node. Since there may be hundreds of neurons for each node, the pulvinar neurons would need to amplify their fan-out via other neurons (a fan-out of 100,000 for pulvinar neurons is probably too large to be plausible). This could possibly be subserved by neurons residing in the deeper layers (5 and 6) of the cortex, as proposed previously by Van Essen and Anderson (1990). Control might then be implemented in a hierarchical fashion, with each pulvinar neuron specifying how information is routed between nodes, and cortical control neurons specifying how information is routed between the neurons belonging to each node.

Bottom-up control sources (saliency map)

As discussed in the previous chapter, the control neurons may be driven by bottom-up, or low-level signals—such as "pop-out" in motion, color, texture, etc.—in order to direct the window to salient regions of the input. Since each of the saliency measures

[3]The pulvinar has somewhat lower neuronal density than the LGN, but also is several times larger. Since the LGN contains $\sim 10^6$ projection neurons, this would constitute a reasonable lower bound for the number of neurons in the pulvinar.

may be computed in a separate cortical area, it would be advantageous to fuse them together into a single representation of saliency—such as the "saliency map" proposed by Koch and Ullman (1985). Two possible anatomical substrates for such a saliency map are the posterior parietal complex and the superior colliculus.

The posterior parietal complex (PP) is known to play an important role in attentional processes. Some studies have reported that neurons in this area show an enhanced response to attended targets within their receptive fields, even when no eye movements are made (Bushnell et al., 1981). Others have reported a 3-fold enhancement for *unattended* targets when the animal is in an attentive state (Mountcastle et al., 1981), or even a relative suppression for attended targets as opposed to unattended targets (Robinson et al., 1991; Steinmetz et al., 1992). In addition, lesion studies show that damage to the parietal lobe in humans hinders the ability of other objects in the field of view to attract the attentional window away from the currently attended location (Posner et al., 1984). Taken together, these results suggest that PP may be representing the locations of potential attentional targets, as opposed to targets already being attended. This is exactly the property we would expect of a saliency map. If this is the case, then these neurons would drive the control neurons in the pulvinar which compete to select the locus of the window of attention.

This proposal contains at least two potential weaknesses, however. One possible drawback is that PP neurons typically have relatively long latencies—$\sim$ 100 ms (Robinson et al., 1978; Duhamel et al., 1992)—which is hard to reconcile with psychophysical data that imply that attention takes $\sim$ 50 ms to move to a new location in the visual field (Saarinen and Julesz, 1991; Nakayama and Mackeben, 1989). A possible solution to this dilemma is that the superior colliculus may supplement PP by acting as a crude saliency map, but with a quicker response time due to its direct retinal input (the latency of neurons in the superficial layers of the superior colliculus is in the range of 40-50 ms; Goldberg and Wurtz, 1972). The other drawback of using PP as a saliency map is that the currently available anatomical data seem to offer relatively few direct pathways by which PP could influence those pulvinar neurons that would be able to modulate connection strengths in the "form" pathway, since PP and V4/IT connect with rather segregated portions of the pulvinar (Baleydier and Morel, 1992). However, there do exist indirect pathways, such as through the superior colliculus, that may provide viable alternatives. Another possibility of course is that the salience measures made along different feature dimensions in different cortical areas could drive the control neurons in the pulvinar directly via cortico-fugal pathways.

Recognition-guided control sources

During recognition, top-down influences will need to take over to refine the position and size of the attentional window for object matching, as depicted in Figures 2.15 and 2.21. The pulvinar would thus need to alternate between bottom-up and top-down sources of input as attention moves from one object to the next. Top-down guidance during recognition would presumably be propagated to lower cortical areas via cortico-cortical feedback pathways from IT, or IT may influence control neurons in the pulvinar directly via its diffuse projections to many nuclei within the pulvinar. Alternatively, IT could supply top-down guidance primarily to *cortical* control neurons via the feedback pathways. Under this scenario then, the pulvinar's role would be analogous to that of a general in an army—coarsely specifying a plan of action, which the cortical control neurons refine into a concise remapping under top-down, or object-based guidance from IT.

3.3 Gating mechanisms

In order for the control neurons to modulate connection strengths we must postulate some possible neuronal gating mechanisms. This is not at all a new idea, as neural gating mechanisms are believed to play an important role in many aspects of nervous system function. For example, the extent to which a noxious stimulus is perceived as painful varies greatly as a function of one's emotional state and other external factors. This is subserved at least in part by gating mechanisms in the spinal cord, where descending fibers from the raphe nuclei form part of a control system that modulates pain transmission via presynaptic inhibition in the dorsal horn (Fields and Basbaum, 1978). Gating mechanisms are also thought to play an important role in sensori-motor coordination; for example, there are many instances in which spinal cord central pattern generators gate sensory inputs according to the phase of the movement cycle in which the input occurs (Sillar, 1991). A somewhat different form of gating seems to take place in the LGN, where thalamic relay cells exhibit two distinct response modes: a *relay* mode, in which cells tend to more or less faithfully replicate retinal input, and a non-relay *burst* mode, in which cells burst in a rhythmic pattern that bears little resemblance to the retinal input (Sherman and Koch, 1986). In this instance, the reticular nucleus of the thalamus is thought to be the source of the signal that switches the LGN into the non-relay burst mode.

Although there is as yet no explicit evidence for gating mechanisms in the visual cortex, there are several possible biophysical mechanisms that would allow control neurons to gate synapses along the V1-IT pathway. Pre-synaptic inhibition, as in the spinal cord, would probably provide the most localized gating effect. However, to date there exists no morphological evidence for this type of synapse in the visual cortex (Berman et al., 1992). Postsynaptically, a control neuron could decrease or possibly nullify the efficacy of a cortico-cortical synapse via shunting inhibition. Evidence for this type of mechanism playing a role in orientation or direction tuning is mixed, with some for (Volgushev et al., 1992; Pei et al., 1992) and some against (Douglas et al., 1988). Another possible post-synaptic gating mechanism could be realized via the combined voltage- and ligand-gated NMDA receptor channel, which has been shown to play an important role in normal visual function (Nelson and Sur, 1992; Miller et al., 1989). In this case, a control neuron could effectively boost the gain of a cortico-cortical synapse by locally depolarizing the membrane in the vicinity of the synapse. Also, there exist voltage-gated Ca^{++} channels in dendrites (Llinas, 1988) that could provide non-linear coupling between inputs. Evidence for non-linear interactions of this type have been reported for synaptic inputs into layer 1 of neocortex (Cauller and

Connors, 1992). All of these mechanisms, and possibly others, offer a multiplicative-type effect that is suitable for gating information flow through the cortex (see also Koch and Poggio, 1992).

Under an inhibitory gating scheme, such as shunting or pre-synaptic inhibition, the control neurons would need to become active only when attention is actively engaged on an object. The finer the resolution desired within the window of attention, the more the control neurons would need to be engaged. The absence of any activity on the control neurons would correspond to the all-connections-open, or inattentive state, in which neurons in IT would exhibit the very large receptive fields observed in anesthetized or inattentive animals (Gross et al., 1972; Desimone et al., 1984).

Under an excitatory gating scheme, such as via NMDA receptors, one would need to hypothesize the existence of a gain control mechanism working in concert with the control neurons. When no control signals are provided, cortical input would be rather weak, and the firing threshold of pyramidal cells should be lowered to let all information through. When control signals are present to boost the gain of individual synapses, however, the threshold should be raised. This way, the unboosted synapses will be essentially suppressed to a relatively low strength. Threshold adjustment could perhaps be subserved by chandelier cells, which make strong inhibitory connections exclusively onto the axon initial segment of pyramidal cells (Douglas and Martin, 1990b). Evidence that gain control mechanisms indeed exist in visual cortex has been established in previous physiological studies (Ohzawa et al., 1982; Pettet and Gilbert, 1992).

From a computational viewpoint, gating of inputs within individual dendrites provides a much higher degree of flexibility than would merely gating the outputs of pyramidal cells. Since the output of a pyramidal cell may branch to several cortical areas and make synaptic connections to a multitude of neurons, any modulation of the cell's output will simply be duplicated at all these subsequent input points. Gating inputs within the dendrites, on the other hand, allows the non-linear computation of many intermediate results ($\sum_k c_k \Gamma_{ijk} I_j^{in}$) within the post-synaptic membrane, which can then be summed together within a single cell. This results in a computational structure that is orders of magnitude richer (Mel, 1992), and provides a higher degree of flexibility in sculpting patterns in connection space (cf. Fig. 2.3). The demonstrable computational advantage of dendritic gating mechanisms for visual processing motivates the need to specifically look for such mechanisms experimentally. (See also Desimone, 1992, for a discussion of output vs. input gating mechanisms.)

Chapter 4

Discussion

We shall see in this chapter that the proposed neurobiological correlates of the model lead to a number of neurophysiological, neuroanatomical, and psychophysical predictions that can be tested experimentally. We shall also describe what distinguishes this model from other network models that have been proposed for visual attention and invariant pattern recognition, and how the model can be generalized to be understood in a broader context. Finally, we shall briefly examine some of the unresolved issues that remain as topics for future research.

4.1 Predictions

Neurophysiology

The most obvious prediction of the dynamic routing circuit model is that the receptive fields of cortical neurons should change their position or size as attention is shifted or rescaled. This effect should be especially pronounced in higher cortical areas. Some support for this prediction comes from the neurophysiological findings of Moran and Desimone (1985) in areas V4 and IT of primate visual cortex. As schematized in Figure 4.1, they found that if two bar-shaped stimuli were placed within the classical receptive field (CRF) of a V4 cell, and the animal was trained to attend to only one of them, then the cell's response to the unattended stimulus was substantially attenuated. This is what one one would expect from our routing circuit, since the pathways between the cell and the unattended stimulus would be effectively disabled in this case (Fig. 4.1*c*). They also found that the V4 cell responded to an unattended stimulus anywhere within its CRF when the animal attended a stimulus outside the CRF. This effect is also predicted by the model, because once a V4 cell lies outside the region of interest in V4 it no longer needs to restrict its inputs (Fig. 4.1*d*). Indeed, other targets of V4, such as those in the posterior parietal cortex, would presumably be interested in the information from regions lying outside of the attentional beam.

While Moran and Desimone's findings offer some support for attentional modulation effects predicted by the model, they did not attempt to map receptive fields under different attentional conditions with any precision; thus, their results do not address the more specific effects predicted by the model. For a cell in one of the

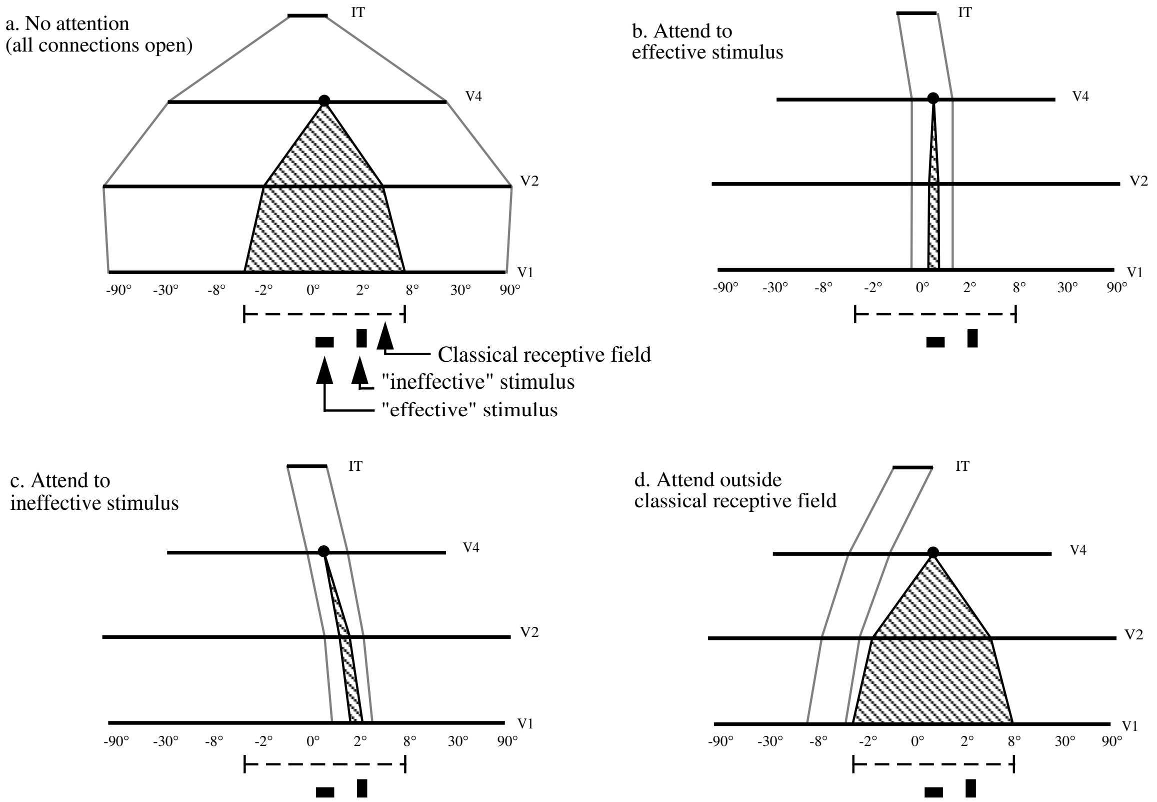

Figure 4.1: **The dynamic routing circuit interpretation of the Moran and Desimone (1985) experiment.**
The node in layer V4 indicates the cell under scrutiny. The hashed region indicates those connections to the cell that are enabled; the others are disabled. The bounds of the window of attention in each area are shown by the stippled lines. (a) In the non-attentive state, all connections will be open and the effective stimulus can excite the cell anywhere within its classical receptive field. (b) When attending to the effective stimulus, the cell's response should be unaltered since the neural pathways to the stimulus are still open. (c) When attending to the ineffective stimulus, the cell's response should decrease substantially since the neural pathways to the effective stimulus are gated out. (d) When attending outside the cell's classical receptive field, there is no need to gate the cell's inputs since it is no longer taking part in the process of routing information within the window of attention.

intermediate stages of a particular scale, one would expect a cortical receptive field to shift as the attentional window is translated, or to expand or shrink somewhat as the attentional window is made slightly larger or smaller respectively. In the highest cortical stages, beyond scale selection, we would expect to see dramatic size changes, as well as a shift in the spatial frequency tuning, for large changes in the size of the attentional window. These predictions can be tested by giving an animal a task that forces it to attend to a region of a specific size and location, and then probing the receptive field with a neutral (behaviorally irrelevant) stimulus to measure its extent. Preliminary results using such a paradigm suggest that the receptive fields of V4 cells do indeed translate toward attentional foci in or near the classical receptive field (Connor et al., 1993). However, the extent of the observed shift is modest (at most about 2° for an approximate 5° shift in window position). One would not expect a V4 cell to cover the entire shift, since it is situated in an intermediate stage of routing (see Figure 4.1), but still for a cell in the third layer of a three-stage routing circuit the average shift should come out to about half the total window shift. The lack of shift could possibly be due to cell relaxing back to its all-connections-open state, since due to various experimental considerations the receptive field is probed 200 ms after the presumed attentional fixation and the cell's response is integrated over a 300-500 ms period. Also, there seems to be little appreciable size change when the attended object is quadrupled in size, but since this is an intermediate area, it is unclear what if any change would be expected for such a large size change outside the scope of a routing circuit for a particular scale.

