Large-Scale Neuronal Theories of the Brain

edited by Christof Koch and Joel L. Davis

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

1994

2 A Critique of Pure Vision[1]

Patricia S. Churchland, V. S. Ramachandran, and Terrence J. Sejnowski

INTRODUCTION

Any domain of scientific research has its sustaining orthodoxy. That is, research on a problem, whether in astronomy, physics, or biology, is conducted against a backdrop of broadly shared assumptions. It is these assumptions that guide inquiry and provide the canon of what is reasonable—of what "makes sense." And it is these shared assumptions that constitute a framework for the interpretation of research results. Research on the problem of how we see is likewise sustained by broadly shared assumptions, where the current orthodoxy embraces the very general idea that the business of the visual system is to create a detailed replica of the visual world, and that it accomplishes its business via hierarchical organization and by operating essentially independently of other sensory modalities as well as independently of previous learning, goals, motor planning, and motor execution.

We shall begin by briefly presenting, in its most *extreme* version, the conventional wisdom. For convenience, we shall refer to this wisdom as the Theory of Pure Vision. We then outline an alternative approach, which, having lurked on the scientific fringes as a theoretical possibility, is now acquiring robust experimental infrastructure (see, e.g., Adrian 1935; Sperry 1952; Bartlett 1958; Spark and Jay 1986; Arbib 1989). Our characterization of this alternative, to wit, *interactive vision*, is avowedly sketchy and inadequate. Part of the inadequacy is owed to the nonexistence of an appropriate vocabulary to express what might be involved in interactive vision. Having posted that caveat, we suggest that systems ostensibly "extrinsic" to literally seeing the world, such as the motor system and other sensory systems, do in fact play a significant role in what is literally seen. The idea of "pure vision" is a fiction, we suggest, that obscures some of the most important computational strategies used by the brain. Unlike some idealizations, such as "frictionless plane" or "perfect elasticity" that can be useful in achieving a core explanation, "pure vision" is a notion that impedes progress, rather like the notion of "absolute downness" or "indivisible atom." Taken individually, our criticisms of "pure vision" are neither new nor convincing; taken collectively in a computational context, they make a rather forceful case.

These criticisms notwithstanding, the Theory of Pure Vision together with the Doctrine of the Receptive Field have been enormously fruitful in fostering research on functional issues. They have enabled many programs of neurobiological research to flourish, and they have been crucial in getting us to where we are. Our questions, however, are not about past utility, but about future progress. Has research in vision now reached a stage where the orthodoxy no longer works to promote groundbreaking discovery? Does the orthodoxy impede really fresh discovery by cleaving to outdated assumptions? What would a different paradigm look like? This chapter is an exploration of these questions.

PURE VISION: A CARICATURE

This brief caricature occupies one corner of an hypothesis-space concerning the computational organization and dynamics of mammalian vision. The core tenets are logically independent of one another, although they are often believed as a batch. Most vision researchers would wish to amend and qualify one or another of the core tenets, especially in view of anatomical descriptions of backprojections between higher and lower visual areas. Nevertheless, the general picture, plus or minus a bit, appears to be rather widely accepted—at least as being correct in its essentials and needing at most a bit of fine tuning. The approach outlined by the late David Marr (1982) resembles the caricature rather closely, and as Marr has been a fountainhead for computer vision research, conforming to the three tenets has been starting point for many computer vision projects.[2]

1. *The Visual World.* What we see at any given moment is in general a fully elaborated representation of a visual scene. The goal of vision is to create a detailed model of the world in front of the eyes in the brain. Thus Tsotsos (1987) says, "The goal of an image-understanding system is to transform two-dimensional data into a description of the three-dimensional spatio-temporal world" (p. 389). In their review paper, Aloimonos and Rosenfeld (1991) note this characterization with approval, adding, "Regarding the central goal of vision as scene recovery makes sense. If we are able to create, using vision, an accurate representation of the three-dimensional world and its properties, then using this information we can perform any visual task" (p. 1250).

2. *Hierarchical Processing.* Signal elaboration proceeds from the various retinal stages, to the LGN, and thence to higher and higher cortical processing stages. At successive stages, the basic processing achievement consists in the extraction of increasingly specific features and eventually the integration of various highly specified features, until the visual system has a fully elaborated representation that corresponds to the visual scene that initially caused the retinal response. Pattern recognition occurs at that stage. Visual leaning occurs at later rather than earlier stages.

3. *Dependency Relations.* Higher levels in the processing hierarchy depend on lower levels, but not, in general, vice versa. Some problems are early (low level) problems; for example, early vision involves determining what is an edge, what correspondences between right and left images are suitable for stereo, what principle curvatures are implied by shading profiles, and where there is movement (Yuille and Ullman 1990). Early vision does not require or depend on a solution to the problems of segmentation or pattern recognition or gestalt.[3]

Note finally that the caricature, and, most especially, the "visual world" assumption of the caricature, gets compelling endorsement from common sense. From the vantage point of how things seem to be, there is no denying that at any given moment we seem to see the detailed array of whatever visible features of the world are in front of our eyes. Apparently, the world is there to be seen, and our brains do represent, in essentially all its glory, what is there to be seen. Within neuroscience, a great deal of physiological, lesion, and anatomical data are reasonably interpretable as evidence for some kind of hierarchical organization (Van Essen and Anderson 1990). Hierarchical processing, moreover, surely seems an eminently sensible engineering strategy—a strategy so obvious as hardly to merit ponderous reflection. Thus, despite our modification of all tenets of the caricature, we readily acknowledge their prima facie reasonableness and their appeal to common sense.

INTERACTIVE VISION: A PROSPECTUS

What is vision for? Is a perfect internal recreation of the three-dimensional world really necessary? Biological and computational answers to these questions lead to a conception of vision quite different from pure vision. Interactive vision, as outlined here, includes vision with other sensory systems as partners in helping to guide actions.

1. *Evolution of Perceptual Systems.* Vision, like other sensory functions, has its evolutionary rationale rooted in improved motor control. Although organisms can of course see when motionless or paralyzed, the visual system of the brain has the organization, computational profile, and architecture it has in order to facilitate the organism's thriving at the four Fs: feeding fleeing, fighting, and reproduction. By contrast, a pure visionary would say that the visual system creates a fully elaborated model of the world in the brain, and that the visual system can be studied and modeled without worrying too much about the nonvisual influences on vision.

2. *Visual Semiworlds.* What we see at any given moment is a partially elaborated representation of the visual scene; only immediately relevant information is explicitly represented. The eyes saccade every 200 or 300 msec, scanning an area. How much of the visual field, and within that, how much of the foveated area, is represented in detail depends on many factors, including the animal's interests (food, a mate, novelty, etc.), its long- and short-term goals, whether the stimulus is refoveated, whether the stimulus is simple or complex, familiar or unfamiliar, expected or unexpected, and so on. Although unattended objects may be represented in some min-

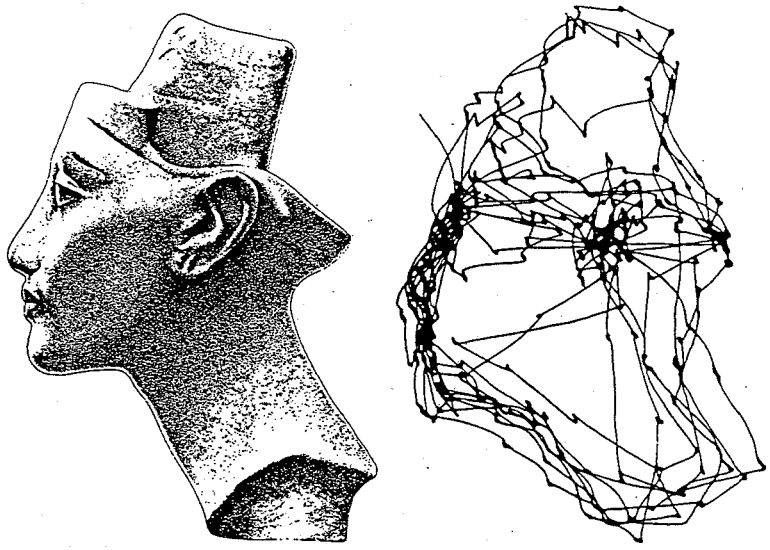


Figure 2.1 The scan path of saccadic eye movements made by a subject viewing the picture. (Reprinted with permission from Yarbus 1967.)

imal fashion (sufficient to guide attentional shifts and eye movements, for example) they are not literally seen in the sense of "visually experienced."

3. *Interactive Vision and Predictive Visual Learning.* Interactive vision is exploratory and predictive. Visual learning allows an animal to predict what will happen in the future; behavior, such as eye movements, aids in updating and upgrading the predictive representations. Correlations between the modalities also improve predictive representations, especially in the murk and ambiguity of real-world conditions. Seeing an uncommon stimulus at dusk such as a skunk in the bushes takes more time than seeing a common animal such as a dog in full light and in full, canonical view. The recognition can be faster and more accurate if the animal can make exploratory movements, particularly of its perceptual apparatus, such as whiskers, ears, and eyes. There is some sort of integration across time as the eyes travel and retravel a scan path (figure 2.1), foveating again and again the significant and salient features. One result of this integration is the strong but false introspective impression that at any given moment one sees, crisply and with good definition, the whole scene in front of one. Repeated exposure to a scene segment is connected to greater elaboration of the signals as revealed by more and more specific pattern recognition [(e.g., (1) an animal, (2) a bear, (3) a grizzly bear with cubs, (4) the mother bear has not yet seen us].

4. *Motor System and Visual System.* A pure visionary typically assumes that the connection to the motor system is made only after the scene is fully elaborated. His idea is that the decision centers make a decision about what to do on the basis of the best and most complete representation of the external world. An interactive visionary, by contrast, will suggest that motor assembling begins on the basis of preliminary and minimal analysis. Some motor decisions, such as eye movements, head movements, and keeping the rest of the body motionless, are often made on the basis of minimal analysis precisely in order to achieve an upgraded and more fully elaborated visuomotor representation. Keeping the body motionless is not doing nothing, and may be essential to getting a good view of shy prey. A very simple reflex behavior (e.g., nociceptive reflex) may be effected using rather minimal analysis, but planning a complex motor act, such as stalking a prey, may require much more. In particular, complex acts may require an antecedent "inventorying" of sensorimotor predictions: what will happen if I do a, b, and g; how should I move if the X does p, and so forth.

In computer science, pioneering work exploring the computational resources of a system whose limb and sensor movements affect the processing of visual inputs is well underway, principally in research by R. Bajcsy (1988), Dana Ballard (Ballard 1991; Ballard et al. 1992; Ballard and Whitehead, 1991; Whitehead and Ballard, 1991, Randall Beer (1990) and Rodney Brooks (1989). Other modelers have also been alerted to potential computational economies, and a more integrative approach to computer vision is the focus of a collection of papers, *Active Vision* (1993), edited by Andrew Blake and Alan Yuille.

5. *Not a Good-Old-Fashioned Hierarchy Recognition.* The recognition (including predictive, what-next recognition) in the real-world case depends on richly recurrent networks, some of which involve recognition of visuomotor patterns, such as, roughly, "this critter will make a bad smell if I chase it," "that looks like a rock but it sounds like a rattlesnake, which might bite me." Consequently, the degree to which sensory processing can usefully be described as hierarchical is moot. Rich recurrence, especially with continuing multicortical area input to the thalamus and to motor structures, appears to challenge the conventional conception of a chiefly unidirectional, low-to-high processing hierarchy. Of course, temporally distinct stages between the time photons strike the retina and the time the behavior begins do exist. There are, as well, stages in the sense of different synaptic distances from the sensory periphery and the motor periphery. Our aim is not, therefore, to gainsay stages per se, but only to challenge the more theoretically emburdened notion of a strict *hierarchy.* No obvious replacement term for "hierarchy" suggests itself, and a new set of concepts adequate to describing interactive systems is needed. (Approaching the same issues, but from the perspective of neuropsychology, Antonio Damasio also explores related ideas [see Damasio 1989 b,d]).

6. *Memory and Vision.* Rich recurrence in network processing also means that stored information from earlier learning plays a role in what the animal literally sees. A previous encounter with a porcupine makes a difference to how a dog sees the object on the next encounter. A neuroscientist and a rancher do not see the same thing in figure 2.2. The neuroscientist cannot

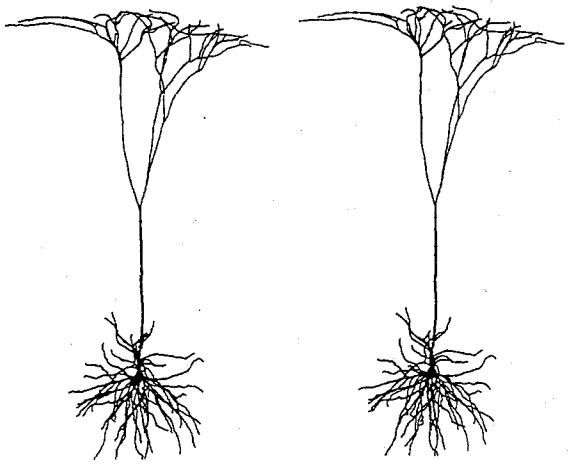


Figure 2.2 Stereo pair of a reconstructed layer five pyramidal neuron from cat visual cortex (courtsey of Rodney Douglas). The apical dendrite extends through the upper layers of the cortex and has an extensive arborization in layer 1. This neuron can be fused by placing a sheet of cardboard between the two images and between your two eyes. Look "through" the figure to diverge your eyes sufficiently to bring the two images into register. The basal dendrites, which receive a majority of the synapses onto the cell, fill a ball in three-dimensional space. Apical dendritic tufts form clusters.

help but see it as a neuron; the rancher wonders if it might be a kind of insect. A sheep rancher looking over his flock recognizes patterns, such as a ewe with lambing troubles, to which the neuroscientist is utterly blind. The latency for fusing a Julesz random-dot stereogram is much shorter with practice, even on the second try. Some learning probably takes place even in very early stages.

7. *Pragmatics of Research.* In studying nervous systems, it seems reasonable to try to isolate and understand component systems before trying to see how the component system integrates with other brain functions. Nevertheless, if the visual system is intimately and multifariously integrated with other functions, including motor control, approaching vision from the perspective of sensorimotor representation and computations may be strategically unavoidable. Like the study of "pure blood" or "pure digestion," the study of "pure vision" may take us only so far.