Another physiological prediction of the model is that lesions to the pulvinar, the hypothesized control center, should dramatically degrade attention and pattern recognition abilities. While there is substantial evidence linking pulvinar lesions to attentional defects (Desimone et al., 1990; Bender, 1988; Rafal and Posner, 1987), some pattern recognition abilities appear to be relatively unimpaired by pulvinar lesions (Bender and Butter, 1987; Nagel-Leiby et al., 1984; Chalupa et al., 1976; Mishkin, 1972). One possible reason for the apparent sparing of pattern recognition is that the tasks used in these studies generally were very simple, such as distinguishing a large 'N' from a 'Z' (Chalupa et al., 1976). It is conceivable that such a task could be carried out even when the fidelity of the remapping process has been compromised. A more rigorous test using stimuli that demand the full spatial resolution capacity of the window of attention would be better suited to test the effect of pulvinar lesions on recognition abilities. Pulvinar lesions would also be expected to diminish the result found by Moran and Desimone (1985) and Connor et al. (1993), and it would be interesting to repeat these experiments while reversibly deactivating the pulvinar.

The physiological responses to be expected from pulvinar neurons depend on how they are configured to gate information flow in the cortex. In an inhibitory gating scheme, one would expect enhanced responses from pulvinar neurons projecting to areas of the cortex within and immediately surrounding the attentional beam, and little or no response from pulvinar neurons projecting to those areas of the cortex substantially outside the attentional beam. In an excitatory gating scheme, one would expect to find enhanced responses from pulvinar neurons projecting to areas of the cortex within the attentional beam only. Petersen et al. (1985) have reported such an enhancement effect for neurons in the dorsomedial portion of the pulvinar (which is connected with PP), but not in the inferior or lateral portion (which is connected to V1-IT). The lack of enhancement in these latter areas may be due to the fact that the task used in this experiment was very simple (detecting the dimming of a spot of light). Again, a more appropriate task would be one that fully taxes the capacity of the attentional window, as this would require the greatest participation from the control neurons in gating out irrelevant information.

Neuroanatomy

The convergence and wiring constraints discussed in the previous chapter (Section 3.1) suggest that the anatomical divergence of intra-cortical connections should increase by roughly a factor of two in the intermediate stages, and that there should be evidence of progressively larger "modules" at higher cortical areas (Fig. 3.5). While there is considerable evidence in support of progressively greater divergence, patchiness, and modularity in ascending stages of the form pathway (Rockland 1992; Van Essen et al., 1986; Van Essen et al., 1990; Van Essen and DeYoe, 1993; DeYoe and Sisola, 1991; Felleman et al., 1992; Felleman and McClendon, 1991), more quantitative and higher resolution data are needed in order to confirm or contradict the proposed routing architecture. Most interesting of all would be to determine the topography and size (in terms of "sample nodes" of visual field) of the the observed modules in V2, V4, and PIT.

Another anatomical prediction of the model is that the terminations of pulvinar-cortical projections should be suitably positioned for effective modulation of inter-cortical synaptic strengths. The pulvinar is known to project to the output layers (2,3) of V1 and to both the input and output layers (3,4) of extrastriate areas V2, V4, and IT (Ogren and Hendrickson, 1977; Rezak and Benevento, 1979; Benevento and Rezak, 1976). These synapses are suspected to be excitatory since they are of the asymmetric type (in layers 1 and 2, Rezak and Benevento, 1979). However, it is not known whether the pulvinar afferents make synapses with inhibitory interneurons or directly onto the dendrites of pyramidal cells.

Finally, the model predicts that there should exist lateral inhibitory and excitatory connections within the pulvinar in order to enforce the constraint of preserving spatial relationships within the window of attention. This prediction is partially supported by the existence of interneurons within the pulvinar (Ogren and Hendrickson, 1979), but it remains to be seen if the axons of projection neurons have collaterals that spread horizontally within the pulvinar, or to what extent the reticular nucleus of the thalamus might subserve this role.

Psychophysics

The fixed window size at the top layer of the routing circuit implies that the spatial resolution of the window of attention is limited. Thus, a large window of attention should have rather poor spatial resolution, whereas a small window of attention should have rather high spatial resolution. It has been proposed by C. Anderson that the size of this window may be on the order of approximately 30x30 sample nodes, an estimate based on psychophysical studies of spatial acuity and pattern recognition (Van Essen et al., 1991; Campbell, 1985).[1] However, one problem with this analysis is that the critical data were derived from experiments in which visual attention was not explicitly controlled. In particular, most of the experiments had display times long enough to permit multiple shifts of attention (although we doubt that this would have been a major contaminating factor in most cases). On the other hand, those experiments that have been directed at studying the amount of "resources" allocated during visual attention have largely ignored the issue of spatial resolution. For example, various studies have reported evidence for a "zoom lens" model of attention in which the density of processing resources decreases as the size of the attentional window increases (Eriksen and St. James, 1986; Shulman and Wilson, 1987). However, these experiments were not designed to measure spatial resolution explicitly. Also, Verghese and Pelli (1992) have attempted to measure the information capacity of the window of attention, which they conclude to have an upper bound of about 50 bits. However, they studied only two tasks—detecting a non-moving target among moving distractors, or detecting a non-flashing square among flashing squares—neither of which is well suited for measuring spatial resolution. In a more recent study, Farell and Pelli (1993) have reported that localization, but not identification, suffers as the window size increases. The fact that identification is not affected may be attributable to the fact that simple or overlearned forms were used in this task (e.g., black or white checks, or digits among letters), thereby reducing identification to an essentially pre-attentive task. It would be interesting to do this with more complicated, non-overlearned shapes. The fact that localization becomes worse as the window size increases is consistent with the model, and it would be interesting to see quantitatively how the size of the positional errors correlates with attentional window size.

The kind of experiment that would be most useful in determining the resolution of the attentional window would be one that tested pattern discrimination ability as a function of the position, size, and resolution of an object. In this case, the model predicts that performance would drop off sharply once the spatial frequency content of the stimulus exceeded approximately 15x15 cycles per object (as dictated by the Nyquist theorem). However, it is important not to interpret "spatial frequency" literally here: A shape composed of high frequency elements would be perfectly discriminable due to non-linear or texture processing. What is important is the resolution of the overall structure, as illustrated in Figure 4.2.

The model also makes some interesting predictions with regard to the dynamics of visual attention. For example, once a location has been attended to in the visual field it should be difficult to stay there or immediately revisit the site, because the the control neurons and/or saliency map neurons corresponding to that part of the visual field would be transiently inhibited from firing. There is some evidence for such a mechanism, in that involuntary attentional fixations tend to be transient

[1] This prediction shares a basic similarity to Nakayama's (1991) "iconic bottleneck" theory, although his estimate (~ 100 pixels total) is somewhat lower.

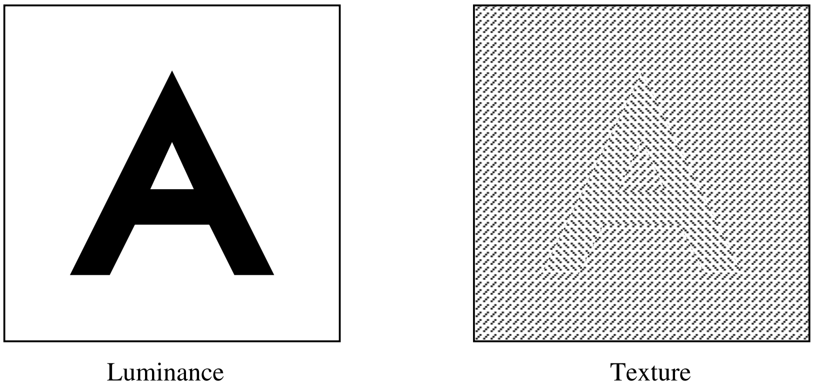

Figure 4.2: **The meaning of "cycles per object."**
An object composed of high frequency texture elements is still recognizable even if the frequency of the elements is beyond 15 cy/window. What is important here is the structure of the 'A' and the amount of spatial detail required to specify it. When this exceeds 15 cy/object, discrimination and/or recognition should suffer.

(Nakayama and Mackeben, 1989) and appear to be inhibited from return (Posner and Cohen, 1984). The amount of time that it takes the attentional window to shift from one location to another would be expected to be roughly independent of the distance between locations. Unlike eye saccades, there is no obvious reason why the control neurons should sequence through all intervening positions of the attentional window. Rather, moving the locus of attention would require merely inhibiting the current control state and activating a new one. This prediction is most consistent with Remington and Pierce's (1984) study showing time-invariant shifts of visual attention, although other studies (e.g., Tsal, 1983) are in disagreement (see also Eriksen and Murphy (1987) and Cave (1991) for a critical commentary on these and other studies). On the other hand, if attention were to actually track a stimulus, then one would indeed expect a smooth transition of activity across the control neurons. It is interesting to note that Cavanagh (1992) has discovered some forms of visual stimuli that produce a motion percept only when tracked with attention. We speculate that the progression of activity across the control neurons is what underlies one's perception of motion in such cases.

4.2 Comparison with other models

Control vs. synchronicity

A number of other models of visual attention and pattern recognition have been proposed that rely on the synchronous firing of neurons in order to effectively change connection strengths (e.g., von der Malsburg and Bienenstock, 1986; Crick, 1984; Crick and Koch, 1990). We contend that a key disadvantage of such approaches is that information about the effective connection-state at any one point in time is not explicitly encoded anywhere in the system. In the routing circuit model, this information is encoded explicitly in the activities of the control neurons, which then allows it to be utilized advantageously in a number of ways.

One way that information about connectivity can be utilized is in constraining the active connections between retinal- and object-based reference frames to be in accordance with a global shift and scale transformation. This constraint is incorporated in our model via the competitive and cooperative interactions among the control neurons (Equation 2.11). During object recognition, this constraint drastically reduces the number of degrees of freedom in matching points between the retinal and object-centered reference frames, because once a few point-to-point correspondences have been established, the number of potential matches between other pairs of points is greatly reduced. In machine vision, this is known as *viewpoint consistency constraint*, and it has proved to be a powerful computational strategy for object recognition systems (Lowe, 1987; Hinton, 1981b).

Another advantage of having knowledge of the active connection state readily available is that the ensemble of control neurons together form a neural code for the current position and size of the window of attention. Therefore, information about the position and size of an object can be obtained by simply reading out the state of the control neurons. In addition, it would also be possible for the control neurons to warp the reference frame transformation in order to form object representations that are invariant to distortion (e.g., hand written digits), in which case information about the particular shape of the object (e.g., its slant or style) could also be preserved. Note that such information is typically lost in networks that utilize feature hierarchies of complex cells (Fukushima, 1980, 1987; LeCun et al., 1990) or Fourier transforms (e.g., Pollen et al., 1971; Cavanagh, 1978, 1985) for forming position-, scale-, and/or distortion-invariant representations.

The routing circuit model can also explain how attention may be directed "at will," or by other modalities, to the extent that those areas of the brain having access to the control neurons (such as parietal cortex) can directly influence where attention is directed. This also provides a convenient format for mediating the access to control among various competing demands. While such forms of top-down control are not impossible to incorporate in models based on synchronicity-gated connections, its implementation would seem to be less straightforward.

Control-based network models

A number of other network models of attention and recognition have also utilized the concept of control neurons for directing information flow. Niebur et al. (1993), Desimone (1992), Laberge (1990; 1992), Ahmad (1992), and Posner et al. (1988), among others, have proposed models that involve the pulvinar as a control site for routing information from a select portion of the visual scene. In addition, Tsotsos (1991) and Mozer (1992) have proposed somewhat more abstract connectionist models that utilize gating units to control attention. However, none of these models explicitly preserve spatial relationships within the window of attention, which is presumed here to be a critical component of the routing process.

Hinton and Lang (1985) and Sandon (1988; 1990) have proposed control-based models that do preserve spatial relationships within the window of attention and share the same basic principle as the model presented here—i.e., remapping object representations from retinal into object-centered reference frames via a third set of units (equivalent to control neurons). Although these models attempt to explain various psychophysical data, they do not contain the necessary level of neurobiological detail to give them strongly predictive value in biology.

Postma et al. (1992) have proposed a neural model based upon the original shifter circuit proposal (Anderson and Van Essen, 1987) to account for translational invariance in visual object priming (Biederman and Cooper, 1992). This model shares many similarities to the routing circuit model, including top-down, or template-driven control, but it differs in the specifics of the gating and control structure. Postma's circuit uses distinct control neurons for each synapse that are locally connected to each other in a "gating lattice" that settles upon a global shift for the circuit. The model also utilizes a series of stages of local, winner-take-all circuits to control the shift, which shares a basic similarity to the hierarchical control scheme proposed here.

4.3 Generalizations of the model

Bayesian interpretation

As discussed in the introduction, the routing circuit model can be viewed as a means for generating separate, independent representations of *what* and *where*, which results in an efficient usage of computational resources. As we shall see here, the energy functional that we use to express the "goal" of the network has an interesting Bayesian interpretation that illustrates another advantage of separating *what* and *where*.