Our perspective is rooted in neuroscience (see also Jeannerod and Decety 1990). We shall mainly focus on three broad questions: (1) Is there empirical plausibility—chiefly, neurobiological and psychological plausibility—to the interactive perception approach? (2) What clues are available from the nervous system to tell us how to develop the interactive framework beyond its nascent stages? and (3) What computational advantage would such an interactive approach have over traditional computational approaches? Under this aegis, we shall raise issues concerning possible reinterpretation of existing neurobiological data, and concerning the implications for the problem of learning in nervous systems. Emerging from this exploration is a general direction for thinking about interactive vision.

IS PERCEPTION INTERACTIVE?

Visual Psychophysics

In the following subsections, we briefly discuss various psychophysical experiments that incline us to favor the interactive framework. In general, these experiments tend to show that whatever stages of processing are really involved in vision, the idea of a largely straightforward hierarchy from "early processes" (detection of lines, shape from shading, stereo) to "later processes" (pattern recognition) is at odds with the data (see also Ramachandran 1986; Nakayama and Shimojo 1992; Zijang and Nakayama 1992).

Are There Global Influences on Local Computation? Subjective Motion Experiments Seeing a moving object requires that the visual system solve the problem of determining which features of the earlier presentation go with which features of the later presentation (also known as the Correspondence Problem). In his work in computer vision, Ullman (1979) proposed a solution to this problem that avoids global constraints and relies only on local information. His algorithm solves the problem by trying out all possible matches and through successive iterations it finds the set of matches that yields the minimum total distance. A computer given certain correspondence tasks and running Ullman's algorithm will perform the task. His results show that the problem can be solved locally, and insofar it is an important demonstration of possibility. To understand how biological visual systems really solve the problem, we need to discover experimentally whether global factors play a role in the system's perceptions. In the examples discussed in this section, "global" refers to broad regions of the visual field as opposed to "local," meaning very small regions such as the receptive fields of cells in the parafoveal region of V1 ($\sim 1°$) or V4 ($\sim 5°$).

1. *Bistable Quartets.* The displays shown in figure 2.3 are produced on a television screen in fast alternation—the first array of dots (A: coded as filled), then the second array of dots (B: coded as open), then A then B, as in a moving picture. The brain matches the two dots in A with dots in B, and subjects see the dots moving from A position to B position. Subjects see either horizontal movement or vertical movement; they do not see diagonal movement. The display is designed to be ambiguous, in that for any given A dot, there is both a horizontal B dot and also a vertical B dot, to which it could correspond. Although the probability is 0.5 of seeing any given A–B pair oscillating in a given direction, in fact observers always see the set of dots moving as a group—they all move vertically or all move horizontally (Ramachandran and Anstis, 1983). Normal observers do not see a mixture of some horizontal and some vertical movements. This phenomenon is an instance of the more general class of effects known as motion capture, and

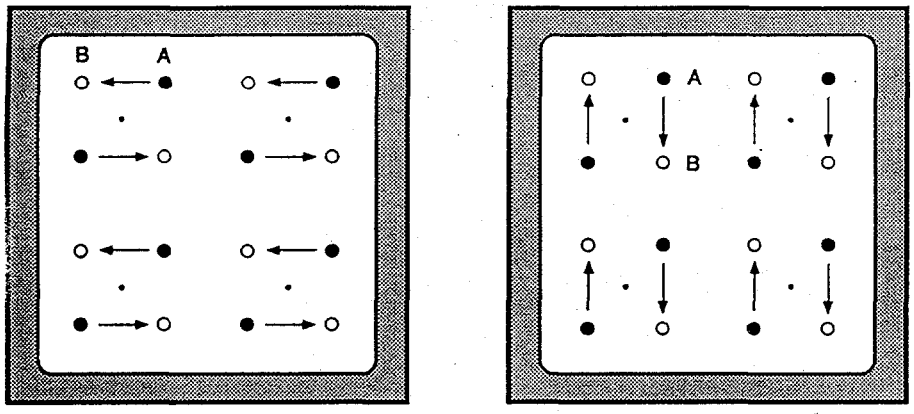


Figure 2.3 Bistable quartets. This figure shows that when the first array of dots (represented by filled circles, and indicated by A in the top left quartet) alternate with the second array of dots (represented by open circles, indicated by B in the top left quartet). Subjects see either all vertical or all horizontal oscillations. Normal observers do not see a mixture of some horizontal and some vertical movements, nor do they see diagonal movement. (Based on Ramachandran and Anstis 1983)

it strongly suggests that global considerations are relevant to the brain's strategy for dealing with the correspondence problem. Otherwise, one would expect to see, at least some of the time, a mix of horizontal and vertical movements.

2. *Behind the Occluder* (figure 2.4). Suppose both the A frame and the B frame contain a shaded square on the righthand side. Now, if all dots in the A group blink off and only the uppermost and lowermost dots of the B1 group blink on, subjects see all A dots, move to the B1 location, including the middle A dot, which is seen to move behind the "virtual" occluder. (It works just as well if the occluder occupies upper or lower positions.) If, however, A contains only one dot in the middle position on the left plus the occluding square on its right, when that single dot merely blinks off, subjects do not see the dot move behind the occluder. They see a square on the right and a blinking dot on the left. Because motion behind the occluder is seen in the context of surrounding subjective motion but not in the context of the single dot, this betokens the relevance of surrounding subjective motion to subjective motion of a single spot. Again, this suggests that the global properties of the scene are important in determining whether subjects see a moving dot or a stationary blinking light (Ramachandran and Anstis 1986).

3. *Cross-Modal Interactions*. Suppose the display consist of a single blinking dot and a shaded square (behind which the moving dot could "hide"). As before, A and B are alternately presented—first A (dot plus occluder), then B (occluder only), then A, then B. As noted above, the subject sees no motion (figure 2.4 III). Now, however, change conditions by adding an auditory stimulus presented by earphones. More exactly, the change is this:

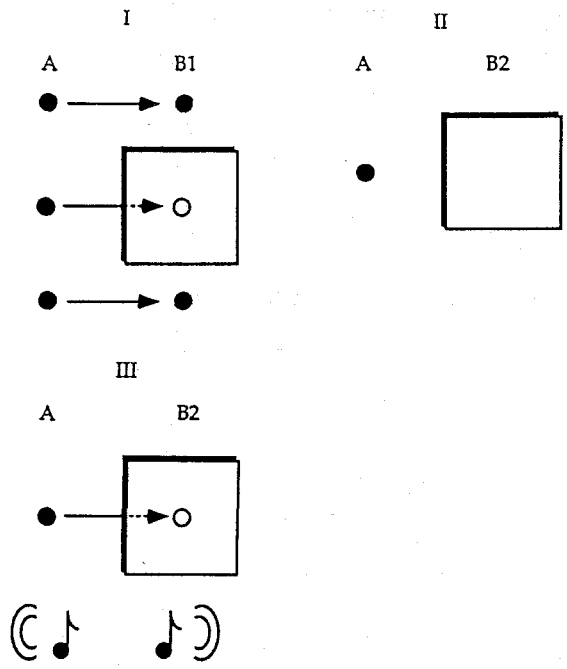


Figure 2.4 This figure shows the stimuli used to elicit the phenomenon of illusory motion behind an occluder. When the occluder is present, the subjects perceive all the dots move to the right, including the middle left dot, which is seen to move to the right and behind the square. In the absence of the occluder, the middle dot appears to move to the upper right. When the display is changed so that only the middle dot remains while upper and lower dots are removed, the middle dot is seen to merely blink off and on, but not to move behind the occluder. When, however, a tone is presented in the left ear simultaneously with the dot coming on, and in the right ear simultaneously with the dot going off, subjects do see the single dot move behind the occluder. (Based on Ramachandran and Anstis 1986)

Simultaneous with the blinking on of the light, a tone is sounded in the left ear; simultaneous with the blinking off, a tone is sounded in the right ear. With the addition of the auditory stimulus, subjects do indeed see the single dot move to the right behind the occluder. In effect, the sound "pulls" the dot in the direction in which the sound moves (Ramachandran, Intriligator, and Cavanaugh, unpublished observations). In this experiment, the cross-modal influence on what is seen is especially convincing evidence for some form of interactive vision as opposed to a pure, straight through, noninteractive hierarchy. (A weak subjective motion effect can be achieved when the blinking of the light is accompanied by somatosensory left–right vibration stimulation to the hands. Other variations on this condition could be tried.)

It comes as no surprise that visual and auditory information is integrated at some stage in neural processing. After all, we see dogs barking and drummers drumming. What is surprising in these results is that the

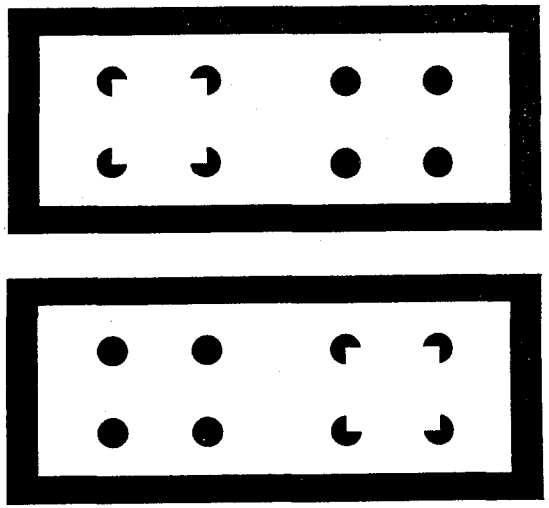


Figure 2.5 Two frames in an apparent motion display. The four Pacmen give rise to the perception of an occluding square that moves from the left circles to the right circles.

auditory stimulus has an effect on a process (motion correspondence) that pure vision orthodoxy considers "early." In this context it is appropriate to mention also influence in the other direction—of vision on hearing. Seeing the speaker's lips move has a significant effect on auditory perception and has been especially well documented in the McGurk effect.

4. *Motion Correspondence and the Role of Image Segmentation.* Figure 2.5 shows two frames of a movie in which the first frame has four Pacmen on the left, and the second has four Pacmen on the right. In the movie, the frames are alternated, and the disks are in perfect registration from one frame to the next. What observers report seeing is a foreground opaque square shifting left and right, occluding and revealing the four black disks in the background. Subjects never report seeing pacmen opening and closing their mouths; they never report seeing illusory squares flashing off and on. Moreover, when a template of this movie was then projected on a regular grid of dots, the dots inside the subjective square appeared to move with the illusory surface even though they were physically stationary (figure 2.6). "Outside" dots did not move (Ramachandran 1985).

These experiments imply that the human visual system does not always solve the correspondence problem independently of the segmentation problem (the problem of what features are parts belonging to the same thing), though pure visionaries tend to expect that solving segmentation is a late process that kicks in after the correspondence problem is solved. Subjects' overwhelming preference for the "occluding square" interpretation over the "yapping Pacmen" interpretation indicates that the solution to the

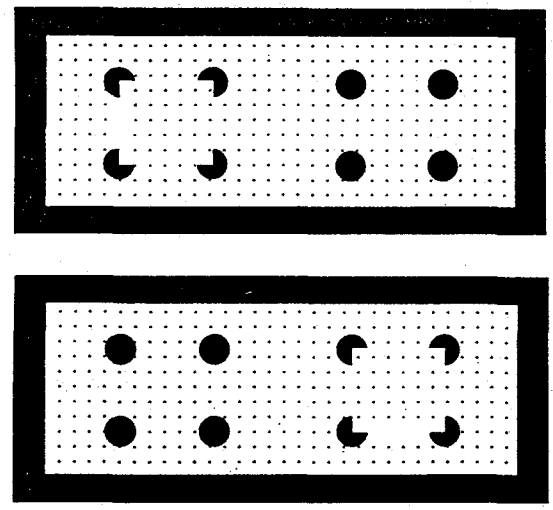


Figure 2.6 When dots are added to the background of figure 2.5, those dots internal to the occluding square appear to move with it when it occludes the right side circles. The background dots, however, appear stationary. (Based on Ramachandran 1985)

segmentation problem itself involves large-scale effects dominating over local constraints. If seeing motion in this experiment depended on solving the correspondence problem at the local level, then presumably yapping Pacmen would be seen. The experiment indicates that what are matched between frames are the larger scale and salient features; the smaller scale features are pulled along with the global decision.

Are the foregoing examples really significant? A poo-pooing strategy may downplay the effects as minor departures ("biology will be biology"). To be sure, a theory can always accommodate any given "anomaly" by making some corrective adjustment or other. Nevertheless, as anomalies accumulate, what passed as corrective adjustments may come to be deplored as ad hoc theory-savers. A phenomenon is an anomaly only relative to a background theory, and if the history of science teaches us anything, it is that one theory's anomaly is another theory's prototypical case. Thus "retrograde motion" of the planets was an anomaly for geocentric cosmologists but a typical instance for Galileo; the perihelion advance of Mercury was an anomaly for Newtonian physics, but a typical instance for Einsteinian physics. Any single anomaly on its own may not be enough to switch investment to a new theoretical framework. The cumulative effect of an assortment of anomalies, however, is another matter.

Can Semantic Categorization Affect Shape-from-Shading? Helmholtz observed that a hollow mask presented from the "inside" (the concave view, with the nose extending away from the observer) about 2 m from

the observer is invariably perceived as a convex mask with features protruding (nose coming toward the observer). In more recent experiments, Ramachandran (1988) found that the concave mask continues to be seen as convex even when is it is illuminated from below, a condition that often suffices to reverse a perception of convexity to one of concavity. This remains true even when the subject is informed about the direction of illumination of the mask. Perceptual persistence of the convex mask as a concave mask shows a strong top-down effect on an allegedly early visual task, namely determining shape from shading.

Does this perceptual reversal of the hollow mask result from a generic assumption that many objects of interest (nuts, rocks, berries, fists, breasts) are usually convex or that faces in particular are typically convex? That is, does the categorization of the image as a face override the shading cues such that the reversal is a very strong effect? To address this question, Ramachandran, Gregory, and Maddock (unpublished observations) presented subjects with two masks: one is right side up and the other is upside down. Upside-down faces are often poorly analyzed with respect to features, and an upside-down mask may not be seen as having facial features at all. In any case, upright faces are what we normally encounter. In the experiment, subjects walk slowly backward away from the pair of stimuli, starting at 0.5 m, moving to 5.0 m. At a close distance of about 0.5 m subjects correctly see both inverted masks as inverted (concave). At about 1 m, subjects usually see the upright mask as convex; the upside-down mask, however, is still seen as concave until viewing distance is about 1.5–2.0 m, whereupon subjects tend to see it too as convex. The stimuli are identical save for orientation, yet one is seen as concave and the other as convex. Hence this experiment convincingly illustrates that an allegedly "later" process (face categorization) has an effect on an allegedly "earlier" process (the shading predicts thus and such curvatures) (figure 2.7).

Can Subjective Contours Affect Stereoscopic Depth Perception? Stereo vision has been cited (Poggio et al. 1985) as an early vision task, one that is accomplished by an autonomous module prior to solving segmentation and classification. That we can fuse Julesz random dot stereograms to see figures in depth is evidence for the idea that matching for stereo can be accomplished with matching of local features only, independently of global properties devolving from segmentation or categorization decisions. While the Julesz stereogram is indeed a stunning phenomenon, the correspondence problem it presents is entirely atypical of the correspondence problem in the real world. The logical point here should be spelled out: "Not always dependent on *a*" does not imply "Not standardly dependent on *a*," let alone, "Never dependent on *a*." Hence the question remains whether in typical real world conditions, stereo vision might in fact make use of top-down, global information. To determine whether under some conditions the segmentation data might be used in solving the correspondence problem, Ramachandran (1986) designed stereo pairs

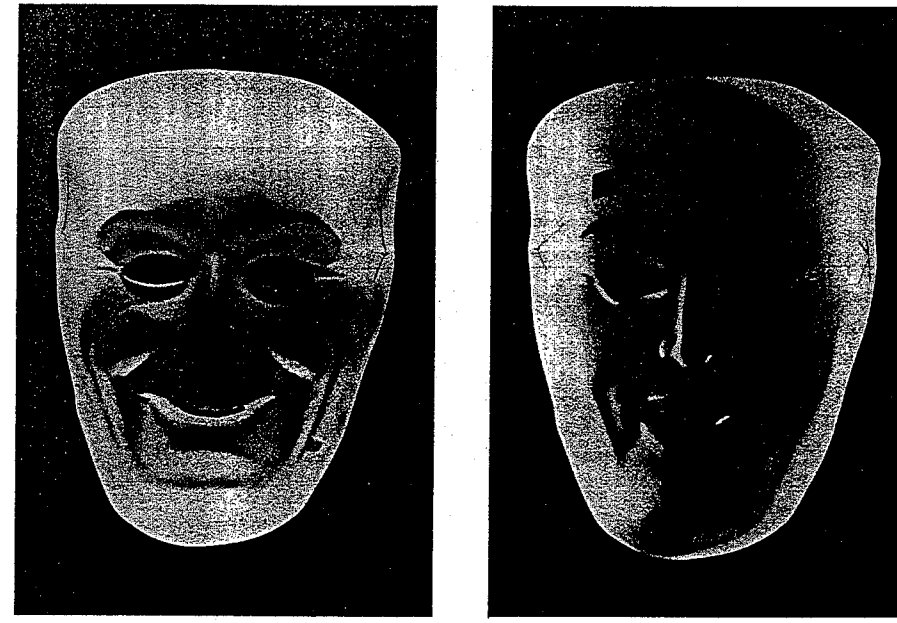


Figure 2.7 The hollow mask, photographed from its concave orientation (as though you are about to put it on). In (*A*) the light comes from above; in (*B*) light comes from both sides.

where the feature that must be matched to see stereoptic depth is some high-level property. The choice was subjective contours, allegedly the result of "later" processing (figure 2.8).

In the monocular viewing condition, illusory contours can be seen in any of the four displays (above). The top pair can be stereoscopically fused so that one sees a striped square standing well in front of a background consisting of black circles on a striped mat. The bottom pair can also be stereoptically fused. Here one sees four holes in the striped foreground mat, and through the holes, well behind the striped mat, one sees a partially occluded striped square on a black background. These are especially surprising results, because the stripes of the perceived foreground and the perceived background are at zero disparity. The only disparity that exists on which the brain can base a stereo depth perception comes from the subjective contour.

According to pure vision orthodoxy, perceiving subjective contours is a "later" effect requiring global integration, in contrast to finding stereo correspondences for depth, which is considered an "earlier" effect. This result, however, appears to be an example of "later" influencing—in fact enabling—"earlier." It should also be emphasized that the emergence of qualitatively different percepts (lined square in front of disks versus lined square behind portholes) cannot be accounted for by any existing stereo algorithms that standardly predict a reversal in sign of perceived depth only

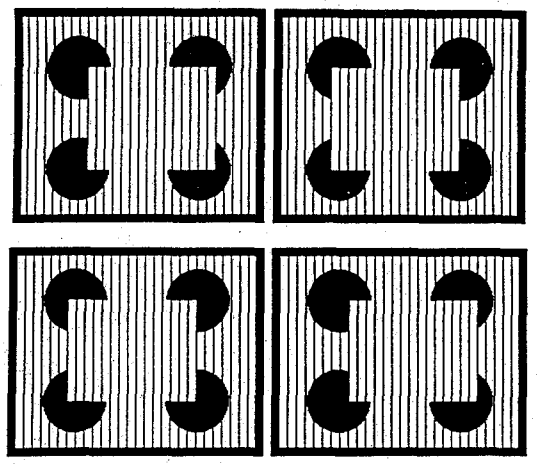


Figure 2.8 By fusing the upper stereo pairs, one sees a striped square standing well fore of a background consisting of black circles on a striped mat. By fusing the bottom pair, one sees four holes in the striped foreground mat, and through the holes, a partially occluded striped square on a black background. (This assumes fusion by divergence. The opposite order is available to those who fuse by convergence.) In both cases, the stripes are at zero disparity. (Based on Ramachandran 1986)

if the disparities are reversed. At the risk of repetition, we note again that in figure 2.8 (top and bottom), the lines are at zero disparity (Ramachandran 1986; Nakayama and Shimojo 1992).

Can Shape Recognition Affect Figure-Ground Relationships? Figure-ground identification is generally thought to precede shape recognition, but recent experiments using the Rubin vase/faces stimulus demonstrate that shape recognition can contribute to the identification of figure-ground (Peterson and Gibson, 1991).

Does the discovery of cells in V1 and V2 that respond to subjective contours (see below, p. 45) mean that detecting subjective contours is an early achievement after all? Not necessarily. The known physiological facts are consistent both with the "early effects" possibility as well as with a "later effect backsignaled" possibility. Further neurobiological and modeling experiments will help answer which possibility is realized in the nervous system.

Visual Attention

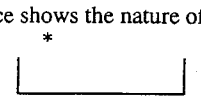
An hypothesis of interactive vision claims that the brain probably does not create and maintain a visual world representation that corresponds detail-by-detail to the visual world itself. For one thing, it need not, since the world itself is highly stable and conveniently "out there" to be sampled and resampled. On any given fixation, the brain can well make do with a partially elaborated representation of the world (O'Regan 1992; Ballard 1991; Dennett 1992). As O'Regan (1992) puts it, "the visual environment functions as a sort of outside memory store."

For another thing, as some data presented below suggest, the brain probably does not create and maintain a picture-perfect world representation. We conjecture that the undeniable feeling of having whole scene visual representation is the result mainly of (1) repeated visual visits to stimuli in the scene, (2) short-term semantic memory on the order of a few seconds that maintains the general sense of what is going on without creating and maintaining the point-by-point detail, (3) the brain's "objectification" of sensory perception such that a signal processed in cortex is represented as being about an object in space, i.e., feeling a burn on the hand, seeing a skunk in the grass, hearing a train approaching from the north, etc., and (4) the predictive dimension of pattern recognition, i.e., recognizing something as a burning log involves recognizing that it will burn my hand if I touch it, that smokey smells are produced, that water will quench the fire, that sand will smother it, that meat tastes better when browned on it, that the fire will go out after a while, and so on and on.

Evidence supporting the "partial-representation per glimpse" or "semi-world" hypothesis derives from research using on-line computer control to change what is visible on a computer display as a function of the subject's eye movements. When major display changes are made during saccades, those changes are rarely noticed, even when they involve bold alterations of color of whole objects, or when the changes consist in removal, shifting about, or addition of objects such as cars, hats, trees, and people (McConkie 1990). The exception is when the subject is explicitly paying attention to a certain feature, watching for a change.

Many careful studies using text-reading tasks elegantly support the "partial-representation per glimpse" hypothesis. These studies use a "moving window paradigm" in which subjects read a line of text that contains a window of normal text surrounded fore and aft by "junk" text. As readers move their eyes along the line, the window moves with the eyes (McConkie and Rayner 1975; Rayner et al. 1980; O'Regan 1990) (figure 2.9). The strategy is to discover the spatial extent of the zone from which useful information is extracted on a given fixation by varying the size of the window and testing using reading rate and comprehension measures. This zone is called the "perceptual" or "attentional" span. If at a given window width reading rate or comphrehension declines from a reader's baseline, it is presumed that surrounding junk text has affected reading, and hence that reader's attentional span is wider than the size of the window. By finding the smallest width at which reading is unaffected, a reader's attentional span can be quite precisely calibrated.

In typical subjects, reading text the size you are now reading, the attentional span is about 17–18 characters in width, and it is asymmetric about the point of fixation, with about 2–3 characters to the left of fixation and about 15 characters to the right. On the other hand, should you be reading



Maximum Perceptual Span

2-3 character spaces left (beginning of current word).

15 character spaces right (2 words beyond current word).

Figure 2.9 The attentional ("perceptual") span is defined as that zone from which useful information can be extracted on a given fixation. Fixation point is indicated by an asterisk. This displays the width of the attentional span and the asymmetry of the span (Courtesy John Henderson)

Hebrew instead of English, and hence traveling from page right to page left, the attention span will be about 2–3 characters to the right and 15 to the left (Pollatsek et al 1981), or reading Japanese, in which case it is asymmetric in the vertical dimension (Osaka and Oda 1991). This means that subjects read as well when junk text surround the 17–18 character span as when the whole line is visible, but read less well if the window is narrowed to 14 or 12 characters. At 17–18 character window width, the surrounding junk text is simply never noticed. Interestingly, it remains entirely unnoticed even when the reading subject is told that the moving window paradigm is running (McConkie 1979; O'Regan 1990; Henderson 1992).

Further experiments using this paradigm indicate that a shift in visual attention precedes saccadic eye movement to a particular location, presumably guiding it to a location that low-level analysis deems the next pretty good landing spot (Henderson et al. 1989). Henderson (1993) proposes that visual attention binds; inter alia, it binds the visual stimulus to a spatial location to enable a visuo-motor representation that guides the next motor response. When the fovea has landed, some features are seen.

Experiments along very different lines suggest that the information capacity of attention per glimpse is too small to contain a richly detailed whole-scene icon. Verghese and Pelli (1992) report results concerning the amount of information an observer's attention can handle. Based on their results, they conclude that the capacity of the attention mechanism is limited to about 44 ± 15 bits per glimpse. Preattentive mechanisms (studied by Treisman and by Julesz) presumably operate first, and operate in parallel. Verghese and Pelli calculated that the preattentive information capacity is much greater—about 2106 bits. The attentional mechanism, in contrast to the preattentive mechanism, they believe to be low capacity. (Verghese and Pelli define a preattentive task as "one in which the probability of detecting the target is independent of the number of distracter elements" and an attentive task as one in which "the probability of detecting the target is inversely proportional to the number of elements in the display" [p. 983].)

Verghese and Pelli ran two subjects on a number of attention tasks of varying difficulty, and compared results across tasks. In a paradigm they call "finding the dead fly," subjects are required to detect the single stationary spot among moving spots. The complementary task of finding the live fly—the moving object among stationary objects—is a preattentive task in which the target "pops out." They note that their calculation of 44 ± 15 bits is consistent with Sperling's (1960) estimate of 40 bits for the iconic store. In Sperling's technique, an array of letters was flashed to the observer. He found that subjects could report only part of the display, roughly 9 letters (= 41 bits).

There are important dependencies between visual attention, visual perception, and iconic memory. To a first approximation: (1) if you are not visually attending to *a* then you do not see *a* (have a visual experience of *a*), and (2) if you are not attending to *a* and you do not have a visual experience of *a*, then you do not have iconic memory for *a*. Given the limited capacity of visual attention, these assumptions imply that the informational capacity of visual perception (in the rough and ready sense of "literally seeing") is approximately as small (see also Crick and Koch 1990b).

Nevertheless, some motor behavior—and goal-directed eye movement in particular—apparently does not require conscious perception of the item to which the movement is directed, but does require some attentional scanning and some parafoveal signals that presumably provide coarse, easy to extract visual cues. During reading a saccade often "lands" the fovea near the third letter of the word (close enough to the "optimal" viewing position of the word), and small correction saccades are made when this is not satisfactory. This implies that the eyes are aiming at a target, and hence that at least crude visual processing has guided the saccade (McConkie et al. 1988; Rayner et al. 1983; Kapoula 1984).

In concluding this section, we emphatically note that what we have discussed here is only a small part of the story since, as Schall (1991) points out, orienting to a stimulus often involves more than eye movements. It often also involves head and whole body movements.

Considerations from Neuroanatomy

The received wisdom concerning visual processing envisages information flows from stage to stage in the hierarchy until it reaches the highest stage, at which point the brain has a fully elaborated world model, ready for motor consideration. In this section, we shall draw attention to, though not fully discuss, some connectivity that is consistent with a loose, interactive hierarchy but casts doubt on the notion of a strict hierarchy. We do of course acknowledge that so far these data provide only suggestive signs that the interactive framework is preferable. (For related ideas based on back-projection data in the context of neuropsychological data, see Damasio 1989b and Van Hoesen 1993.)

Backprojections (Corticocortical) Typically in monkeys, forward axon projections (from regions closer in synaptic distance to the sensory periphery to regions more synaptically distant; e.g., V2 to V4) are equivalent to or outnumbered by projections back (Rockland and Pandya 1979; Rockland and Virga 1989; Van Essen and Maunsell 1983; Van Essen and Anderson 1990). The reciprocity of many of these projections (P to Q and Q to P) has been documented in many areas, including connections back to the LGN (figure 2.10). It has begun to emerge that some backprojections, however, are not merely reciprocating feedforward connections, but appear to be widely distributed, including distribution to some areas from which they do not receive projections. Thus Rockland reports (1992a,b) injection data showing that some axons from area TE do indeed project reciprocally to V4, but sparser projections were also seen to V2 (mostly layer 1, but some in 2 and 5) and V1 (layer 1). These TE axons originated mainly in layers 6 and 3a (figure 2.11).