The total energy functional of the network, in its most general form, is

$$E_{total} = -\beta_1 \, \mathbf{I}^{out}\mathbf{V} - \beta_2 \, \mathbf{VTV} - \beta_3 \, \mathbf{cUc}. \tag{4.1}$$

Note that we have switched into vector notation here to eliminate the bulky summation signs. The first term of Equation 4.1 measures the similarity between the output of the routing circuit, $\mathbf{I}^{out}$, and the "desired" output, $\mathbf{V}$. In blob search mode, $\mathbf{V}$ is set equal to the blob function, $\mathbf{G}$, and in recognition mode $\mathbf{V}$ is the output of the associative memory. The second term is the associative memory energy term, which only takes effect in recognition mode; in blob search mode, $\beta_2 = 0$. The third term is the constraint term for the control neurons, $\mathbf{c}$, that enforces a single scale and position of the window of attention. We can write $\mathbf{I}^{out}$ in terms of the input image, $\mathbf{I}^{in}$, and the control neurons as

$$\mathbf{I}^{out} = \mathbf{c\Gamma I}^{in} \tag{4.2}$$

where $\mathbf{c}$ and $\mathbf{\Gamma}$ are a concatenation of the control neurons and coupling coefficients, Γ^l_{ijk}, in all stages of the routing circuit. Thus, we can write E_{total} as

$$E_{total} = -\beta_1 \, \mathbf{Vc\Gamma I}^{in} - \beta_2 \, \mathbf{VTV} - \beta_3 \, \mathbf{cUc}. \tag{4.3}$$

Now, in plain English we can say that the goal of our model network is to infer the position, size, and identity of objects in an image, given the image data. Or, in probabilistic terms, we can say that that we wish to maximize

$$P(\text{WHAT},\text{WHERE}|\text{IMAGE}) \tag{4.4}$$

where WHAT denotes the identity of the object, WHERE denotes the position and size of the object, and IMAGE denotes the image data. Expanding Equation 4.4 according to Bayes rule gives us

$$\begin{aligned} P(\text{WHAT},\text{WHERE}|\text{IMAGE}) &\propto P(\text{IMAGE}|\text{WHAT},\text{WHERE})P(\text{WHAT},\text{WHERE}) \\ &= P(\text{IMAGE}|\text{WHAT},\text{WHERE})P(\text{WHAT})P(\text{WHERE})\,. \end{aligned} \tag{4.5}$$

The last step assumes that WHAT and WHERE are statistically independent (i.e., that any given object is equally likely to appear at any location and size), which is going to be true for the most part.

In our network, the IMAGE is $\mathbf{I}^{in}$, WHAT is expressed in the ensemble of activity in $\mathbf{V}$, and WHERE is expressed in the ensemble of activity in $\mathbf{c}$. Thus, if we make the following equivalences via the Gibb's distribution

$$P(\text{WHAT},\text{WHERE}|\text{IMAGE}) \propto e^{-\beta E_{total}} \tag{4.6}$$

$$P(\text{IMAGE}|\text{WHAT},\text{WHERE}) \propto e^{\beta_1 V c \Gamma I^{in}} \tag{4.7}$$

$$P(\text{WHAT}) \propto e^{\beta_2 VTV} \tag{4.8}$$

$$P(\text{WHERE}) \propto e^{\beta_3 cUc} \tag{4.9}$$

then we can see that Equation 4.3 is just the logarithm of Equation 4.5, and so minimizing the energy function, E_{total}, will also tend to maximize $P(\text{WHAT},\text{WHERE}|\text{IMAGE})$.

The terms $P(\text{WHAT})$ and $P(\text{WHERE})$ are the "priors," which provide the prior probability of WHAT and WHERE before we know anything about the image. The term $e^{\beta_2 VTV}$ gives a high probability to any state of the $\mathbf{V}$ corresponding to a known object, and the term $e^{\beta_3 cUc}$ gives high probability to any state of the $\mathbf{c}$ corresponding to a unique position or size of the window of attention. The term $P(\text{IMAGE}|\text{WHAT},\text{WHERE})$ is the "likelihood," which expresses how likely the IMAGE could have arisen from a particular WHAT and WHERE. The term $e^{\beta_1 V c \Gamma I^{in}}$ gives a high probability to that state of the $\mathbf{c}$ and $\mathbf{V}$ that can "explain" $\mathbf{I^{in}}$ (measured by taking the inner product $\mathbf{Vc\Gamma} \cdot \mathbf{I^{in}}$).

Finding the WHAT and WHERE that maximizes the posterior, $P(\text{WHAT},\text{WHERE}|\text{IMAGE})$, requires optimizing over a huge search space. The strategy we adopt in the routing circuit for dealing with this dilemma is to first set $\beta_2 = 0$ (no recognition) and let $\mathbf{c}$ evolve while holding $\mathbf{V} = \mathbf{G}$. This essentially moves $\mathbf{c}$ into a good initial state. From this point, we turn on β_2 and let $\mathbf{V}$ evolve. In terms of the probabilistic framework, the network hill climbs on $P(\text{WHAT},\text{WHERE}|\text{IMAGE})$ initially along the WHERE axes, and then along the WHAT axes. The reason we can do this is that the statistical independence of WHAT and WHERE allows the use of rather primitive, pre-attentive measures to guess the WHERE without knowing WHAT. More generally though, the pre-attentive measures can also help guess WHAT. For example, measures such as color, texture, or convexity may be able to narrow down the class of objects to search among. In this framework, then, "attention" can be understood as a heuristic that exploits the statistical independence of WHAT and WHERE in order to make an extremely computationally intensive problem tractable with limited resources.

The big picture

How are the various "snapshots" obtained by the window of attention eventually incorporated in order to form an overall percept of a scene? One possibility, initially suggested by Hinton (1981b), is that a compact representation of each object may be maintained in the form of the activities on a set of neurons within a "scene buffer." My own rendition of this scheme is illustrated in Figure 4.3. One can essentially think of the scene buffer as a spatially indexed RAM (random access memory). Each attentional fixation writes its contents into a different part of the buffer, depending on the position and size of the attentional window as well as the orientation of the eyes, head, and body with respect to the environment. (See also Baron, 1987, for another variation on this theme.)

An important property of the model advanced here is that all the features belonging to an object, such as the eyes, nose, and mouth of a face, are bound together as an entity, and that a symbolic code for the object is entered into the spatial buffer. An alternative scheme that may be favored by a "feature *Gemisch*ist" would be to collect each feature with an attentional snapshot independently, and then enter these into the scene buffer with the proper spatial relationships. This would avoid the earlier mentioned problem of confusing the spatial relationships of features (Fig. 1.3), but it would not be very efficient because it would require accurate storage of the pointers for each feature of an object. The advantage of the scheme proposed here is that the spatial relationships that recur repeatedly in the visual world are encoded explicitly (presumably by a population of cells in IT). The spatial buffer is then reserved only for essentially non-predictable structure. For example, Joe may only rarely appear with his car—indeed, he appears at home, in the lab, riding a bicycle, etc. What is the same in all these situations are the spatial relationships among the features of Joe's face. But if Joe were to recur repeatedly with some other object, say, a particular shirt or style of clothing, then one might eventually adopt an explicit representation, or a new code, for this combination. Subsequent presentations of this combination would then result in the code being entered into the scene buffer, instead of the individual parts. The strategy advocated here, then, is that the natural redundancies in the input should be captured explicitly, leaving the dynamic representation of the spatial relationships (in the spatial buffer) for those combinations that rarely or spontaneously occur in our everyday visual experience.

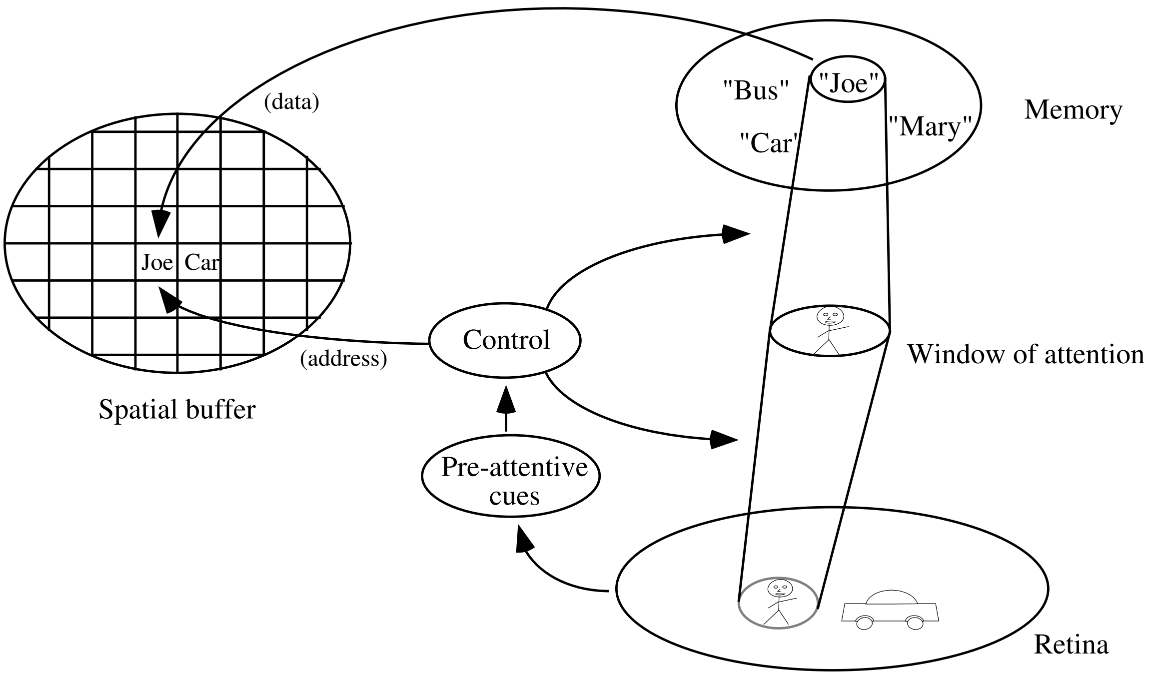

Figure 4.3: **The big picture.**
Pre-attentive cues drive the control neurons, which direct attention to a restricted region of visual space and "object space." Once the object has been recognized, its label is written into the spatial buffer at the address indexed by the control neurons.

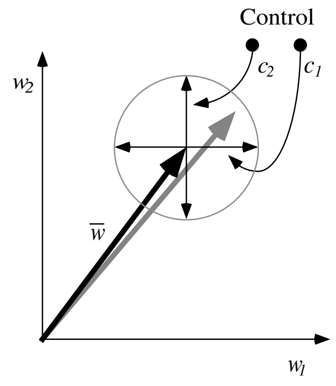

Figure 4.4: **A more general way of viewing control.**
A weight vector with two components, w_1 and w_2, is shown. Control neurons c_1 and c_2 modulate each of these components, respectively, to dynamically change the weight vector. Thus, the weight vector may be able to occupy any region within the circular outline in order to optimize the network for the particular input and task at hand.

Control as a general strategy for neural computation

Another of the general themes advances in this thesis is that the utilization of explicit control neurons is a useful computational principle for visual processing. This principal may be employed by the brain in other domains as well. A different perspective of dynamic control is illustrated in Figure 4.4. In most "static" neural network models, the output of a neuron is computed by forming the inner product of a weight vector, $\overline{w}$, with the inputs to the neuron, and then passing the result through a non-linearity. The weight vector may change on a slow time scale in order to optimize the network for performing a certain task, but typically $\overline{w}$ remains fixed over the relatively short time in which the task is actually performed (e.g., < 1 sec). By having control neurons available to modify $\overline{w}$ on a short time scale, the computation being carried out by the network can be dynamically reconfigured and optimized for the particular task at hand. This added degree of flexibility reduces the neural resources required for solving a complicated task, since it is no longer necessary to have dedicated, specialized networks with fixed connections to deal with each variation of a task.

A particularly interesting example is in the motor system. As Lashley (1942) pointed out long ago, somehow we are able to route prototypical motor commands to any set of limbs (he termed this "motor equivalence"). For example, one may write with either hand, or even with the mouth, and the style of writing is more or less preserved. This would seem to imply that the trajectories for writing the name may be stored in a canonical reference frame without reference to a particular limb, and that these trajectories are dynamically routed out to any desired set of actuators. This problem could essentially be considered the inverse of the sensory routing problem (see Van Essen et al., 1994).

4.4 Unresolved issues

The dynamic routing circuit as described in this paper is intended as a "zero-th order" model, and as such many details have been neglected or oversimplified. Here we outline some of the more important unresolved issues that remain as topics for future research.

Features instead of pixels

As already noted, one key neurobiological characteristic neglected in the present model is the known preponderance of feature selective cells in the visual cortex. V1, for example, is known to contain cells tuned for various orientations, and V2 and V4 contain cells that seem to be tuned for more complex stimuli (von der Heydt and Peterhans, 1989; Gallant et al., 1993). It has been assumed here that this will not significantly affect the general nature of the routing process, but in order to make our predictions more precise, it would be helpful to have a better idea of what features are computed. For example, what is role of position invariance in complex cells? Also, dynamic routing need not necessarily be restricted to the space domain, but could work over feature domains as well.

Feedback pathways

Although the model relies upon feedback pathways for top-down control during recognition, the information processing role of these pathways has more or less been ignored. Mumford (1992) has sketched a theory proposing that the role of these feedback pathways is to relay the interpretations of higher cortical areas to lower cortical areas in order to verify the high-level interpretation of a scene. Such a mechanism would obviously be of use for Step 4 of our proposed strategy for an autonomous visual system. Under this scenario, it would be necessary to route information flow within the feedback pathways as well in order to ensure that the high-level interpretation is matched against the appropriate region within the cortical area below (i.e., within the window of attention). Another possible attentional role for information flow in the feedback pathways may be to refine the tuning characteristics of lower-level cortical cells based upon the interpretations made in higher cortical areas (see, for example, Tsotsos, 1991).

Pop-out in multiple dimensions

In the simple autonomous visual system we have proposed, "blobs" were the only salient features used to attract the window of attention. Presumably, other salience measures—such as pop-out due to motion or texture gradients—would provide a much richer and more robust system for pre-attentively guessing the size and position of potential objects. It will also be of interest to see how these measures could be used to position the window in object space, as illustrated in Figure 4.3.

Rotation and warp

Our model accounts for how reference frames can be shifted and rescaled, but it does not address rotation and other distortions (e.g. hand-written characters). The ability to rotate or warp reference frames could probably be included in the model without much difficulty, since this would just involve another form of routing. Moreover, for foveated objects the log-polar representation in V1 would convert rotations into approximate linear shifts on the cortex (Schwartz, 1977), which may facilitate the routing.

3D objects

How are 3D objects represented neurally, and how is information in the retinal reference frame transformed to match this representation? One possibility, as advanced by Poggio and Edelman (1990), is that 3D objects are actually represented by a few characteristic two-dimensional views, and that a match to the retinal representation is achieved by interpolating among these views. In this case, the routing circuit would be required to properly reposition and rescale the object so that the interpolation could take place.