Diffuse Ascending Systems In addition to the inputs that pass through the thalamus to the cortex, there are a number of afferent systems that arise in small nuclei located in the brainstem and basal forebrain. These systems include the locus coeruleus, whose noradrenergic axons course widely throughout the cortical mantle, the serotonergic raphe nuclei, the ventral tegmental area, which sends dopamine projections to the frontal cortex, and cholinergic inputs emanating from various nuclei, including the nucleus basalis of Meynert. These systems are important for arousal, for they control the transition from sleep to wakefulness. They also provide the cortex with information about the reward value (dopamine) and salience (noradrenaline) of sensory stimuli. Another cortical input arises from the amygdala, which conveys information about the affective value of sensory stimuli to the cortex, primarily to the upper layers. Possible computational utility for these diffuse ascending system will be presented later.

Corticothalamic Connections Sensory inputs from the specific modalities project from the thalamus to the middle layers (mainly layer 4) of the cortex. Reciprocal connections from each cortical area, mainly originating in deep layers, project back to the thalamus. In visual cortex of the cat it is

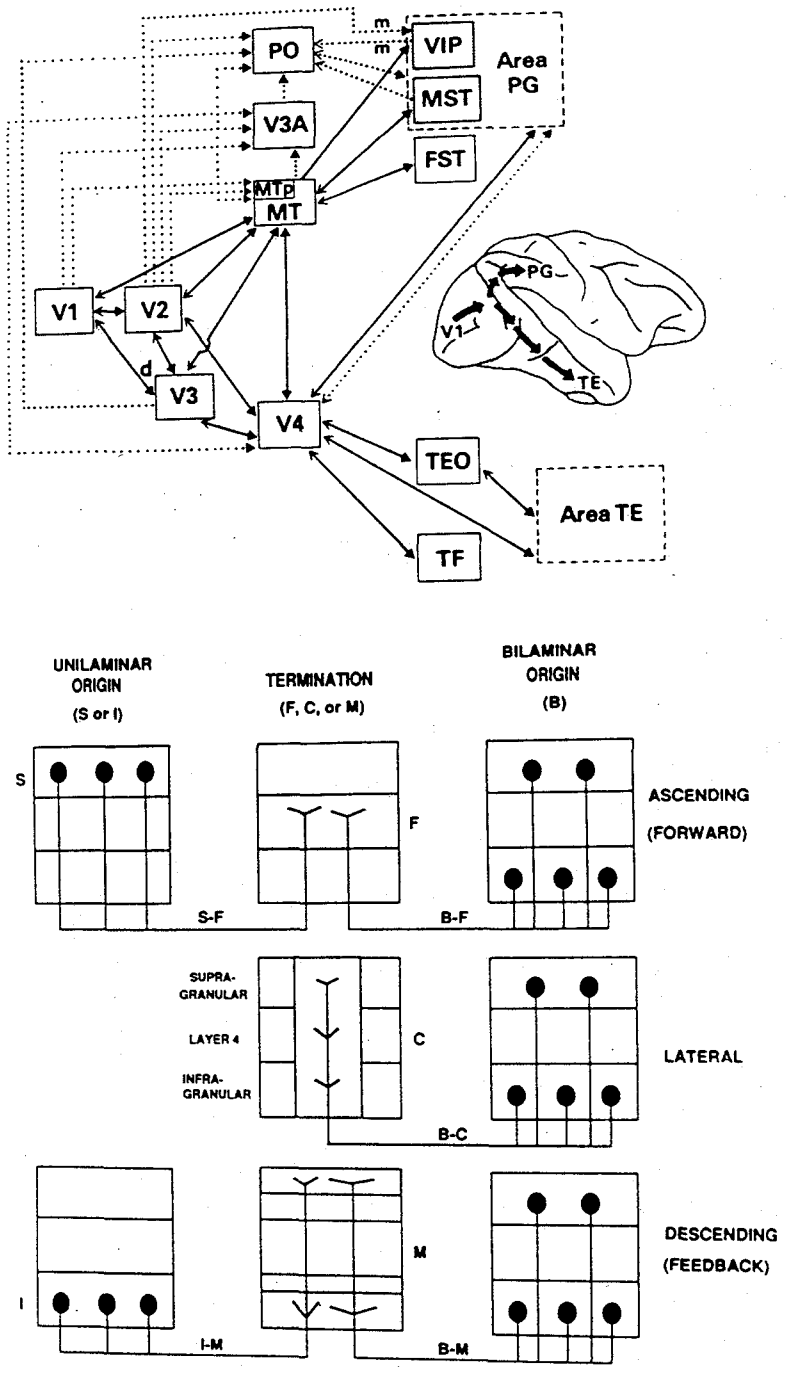


Figure 2.10 (*Top*) Schematic diagram of some of the cortical visual areas and their connections in the macaque monkey. Solid lines indicate projections involving all portions of the visual field representation in an area; dotted lines indicate projections limited to the representation of the peripheral field. Heavy arrowheads indicate forward projections; light arrowheads indicate backward projections. (Reprinted with permission from Desimone and Ungerleider 1989) (*Bottom*) Laminar patterns of cortical connectivity used for making "forward" and "backward" assignments. Three characteristic patterns of termination are indicated in the central column. These include preferential termination in layer 4 (the F pattern), a columnar (C) pattern involving approximately equal density of termination in all layers, and a multilaminar (M) pattern that preferentially avoids layer 4. There are also characteristic patterns for cells of origin in different pathways. Filled ovals, cells bodies; angles, axon terminals. (Reprinted with permission from Felleman and Van Essen 1991)

known that the V1 projections back to the LGN of the thalamus outnumber thalamocortical projections by about 10:1.

Corticofugal projections have collaterals in the reticular nucleus of the thalamus. The reticular nucleus of the thalamus is a sheet of inhibitory neurons, reminiscent of the skin of a peach. Both corticothalamic axons as well as thalamocortical projection neurons have excitatory connections on these inhibitory neurons whose output is primarily back to the thalamus. The precise function of the reticular nucleus remains to be discovered, but it does have a central role in organizing sleep rhythms, such as spindling and delta waves in deep sleep (Steriade et al. 1993b).

Connections from Visual Cortical Areas to Motor Structures Twenty-five cortical areas (cat) project to the superior colliculus (SC) (Harting et al. 1992). These include areas 17, 18, 19, 20a, 20b, 21a, and 21b. Harting et al. (1992) found that the corticotectal projection areas 17 and 18 terminate exclusively in the superficial layers, while the remaining 23 areas terminate more promiscuously (figure 2.12). The SC has an important role in directing saccadic eye movements, and, in animals with orientable ears, ear movements.

Nearly every area of mammalian cortex has some projections to the striatum, with some topological preservation. Although the functions of the striatum are not well understood, the correlation between striatal lesions and severe motor impairments is well known, and it is likely that the striatum has an important role in integrating sequences of movements. Lesion studies also indicate that some parts of the striatum are relevant to producing voluntary eye movements, as opposed to sensory-driven or reflex eye movements. It appears that the striatum can veto some reflexive responses via an inhibitory effect on motor structures, whereas voluntary movements are facilitated by disinhibitory striatal output to motor structures.

What is frustrating about this assembly of data, as with neuroanatomy generally, is that we do not really know what it all means. The number of neurons and connections is bewildering, and the significance of projections to one place or another, of distinct cell populations, and so on, is typically puzzling. (See Young 1992 for a useful startegy for clarifying the significance.) Neuroanatomy is, nonetheless, the observational hard-

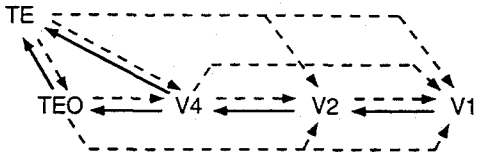


Figure 2.11 Schematic diagram of the feedforward connections (solid lines) and backprojections (broken lines) in the monkey. What is especially striking is that fibers from visual cortical areas TE (inferior temporal cortex) and TEO (posterior to TE and anterior to V4) project all the way back to V2 and V1. (Based on Rockland et al. 1992.)

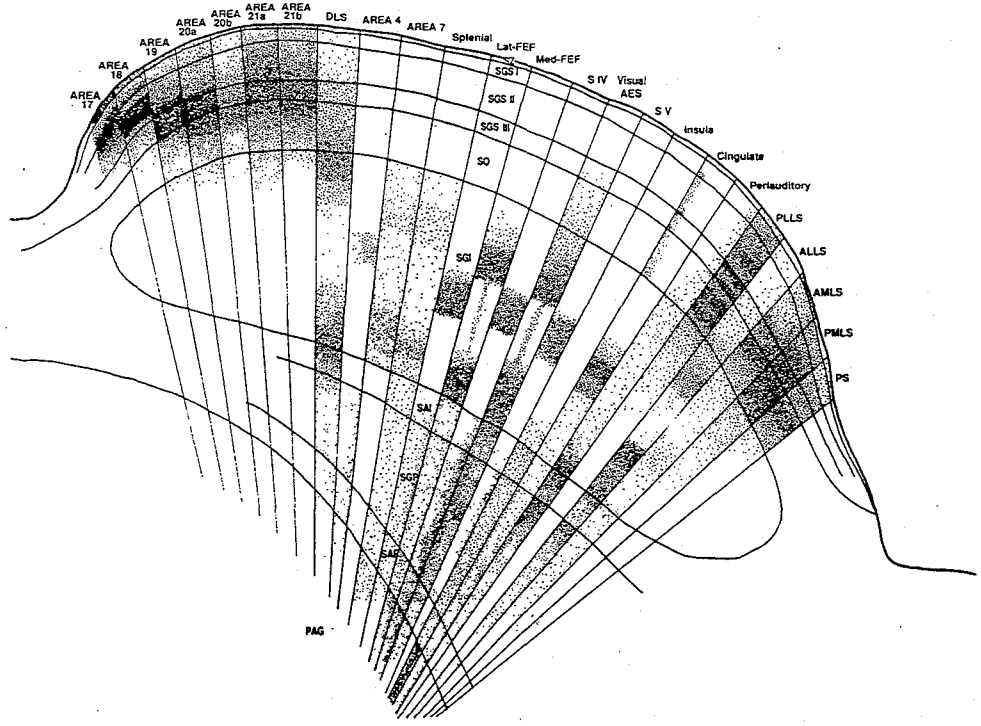


Figure 2.12 Summary diagram in the sagittal plane of the superior colliculus (SC) showing the laminar and sublaminar distribution of axons from cortical areas to the SC in the cat, as labeled above each sector. (Reprinted with permission from Harting et al. 1992)

pan for neuroscience, and the data can be provocative even when they are not self-explanatory. The prevalence and systematic character of feedback loops are particularly provocative, at least because such loops signify that the system is dynamic—that it has time-dependent properties. Output loops back to affect new inputs, and it is possible for a higher areas to affect inputs of lower areas. The time delays will matter enormously in determining what capacities the system display.

The second point is that all cortical visual areas, from the lowest to the highest, have numerous projections to lower brain centers, including motor-relevant areas such as the striatum, superior colliculus, and cerebellum. The anatomy is consistent with the idea that motor assembly can begin even before sensory signals reach the highest levels. Especially for skilled actions performed in a familiar context, such as reading aloud, shooting a basket, and hunting prey, this seems reasonable. Are the only movements at issue here eye movements? Probably not. Distinguishing gaze-related movements from extra-gaze movements is anything but straightforward, for the eyes are in the head, and the head is attached to the rest of the body. Foveating an object, for example, may well involve

movement of the eyes, head, and neck—and on occasion, the entire body. Watching Michael Jordan play basketball or a group of ravens steal a caribou corpse from a wolf tends to underscore the integrated, whole-body character of visuomotor coordination.

Considerations from Neurophysiology

In keeping with the foregoing section, this section is suggestive rather than definitive. It is also a bit of a fact salad, since at this stage the evidence does not fit together into a tight story of how interactive vision works. Such unity as does exist is the result of to the data's constituting evidence for various interactions between so-called "higher" and "lower" stages of the visual system, and between the visual and nonvisual systems. (see also Goldman-Rakic 1988; Van Hoesen 1993.)

Connections from Motor Structures to Visual Cortex Belying the assumption that the representation of the visual scene is innocent of nonvisual information, certain physiological data show interactive effects even at very early stages of visual cortex. For example, the spontaneous activity of V1 neurons is suppressed according to the onset time of saccades. The suppression begins about 20–30 msec after the saccade is initiated, and lasts about 200 msec (Duffy and Burchfield 1975). The suppression can be accomplished only by using oculomotor signals, perhaps efference copy, and hence this effect supports the interactive hypothesis. Neurons sensitive to eye position have been found in the LGN (Lal and Friedlander 1989), visual cortical area V1 (Trotter et al. 1992; Weyand and Malpeli 1989), and V3 (Galleti and Battaglini 1989). Given the existence and causal efficacy of various nonvisual V1 signals, Pouget et al. (1993) hypothesized that visual features are encoded in egocentric (spatiotopic) coordinates at early stages of visual processing, and that eye-position information is used in computing where in egocentric space the stimulus is located. Their network model demonstrates the feasibility of such a computation when the network takes as input both retinal and eye-position signals.

Consider also that a few V1 cells and a higher percentage of V2 cells show an enhanced response to a target to which a saccade is about to be made (Wurtz and Mohler 1976). Again, these data indicate some influence of motor system signals, specifically motor planning signals, on cells in early visual processing. As further evidence, note that some neurons in V3A show variable response as a function of the angle of gaze; response was enhanced when gaze was directed to the contralateral hemifield (Galletti and Battaglini 1989).

Inferior Parietal Cortex and Eye Position Caudal inferior parietal cortex (IPL) has two major subdivisions: LIP and 7a (figure 2.13). LIP is directly connected to the superior colliculus, the frontal eye fields. Area 7a has a different connectivity: mainly polymodal cortex, limbic, and some prestriate.

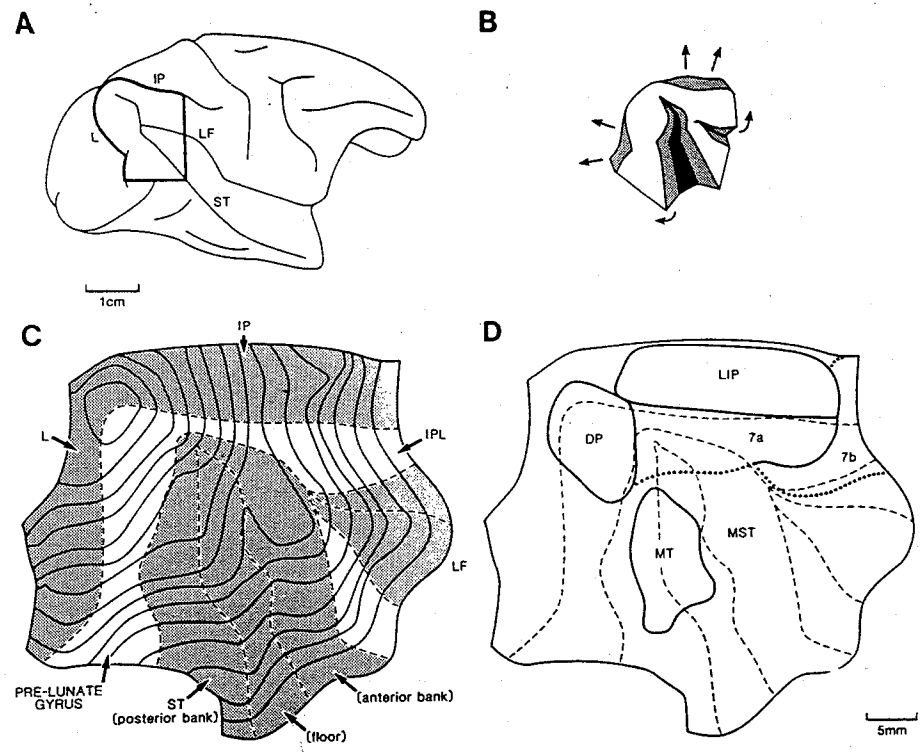


Figure 2.13 Parcellation of inferior parietal lobule and adjoining dorsal aspect of the prelunate gyrus. The cortical areas are represented on flattened reconstructions of the cortex. (*A*) Lateral view of the monkey hemisphere. The darker line indicates the area to be flattened. (*B*) The same cortex isolated from the rest of the brain. The stippled areas are cortex buried in sulci, and the blackened area is the floor of the superior temporal sulcus. The arrows indicate movement of local cortical regions resulting from mechanical flattening. (*C*) The completely flattened representation of the same area. The stippled areas represent cortical regions buried in sulci and the contourlike lines are tracings of layer IV taken from frontal sections through this area. (*D*) Locations of several of the cortical fields. The dotted lines indicate borders of cortical fields that are not precisely determinable. IPL, inferior parietal lobule; IPS, intraparietal sulcus; LIP, lateral intraparietal region; STS, superior temporal sulcus. (Reprinted with permission from Andersen 1987).

LIP responses are correlated with execution of saccadic eye movements; area 7a cells respond to a stimulus at a certain retinal location, but modulated by the position of the eye in the head (Zipser and Andersen 1988). Hardy and Lynch (1992) report that both LIP and area 7a receive the majority of their thalamic inputs from distinct patches in the medial pulvinar nucleus (figure 2.14).

Illusory Contours and Figure Ground In 1984 von der Heydt et al. reported that neurons in visual area V2 of the macaque will respond to il-

case 1

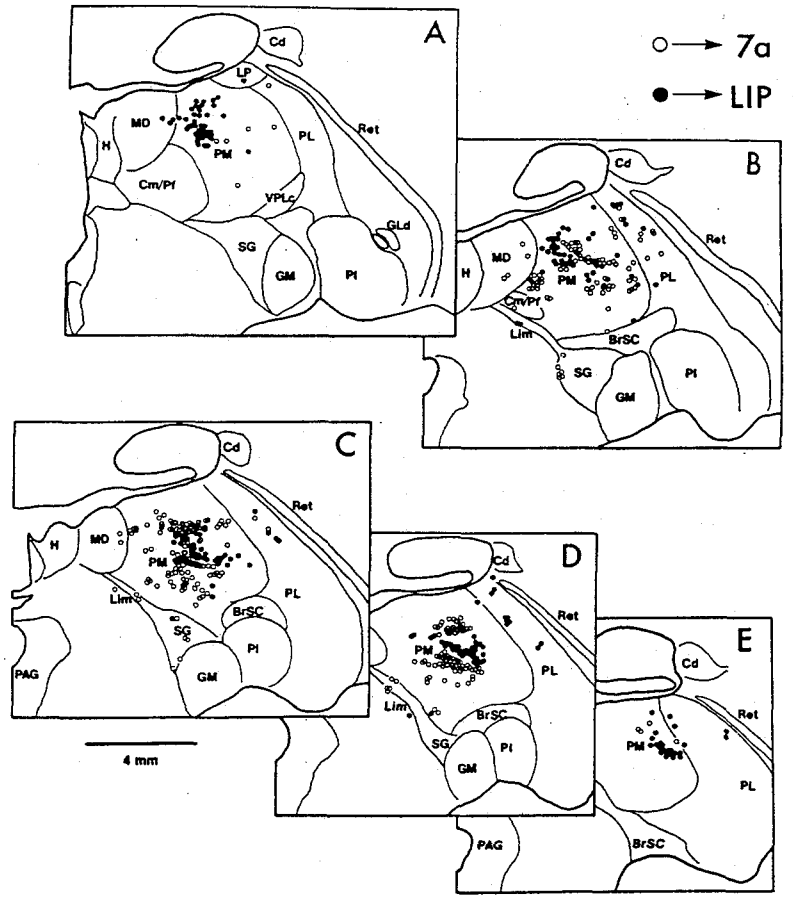


Figure 2.14 Diagram of sections of the thalamus showing the distribution of retrogradely labeled neurons within the thalamus resulting from inferior parietal lobule (IPL) injections in parietal areas 7a (open circles) and lateral intraparietal region (LIP) (filled circles). A single injection (1.2 μ) of the fluorescent dye DY (diamidino-dihydrochloride yellow) was made in 7a and a single injection (0.5 μ) of fast blue in LIP. Each individual symbol denotes a singled labeled neuron. The densest labeling is in medial pulvinar nucleus of the thalamus (PM), with LIP and 7a showing a distinct projection pattern. BrSC, brachium of the superior colliculus; Cd, caudate nucleus; Cm/Pf, centromedian and parafascicular nuclei; GM, medial geniculate nucleus, pars parvicellularis; H, habenula; Lim, nucleus limitans; MD, mediodorsal nucleus; PAG, periaqueductal gray; PL, lateral pulvinar nucleus; Ret, reticular nucleus. (Reprinted with permission from Hardy and Lynch 1992)

lusory contours. More recently, Grosof et al. (1992) report that some orientation selective cells in V1 respond to a class of illusory contours. As noted earlier, these data are consistent with the possibility that low-level response depends on higher level operations whose results are backprojected to lower levels. This is lent plausibility by the facts that detection of illusory contours will depend on previous operations involving interpolation across a span of the visual field.

Neurons in area MT respond selectively to direction of motion but not to wavelength. Nonetheless, color can have a major effect on how these cells respond by virtue of how the visual stimulus is segmented (Dobkins and Albright 1993).

Cross-Modal Interactions The responses of cells in V4 to a visual stimulus can be modified by somatosensory stimuli (Maunsell et al. 1991). Fuster (1990) has shown similar task-dependent modifications for cells in somatosensory cortex, area S1.

Dynamic Mapping in Exotropia

In this section we discuss an ophthalmic phenomenon observed in human subjects. Conventionally, this is a truly surprising phenomenon, and it seems to demonstrate that processing as early as V1 can be influenced by top-down factors. The phenomenon has not been well studied, to say the least, and much more investigation is required. Nevertheless, we mention it here partly because it is intriguing, but mainly because if the description below is accurate, then we must rethink the Pure Vision's conventional assumptions about the Receptive Field.

Exotropia is a form of squint in which both eyes are used when fixated on small objects close by (e.g., 12 in from the nose) but when looking at distant objects, the *squinting* eye deviates outward by as much as 45° to 60°. Curiously, the patient does not experience double vision—the deviating eye's image is usually assumed to be *suppressed*. It is not clear, however, at what stage of visual processing the suppression occurs.

Ophthalmologists have claimed that, contrary to expectations, in a small subset of these patients, *fusion* occurs not only during inspection of near objects, but also when the squinting eye deviates. This phenomenon, called *anomalous retinal correspondence* or ARC, has been reported frequently in the literature of ophthalmology and orthoptics. The accuracy of the reports, and hence the existence of ARC, has not always been taken seriously, since ARC implies a rather breathtaking lability of receptive fields. Clinicians and physiologists raised in the Hubel–Wiesel tradition usually take it as basic background fact that (1) binocular connections are largely established in area 17 in early infancy and that (2) binocular *fusion* is based exclusively on anatomical correspondence of inputs in area 17. For instance, if a squint is surgically induced in a kitten or an infant monkey, area 17 displays a complete loss of binocular cells (and two populations of monocular cells) but the maps of the two eyes never change. No apparent compensation such as *anomalous correspondence* has been observed in area 17 and this has given rise to the conviction that it is highly improbable that an ARC phenomenon truly exists.

On the possibility that there might be more to the ARC reports, Ramachandran, Cobb, and Valente (unpublished) recently studied two patients who had intermittent exotropia. These patients appeared to fuse images both during near vision and during far vision—when the left eye deviated outward—a condition called *intermittent exotropia with anomalous correspondence.* To determine whether the patients do indeed have two (or more) separate binocular *maps* of the world, Ramachandran, Cobb, and Valenti devised a procedure that queried the alignment of the subject's afterimages where the afterimage for the right eye was generated independently of the afterimage for the left eye. Here is the procedure: (1) The subject (with squint) was asked to shut one eye and to fixate on the bottom of a vertical slit-shaped window mounted on a flashgun. A flash was delivered to generate a vivid monocular afterimage of the slit. The subject was then asked to shut this eye and view the top of the slit with the other eye (and a second flash was delivered). (2) The subject opened both eyes and viewed a dark screen, which provided a uniform background for the two afterimages.

The results were as follows: (1) The subject (with squint) reported seeing afterimages of the two slits that were perfectly lined up with each other, so long as the subject was deliberately verging within about arm's length. (2) On the other hand, if the subject relaxed vergence and looked at a distant wall (such that the left eye deviated), the upper slit (from the anomalous eye) vividly appeared to move continuously outward so that the two slits became misaligned by several degrees. Then this experiment was repeated on two normal control subjects and it was found that no misalignment of the slits occurred for any ordinary vergence or conjugate eye movements. Nor could misalignments of the slits be produced by passively displacing one eyeball in the normal individuals to mimic the exotropia. It appears that eye position signals from the deviating eye selectively influence the egocentric localization of points for that eye alone.

In the next experiment, a light point was flashed for 150 msec either to the right eye alone or the left eye alone; the subject's task was merely to point to the location of the light point. Subjects became quite skilled at deviating their anomalous eye by between about 1° and 40°, and the afterimage alignment technique could be used to calibrate the deviation. Tests were made with deviations between 1° and 15°. It was found that regardless of the degree of deviation of the anomalous eye, and regardless of which eye was stimulated, subjects made only marginally more errors than normal subjects in locating the light point. Is the remapping sufficiently fine-grained to support stereopsis? Testing for accuracy of stereoptic judgments using ordinary stereograms under conditions of anomalous eye deviations between 1° and 12°, Ramachandran, Cobb, and Valenti found that disparities as small as 20 min of arc could be perceived correctly even though the anomalous eye deviated by as much as 12°. Even when the half-images of the two eyes were exciting noncorresponding retinal points separated by 12°, very small retinal disparities could be detected.

Ramachandran and his colleagues have dubbed this phenomenon *dynamic anomalous correspondence.* Their results suggest that something in the ARC reports is genuine, with a number of implications.

First, binocular correspondence can change continuously in *real time* in a single individual depending on the degree of exotropia. Hence, binocular correspondence (and *fusion*) cannot be based exclusively on the anatomical convergence of inputs in area 17. The relative displacement observed between the two afterimages also implies that the *local sign* of retinal points (and therefore binocular correspondence) must be continuously updated as the eye deviates outward.

Second, since the two slits would always be *lined up* as far as area 17 is concerned, the observed misalignment implies that feedback (or feedforward) signals from the deviating eye must somehow be extracted separately for each eye and must then influence the egocentric location of points selectively for that eye alone. This is a somewhat surprising result, for it implies that time *remapping* of egocentric space must be done very early—before the *eye of origin* label is lost—i.e., before the cells become completely binocular. Since most cells anterior to area 18 (e.g., MT or V4) are symmetrically binocular we may conclude that the correction must involve interaction between reafference signals and the output of cells as early as 17 or 18.

Nothing in the psychophysical results suggests what the mechanism might be by which these interactions occur. Whatever the ultimate explanation, however, the results do imply that even as simple a perceptual process as the localization of an object in X/Y coordinates is not strictly and absolutely a *bottom-up* process. Even the output of early visual elements—in this case the monocular cells of area 17—can be strongly modulated by back projections from eye movement command centers.

If indeed a complete remapping of perceptual space can occur selectively for one eye's image simply in the interest of preserving binocular correspondence, this is a rather remarkable phenomenon. It would be interesting to see if this remapping process can be achieved by algorithms of the type proposed by Zipser and Andersen (1988) for parietal neurons or by *shifter-circuits* of the kind proposed by Anderson and Van Essen (1987; see also chapter 13).

COMPUTATIONAL ADVANTAGES OF INTERACTIVE VISION

So far we have discussed various empirical data that lend some credibility to an interactive-vision approach. But the further question is this: Does it make sense computationally for a nervous system to have an interactive style rather than a hierarchical, modular, modality-pure, and motorically unadulterated organization? In this section, we briefly note four reasons, based on the computational capacities of neural net models, why evolution might have selected the interactive modus operandi in nervous systems. As more computer models in the interactive style are developed and explored, additional factors, for or against, may emerge. The results from

neural net models also suggest experiments that could be run on real nervous systems to reveal whether they are in fact computationally interactive.