Learning

Although the model we have presented here is neurobiologically plausible in terms of the number of neurons, connectivity, and computational mechanisms required, it remains to be seen whether such a system can self-organize or fine tune itself with experience, beginning with only roughly appropriate connections. A hint as to how this may be accomplished has been described by Foldiak (1991), who has demonstrated how a complex cell can learn translation invariance using the objective function of "perceptual stability." In our model, perceptual stability would be desired in IT, and the control neurons would need to learn how to configure themselves to maintain a stable percept as an attended object moves or changes size on the retina. More generally, there is a clear need to devise learning rules for networks with control-like structures, or three-way interactions, rather than simple perceptron-type networks with two-way interactions only. Recent work in this direction by Lee and Olshausen (1994) has shown that networks in which inputs interact in a local, nonlinear fashion are capable of learning higher-order regularities—such as disparity—using a local Hebb rule. It is conceivable that such learning rules may be extended to the control networks proposed here.

Chapter 5

Conclusions

In order for us to make sense of the visual world, the brain must be capable of forming object representations that are invariant with respect to the dramatic fluctuations occurring on the retina. We have demonstrated here how this feat may be accomplished by simplified, model neural circuits that are largely consistent with our current knowledge of neurophysiology and neuroanatomy.

While the model has ignored the exact nature of the representation of visual information—for example, the preponderance of feature selective cells in the cortex—it still retains its essential predictive value. Feature processing will certainly affect the picture; but the need for routing still exists, because there is no known theory by which feature processing *per se* can take care of the invariance problem. Thus, the dynamic effects predicted by this model will be expected to be evidenced in some form if routing is indeed employed by the visual cortex.

Besides generating some basic predictions, the model also provides us with a more concrete understanding of what attention is (or what it may be). The action of attention in our model neural circuit can be understood as exploiting the statistical independence of *what* and *where* to make an otherwise enormous, computationally intensive problem tractable with limited resources. More generally, such mechanisms could be used to take advantage of statistical independence along other dimensions as well, such as motion and color. Within this framework, attention can be seen not merely as a phenomenon, but rather as a computational strategy for dealing with complex visual tasks.

It is interesting (and sometimes amusing) to look back at the early theories, such as those of Lashley, and Pitts and McCulloch, because in certain respects their ideas seem ludicrous by today's standards. It is certainly possible that the theory proposed here will be viewed in retrospect with equal disbelief. However, to date there are no alternative, comparably detailed models that suggest a means for forming invariant representations in a manner that is more parsimonious with our current understanding of the brain. Only after the proper experimental evidence is collected can we alter our conviction in the model. As these experiments are carried out, the results will either help to increase our confidence in the model, or will suggest where it is wrong and how it might be revised. It is this combined process of computational modeling and experimentation that eventually will lead us to understand how visual attention and recognition are actually implemented in the brain.

Appendix A

Derivation of autonomous control dynamics

A.1 Blob search

The total energy functional we wish to minimize is

$$E_{total} = E_{blob} + \beta\, E_{constraint}\ , \tag{A.1}$$

where E_{blob} and $E_{constraint}$ are defined in Equations 2.10 and 2.11, and β is a constant determining the relative contribution of the constraint term. Letting c_k follow the gradient of this functional, we obtain

$$\begin{aligned} \frac{dc_k}{dt} &= -\eta\, \frac{\partial E_{total}}{\partial c_k} \\ &= -\eta\, \frac{\partial E_{blob}}{\partial c_k} - \eta\beta\, \frac{\partial E_{constraint}}{\partial c_k}\ , \end{aligned} \tag{A.2}$$

where η is a constant determining the rate of gradient descent.

As it stands, c_k is unbounded; hence E_{blob} and $E_{constraint}$ will also be unbounded and the network will not be guaranteed to converge. We can ameliorate this problem by letting c_k be a monotonically increasing function of another analog variable, u_k, that actually follows the gradient. That is,

$$c_k = \sigma(u_k) \tag{A.3}$$

$$\frac{du_k}{dt} = -\eta \frac{\partial E_{total}}{\partial c_k} \tag{A.4}$$

$$\sigma(x) = [1 + \exp(-\lambda x)]^{-1}\ . \tag{A.5}$$

This has the effect of limiting c_k to the interval $[0, 1]$, but since we know *a priori* that the desired minimum of E_{blob} and $E_{constraint}$ lies in this range, the limitation does not present a problem.

Taking the derivative of E_{blob} and $E_{constraint}$ with respect to c_k yields

$$\frac{\partial E_{blob}}{\partial c_k} = -\sum_i \sum_j G_i\, \Gamma_{ijk} I_j^{in} \tag{A.6}$$

$$\frac{\partial E_{constraint}}{\partial c_k} = -\sum_l U_{kl}\, c_l \tag{A.7}$$

and so the dynamical equation for u_k is thus

$$\frac{du_k}{dt} = \eta \sum_i \sum_j G_i\, \Gamma_{ijk}\, I_j^{in} + \eta\beta \sum_l U_{kl}\, c_l\,. \tag{A.8}$$

One remaining problem is that u_k must be computed via pure integration, which may cause implementation difficulties. We can convert the integrator to a more biologically plausible leaky integrator by adding to E_{total} the term

$$E_{leak} = \sum_k \int_{0.5}^{c_k} \sigma^{-1}(c)\, dc\,. \tag{A.9}$$

The total energy functional is now defined as

$$E_{total} = E_{blob} + \beta\, E_{constraint} + \alpha\, E_{leak}\,, \tag{A.10}$$

where the constant α determines the relative contribution of E_{leak}. (The effect of adding this term is discussed in Hopfield's 1984 paper. It essentially pushes c_k slightly away from 0 and 1.0, depending on the value of α and λ.)

Taking the derivative of E_{leak} with respect to c_k yields

$$\frac{\partial E_{leak}}{\partial c_k} = u_k \tag{A.11}$$

and so the final dynamical equation for c_k is now

$$\begin{aligned} c_k &= \sigma(u_k) \\ \frac{du_k}{dt} + \tau^{-1}\, u_k &= \eta \sum_i \sum_j G_i\, \Gamma_{ijk}\, I_j^{in} + \eta\beta \sum_l U_{kl}\, c_l\,, \end{aligned} \tag{A.12}$$

where the time constant, τ, is defined as $\frac{1}{\eta\alpha}$.

A.2 Recognition

Now the total energy functional is

$$E_{total} = E_{mem} + \beta\, E_{constraint} + \alpha\, E_{leak}\,, \tag{A.13}$$

where E_{mem} is defined as in Equation 2.14. Note that Equation A.13 is just the same as Equation A.10, except with E_{blob} replaced by E_{mem}.

Taking the derivative of E_{mem} with respect to c_k yields

$$\frac{\partial E_{mem}}{\partial c_k} = -\sum_i \sum_j V_i\, \Gamma_{ijk}\, I_j^{in} \tag{A.14}$$

and so the new dynamical equation for c_k is thus

$$\begin{aligned} c_k &= \sigma(u_k) && \text{(A.15)} \\ \frac{du_k}{dt} + \tau^{-1}\, u_k &= \eta \sum_i \sum_j V_i\, \Gamma_{ijk}\, I_j^{in} + \eta\beta \sum_l U_{kl}\, c_l\,. && \text{(A.16)} \end{aligned}$$

Note that this result is just the same as Equation A.12, with the exception that G_i is replaced with V_i.

Appendix B

Open questions about spatial frequency tuning

The multiscale stack model proposed in Chapter 3 raises a number of interesting issues with regard to the nature of the multiscale representation of visual space within V1. What is the range of peak spatial-frequency tuning observed among cells in foveal V1? Is the coverage—i.e., the number of cells at each spatial frequency and their relative spacing—sufficient to preserve spatial information at a level of detail that is consistent with our perceptual capabilities? This appendix examines the available evidence and points out some lurking mysteries that will need to be resolved in future experiments.

B.1 Conflicting data

De Valois et al. (1982) and Tootell et al. (1988) report that the peaks of the spatial-frequency tuning curves of foveal V1 cells are in the range of 1-10 cy/deg, with the greatest number of cells in the range of 4-8 cy/deg. Few if any cells in their study have a peak spatial-frequency at 16 cy/deg or above.

On the other hand, Parker and Hawken (1988) report that the majority of foveal V1 cells can be fit by a difference-of-difference-of-Gaussians function with a central, excitatory zone of about 2-4 minutes in diameter. The number of cells with larger central diameters drops of rapidly, with few if any larger than 20 minutes. The cells with the smallest central diameters (2-4 minutes) would presumably have their peak spatial-frequency in the range of 8-15 cy/deg, while the cells with the largest diameters (20 minute) would be centered around 1.5 cy/deg (Hawken and Parker, 1987; this assumes the spacing between the peaks of the inhibitory flanks is about twice the central diameter).

The results of Parker and Hawken seem to be shifted upwards by about an octave from the results of the Tootell and De Valois groups. The apparent discrepancy between the data is not addressed by Parker and Hawken, although both experiments are in anesthetized monkeys and utilize similar methods. Some potential reasons for this difference are given below.

B.2 How do we see so clearly?

We can essentially consider each cortical cell as a sample node conveying certain properties (e.g., spatial-frequency, orientation) within a local region of visual space. In order to represent information within a certain spatial-frequency band without loss, cortical cells should be spaced by an amount proportional to their peak spatial-frequency, as dictated by the Nyquist sampling theorem. Low spatial-frequency cells on the order of 1 cy/deg would be expected to be spaced by about 0.5 degrees or less, while high spatial-frequency cells on the order of 15 cy/deg would be expected to be spaced by about 0.03 degrees or less. In a 1D array, there should be twice as many cells for each octave increase in spatial-frequency; in 2D, there should be a fourfold increase for each octave.

The fact that the De Valois and Tootell groups find the most number of cells in the range of 4-8 cy/deg is disturbing, because it implies that the representation of spatial structure at 8 cy/deg or above is incomplete or non-existent. Grating detection tasks or simple hyperacuity tasks could still be carried out at high frequencies by utilizing information in the tails of the low frequency tuning curves, but pattern recognition or other tasks requiring a veridical representation of spatial structure at 8 cy/deg or above will suffer. To make this more concrete, Figure B.1 illustrates how a 12-point Times-font 'A' would appear on the retinal sampling lattice in the fovea when viewed at a distance of two feet from the eye, and how it would appear when subsequently filtered by hypothetical cortical cells centered at different spatial frequencies. Needless to say, our perception of this shape corresponds more closely with those filters centered above 8 cy/deg, and not with those reported by the De Valois and Tootell groups. Not only can we recognize the letter, but we are also perfectly capable of discerning the details pertaining to its font, style, etc. The apparent discrepancy becomes even more obvious when one considers that faces can be recognized when they are reduced to a size spanning as few as 19x19 cones (Campbell, 1985). Such faces would appear as little more than amorphous blobs when filtered at 8 cy/deg or below. Thus, there is clearly something awry with the range of peak spatial-frequencies reported by the De Valois and Tootell groups. The data of Parker and Hawken is certainly more consistent with our perceptual capabilities, but one would still expect to find four times as many cells with widths of 2 minutes than of 4 minutes, and the ratio is only 1 : 1 at best.

Thus, we are left with a discrepancy not only between the existing data, but also between the data and our perceptual capabilities. [1] One possible explanation for the discrepancy is that the receptive fields of V1 cells are dynamic, as demonstrated by Pettet and Gilbert (1992); each cell may have a very large "potential" receptive field, and depending on the nature of the visual stimuli and the task at hand the cell's receptive field may either expand or contract. It is conceivable that putting the animal into an anesthetized state and exposing it only to gratings in a darkened

[1] While the neurophysiological data have been collected on monkeys, it can be assumed that the resolution of their visual perception is nearly equivalent to ours, since their foveal acuity is about 3/4 that of humans (Merigan and Katz, 1990) and their contrast sensitivity functions are practically identical to those of humans (De Valois and De Valois, 1988).

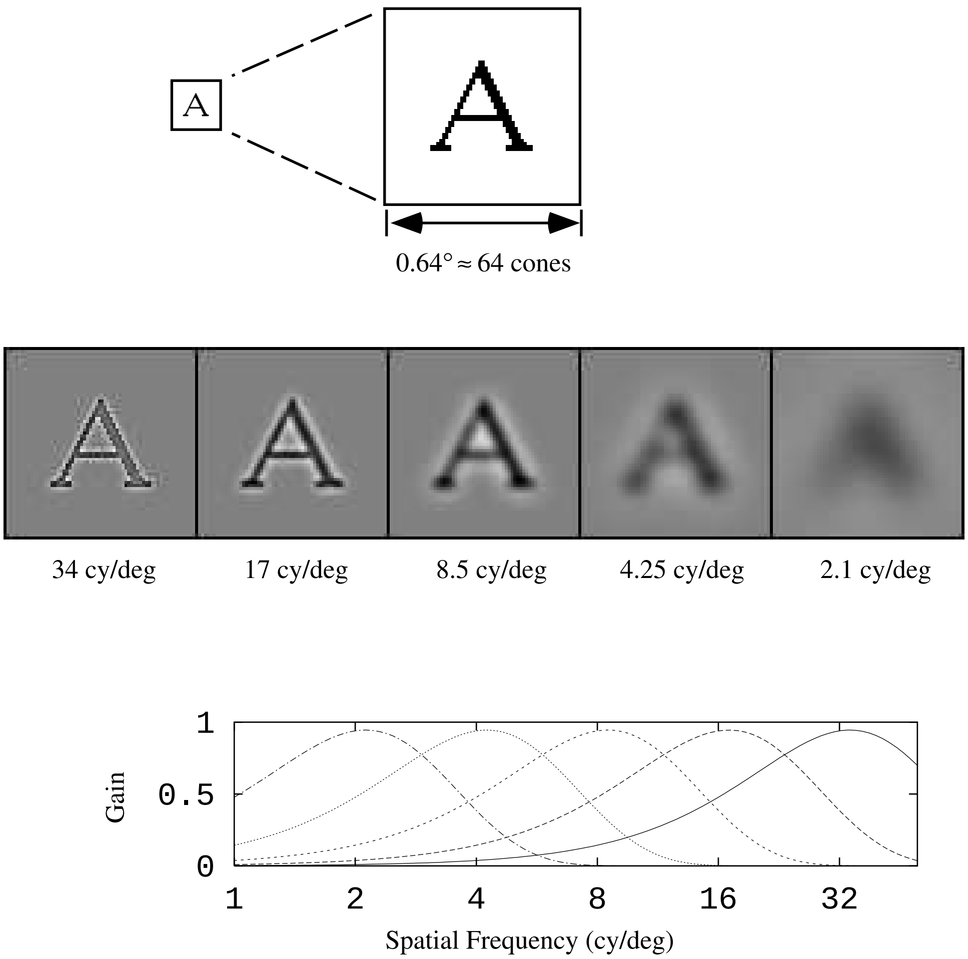

Figure B.1: **Relation of spatial-frequency tuning to perception.** Viewing the small 'A' in the box at upper left from a distance of two feet will result in it being projected onto an array of about 64x64 cones. Shown below this are the results of filtering this image with a difference-of-Gaussians filter centered at various peak spatial frequencies, as shown in the plot. Each filter has approximately a 1 octave bandwidth, as with the cells in V1.

room allows the receptive fields to expand somewhat, and thus lowers their peak spatial-frequency tuning. It would be desirable then to repeat these assays in awake animals. Another consideration is that the stimuli used to locate and isolate a cell may bias the experimenter to record from particular types of cells. For example, it is typical in these experiments to use fairly large bars to locate a cell, and these stimuli will preferentially excite lower frequency tuned cells. Interestingly, Parker and Hawken also used a variety of high-frequency stimuli to locate their cells, such as lines and spots, which may be why they found more of the small, high frequency cells. It is also important to keep in mind that the low spatial-frequency cells will fire most often and most prolonged in response to a bar passing over their receptive fields, whereas the high frequency cells with small receptive fields will register only blips of activity by comparison, and will thereby be harder to find. In an analogous situation in the hippocampus, where cells are believed to form a sparse population code, new recording techniques using "stereotrodes" have been successfully used to reliably record from many of the cells that fire only rarely (Wilson and McNaughton, 1993). It would be highly desirable then to utilize these same techniques to determine the number and range of spatial-frequency tuned cells in the visual cortex.