Figure-Ground Segmentation and Recognition Are More Efficiently Achieved in Tandem Than Strictly Sequentially Segmentation is a difficult task, especially when there are many objects in a scene partially occluding one another. The problem is essentially that global information is needed to make decisions at the local level concerning what goes with what. At lower levels of processing such as V1, however, the receptive fields are relatively small and it is not possible locally to decide which pieces of the image belong together. If lower levels can use information that is available at higher levels, such as representation of whole objects, then feedback connections could be used to help tune lower levels of processing. This may sound like a chicken-and-egg proposal, for how can you recognize an object before you segment it from its background? Just as the right answer to the problem "where does the egg come from" is "an earlier kind of chicken," so here the the answer is "use partial segmentation to help recognize, and use partial recognition to help segment." Indeed, interactive segmentation–recognition may enable solutions that would otherwise be unreachable in short times by pure bottom-up processing.

It is worth considering the performance of machine reading of numerals. The best of the "pure vision" configured machines can read numerals on credit card forms only about 60% of the time. They do this well only because the sales slip "exactifies" the data: numerals must be written in blue boxes. This serves to separate the numerals, guarantee an exact location, and narrowly limit the size. Carver Mead (in conversation) has pointed out that the problem of efficient machine reading of zip codes is essentially unsolved, because the preprocessing regimentations for numeral entry on sales slips do not exist in the mail world. Here the machine readers have to face the localization problem (where are the numerals and in what order?) and the segmentation problem (what does a squiggle belong to?) as well as the recognition problem (is it a 0 or a 6?).

Conventional machines typically serialize the problem, addressing first the segmentation problem and then, after that is accomplished, addressing the recognition problem. Should the machine missolve or fail to solve the first, the second is doomed. In the absence of strict standardization of location, font, size, relation to other numerals, relation of zip code to other lines, and so forth, classical machines regularly fumble the segmentation problem. Unlike engineers working with the strictly serial problem-design, Carver Mead and Federico Faggin (in conversation) have found that if networks can address segmentation and recognition in parallel, they well outperform their serial competitors.

The processing of visual motion is another example of how segregation may proceed in parallel with visual integration. Consider the problem of trying to track a bird flying through branches of a tree; at any moment only parts of the bird are visible through the occluding foliage, which may itself be moving. The problem is to identify fleeting parts of the bird that may be combined to estimate the average velocity of the bird and to keep this information separate from information about the tree. This is a global problem in that no small patch of the visual field contains enough information to unambiguously solve the segregation problem. However, area MT of the primate visual cortex has neurons that seem to have "solved" this segregation problem. A recent model of area MT that includes two parallel streams, one that selects regions of the visual field that contain reliable motion information, and another that integrates information from that region, exhibits properties similar to those observed in area MT neurons (Nowlan and Sejnowski 1993). This model demonstrates that segmentation and integration can to some extent be performed in parallel at early stages of visual processing.

It would not be surprising if evolution found the interactive strategy good for brains. So long as the segmentation problem is partially solved, a good answer can be dumped out of the visual "pipeline" very quickly. When, however, the task is more difficult, iterations and feedback may be essential to drumming up an adequate solution. To speed up processing in the difficult cases—which will be the rule, not the exception, in real-world vision—the system may avail itself of learning. If, after frequent encounters, the brain learns that certain patterns typically go together, thereafter the number of iterations needed to find an adequate solution is reduced (Sejnowski 1986). Humans probably "overlearn" letter and word patterns, and hence seasoned readers are faster and more accurate than novice readers. Even when text is degraded or partially occluded, a good reader may hardly stumble.

Movement (of Eye, Head, Body) Makes Many Visual Computations Simpler A number of reasons support this point. First, the smooth pursuit system for tracking slowly moving objects supports image stability on the retina, simplifying the tasks of analyzing and recognizing. Second, head movement during eye fixation yields cues useful in the task of separating figure from ground and distinguishing one object from another. Motion parallax (the relative displacement of objects caused by change in observer position) is perhaps the most powerful cue to the relative depth of objects (closer objects have greater relative motion than more distant objects), and it continues to be critical for relative depth judgments even beyond about 10 m from the observer, where stereopsis fades out. Head bobbing is common behavior in animals, and a visual system that integrates across several glimpses to estimate depth has computational savings over one that tries to calculate depth from a single snapshot.

Another important cue is optical flow (figure 2.15). When an animal is running, flying, or swimming, for example, the speed of an image moving radially on the retina is related to the distance of the object from the observer.[4] This information allows the system to figure out how fast it is closing in on a chased object, as well as how fast a chasing object is closing

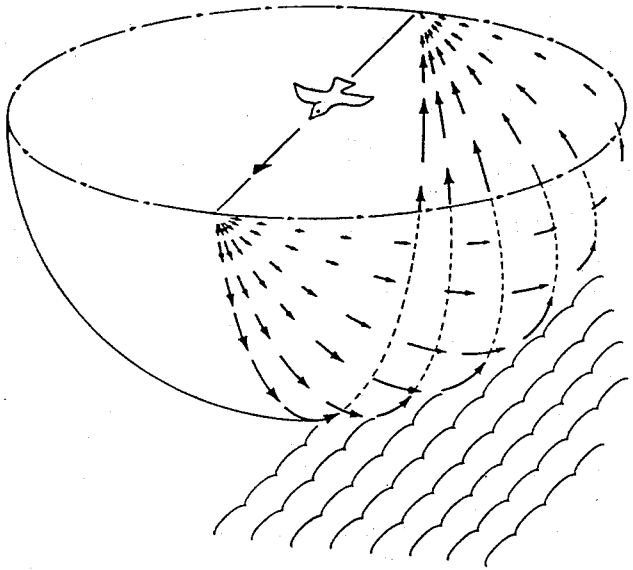


Figure 2.15 Optical flow represented by a vector field around a flying bird provides information about self-movement through the environment. (From Gibson 1966)

in. Notice that any of these movements (eyeball, head, and whole body) on its own means computational economies. In combination, the economies compound.

There are many more example of how the self-generated movement can provide solutions to otherwise intractable problems in vision (Ballard 1991; Blake and Yuille 1992).

The Self-Organization of Model Visual Systems during "Development" Is Enhanced by Eye-Position Signals An additional advantage of interactive vision is its role in the construction of vision systems. Researchers in computer vision often reckon—and bemoan—the cost of "hand" building vision systems, but rarely consider the possibility of growing a visual system. Nature, of course, uses the growing strategy, and relies on genetic instructions to create neurons with the right set of components. In addition, interactions between neurons as well as interactions between the world and neurons, are critical in getting networks of neurons properly wired up. Understanding the development of the brain is perhaps as challenging a problem as that of understanding the function of the brain, but we are beginning to figure out some of the relevant factors, such as position cues, timing of gene expression, and activity-dependent modifications. Genetic programming has been explored as an approach to solving some construction problems, but value of development as an an intermediary between genes and phenotype is only beginning to be appreciated in the computational community (Belew 1993).

Most activity-dependent models of development are based on the Hebb rule for synaptic plasticity, according to which the synapse strengthens when the presynaptic activity is correlated with the postsynaptic activity (Sejnowski and Tesauro 1989). Correlation-based models for self-organization of primary visual cortex during development have shown that some properties of cortical cells, such as ocularity, orientation, and disparity, can emerge from simple Hebbian mechanisms for synaptic plasticity (Swindale 1990; Linsker 1986; Miller and Stryker 1990; Berns et al. 1993). Hebbian schemes are typically limited in their computational power to finding the principal components in the input correlations. It has been difficult to extend this approach to a hierarchy of increasingly higher-order response properties, as found in the extrastriate areas of visual cortex. One new approach is based on the observation that development takes place in stages. There are critical periods during which synapses are particularly plastic (Rauschecker 1991), and there are major milestones, such as eye opening, that change the nature of the input correlations (Berns et al. 1993).

Nature exploits additional mechanisms in the developing brain to help organize visual pathways. One important class of mechanisms is based on the interaction between self-generated actions and perception, along the lines already discussed in the previous section. Eye movement information in the visual cortex during development, when combined with Hebbian plasticity, may be capable of extracting higher-order correlations from complex visual inputs. The correlation between eye movements and changes in the image contains information about important visual properties. For example, correlation of saccadic eye movements with the response of a neuron can be used in a Hebbian framework to develop neurons that respond to the direction of motion. At still higher levels of processing, eye movement signals that direct saccades to salient objects can be used as a reward signal to build up representations of significant objects (Montague et al. 1993). This new view provides eye-movement signals with an important function in visual cortex both during development and in the adult.

The plasticity of the visual cortex during the critical period is modulated by inputs from subcortical structures that project diffusely throughout the cortex (Rauschecker 1991). The neurotransmitters used by these systems often diffuse from the release sites and act at receptors on neurons some distance away. These diffuse ascending systems to the cerebral cortex that are used during development to help wire up the brain are also used in the adult for signaling reward and salience. The information carried by neurons in these systems is rather limited: there are relatively few neurons compared to the number in the cortex, they have a low basal firing rate, and changes in their firing rates occur slowly. This is, however, just the sort of information that could be used to organize and regulate information storage throughout the brain, as shown below.

Interactive Perception Simplifies the Learning Problem A difficulty facing conventional reinforcement learning is this: Assuming the brain creates and maintains a picture-perfect visual scene at each moment, how does the brain determine which, among the many features and objects it recently perceived, are the ones relevant to the reward or punishment? An experienced animal will have a pretty good idea, but how does its experience get it to that stage? How does the naive brain determine which "stimulus" in the richly detailed stimulus array gets the main credit when a certain response brings a reward? How does the brain know what synapses to strengthen?

This "relevance problem" is even more vexing as there are increases in the time delay between the stimulus and the reinforcement. For then the stimulus array develops over time, getting richer and richer as time passes. The correlative problem of knowing which movement among many movements made was the relevant one is likewise increasingly difficult as the delay increases between the onset of various movements and the rewarding or punishing outcome.[5] These questions involve considerations that go well beyond the visual system, and include parts of the brain that evaluate sensory inputs.

Suppose that evolution has wired the brain to bias attention as a function of how the species makes its living, and that the neonate is tuned to attend to some basic survival-relevant properties. The evolutionary point legitimates the assumption that an attended feature of the stimulus scene is more likely to be causally implicated in producing the conditions for the reward, and assuming that the items in iconic and working memory are, by and large, items previously attended to, then the number of candidate representations to canvas as "relevant" is far smaller than those embellishing a rich-replica visual world representation. Granted all these assumptions, the credit assignment problem is far more manageable here than in the pure vision theoretical framework (Ballard 1991).

By narrowing the number of visuomotor trajectories that count as salient, attention can bias the choice of synapses strengthen. Selective strengthening of synapses of certain visuomotor representations "spotlit" by attention is a kind of hypothesis the network makes. It is, moreover, an hypothesis the network tests by repeating the visuomotor trajectory. Initially the network will shift attention more or less randomly, save for guidance from startle responses and other reflexive behavior. Given that attention downsizes the options, and that the organism can repeatedly explore the various options, the system learns to direct attention to visual targets that it has learned are "good bets" in the survival game. This, in turn, contributes to further simplifying the learning problem in the future, for on the next encounter, attention will more likely be paid to relevant features than to irrelevant features, and the connections can be up-regulated or down- regulated as a result of reward or lack of same (the above points are from Whitehead and Ballard 1990, 1991; see also Grossberg 1987).

A robust property of animal learning is that responses reinforced by a reward are likely to be produced again when relevantly similar conditions obtain. This is the starting point for behavioral studies of operant conditioning in psychology (Rescorla and Wagner 1972; Mackintosh 1974), neuroscientific inquiry into the reward systems of the brain (Wise 1982) and engineering exploration of the principles and applications of reinforcement learning theory (Sutton and Barto 1981, 1987). Both neuroscience and computer engineering draw on the vast and informative psychological literature describing the various aspects of reinforcement learning, including such phenomena as blocking, extinction, intermittent versus constant reward, cue ranking, how time is linked to other cues, and so forth. The overarching aim is that the three domains of experimentation will link up and yield a unified account of the scope and limits of the capacity and of its underlying mechanisms (see Whitehead and Ballard 1990; Montague et al. 1994).

Detailed observations of animal foraging patterns under well quantified conditions indicate that animals can display remarkably sophisticated adaptive behavior. For example, birds and bees quickly adopt the most efficient foraging pattern in "two-armed bandit" conditions (a: high-payoff when a "hit" and "hits" are infrequent; b: low payoff when a "hit" and "hits" are frequent) (see Krebs et al. 1978; Gould 1984; Real et al. 1990; Real 1991). Cliff-dwelling rooks learn to bombard nest-marauders with pebbles (Griffin 1984). A bear learns that a bluff of leafy trees in a hill otherwise treed with pines means a gully with a creek, and a creek means rocks under which crawfish are often living, and that means tasty dinner.

The questions posed in the previous section concerning reinforcement learning, along with the dearth of obvious answers, have moved some cognitive psychologists (e.g., Chomsky 1965, 1980; Fodor 1981) to conclude that reinforcement learning cannot be a serious contender for the sophisticated learning typical of cognitive organisms. Further skepticism concerning reinforcement learning as a cognitive contender derives from neural net modeling. Here the results shows that neural nets trained up by available reinforcement procedures scale poorly with the number of dimensions of the input space. In other words, as a net's visual representation approximates a rich replica of the real world, the training phase becomes unrealistically long. Consequently, computer engineers often conclude that reinforcement learning is impractical for most complex task domains. According to some cognitive approaches (Fodor, Pylyshyn, etc.), suitable learning theories must be "essentially cognitive," meaning, roughly, that cognitive learning consists of logic-like transformations over language-like representations. Moreover, the theory continues, such learning is irreducible to neurobiology. (For a fuller characterization and criticism of this view, see Churchland 1986.)

By contrast, our hunch is that much cognitive learning may well turn out

to be explainable as reinforcement learning once the encompassing details of the rich-replica assumption no longer inflate the actual magnitude of the "relevance problem."

Natural selection and reinforcement learning share a certain scientific appeal; to wit, neither presupposes an intelligent homunculus, an omniscient designer, or a miraculous force—both are naturalistic, as opposed to supernaturalistic. They also share reductionist agendas. Thus, as a macrolevel phenomenon, reinforcement learning behavior is potentially explainable in terms of micromechanisms at the neuronal level. And we are encouraged to think so because the Hebbian approach to mechanisms for synaptic modification underlying reinforcement learning looks very plausible. These general considerations, in the context of the data discussed earlier, suggest that the skepticism concerning the limits of reinforcement learning should really be relocated to the background assumption—the rich-replica assumption. Consequently, the question guiding the following discussion is this: What simplifications in the learning problem can be achieved by abandoning Pure Vision's rich-replica assumption? How much mileage can we get out of the reinforcement learning paradigm if we embrace the assumption that the perceptual representations are semiworld representations consisting of, let us say, goal-relevant properties? How might that work?

Using an internal evaluation system, the brain can create predictive sequences by rewarding behavior that leads to conditions that in turn permit a further response that will produce an external reward, that is, sequences where one feature is a cue for some other event, which in turn is a cue for a further event, which is itself a cue for a reward. To get the flavor, suppose, for example, a bear cub chances on crawfish under rocks in a creek, whereupon the crawfish/rocks-in-water relationship will be strengthened. Looking under rocks in a lake produces no crawfish, so the crawfish/rocks-in-lake relationship does not get strengthened, but the crawfish/rocks-in-creek relationship does. Finding a creek in a leafy-tree gully allows internal diffusely-projecting modulatory systems, such as the dopamine system, to then reward associations between creeks and leafy-trees-in-gullies, even in the absence of a external reward between creeks and leafy-trees-in-gullies.

Given such an internal reward system, the brain can build a network replete with predictive representations that inform attention as to what is worth looking at given one's interests ("that big dead tall tree will probably have hollows in it, and there will probably be a blue-bird's nest in one cavity, and that nest might have eggs and I will get eggs to eat"). To a first approximation, a given kind of animal comes to have an internal model of its world; that is, of its relevant-to-my-life-style world, as opposed to a world-with-all-its-perceptual-properties. For bears, this means attending to creeks and dead trees when foraging, and not noticing much in anything about rocks at lake edges, or sunflowers in a meadow. All of which then makes subsequent reinforcement tasks and the delimiting of what is relevant that much easier. (To echo the school marm's saw, the more you know about the world, the better the questions you can ask of it and the faster you learn.)

A neural network model of predictive reinforcement learning in the brain roughly based on a diffuse neurotransmitter system has been applied to the adaptive behavior of foraging bumble bees (Montague et al. 1994). This is an especially promising place to test the semiworld hypothesis, for it is an example in which both the sensory input and the motor output can be quantified, the animal gets quantifiable feedback (sugar reward) , and something of the physiology of the reward system, the motor system, and the visual system in the animal's brain has been explored. Furthermore, bee foraging behavior has been carefully studied by several different research groups, and there are lots of data available to constrain a network model.

Bees decide which flowers to visit according to past success at gathering nectar, where nectar volume varies stochastically from flower to flower (Real 1991). The cognitive characterization of the bees' accomplishments involves applications of computational rules over representations of the arithmetic mean of rewards and variance in reward distributions. On the other hand, according to the Dayan–Montague reinforcement hypothesis (Montague et al. 1994), when a bee lands on a flower, the actual reward value of the nectar collected by the bee is compared (more? or less? or right on?) with the reward that its brain had predicted, and the difference is used to improve the prediction of future reward using predictive Hebbian synapses. Dayan and Montague propose that the very same predictive network is used to bias the actions of the bee in choosing flowers. Using this nonhomuncular, nondivine, naturalistic learning procedure, the model network accurately mimics the foraging behavior of real bumblebees (figure 2.16).

That such a simple, "dumb" organization can account for the apparent statistical cunning of bumblebees is encouraging, for it rewards the hunch that much more can be got out of a reinforcement learning paradigm once the "pure vision" assumption is replaced by the "interactive-vision-cum-predictive-learning" assumption. As we contemplate extending the paradigm from bees to primates, it is also encouraging that similar diffuse neurotransmitter systems are found in primates where there is evidence that some of them are involved in predicting rewards (Ljunberg et al. 1992).

Bees successfully forage, orient, fly, communicate, houseclean and so forth—and do it all with fewer than 10^6 neurons (Sejnowski and Churchland 1992). Human brains, by contrast, are thought to have upward of 10^{12} neurons. Although an impressively long evolutionary distance stretches between insects and mammals, what remains constant is the survival value of learning cues for food, cues for predators and so forth. Consequently, conservation of the diffuse, modulatory, internal reward system makes good biological sense. What is sensitive to the pressure of natural selection is additional processors that permit increasingly subtle, fine-grained, and long-range predictions—always, of course, relevant-to-my-thriving predictions. That in turn may entail making better and better classifications

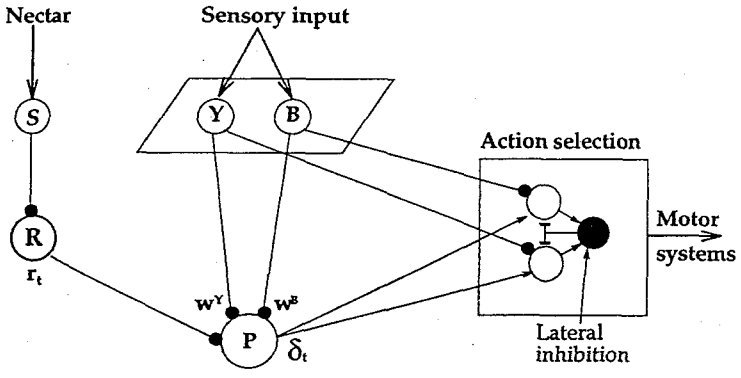


Figure 2.16 Neural architecture for a model of bee foraging. Predictions about future expected reinforcement are made in the brain using a diffuse neurotransmitter system. Sensory input drives the units B and Y representing blue and yellow flowers. These units project to a reinforcement neuron P through a set of plastic weights (filled circles w^B and w^Y) and to an action selection system. S provides input to R and fires while the bee sips the nectar. R projects its output r_t through a fixed weight to P. The plastic weights onto P implement predictions about future reward and P's output is sensitive to temporal changes in its input. The outputs of P influence learning and also the selection of actions such as steering in flight and landing. Lateral inhibition (dark circle) in the action selection layer performs a winner-takes-all. Before encountering a flower and its nectar, the output of P will reflect the temporal difference only between the sensory inputs B and Y. During an encounter with a flower and nectar, the prediction error δ_t is determined by the output of B or Y and R, and learning occurs at connections w^B and w^Y. These strengths are modified according to the correlation between presynaptic activity and the prediction error δ_t produced by neuron P. Before encountering a flower and its nectar, the output of P will reflect the temporal difference only between the sensory inputs B and Y. During an encounter with a flower and nectar, the prediction error δ_t is determined by the output of B or Y and R, and learning occurs at connections w^B and w^Y. These strengths are modified according to the correlation between presynaptic activity and the prediction error δ_t produced by neuron P. Simulations of this model account for a wide range of observations of bee preference, including aversion for risk. (From Montague et al. 1994)

(relative to the animals' lifestyle), as well as more efficient and predictively sound generalizations (relative to the animals' life-style).

To a first approximation, cortical enlargement was driven by the competitive advantage accruing to brains with fancier, good-for-me-and-my-kin predictive prowess, where the structures performing those functions would have to be knit into the reward system. Some brand new representational mechanisms may also have been added, but the increased "intelligence" commonly associated with increased size of the cortical mantle may be a function chiefly of greater predictive-goal-relevant representational power, not to greater representational power per se. Whether some property of the world is visually represented depends on the representation's utility in the predictive game, and for this to work, the cortical representational structures must be plastic and must be robustly tethered to the diffusely projecting systems. World-perfect replicas, unhitched from the basic engines of reward and punishment, are probably more of a liability than an advantage—they are likely to be time-wasters, space-wasters, and energy-wasters.

On this approach, various contextual aspects of visual perception, such as filling in, seeing the dot move behind the occluder, cross-modal effects, and plasticity in exotropia, can be understood as displaying the predictive character of cortical processing.

CONCLUDING REMARKS

A well-developed geocentric astronomy was probably an inevitable forerunner to modern astronomy. One has to start with what seems most secure and build from there. The apparent motionless of the earth, the fixity of the stars, and the retrograde motions of the planets were the accessible and seemingly secure "observations" that grounded theorizing about the nature of the heavens. Such were the first things one saw—saw as systematically and plainly as one saw anything. The geocentric hypothesis also provided a framework for the very observations that eventually caused it to be overhauled.

In something like the same way, the Theory of Pure Vision is probably essential to understanding how we see, even if, as it seems, it is a ladder we must eventually kick out from under us. The accessible connectivity suggests a hierarchy, the most accessible and salient temporal sequence is sensory input to the transducers followed by output from the muscles, the accessible response properties of single cells show simple specificities nearer the periphery and greater complexity the further from the periphery, and so on. Such are the grounding observations for a hierarchical, modular, input–output theory of how we see.

But there are nagging observations suggesting that the brain is only grossly and approximately hierarchical, that input signals from the sensory periphery are only a part of what drives "sensory" neurons, that ostensibly later processing can influence earlier processes; that motor business can influence sensory business, that processing stages are not much like assembly line productions, that connectivity is nontrivially back as well as forward, etc. Some phenomena, marginalized within the Pure Vision framework, may be accorded an important function in the context of a heterarchical, interactive, space-critical and time-critical theory of how we see. Consider, for example, spontaneous activity of neurons, so-called "noise" in neuronal activity, nonclassical receptive field properties, visual system learning, attentional bottlenecks, plasticity of receptive field properties, time-dependent properties, and backprojections.

Obviously visual systems evolved not for the achievement of sophisticated visual perception as an end in itself, but because visual perception can serve motor control, and motor control can serve vision to better serve motor control, and so on. What evolution "cares about" is who survives, and that means, basically, who excels in the four Fs: feeding, fleeing, fight-

ing, and reproducing. How to exploit that evolutionary truism to develop a theoretical framework that is, as it were, "motocentric" rather than "visuocentric" we only dimly perceive. (see also Powers 1973; Bullock et al. 1977; Llinas 1987; Llinas 1991; Churchland 1986). In any event, it may be worth trying to rethink and reinterpret many physiological and anatomical results under the auspices of the idea that perception is driven by the need to learn action sequences to be performed in space and time.

ACKNOWLEDGMENTS

We are especially indebted to Dana Ballard, Read Montague, Peter Dayan, and Alexandre Pouget for their ideas concerning reinforcement learning and vision, and for their indefatigable discussions and enthusiasm. We are grateful to Tony Damasio for valuable discussion on the role of back-projections, the memory-ladenness of perception, and the emotion dependence of planning and deciding, which have become essential elements in how we think about brain function. We are grateful to A. B. Bonds for reading an early version of this chapter, for providing wonderfully direct and invaluable criticism, and for the Bonding—an incomparable extravaganza of neuroscientific fact, engineering cunning, and Tennessee allegory. We thank also Richard Gregory for discussion about the visual system, Christof Koch and Malcolm Young for helpful criticism, and Francis Crick for criticism and valuable daily discussion. Michael Gray provided helpful assistance with figures and references. P. S. Churchland is supported by a MacArthur fellowship; V. S. Ramachandran is supported by the Air Force Office of Scientific Research and the Office of Naval Research; T. J. Sejnowski is an investigator with the Howard Hughes Medical Institute and is also supported by the Office of Naval Research.

NOTES

1. With apologies to Immanuel Kant.

2. For further research along these lines see for example, Ullman and Richards (1984), Poggio et al. (1985), and Horn (1986). For a sample of current research squarely within this tradition, see, for example, a recent issue of Pattern Analysis and Machine Intelligence.

3. Poggio et al. (1985) say: "[Early vision] processes represent conceptually independent modules that can be studied, to a first approximation, in isolation. Information from the different processes, however, has to be combined. Furthermore, different modules may interact early on. Finally, the processing cannot be purely "bottom-up": specific knowledge may trickle down to the point of influencing some of the very first steps in visual information processing" (p. 314). Although we agree that this is a step in the right direction, we shall argue that "trickle" does not begin to do justice to the cascades of interactivity.

4. These brief comments give no hint of the complexities of optic flow cues and their analysis. For discussion, see Cutting (1986).

5. In the case of food-aversion learning the delay between ingested food and nausea may be many hours.

REFERENCES

Adrian, E. D. (1935) The Mechanism of Nervous Action. Philadelphia: University of Pennsylvania Press.

Aloimonos, Y. and A. Rosenfield (1991). Computer vision. Science. 253: 1249-1254.

Andersen, R. A. (1987). The role of the inferior parietal lobule in spatial perception and visual-motor integration. In: The Handbook of Physiology. Section 1: The Nervous System, Volume IV. Higher Functions of the Brain, Part 2. Bethesda, MD: American Physiological Society. pp. 483-518.

Andersen, R. A., R. M. Siegel and G. K. Essick. (1987). Neurons of area 7 activated by both visual stimuli and oculomotor behavior. Experimental Brain Research 67: 316-322.

Arbib, Michael A. (1989) The Metaphorical Brain 2: Neural Networks and Beyond. New York, N.Y.: Wiley.

Belew, R. K. (1993) Interposing on ontogenetic model between genetic algorithms and neural networks. In: Advance in Neural Information Processing Systems 5, Ed. Hanson, S. J., Cowan, J. D., and Giles, C. L. San Mateo, California: Morgan Kaufmann Publishers. pp. 99-106.

Bajcsy, R. (1988). Active perception. Proceedings of the IEEE. 76: 996-1005.

Ballard, D. H. (1991) Animate vision. Artificial Intelligence 48: 57-86.

Ballard, D. H., M. M. Hayhoe, L. Feng, S. D. Whitehead (1992). Hand-eye co-ordination during sequential tasks. Philosophical Transactions of the Royal Society of London B. 337: 331-339.

Ballard, D. H. and Whitehead, S. D. (1990) Active perception and reinforcement learning. Neural Computation 2: 409-419.

Bartlett, R. C. (1958) Thinking. New York: Basic Books.

Beer, R. J. (1990). Intelligence as Adaptive Behavior. New York: Academic Press.

Berns et al (1993). A correlational model for the development of disparity selectivity in visual cortex that depends on prenatal and postnatal phases. Proceedings of the National Academy of Sciences USA (in press).

Blake, A. and Yuille, A. (Eds.) (1992) Active vision, Cambridge, Mass.: MIT Press.