B.3 Very low-frequency cells

The observation of very low-frequency cells (≤ 1 cy/deg) in foveal V1 is likely to be correct, since it is hard to conceive of ways this measurement could have been made unless a cell actually does possess the capability to integrate information in this way. How these cells achieve such low frequency tuning is an interesting question, since they would need to integrate information over a diameter of about 12 mm on the cortical surface, and the largest observed diameter of cortical axonal arbors is on the order of 6-8 mm (Lund, 1988; Gilbert and Wiesel, 1989). One possibility is that these cells integrate over a large region by collecting the responses of cells two or three synapses removed via intermediary cells acting as relays. Another possibility is that they may expand the range of their connectivity via reciprocal pathways to V2. In any case, it would be interesting to compare the extent of the input field for low vs. high spatial-frequency cells, perhaps by using viruses that are transported retrogradely and transynaptically.

Appendix C

Details of pulvinar anatomy and physiology

C.1 An overview of the pulvinar

The pulvinar occupies the posterior 2/5 of the thalamus and is the largest nucleus of the thalamus in man. It has evolved along with the primate brain, growing in size with increasingly complex subdivisions as the extrastriate visual cortex becomes more complex. Figure C.1 shows a first-pass attempt at constructing a comprehensive connectivity diagram of the primate pulvinar and its relationship with the cortex and other brain structures. Roughly, the pulvinar can be subdivided into four separate subnuclei on cytoarchitectural grounds: inferior (PI), lateral (PL), medial (PM), and oral (PO). The connectivities shown have been accumulated from many different anatomical studies. Recently, a graphical anatomical database of pulvinar connectivity has been constructed by Press and Olshausen (1993) that allows these connections to be analyzed and compared to each other in a detailed fashion.

There are at least two distinct maps of visual space in the pulvinar, which are illustrated in Figure C.2. One map occupies all of PI and some of PL and is referred to as the PI/PL map. The other map lies entirely within PL and is referred to as the PL map. (Notice that the PL map wraps around the PL/PI map the same way that V2 wraps around V1! - i.e., by splitting the map along horizontal meridian.) A separate region within PL lying dorsal and medial to the PL map is referred to as Pdm, which does not seem to exhibit much topography - or at least no map has been made of it. The RF sizes in the PI/PL and PL maps as a function of eccentricity are comparable to those in V2. RF sizes in Pdm are a mix of very large and normal (V2-like) sizes.

C.2 Doubts about pulvinar

The routing circuit model proposes that the pulvinar is a major source of control signals for gating intra-cortical connection strengths along the occipito-temporal pathway. During visual attention, this allows information from a windowed region of

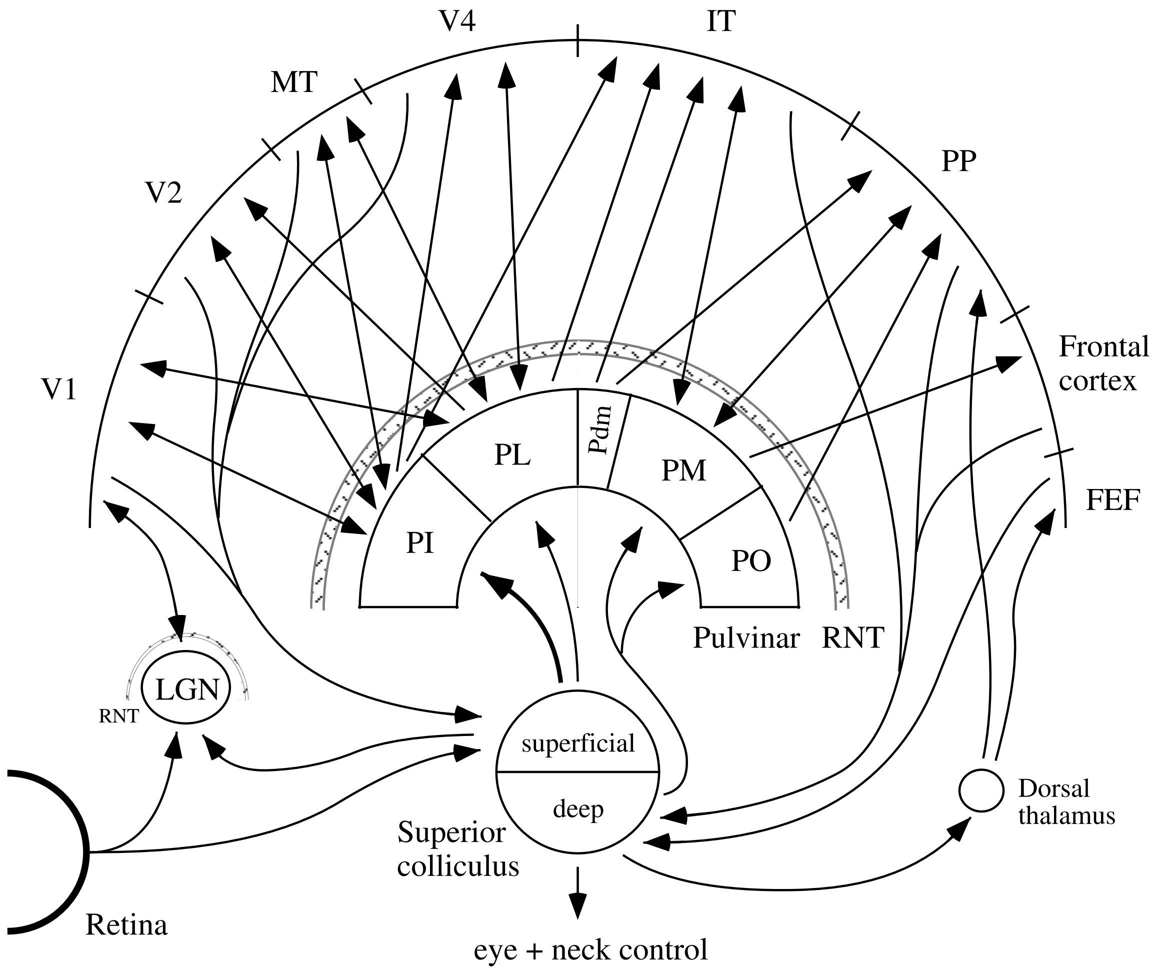

Figure C.1: **A schematic of the known connections of the pulvinar.** These connections were compiled from the following references: Baleydier and Morel (1992), Hardy and Lynch (1992), Yeterian and Pandya (1991), Schmahmann and Pandya (1990), Robinson and McClurkin (1989), Benevento and Davis (1977), Benevento and Rezak (1976), Trojanowski and Jacobsen (1976), Ogren and Hendrickson (1977). Anatomical distinctions within subnuclei of the pulvinar are not shown in this diagram (e.g., the projections from PM to IT and PP arise from distinctly different zones within PM, and they intermingle only at the PM/PL border). The pulvinar anatomy database of Press and Olshausen (1993) allows for a more detailed viewing of the topography of connections revealed by these and other studies.

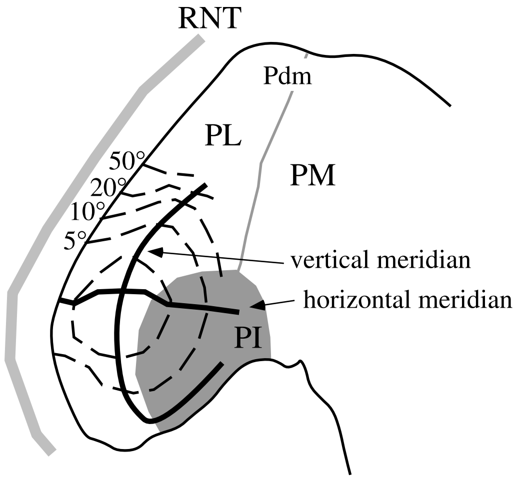

Figure C.2: **The two maps of visual space within the pulvinar.** Shown is a coronal cross-section of the pulvinar. The PI/PL map fills all of PI and extends a bit into PL. The PL map lies entirely within PL. (Adapted from Robinson and McClurkin (1989) and Bender (1981).)

retinal inputs to be selectively routed to higher cortical areas concerned with pattern analysis and recognition (namely, IT). Thus, one may ask, why is the pulvinar not just interconnected to areas in the occipito-temporal pathway, but to nearly every area within the visual cortex and to many other parts of the cerebral cortex as well? Indeed, this massive interconnectivity with such diverse areas may seem more suggestive of a general integrative role for the pulvinar, rather than the very specific role we propose for it in pattern recognition. In fact, its connectivity is reminiscent to that of the claustrum, which some people have proposed as an area for integrating information across multiple modalities. Perhaps, then, the pulvinar acts something like a blackboard - a la Mumford (1992) - where various areas of the cortex "write" their interpretation of the world, and this information gets fed back to cortex to further enhance or nullify the interpretation.

Some of the physiological studies of the pulvinar are no less discouraging. The apparent lack of any distinct pattern recognition or visual search deficits after pulvinar lesions, such as in the experiments of Bender (1988), would seem to cast doubt on the specific role we propose for the pulvinar in pattern recognition. In fact, it is the lesions to SC that seem to have the effect we would expect from lesions to pulvinar! Also, Gattass has reported that stimulating pulvinar does not result in any behavioral modulation in an attentional task (personal communication), whereas stimulating SC does. Even in the physiological experiments of Petersen et al. (1985), attentional enhancement was obtained only in Pdm, which is interconnected with PP, and not in PL or PI, which interconnect to the occipito-temporal pathway. Perhaps then the enhancement observed here is merely a reflection of activity in PP, which has been previously reported to show such attentional enhancements.

In addition, the activity of pulvinar neurons seems to be commonly enhanced or inhibited as a function of eye position or eye movements. This raises the possibility that the pulvinar may be more involved in bringing information about eye movements to bear on cortical visual processing, rather than directing covert visual attention.

A first glance at this evidence might easily lead one to the conclusion that the pulvinar has little or no role in directing attention. But the brain is a complex beast, and so deducing its function requires a deeper analysis of all the factors involved in these and other experiments (see Section C.5).

C.3 The case for pulvinar

If we make the assumption that the process of visual attention involves control of information flow in the visual cortex (which every model of attention does), then we might suspect that some area (or areas) of the brain would serve as the controller of attention. The properties we would expect of such a controller are that

A. it should be able to integrate, or bring together, information over a large region of visual space in order to be informed of what's "out there" to attend to;

B. it should be able to arbitrate, or mediate, the various demands for visual attention and choose *one* locus of the visual field for focusing attention (assuming for now that we attend to only one locus at a time);

C. it should be able to affect information flow in the cortex in such a way that information within the chosen attentional window is processed preferentially.

The pulvinar seems to have the capacity to satisfy each of these important properties. Cells in Pdm and PM tend to have very large receptive fields, which would imply that these subnuclei could bring together information over a large region of visual space (thus satisfying **A**). On the other hand, neurons within PI and PL tend to have fairly localized receptive fields, which probably implies more local processing in these areas (although cells with extremely large, diffuse RF's are scattered throughout PI and PL). Such a mixture of large and small RF's is actually what one would expect of a hierarchical routing circuit, since control neurons in the initial stages would receive their input from V1 and V2, whereas control neurons in the top stages would receive their input from V4 and IT, and hence have larger RF's). The fact that PI and PL contain nearly equal RF sizes may or may not be consistent with this scheme, but it is hard to say at this stage of formulation of the model.

Property **B** could be satisfied by the pulvinar since a relatively large portion of the visual field is brought within the extent of horizontal connections within the pulvinar. For example, in the inferior pulvinar, 50° of visual space is traversed in about 3 mm. Given that the arbors of interneuron dendritic fields are on the order of 600 microns in diameter, this would provide a good degree of crosstalk among large parts of visual space, thus facilitating mediation. This crosstalk would become even more global in Pdm and PM, which project to IT. Moreover, a study in *Galago* (Conley and Diamond, 1990) has shown that pulvinar neurons (presumably in the homologue of the PL/PM area) project very diffusely into the RNT, which is in stark contrast to the very localized projections of LGN afferents into the RNT. This combined with the strongly inhibitory nature of the RNT would provide a suitable mechanism for a global winner-take-all function, thus enabling the pulvinar to choose one locus for attention. Note that such diffuse connectivity with the RNT is inconsistent with the notion of a blackboard, since it would tend to screw-up the topology of any pattern of activation fed into it.

Finally, the fact that the pulvinar projects to all areas in the occipito-temporal pathway means that it could perform the requisite modulation of information flow (property **C**). The fact that the pulvinar projects to other visual areas outside the occipito-temporal pathway, such as MT, may be to focus attention for other aspects of visual processing—such as motion. Indeed, our current focus on pattern vision is probably too restrictive. The broader role proposed for routing by Van Essen and Anderson (1990), in which control operates in the "M stream" in addition to the form pathway, is more parsimonious with the observed connections between the pulvinar and other visual areas. The connections with areas outside the visual cortex, such as frontal cortex, may play the role of informing other brain regions of the current focus of attention, or allowing them to control where attention is focused.

It is also worth noting that nearly every neurobiological theory of visual attention proposes nearly identical roles for the pulvinar. For example, Posner et al. (1988), on the basis of lesion studies, assign the pulvinar with the operation of "engaging" attention, which would presumably mean implementing the modulation of cortical activity. Laberge (1990), on the basis of his PET study, assigns the pulvinar with the task of filtering out distracting information in a cluttered field. And Niebur et al. (1993) propose the pulvinar as the site of origination of oscillatory activity in the cortex for gating connections. It may well be that the commonality in all these theories is an artifact of the sociology of science, but nevertheless it interesting that no one else has yet proposed a better alternative for control of visual attention.

In any case, the increasing size of the pulvinar throughout evolution would seem to imply that it plays some sort of an important role in visual function. Otherwise, the cost of wiring and neuronal resources would surely dictate its demise.

C.4 Other alternatives

What are the alternatives to the pulvinar as a control site? Desimone, on the basis of his lesion study, has suggested a distributed control system. However, what exactly he means by this is unclear. I have trouble with the concept of having more than one controller, because if you did, you would then have to posit the existence of another controller to decide which of the control systems takes precedence over the other. At this point you are back to having a single control site. Posner also has proposed a sort of distributed control system, with PP disengaging attention, SC moving attention, and pulvinar engaging attention. However, even this scheme centralizes the process of engaging - which is the quintessential operation of attention - in the pulvinar.

One form of distributed control that does make sense is a hierarchical control scheme, such as in the hierarchical routing circuit of Figure 2.19. In this scheme, the top-stage control neurons would serve as the global "master controller," and the bottom-stage control neurons would control information flow over more localized regions of visual space. Thus, if control neurons near the top stage were disabled, as presumably was done with the PL lesions of Desimone et al. (1990), one would observe an asymmetry for global vs. local attentional effects, just as Desimone did in his study (i.e., there was no impairment for distractors within a hemifield, but distractors in the opposite hemifield did produce a decrement in performance). On the other hand, if control nodes near the bottom were disabled, one would be able to focus attention only roughly (without much precision) for a limited region of visual space.

Alternatively, the control neurons may be distributed between the pulvinar and the cortex. In this scheme, the pulvinar control neurons may set up the more global context for attention (rough position and size) at each stage of the routing circuit, and then control neurons in the cortex (presumably in layer 6, as described by Van Essen and Anderson (1990)) would provide refinements. (Such a distribution of control circuitry would indeed make sense, because the function of control is so important that it would be risky to put all the eggs in one basket.) It is conceivable under such a scheme that lesioning the pulvinar control neurons would have little effect on recognition when a single object is presented alone in the visual field, since the cortical control neurons themselves may be able to handle the routing for such a simple scene. The pulvinar control neurons may only be necessary to filter out other information when multiple objects are present.

It seems unlikely that the parietal cortex could play a role in controlling information flow because it is so lacking in connectivity to the occipito-temporal pathway. Also, it does not appear to have the necessary structure for choosing a single locus for attention, such as the RNT, although this remains an unknown. The superior colliculus also does not appear as a good candidate for control for the same reasons.

C.5 Which data are important?

The routing circuit model was formulated to solve a specific problem - namely, the recognition of objects in a complex environment independent of translation and scale on the retina. We have hypothesized that the process of visual attention provides a solution to this problem by reducing the amount of incoming information to a manageable size and by bringing this information to the pattern recognition system in the correct format (i.e., normalized for position and scale). In the psychophysics community, on the other hand, attention is largely viewed as a phenomenon - something that gives you faster reaction times or better performance when you are pre-cued to a particular location (or feature dimension). This view naturally leads to differences between the way attention is commonly investigated and the way one would investigate attention if testing our model. However, we must make do with the data that are available, and so we need some means of judging which data to weigh seriously and which to weigh less seriously.

Since the model is directed at the task of pattern recognition, then those experiments involving the processing of patterns, such as the experiments of Bergen and Julesz (1983) and Triesman (1988), are probably ones that we can most directly relate to our model. On the other hand, those experiments involving such tasks as detecting the onset or dimming of a spot of light, or color or orientation discrimination, are more difficult to relate to our model. Although these tasks may qualify as "attention-related" since certain stimuli can be processed quicker or more efficiently when computational resources are focused on them, they can be solved with much simpler models that do not require a routing circuit for shifting and scaling information. It may well be that the capabilities of our model could transfer to such tasks, but the outcome of these experiments, either way, cannot be weighed as serious evidence for or against the model.

Lesion studies, in particular, need to be interpreted with extreme caution. Previous lesion studies of other parts of the brain have revealed that it is a highly distributed, redundant system. This makes it very difficult to discount a particular area taking part in some function just because one doesn't obtain a substantial behavioral effect. For example, lesions of SC or FEF alone seem to have only a mild effect on eye movements, but simultaneous lesions to both areas produce a dramatic effect (eye movements are virtually abolished). Also, as Ungerleider and Mishkin have pointed out, initial lesion studies of the prestriate cortex seemed to indicate that it was not the principal pathway between V1 and IT, but later studies made it clear that sparing even small amounts of the prestriate cortex left viable pathways between V1 and IT that enabled the animals to still perform object recognition tasks. Even prosopagnosics seem to retain more or less normal visual function in most respects, except that they may occasionally confuse one object for another.

In this context, I would evaluate each of the physiological and lesion studies of the pulvinar to date as follows:

- Petersen et al.'s (1985) physiology experiment utilizes the task of detecting the dimming of a spot of light. It is assumed that the spot of light was attended if its dimming was correctly detected. This task is unrelated to pattern recognition and can be solved by many other means besides the dynamic routing circuit, so we would be hard pressed to use the outcome as direct evidence. They find an enhancement in 50% of the cells recorded in Pdm when the attended spot of light falls within its RF. Attentional enhancements are absent in PL and PI, but one does find cells here, and in Pdm, that are modulated by saccades. Cells in Pdm are connected with both IT and PP. Perhaps the Pdm cells are exerting an attentional effect on IT, or on PP? Or perhaps the Pdm cells are merely reflecting the enhancement of PP cells - i.e., they are an effect, and not the cause, of attention.

- Petersen et al.'s (1987) neuropharmacological experiment utilizes the task of detecting the onset of a spot of light in pre-cued and un-cued conditions. Reaction times are faster in the pre-cued condition than the un-cued condition. Again, this task is unrelated to pattern recognition, so it provides relatively weak evidence. Injecting GABA agonists and antagonists into Pdm seems to respectively slow down or speed up shifts of attention, insofar as reaction times get slower and faster. The results of this experiment would seem to support a causal role for Pdm in visual attention, although whether the effect is exerted on PP or IT cannot be resolved. Presumably, either of these areas could be involved in performing this task.

- Rafal and Posner's (1987) lesion experiment also utilizes the task of detecting the onset of a spot of light (actually, an asterix). Subjects were precued to one of two locations on opposite sides of the visual field and asked to respond as quickly as possible when an asterix was presented in one of the locations. The subjects with thalamic lesions (i.e., pulvinar + presumably other parts of the thalamus) consistently had slower reaction times on the contralesional side, although the improvement obtained with increasing SOA times was the same for both sides. This is in contrast to what one obtains with SC lesions, in which targets presented on the contralesional side result in a much slower improvement with increasing SOA times, suggesting an impairment in the "move" operation of attention. Thus, it is concluded that patients with pulvinar lesions are moving their attention OK, but not "engaging" it as thoroughly as without the lesion. Although it is noted that all the thalamic lesioned patients had "no clinical evidence of visual impairment," it is still hard to rule out the possibility that the thalamic lesions merely screw up the visual system in some general way unrelated to attention.

- Desimone et al.'s (1990) lesion experiment utilizes the task of color discrimination. The animal is pre-cued to a certain location in the visual field and must make a judgement of the color of a bar subsequently presented at that location (maintaining fixation) with and without a distractor. Although this task does not involve pattern recognition, it does at least involve a distractor stimulus - albeit one - that would seem to force the animal to process one location preferentially over the other. They find that complete deactivation of PL (the posterior portion, connected to V4 and IT) of one hemisphere results in an increase in errors ($\sim$ 30%) when a distractor is present in the opposite hemifield. One does not get this effect, however, without a distractor, or when the distractor is placed in the same hemifield as the target. On the other hand, local deactivation of the superior colliculus does produce an increase in errors when the target and distractor are in the same hemifield. Thus, it would seem that both pulvinar and SC play some role in directing information processing to a particular part of the visual scene; but note the only way that SC could actually affect cortical information flow is through the pulvinar, or perhaps via its direct connection to LGN.

- Chalupa et al.'s (1976) tachistoscopic pattern discrimination experiment is getting closer to an appropriate experiment for testing the model, in that he utilizes spatial patterns with brief presentation times. They find that lesions to PI result in a dramatic impairment in pattern discrimination ability, whereas lesions to PL and PM do not. However, his lesions to these latter two areas are incomplete. Also, the patterns being discriminated are an "N" vs. a "Z", which are probably not substantially complex to tax the monkey's pattern recognition system. Bender attempted to repeat Chalupa's experiment, however since he used cynomolgus monkey's, which are slower learners of visual discriminations, he had to scale up the patterns so they were 20°x20° in size! So it does not come as a great surprise that his lesions produced no effect. Both Chalupa and Bender utilized RF lesions, which damage fibers of passage. Thus, it is difficult to say whether Chalupa's effect is due to destroying the cortico-tectal pathway or the pulvinar itself.

- Bender and Butter's (1987) study also included a visual search task in which the monkey scanned a display (with eye movements) to look for a particular target item - for example, a circle - placed in a field of distracting items - squares, triangles, crosses, etc. The patterns were 2°-4° in size. Although this task relies on overt instead of covert attention, it does require the animal to discriminate among patterns. Bender's lesions produced massive bilateral damage to PI and PL - and in one monkey much of PM too. However, only a very mild increase in errors was observed. This result is not inconsistent with our model, since it is conceivable that the animal could simply be homing in on a particular feature in the target, and hence there would be no need for a routing circuit. However, it is still a bit disturbing that a stronger effect was not observed.

- Nagel-Leibey et al.'s (1984) lesion study utilizes the task of discriminating a plus vs. square pattern. Both RF and kainic acid lesions are used (the latter leaves the cortico-tectal fibers of passage intact). Both lesions result in mild deficits in the retention of the ability to learn these tasks. The lesions involved large portions of PI and PL. The effect obtained here seems to be suggestive of some role for the pulvinar in pattern discrimination, however the patterns used here are still quite simple. Nor is the task particularly demanding of attention (long presentation times, no distractors, etc.).

- Laberge and Buchsbaum's (1990) PET study utilizes the task of discriminating the shape of a target item when it is surrounded by a field of distractors vs. when it is displayed alone. This is probably the best designed task of any experiment to date for testing the model, in that it involves both pattern discrimination and the filtering out of irrelevant information. They report that the pulvinar shows a higher differential activation when the target item is surrounded by distractors than when it is displayed alone. However, his data are less than convincing. Only the left pulvinar shows enhanced activity when the distractors are present. In addition, V1 shows enhanced activity too when the distractors are present. Thus, it could well be the case that the presence of the distractors causes more activation overall in V1, which then excites the pulvinar differentially. This experiment would have been more compelling had he done a control in which the distractor stimuli were viewed passively, as opposed to attentively.

- Finally, it should be noted that Robinson et al. (1986;1990) have carried out several studies investigating the responses of pulvinar neurons in relation to eye movements. Most cells are enhanced before an eye movement, irrespective of the direction of the eye movement; others are selective for direction. Some cells are suppressed during eye movements, others not. Also, some cells seem to show activity signaling the end of a saccade. It has been determined that most of these effects are due to an extraretinal signal - i.e. proprioceptive information. In general, it makes sense that the control for covert attention would somehow have to be coordinated with eye movements, but since the model does not yet address this issue it is hard to say what one would expect of pulvinar neurons with respect to eye movements. At this stage, my best guess is that the general enhancement effect observed before and after saccades could possibly be a way of resetting the pulvinar before and after an eye movement (this effect is also found in V1 and other visual cortical areas). The suppression during eye movements would also be desirable in order to filter out irrelevant information as the eye sweeps over a scene (such an effect has also been observed in the LGN).

C.6 Design of experiments

Consider the problem a monkey faces in recognizing another monkey in the jungle, among the many other objects—coconuts, flowers, foliage, etc.—that clutter the natural visual environment. This task demands enormous amounts of sensory information processing, yet it probably comes quite naturally and effortlessly to the monkey. Now compare this with the situation a monkey typically faces in any one of the above experiments—for example, detecting the dimming of a spot of light in the periphery, or judging whether the orientation or color of two bars is the same or different. These latter tasks demand trivial amounts of visual computation by comparison, yet I would guess that they are conceptually much more difficult for the animal to perform.

The model we have proposed is meant to address a problem of the former type—i.e., computationally intensive, yet effortless. Thus, experiments designed to test the routing circuit hypothesis should be designed along these lines. The task used in the experiment should be *computationally* challenging, but not necessarily conceptually difficult for the animal to perform. Indeed, designing tasks that naturally fit an animals capabilities would be an advantage in training. An example might be to train an animal to discriminate among two or more complex shapes - for example, different faces - and then test the animals performance with the same shapes presented at different sizes and locations on the retina, or among a field of distracting items. If an animal could still perform such a task with extensive pulvinar lesions, or if no modulation of activity were observed in pulvinar neurons during such a task, then we would have to reconsider what role, if any, the pulvinar plays in visual attention.

Another important consideration in the design of experiments is to determine where data are most needed to test a certain theory. Here, it would be helpful to have an alternative theory against which to judge the evidence. Finding data that do not fit one particular model will tend to lower the likelihood of that model somewhat, but it would be even nicer if we could show that the data better fit an alternative model. Unfortunately, few successful models exist that are designed to solve the same problem as ours (see Section 1.2). In order to qualify as a valid neurobiological theory, one must be able to show (a) that the method actually works for recognizing shapes over wide ranges of position and scale, (b) that it does so in a manner consistent with known neuroanatomy, neurophysiology and psychophysics, and (c) that it makes solid, experimentally testable predictions about what one would expect to find in various cortical areas.

On the other hand, there are plenty of other models of attention in general that could be tested against ours. The model of Niebur et al. (1993), for example, would predict 40Hz oscillatory activity in the pulvinar and within attended regions of V1, and non-oscillatory activity during states of inattention. Their model also predicts shifts and enlargements of V4 receptive fields, but our model would predict that an RF would cover only a fraction of the attentional window, whereas in their model the RF could easily cover the entire attentional window, since spatial relationships within the attentional window are not preserved. Another consequence of preserving spatial relationships is that the patterns of activation on the cortical surface should translate, or dilate and contract, as attention is shifted and scaled over the scene. This prediction could probably be tested most directly with optical recording techniques, but this method still seems to be in its infancy.

C.7 Conclusion

The available data on the pulvinar leave us with a mixed bag of evidence. All of the physiological and lesion study evidence is rather weak, with most of it for, and some (namely, Bender's search task) against the pulvinar playing an important role in attention and pattern recognition. Thus, the strongest argument for the pulvinar as a control site comes from the way that it is anatomically situated with respect to the rest of the cortex. Until additional data are obtained to the contrary - from experiments designed in the appropriate manner - it would be premature to discount the pulvinar's role in controlling attention with what little negative data exist.

The brain has been wonderfully constructed to perform computational feats that are beyond our most powerful supercomputers. In order to deduce how the brain carries out these operations, it is important to observe the system during periods of intense computation. Simple probes have traditionally been used in order to isolate various aspects of the stimulus, and this approach has been useful at giving us a first cut estimate of what may be going in different brain areas. However, revealing the more complex operations of the brain - such as visual attention and complex pattern recognition - will require experimental designs that better challenge an animal's natural computational capabilities.

References

Ahmad S (1992) VISIT: A neural model of covert visual attention. In: Advances in Neural Information Processing Systems 4 (Moody JE, Hanson SJ, Lippman RP, eds), San Mateo, CA: Kaufmann, pp 420-427.

Anderson CH, Burt PJ, van der Wall GS (1985) Change detection and tracking. SPIE Vol. 579—Intelligent Robots and Computer Vision, 72-78.

Anderson CH, Van Essen DC (1987) Shifter circuits: A computational strategy for dynamic aspects of visual processing. Proceedings of the National Academy of Sciences, USA, 84:6297-6301.

Anderson CH, Van Essen DC (1993) Dynamic neural routeing circuits. In: Visual Search 2 (Brogan D, Gale A, Carr K, eds), London: Taylor & Francis.

Anderson JA, Rosenfeld E (1988) Neurocomputing: Foundations of research. Cambridge, MA: MIT Press. p.29.

Anstis SM (1974) A chart demonstrating variations in acuity with retinal position. Vision Research, 14:589-592.

Atick JJ, Redlich AN (1990) Towards a theory of early visual processing. Neural Computation, 2: 308-320.

Attneave, F (1954) Some informational aspects of visual perception. Psychological Review, 61: 183-193.

Attneave, F (1965) Triangles as ambiguous figures. American Journal of Psychology, 81: 447-453.

Baleydier C, Morel A (1992) Segregated thalamocortical pathways to inferior parietal and inferotemporal cortex in macaque monkey. Visual Neuroscience, 8: 391-405.

Baron RJ (1987) The Cerebral Computer. Erlbaum.

Bender DB (1981) Retinotopic organization of macaque pulvinar. J Neurophysiol, 46: 672-693.

Bender DB (1988) Electrophysiological and behavioral experiments on the primate pulvinar. In: Progress in Brain Research, 75 (Hicks TP, Benedek G, eds), New York: Elsevier, pp 55-65.

Bender DB, Butter CM (1987) Comparison of the effects of superior colliculus and pulvinar lesions on visual search and tachistoscopic pattern discrimination in monkeys. Exp Brain Res, 69: 140-154.

Benevento LA, Rezak M (1976) The cortical projections of the inferior pulvinar and adjacent lateral pulvinar in the Rhesus monkey: An autoradiographic study. Brain Research, 108:1-24.

Benevento LA, Davis B (1977) Topographical projections of the prestriate cortex to the pulvinar nuclei in the macaque monkey: An autoradiographic study, Brain Research, 30: 405-424.

Bergen JR, Julesz B (1983) Parallel versus serial processing in rapid pattern discrimination. Nature, 303:696-698.

Berman NJ, Douglas RJ, Martin KAC (1992) GABA-mediated inhibition in the neural networks of visual cortex. In: Progress in Brain Research, 90 (Mize RR, Marc RE, Silito AM, eds), New York: Elsevier, pp 443-476.

Biederman I, Cooper EE (1992) Evidence for complete translational and reflectional invariance in visual object priming. Perception, 20:585-593.

Buhmann J, Lades M, von der Malsburg C (1990) Size and distortion invariant object recognition by hierarchical graph matching. In: Proceedings of the International Joint Conference on Neural Networks, San Diego, June, pp. 411-416.

Burt PJ, Adelson EH (1983) The Laplacian pyramid as a compact image code. IEEE Transactions on Communications, 31: 532-540.

Bushnell C, Goldberg ME, Robinson DL (1981) Behavioral enhancement of visual responses in monkey cerebral cortex. I. Modulation in posterior parietal cortex related to selective visual attention. Journal of Neurophysiology, 46(4): 755-772.

Campbell FW (1985) How much of the information falling on the retina reaches the visual cortex and how much is stored in the visual memory? In: Pattern Recognition Mechanisms (Chagas C, Gattass R, Gross C, eds), Berlin: Springer, pp 83-95.

Carpenter G, Grossberg S (1987a) A massively parallel architecture for a self-organizing neural pattern recognition machine. Computer Vision Graphics and Image Processing, 37:54-115.

Carpenter GA, Grossberg S (1987b) Invariant pattern recognition and recall by an attentive self-organizing ART architecture in a nonstationary world. In: Proceedings of the IEEE First International Conference on Neural Networks.

Cauller LJ, Connors BW (1992) Functions of very distal dendrites: Experimental and computational studies of layer I synapses on neocortical pyramidal cells. In: Single Neuron Computation (McKenna T, Davis JL, Zornetzer SF, eds), Academic Press, Cambridge, MA, pp 199-229.

Cavanagh P (1978) Size and position invariance in the visual system. Perception, 7:167-177.

Cavanagh P (1985) Local log polar frequency analysis in the striate cortex as a basis for size and orientation invariance. In: Models of the Visual Cortex (Rose D, Dobson VG, eds), New York: Wiley, pp 85-95.

Cavanagh P (1992) Attention-based motion perception. Science, 257:1563-1565.

Cave KR (1991) What makes a spotlight a spotlight? (unpublished manuscript—Dept. of Psychology, Vanderbilt University)

Chalupa LM, Coyle D, Lindsley DB (1976) Effect of pulvinar lesions on visual pattern discrimination in monkeys. Journal of Neurophysiology, 39(2):354-369.

Cherniak C (1990) The bounded brain: Toward quantitative neuroanatomy. Journal of Cognitive Neuroscience, 2(1):58-68.

Conley M, Diamond IT (1990) Organization of the visual sector of the thalamic reticular nucleus in *Galago*. European Journal of Neuroscience, 2(3):211-226.

Connor CE, Gallant JL, Van Essen DC (1993) Effects of focal attention on receptive field profiles in area V4. Soc. Neurosci. Abstr., 19.

Corbetta M, Miezin FM, Dobmeyer S, Shulman GL, Petersen SE (1991) Selective and divided attention during visual discriminations of shape, color, and speed: Functional anatomy by positron emission tomography. The Journal of Neuroscience, 11(8): 2383-2402.

Crick F (1984) Function of the thalamic reticular complex: The searchlight hypothesis. Proceedings of the National Academy of Sciences, 81:4586-4590.

Crick F, Koch C (1990) Towards a neurobiological theory of consciousness. Seminars in the Neurosciences, 2:263-275.

Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. The Journal of Neuroscience, 4(8):2051-2062.

Desimone R, Schein SJ (1987) Visual properties of neurons in area V4 of the macaque: Sensitivity to stimulus form. Journal of Neurophysiology, 57(3): 835-868.

Desimone R, Wessinger M, Thomas L, Schneider W (1990) Attentional control of visual perception: Cortical and subcortical mechanisms. In: Cold Spring Harbor Symp Quant Biol, 55:963-971.

Desimone R (1992) Neural circuits for visual attention in the primate brain. In: Neural Networks for Vision and Image Processing (Carpenter GA, Grossberg S, eds), Cambridge, Mass.: MIT Press, pp 343-364.

De Valois RL, De Valois KK (1988) Spatial Vision. New York: Oxford.

De Valois RL, Albrecht DG, Thorell LG (1982) Spatial frequency selectivity of cells in macaque visual cortex. Vision Res, 22: 545-559.

DeYoe EA, Sisola LC (1991) Distinct pathways link anatomical subdivisions of V4 with V2 and temporal cortex in the macaque monkey. Society for Neuroscience Abstracts, 17: 1282.

Douglas RJ, Martin KAC, Whitteridge D (1988) Selective responses of visual cortical cells do not depend on shunting inhibition. Nature, 332:642-644.

Douglas RJ, Martin KAC (1990a) Neocortex. In: Synaptic Organization of the Brain (Shepard GM, ed), New York: Oxford UP, pp 389-438.

Douglas RJ, Martin KAC (1990b) Control of neuronal output by inhibition at the axon initial segment. Neural Computation, 2:283-292.

Duhamel J, Colby L, Goldberg ME (1992) The updating of the representation of visual space in parietal cortex by intended eye movements. Science, 255:90-92.

Eriksen CW, Murphy TD (1987) Movement of attentional focus across the visual field: A critical look at the evidence. Perception and Psychophysics, 42(3):299-305.

Eriksen CW, St. James JD (1986) Visual attention within and around the field of focal attention: A zoom lens model. Perception and Psychophysics, 40 (4):225-240.

Farell B, Pelli DG (1993) Can we attend to large and small at the same time? Vision Research, 33: 2757-2772.

Felleman DJ, McClendon E (1991) Modular connections between area V4 and temporal lobe area PITv in macaque monkeys. Society for Neuroscience Abstracts, 17: 1282.

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cerebral Cortex, 1(1):1-47.

Felleman DJ, McClendon E, Lin K (1992) Modular segregation of visual pathways in occipital and temporal lobe visual areas in the macaque monkey. Society for Neuroscience Abstracts, 18: 390.

Field, DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Am, A, 4: 2379-2394.

Fields HL, Basbaum AI (1978) Brainstem control of spinal pain-transmission neurons. Annu Rev Physiol, 40:217-248.

Fischer B (1973) Overlap of receptive field centers and representation of the visual field in the cat's optic tract. Vision Research, 13:2113-2120.

Foldiak P (1991) Learning invariance from transformation sequences. Neural Computation, 3:194-200.

Fukushima K (1980) Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics, 36:193-202.

Fukushima K (1987) Neural network model for selective attention in visual pattern recognition and associative recall. Applied Optics, 26(23):4985-4992.

Gallant JL, Braun J, Van Essen DC (1993) Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. Science, 259:100-103.

Gattass R, Sousa APB, Covey E (1985) Cortical visual areas of the macaque: Possible substrates for pattern recognition mechanisms. In: Pattern Recognition Mechanisms (Chagas C, Gattass R, Gross C, eds), Berlin: Springer, pp 1-20.

Gattass R, Desimone R (1991) Attention-related responses in the superior colliculus of the macaque. Soc Neurosci Abstr, 17: 545.

Gattass R, Desimone R (1992) Stimulation of the superior colliculus (SC) shifts the focus of attention in the macaque. Soc Neurosci Abstr, 18: 703.

Gilbert CD, Wiesel TN (1989) Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. The Journal of Neuroscience, 9: 2432-2442.

Goldberg ME, Wurtz RH (1972) Activity of superior colliculus in behaving monkey. I. Visual receptive fields of single neurons. Journal of Neurophysiology, 35:542-559.

Gross CG, Rocha-Miranda CE, Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the macaque. Journal of Neurophysiology, 35:96-111.

Grossberg S (1976) Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. Biological Cybernetics, 23: 121-134.

Hardy SGP, Lynch JC (1992) The spatial distribution of pulvinar neurons that project to two subregions of the inferior parietal lobule in the macaque, Cerebral Cortex, 2: 217-230.

Hawken MJ, Parker AJ (1987) Spatial properties of neurons in the monkey striate cortex. Proc R Soc Lond B, 231: 251-288.

Hinton GE (1979) Some demonstrations of the effects of structural descriptions in mental imagery. Cognitive Science, 3: 231-250.

Hinton GE (1981a) A parallel computation that assigns canonical object-based frames of reference. In: Proceedings of the Seventh International Joint Conference on Artificial Intelligence 2, Vancouver B.C., Canada.

Hinton GE (1981b) Shape representation in parallel systems. In: Proceedings of the Seventh International Joint Conference on Artificial Intelligence 2, Vancouver B.C., Canada.

Hinton GE, Lang KJ (1985) Shape recognition and illusory conjunctions. In: Proceedings of the Ninth International Joint Conference on Artificial Intelligence, Los Angeles.

Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA, 79:2554-2558.

Hopfield JJ (1984) Neurons with graded response have collective computational properties like those of two-state neurons. Proc Natl Acad Sci USA, 81:3088-3092.

Kanerva P (1988) Sparse Distributed Memory. Cambridge, Mass.: MIT Press.

Kennedy H, Bullier J (1985) A double-labeling investigation of the afferent connectivity to corical areas V1 and V2 of the macaque monkey. The Journal of Neuroscience, 5: 2815-2830.

Koch, C. and Poggio, T. (1992) Multiplying with synapses and neurons. In: Single Neuron Computation (McKenna T, Davis JL, Zornetzer SF, eds), Cambridge, MA: Academic, pp 315-345.

Koch C, Ullman S (1985) Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiology, 4:219-227.

Koenderink JJ, van Doorn AJ (1978) Visual detection of spatial contrast; Influence of location in the visual field, target extent and illuminance level. Biological Cybernetics, 30: 157-167.

Kosslyn SM, Schwartz SP (1978) Visual images as spatial representations in active memory. In: Computer Vision Systems (Hanson AR, Riseman EM, eds), New York: Academic Press.

LaBerge D (1990) Thalamic and cortical mechanisms of attention suggested by recent positron emission tomographic experiments. Journal of Cognitive Neuroscience, 2(4): 358-372.

LaBerge D, Buchsbaum MS (1990) Positron emission tomographic measurements of pulvinar activity during an attention task. The Journal of Neuroscience, 10(2):613-619.

LaBerge D, Carter M, and Brown V (1992) A network simulation of thalamic circuit operations in selective attention. Neural Computation, 4: 318-331.

Larsen A, Bundesen C (1978) Size scaling in visual pattern recognition. Journal of Experimental Psychology: Human Perception and Performance, 4: 1-20.

Lashley, KS (1942) The problem of cerebral organization in vision, Biol Symp, 7: 301-322.

LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, Jackel LD (1990) Backpropagation applied to handwritten Zip code recognition. Neural Computation, 1:541-551.

Lee CW, Olshausen BA (1993) A nonlinear hebbian network that learns to detect disparity in random-dot stereograms. (submitted)

Li Z, Atick JJ (1994) Towards a theory of the striate cortex. Neural Computation, 6.

Llinas RR (1988) The intrinsic electrophysiological properties of mammalian neurons: Insights into central nervous system function. Science, 242:1654-1663.

Lowe DG (1987) The viewpoint consistency constraint. International Journal of Computer Vision, 1:57-72.

Lund JS (1988) Anatomical organization of macaque monkey striate visual cortex. Annual Review of Neuroscience, 11: 253-288.

Marr D, Poggio T (1976) Cooperative computation of stereo disparity. Science, 194:283-287.

Marr D (1982) Vision. New York: W.H. Freeman.

Mel, B.W. (1992) NMDA-based pattern discrimination in a modeled cortical neuron. Neural Computation, 4: 502-517.

Merigan WH, Katz LM (1990) Spatial resolution across the macaque retina. Vision Research, 30: 985-991.

Miller KD, Chapman B, Stryker MP (1989) Visual responses in adult cat visual cortex depend on N-methyl-D-aspartate receptors. Proceedings of the National Academy of Sciences, 86:5183-5187.

Mishkin M (1972) Cortical visual areas and their interactions. In: Brain and Human Behavior (A.G. Karczmar, J.C. Eccles, eds), New York: Springer, pp 187-208.

Moran J, Desimone R (1985) Selective attention gates visual processing in the extrastriate cortex. Science, 229:782-784.

Mountcastle VB, Andersen RA, Motter BC (1981) The influence of attentive fixation upon the excitability of the light-sensitive neurons of the posterior parietal cortex. The Journal of Neuroscience, 1(11):1218-1235.

Mozer MC, Behrmann M (1992) Reading with attentional impairments: A brain-damaged model of neglect and attentional dyslexias. In: Connectionist Approaches to Natural Language Processing (Reilley RG, Sharkey NE, eds), Hillsdale, NJ: Erlbaum, pp 409-460.

Mumford D (1992) On the computational architecture of the neocortex. II The role of cortico-cortical loops. Biological Cybernetics, 66:241-251.

Nagel-Leiby S, Bender DB, Butter CM (1984) Effects of kainic acid and radiofrequency lesions of the pulvinar on visual discrimination in the monkey. Brain Res, 300: 295-303.

Nakayama K, Mackeben M (1989) Sustained and transient components of focal visual attention. Vision Res, 29 (11):1631-1647.

Nakayama K (1991) The iconic bottleneck and the tenuous link between early visual processing and perception. In: Vision: Coding and Efficiency (Blakemore C, ed), Cambridge: Cambridge UP, pp 411-422.

Nelson SB, Sur M (1992) NMDA receptors in sensory information processing. Current Opinion in Neurobiology, 2:484-488.

Niebur E, Koch C, Rosin C (1993) An oscillation-based model for the neural basis of attention. Vision Research, 33: 2789-2802.

Nowlan SJ, Sejnowski TJ (1993) Filter selection model for generating visual motion signals. Advances in Neural Information Processing Systems, 5 (Hanson SJ, Cowan JD, Giles CL, eds), San Mateo, CA: Morgan-Kaufmann, pp. 369-376.

Ogren MP, Hendrickson AE (1977) The distribution of pulvinar terminals in visual areas 17 and 18 of the monkey. Brain Research, 137:343-350.

Ogren MP, Hendrickson AE (1979) The structural organization of the inferior and lateral subdivisions of the *Macaca* monkey pulvinar. J Comp Neur, 188:147-178.

Ohzawa I, Sclar G, Freeman RD (1982) Contrast gain control in the cat visual cortex. Nature, 298:266-268.

O'Kusky J, Colonnier M (1982) A laminar analysis of the number of neurons, glia, and synapses in the visual cortex (area 17) of adult macaque monkeys. Journal of Comparative Neurology, 210: 178-290.

Olshausen BA, Anderson CH, Van Essen DC (1993) A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. The Journal of Neuroscience, 13: 4700-4719.

Palmer SE (1983) The psychology of perceptual organization: A transformational approach. In: Human and Machine Vision (Beck J, Hope B, Rosenfeld A, eds), Orlando: Academic, pp 269-339.

Parker AJ, Hawken MJ (1988) Two-dimensional spatial structure of receptive fields in monkey striate cortex. Journal of the Optical Society of America A, 5: 598-605.

Pei X, Volgushev M, Creutzfeldt O (1992) A comparison of directional sensitivity with the excitatory and inhibitory field structure in cat striate cortical simple cells. Perception, 21, supplement 2, p. 26.

Petersen SE, Robinson DL, Keys W (1985) Pulvinar nuclei of the behaving rhesus monkey: Visual responses and their modulation. Journal of Neurophysiology, 54(4):867-886.

Petersen SE, Robinson DL, Morris JD (1987) Contributions of the pulvinar to visual spatial attention. Neuropsychologia, 25(1A):97-105.

Pettet MW, Gilbert CD (1992) Dynamic changes in receptive-field size in cat primary visual cortex. Proc Natl Acad Sci USA, 89:8366-8370.

Pitts W, McCulloch WS (1947) How we know universals: The perception of auditory and visual forms. Bulletin of Mathematical Biophysics, 9:127-147.

Pollen DA, Lee JR, Taylor JH (1971) How does the striate cortex begin the reconstruction of the visual world? Science, 173:74-77.

Poggio T, Edelman S (1990) A network that learns to recognize three-dimensional objects. Science, 343: 263-266.

Posner MI, Cohen Y (1984) Components of visual orienting. In: Attention and Performance X: Control of Language Processes. (Bouma H, Bouwhuis DG, eds), Hillsdale, NJ: Erlbaum, pp 531-556.

Posner MI, Walker JA, Friedrich FJ, Rafal RD (1984) Effects of parietal injury on covert orienting of attention. The Journal of Neuroscience, 4(7):1863-1874.

Posner MI, Petersen SE, Fox PT, Raichle ME (1988) Localization of cognitive operations in the human brain. Science, 240: 1627-1631.

Posner MI, Petersen SE (1990) The attention system of the human brain. Annu Rev Neurosci, 13:25-42.

Postma EO, van den Herik HJ, Hudson PTW (1992) The gating lattice: A neural substrate for dynamic gating. In: CNS*92 Proceedings, July 26-29, San Francisco, California. Kluwer Academic Publishers.

Press WA, Olshausen BA, Van Essen DC (1993) Analyzing connections between the pulvinar and visual cortex: An interactive graphical database. Soc Neurosci Abstr, 19: 331.

Rafal RD, Posner MI (1987) Deficits in human visual spatial attention following thalamic lesions. Proc Natl Acad Sci USA, 84: 7349-7353.

Remington R, Pierce L (1984) Moving attention: Evidence for time-invariant shifts of visual selective attention. Perception and Psychophysics, 35(4):393-399.

Rezak M, Benevento LA (1979) A comparison of the organization of the projections of the dorsal lateral geniculate nucleus, the inferior pulvinar and adjacent lateral pulvinar to primary visual cortex (area 17) in the macaque monkey. Brain Research, 167:19-40.

Robinson DL, McClurkin JW (1989) The visual superior colliculus and pulvinar. In: The Neurobiology of Saccadic Eye Movements (Wurtz and Goldberg, eds), Elsevier Science Publishers BV, pp. 337-359.

Robinson DL, Petersen SE (1992) The pulvinar and visual salience, Trends in Neuroscience, 15(4):127-132.

Robinson DL, Goldberg ME, Stanton GB (1978) Parietal association cortex in the primate: Sensory mechanisms and behavioral modulations. Journal of Neurophysiology, 41(4):910-932.

Robinson DL, Petersen SE, Keys W (1986) Saccade-related and visual activities in the pulvinar nuclei of the behaving rhesus monkey. Exp Brain Res, 62: 625-634.

Robinson DL, McClurkin JW, Kertzman C (1990) Orbital position and eye movement influences on visual responses in the pulvinar nuclei of the behaving macaque. Exp Brain Res, 82: 235-246.

Robinson DL, Bowman EM, Kertzman C (1991) Covert orienting of attention in macaque: II. A signal in parietal cortex to disengage attention. Society for Neuroscience Abstracts, 17: 442.

Rock I (1973) Orientation and form. New York: Academic Press.

Rock I (1988) On Thompson's inverted-face phenomenon (Research Note). Perception, 17:815-817.

Rockland KS (1992) Configuration, in serial reconstruction, of individual axons projecting from area V2 to V4 in the macaque monkey. Cerebral Cortex, 2:353-374.

Rolls ET, Baylis GC (1986) Size and contrast have only small effects on the responses to faces of neurons in the cortex in the superior temporal sulcus of the monkey. Exp Brain Res, 65: 38-48.

Saarinen J, Julesz B (1991) The speed of attentional shifts in the visual field. Proceedings of the National Academy of Sciences, 88:1812-1814.

Sandon PA, Uhr LM (1988) An adaptive model for viewpoint-invariant object recognition. Proceedings of the 10th Annual Conference of the Cognitive Science Society, Montreal, Canada, August, pp. 209-215.

Sandon PA (1990) Simulating visual attention. Journal of Cognitive Neuroscience, 2(3):213-231.

Schmahmann JD, Pandya DN (1990) Anatomical investigation of projections from thalamus to posterior parietal cortex in the rhesus monkey: A WGA-HRP and fluorescent tracer study, The Journal of Comparative Neurology, 295: 299-326.

Schwartz EL (1977) Spatial mapping in the primate sensory projection: Analytic structure and relevance to perception. Biological Cybernetics, 25: 181-194.

Shepard RN, Metzler J (1971) Mental rotation of three-dimensional objects. Science, 171: 701-703.

Sherman SM, Koch C (1986) The control of retinogeniculate transmission in the mammalian lateral geniculate nucleus. Experimental Brain Research, 63:1-20.

Shulman GL and Wilson J (1987) Spatial frequency and selective attention to local and global information. Perception, 16: 89-101.

Sillar KT (1991) Spinal pattern generation and sensory gating mechanisms. Current Opinion in Neurobiology, 1:583-589.

Sperling G, Landy MS, Cohen Y, Pavel M (1985) Intelligible encoding of ASL image sequences at extremely low information rates. Computer Vision Graphics and Image Processing, 31: 335-391.

Steinmetz MA, Connor CE, MacLeod KM (1992) Focal spatial attention suppresses responses of visual neurons in monkey posterior parietal cortex. Society for Neuroscience Abstracts, 18: 148.

Toet A, van Eekhout P, Simons HLJJ, Koenderink JJ (1987) Scale invariant features of differential spatial displacement discrimination. Vision Res, 27: 441-451.

Tootell BH, Silverman MS, Hamilton SL, Switkes E, De Valois RL (1988) Functional anatomy of macaque striate cortex. V. Spatial Frequency. The Journal of Neuroscience, 8: 1610-1624.

Treisman AM (1988) Features and objects: The fourteenth Bartlett memorial lecture. The Quarterly Journal of Experimental Psychology, 40A(2): 201-237.

Treisman AM, Schmidt H (1982) Illusory conjunctions in the perception of objects. Cognitive Psychology, 14: 107-141.

Trojanowski JQ, Jacobson S (1976) Areal and laminar distribution of some pulvinar cortical efferents in rhesus monkey, J Comp Neur, 169: 371-392.

Tsal Y (1983) Movements of attention across the visual field. Journal of Experimental Psychology: Human Perception and Performance, 9:523-530.

Tsotsos JK (1991) Localizing stimuli in a sensory field using an inhibitory attention beam. Technical Report, RBCV-TR-91-37, Dept. of Computer Science, University of Toronto.

Ungerleider LG, Mishkin M (1982) Two cortical visual systems. In: Analysis of Visual Behavior (Ingle DJ, Goodale MA, Mansfield RJW, eds), Cambridge, Mass.: MIT Press, pp 549-586.

Van Essen DC, Newsome WT, Maunsell JHR (1984) The visual field representation in striate cortex of the macaque monkey: Asymmetries, anisotropies, and individual variability. Vision Res, 24: 429-448.

Van Essen DC, Newsome WT, Maunsell JHR, Bixby JL (1986) The projections from striate cortex (V1) to areas V2 and V3 in the macaque monkey: Asymmetries, areal boundaries, and patchy connections. The Journal of Comparative Neurology, 244:451-480.

Van Essen DC, Felleman DJ, DeYoe EA, Olavarria J, Knierim J (1990) Modular and hierarchical organization of extrastriate visual cortex in the macaque monkey. In: Cold Spring Harbor Symp Quant Biol, 55:679-696.

Van Essen DC, Anderson CH (1990) Information processing strategies and pathways in the primate retina and visual cortex. In: An Introduction to Neural and Electronic Networks (Zornetzer SF, Davis JL, Lau C, ed), New York: Academic, pp 43-72. (2nd Edition in press)

Van Essen DC, Olshausen B, Anderson CH, Gallant JL (1991) Pattern recognition, attention, and information bottlenecks in the primate visual system. In: Proc SPIE Conf on Visual Information Processing: From Neurons to Chips, Vol 1473 (Mathur BP, Koch C, eds), Bellingham, WA: SPIE, pp 17-28.

Van Essen DC, Anderson CH, Olshausen BA (1994) Dynamic routing strategies in sensory, motor, and cognitive processing. In: Large Scale Neuronal Theories of the Brain (Koch C, Davis J, eds) MIT Press.

Van Essen DC, DeYoe EA (1993) Concurrent processing in the primate visual cortex. In: The Cognitive Neurosciences (Gazzaniga MS, ed), Cambridge, MA: MIT PRess.

Verghese P, Pelli DG (1992) The information capacity of visual attention. Vision Research, 32(5):983-995.

Volgushev M, Pei X, Vidyasagar TR, Creutzfeldt OD (1992) Orientation-selective inhibition in cat visual cortex: an analysis of postsynaptic potentials. Perception, 21, supplement 2, p. 26.

von der Heydt R, Peterhans E (1989) Mechanisms of contour perception in monkey visual cortex. The Journal of Neuroscience, 9(5):1731-1763.

von der Malsburg C, Bienenstock E (1986) Statistical coding and short-term synaptic plasticity: A scheme for knowledge representation in the brain. In: Disordered Systems and Biological Organization (NATO ASI Series, Vol. F20) (Bienenstock E, Fogelman Soulie F, Weisbuch G, eds), Berlin: Springer, pp 247-272.

Wechsler H, Zimmerman GL (1988) 2-D invariant object recognition using distributed associative memory. IEEE Transactions on Pattern Analysis and Machine Intelligence, 10: 811-821.

Wilson MA, McNaughton BL (1993) Dynamics of the hippocampal ensemble code for space. Science, 261: 1055-1058.

Witkin AP, Terzopoulos D, Kass M (1987) Signal matching through scale space. International Journal of Computer Vision, 1(2):133-144.

Yeterian EH, Pandya DN (1991) Corticothalamic connections of the superior temporal sulcus in rhesus monkeys, Exp Brain Res, 83: 268-284.

Yukie M, Iwai E (1985) Laminar origin of direct projection from cortex area V1 to V4 in the rhesus monkey. Brain Research, 346: 383-386.