Brooks, R. A. (1989) A robot that walks: Emergent behaviors from a carefully evolved network. Neural Computation 1: 253-262.

Bullock, T. H., Orkand, R., & Grinnell, A. (1977). Introduction to Nervous Systems. San Francisco: W. H. Freeman.

Burkhalter, A. (1993) Development of forward and feedback connections between areas V1 and V2 of human visual cortex, Cerebral Cortex 3: 476-487.

Chomsky, N. (1965). Aspects of the Theory of Syntax. Cambridge, Mass.: MIT Press.

Chomsky, N. (1980). Rules and Representations. New York: Columbia University Press.

Churchland, P. S. (1986). Neurophilosophy: Toward a unified science of the mind-brain. Cambridge, MA: MIT Press.

Cutting, J. E. (1986). Perception with an eye for motion. Cambridge, MA: MIT Press.

Damasio, A. R. (1989a). The brain binds entities and events by multiregional activity from convergence zones. Neural Computation. 1: 123-132.

Damasio, A. R. (1989b). Multiregional retroactivation. Cognition. 12: 263-288.

Dennett, D. C. (1992). Consciousness Explained. New York: Little, Brown, & Co.

Desimone, R., & Ungerleider, L. G. (1989). Neural mechanisms of visual processing in monkeys. In: Handbook of Neuropsychology, vol. 2, ed. F. Boller and J. Grafman. Amsterdam: Elsevier.

Dobkins K. R. & Albright T. D. (1993) What happens if it changes color when it moves?: Psychophysical experiments on the nature of chromatic input to motion detectors. Vision Research 33: 1019-1036.

Duffy, F. H., & Burchfield, J. L. (1975). Eye movement-related inhibition of primate visual neurons. Brain Research. 89: 121-132.

Felleman, D. J. and D. C. van Essen (1991). Distributed hierarchical processing in the primate visual cortex. Cerebral Cortex 1: 1-47.
Fodor, J. A. (1981). Representations. Cambridge, Mass.: MIT Press.
Fuster, J. (1990). Inferotemporal units in selective visual attention and short-term memory. Journal of Neurophysiology. 64: 681-697.
Galletti, C., & Battaglini, P. P. (1989). Gaze-dependent visual neurons in area V3A of monkey prestriate cortex. Journal of Neuroscience. 9: 1112-1125.
Gibson, J. J. (1966) The Senses Considered as Perceptual Systems. Boston: Houghton Mifflin.
Goldman-Rakic, P. S. (1988). Changing concepts of cortical connectivity: Parallel distributed cortical networks. In: Neurobiology of Neocortex. Ed. by P. Rakic and W. Singer. New York: Wiley and Sons. 177-202.
Gould, J. L. (1984) Natural history of honey bee learning. In: The Biology of Learning, Eds. P. Marler and H. S. Terrace, pp 149-180, Berlin: Springer-Verlag.
Griffin, D. R. (1984). Animal Thinking. Cambridge, Mass.: Harvard University Press.
Grosof, D. H., Shapley, R. M., & Hawken, M. J. (1992). Macaque striate responses to anomalous contours? Investigative Ophthalmology and Visual Science 33: 1257.
Grossberg, S. (Ed.) (1987) The Adaptive Brain I, Amsterdam: North-Holland.
Hardy, S. G. P., & Lynch, J. C. (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.
Harting, J. K., Updyke, B. V., & Lieshout, D. P. V. (1992). Corticotectal projections in the cat: Anterograde transport studies of twenty-five cortical areas. The Journal of Comparative Neurology. 324: 379-414.
Henderson, J. M. (1992). Visual attention and eye movement control during reading and picture viewing. In: Eye Movements and Visual Cognition: Scene Perception and Reading. Ed. by K. Rayner. New York: Springer-Verlag.
Henderson, J. M. (in press) Visual attention and the perception-action interface. In: Vancouver studies in cognitive science (Vol. V): Problems in perception. Oxford: Oxford University Press.
Henderson, J. M., A. Pollatsek, and K. Rayner (1989). Covert visual attention and extrafoveal information use during object identification Perception and Psychophysics. 45: 196-208.
Hohmann, C. F., K. K. Kwiterovich, M. L. Oster-Granite, and J. T. Coyle. Newborn basal forebrain lesions disrupt cortical cytodifferentiation as visualized by rapid Golgi staining. Cerebral Cortex. 1: 143-157.
Horn, K. B. P. (1986) Robot Vision. MIT Press
Jeannerod, M. and J. Decety (1990). The accuracy of visuo-motor transformation: An investigation into the mechanisms of visual recognition of objects. In: Vision and action: The control of grasping. Ed. by M. A. Goodale. Norwood, N. J. : Ablex Pub. Co.
Kapoula, Z. (1984). Aiming precision and characteristics of saccades. In: Theoretical and applied aspects of eye movement research. ed. A. G. Gale and F. Johnson. Amsterdam: North-Holland.
Krebs, J.R., Kacelnik, A. and Taylor, P. (1978) Test of optimal sampling by foraging great tits, Nature 275, 27-31.
Lal, R. and M. J. Friedlander (1989). Gating of the retinal transmission by afferent eye position and movement signals. Science, 243: 93-96.
Linsker, R. (1986). From basic network principles to neural architecture (series). Proceedings of the National Academy of Sciences USA. 83: 7508-7512, 8390-8394, 8779-8783.
Ljunberg, T., Apicella, P. and Schultz, W. (1992) Responses to monkey dopamine neurons during learning of behavioral reactions. J. Neurophysiology 67, 145-163.
Llinas, R. R. (1987). 'Mindness' as a functional state of the brain. In: Mindwaves. Ed. by C. Blakemore and S. Greenfield. Oxford: Blackwells.
Llinas, R. R. and D. Pare (1991). Of dreaming and wakefulness. Neuroscience. 44: 521-535.

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

Maunsell, J. H. R., G. Sclar, T. A. Nealey, & DePriest, D. D. (1991). Extraretinal representations in area V4 in the macaque monkey. *Visual Neuroscience*. 7: 561-573.
Mackintosh, N. J. (1974) The Psychology of Animal Learning. New York: Academic Press.
McConkie, G. W. (1979). On the role and control of eye movements in reading. In: Processing of visible language. ed. A. Kolers, M. E. Wrolstad, and H. Bouma. New York: Plenum Press.
McConkie, G. W. (1990). Where vision and cognition meet. Paper presented at the HFSP Workshop on Object and Scene Perception. Leuven, Belgium.(@@check with henderson re publication@@@@@)
McConkie, G. W., & Zola, D. (1987). Visual attention during eye fixations in reading. In: Attention and Performance XII. London: Erlbaum
McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations on words. Vision Research. 28: 1107-1118.
Miller, K. D., & Stryker, M. P. (1990). Ocular dominance column formation: mechanisms and models. In: Connectionist Modeling and Brain Function: The Developing Interface. Ed. S. J. Hanson and C. R. Olson. Cambridge, MA: MIT Press
Montague, P. R., P., Nowlan, S. J., Pouget, A. and Sejnowski, T. J. (1993) Using aperiodic reinforcement for directed self-organization during development. In: Advance in Neural Information Processing Systems 5, Hanson, S. J., Cowan, J. D., and Giles, C. L. (Eds.) San Mateo, California: Morgan Kaufmann Publishers. pp. 969-976.
Montague, P. R., Dayan, P. and Sejnowski, T. J. (1994) Foraging in an Uncertain Environment Using Predictive Hebbian Learning. In: Cowan, JD, Tesauro, G and Alspector, J, (Eds.) Advances in Neural Information Processing 6. San Mateo, CA: Morgan Kaufmann
Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. Science. 257: 1357-1362.
Nowlan, S. J. and Sejnowski, T. J. (1993) Filter selection model for generating visual motion signals. In: Advance in Neural Information Processing Systems 5, Ed. Hanson, S. J., Cowan, J. D., and Giles, C. L. San Mateo, California: Morgan Kaufmann Publishers. pp. 369-376.
O'Regan, J. K. (1992) Solving the "real" mysteries of visual perception: The world as an outside memory. Canadian Journal of Psychology.
O'Regan, J. K. (1990) Eye movements and reading. In: Eye movements and their role in visual and cognitive processes. Ed. E. Kowler. Elsevier.
Osaka, N. and K. Oda (1991). Effective visual size necessary for vertical reading during Japanese text processing. Bulletin of the Psychonomic Society. 29: 345-347.
Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. Nature. 317: 314-319.
Pollatsek, A., S. Bolozky, A. D. Well, and K. Rayner (1981). Asymmetries in perceptual span for Israeli readers. Brain and Language. 14: 174-180.
Pouget, A., S. A. Fisher, and T. J. Sejnowski. (1993). Egocentric spatial representation in early vision. Journal of Cognitive Neuroscience. 5: 150-161.
Powers, W. T. (1973). Behavior: The Control of Perception. Chicago: Aldine Pub. Co.
Ramachandran, V. S. (1985). Apparent motion of subjective surfaces. Perception. 14: 127-134.
Ramachandran, V. S. (1986). Capture of stereopsis and apparent motion by illusory contours. Perception and Psychophysics. 39: 361-373.
Ramachandran, V. S. (1988). Perception of depth from shading. Scientific American. 269: 76-83.
Ramachandran, V. S., and S. M. Anstis (1983). Perceptual organization in moving patterns. Nature. 304: 529-531.
Ramachandran, V. S., and S. M. Anstis (1986). Perception of apparent motion. Scientific American. 254: 102-109.
Ramachandran, V. S., S. Cobb, and C. Valenti (unpublished). Plasticity of binocular correspondence, stereopsis, and egocentric localization in human vision.
Rauschecker, J. P. (1991). Mechanisms of visual plasticity: Hebb synapses, NMDA receptors, and beyond. Physiological Reviews 71: 587-615.

Rayner, K., A. D. Well, and A. Pollatsek (1980). Asymmetry of the effective visual field in reading. Perceptual Psychophysics. 27: 537-544.
Rayner, K., Slowiaczek, M. L., Clifton, C., & Bertera, J. H. (1983). Latency of sequential eye movements: Implications for reading. Journal of Experimental Psychology: Human Perception and Performance. 9: 912-922.
Real. L.A. (1991) Animal choice behavior and the evolution of cognitive architecture. Science 253, 980-986.
Real, L. A., S. Ellner, and Hardy, L. D. (1990). Short-term energy maximization and risk aversion in bumblebees: A reply to Possingham. Ecology 71: 1625-1628.
Rescorla, RA and Wagner, AR (1972). A theory of Pavlovian conditioning:
The effectiveness of reinforcement and non-reinforcement. In: AH Black
and WF Prokasy, Ed., Classical Conditioning II: Current Research and Theory, New York, NY: Appleton-Century-Crofts pp 64-69.
Rockland, K. S. (1992a). Laminar distribution of neurons projecting from area V1 to V2 in macaque and squirrel monkeys. Cerebral Cortex. 2: 38-47.
Rockland, K. S. (1992b). Configuration, in serial reconstruction, of individual axons projecting from area V2 to V4 in the macaque monkey. Cerebral Cortex. 2: 353-374.
Rockland, K. S. (In press). The organization of feedback connections from area V2 (18) to V1 (17).
Rockland, K. S and D. N. Pandya (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Research.* 179: 3-20.
Rockland, K. S., Saleem, K. S., and Tanaka, K. (1992). Widespread feedback connections from areas V4 and TEO. Society for Neuroscience Abstracts.
Rockland, K. S. and Virga, A. (1990). Organization of individual cortical axons projecting from area V1 (area 17) to V2 (area 18) in the macaque monkey. *Visual Neuroscience.* 4: 11-28
Schall, J. D. (1991). Neural basis of saccadic eye movements in primates. In: Eye movements. ed. R. H. S. Carpenter. Boca Raton: CRC Press.
Sejnowski, T. J. (1986). Open questions about computation in cerebral cortex. In: Parallel Distributed Processing, Vol. 2. Ed. J. L. McClelland and D. E. Rumelhart. Cambridge, MA: MIT Press.
Sejnowski, T. J. and Churchland, P. S. (1992). Computation in the age of neuroscience. In: A New Era in Computation. Ed. by N Metropolis and G.-C. Rota. Cambridge, Mass.: MIT Press.
Sparks, D. L. and Jay, M. F. (1986) The functional organization of the primate superior colliculus: A motor perspective. Progress in Brain Research 64: 235-241.
Sperling , G. (1960). The information available in brief visual presentations. Psychological Monographs 74: 498.
Sperry, R. W. (1952). Neurology and the mind-brain problem. American Scientist 40: 291-312.
Steriade, M., McCormick, D. and Sejnowski, T.J., (1993) The sleeping and aroused brain: thalamocortical oscillations in neurons and networks. Science (in press).
Sutton, R. S., and A. G. Barto (1981). Toward a modern theory of adaptive networks: expectation and prediction. Psychological Review. 88: 135-170.
Sutton, R. S. and A. G. Barto (1987). Toward a modern theory of adaptive networks: Expectation and prediction. Proceedings of the Ninth Annual Conference of the Cognitive Science Society. Seattle WA.
Swindale, N. V. (1990). Is the cerebral cortex modular? Trends in Neurosciences. 13: 487-492.
Sejnowski, T. J. and Tesauro, G. (1989). The Hebb rule for synaptic plasticity: algorithms and implementations. In: Neural Models of Plasticity: Experimental and theoretical approaches, Ed. J. H. Byrne and W. O. Berry, San Diego, California: Academic Press. pp. 94-103.
Trotter, Y., S. Celibrini, B. Stricanne, S. Thorpe, and M. Imbert (1992). Modulation of neural stereoscopic processing in primate area V1 by the viewing distance. *Science.* 257: 1279-1281.
Tsotsos, J.K. (1987). In: Encyclopedia of Artificial Intelligence, ed. S. Shapiro. New York: Wiley. 389-409.
Ullman, S. (1979). The Interpretation of Visual Motion. Cambridge, MA: MIT Press.

Ullman, S. and W. Richards (1984). Image Understanding. Norwood, New Jersey: Ablex Publishing Corporation, 1984

Van Essen, D. and J. H. R. Maunsell (1983). Hierarchical organization and functional streams in the visual cortex. Trends in Neurosciences. 6: 370-375.

Van Essen, D. and C. H. Anderson. (1990) Information processing strategies and pathways in the primate retina in visual cortex. In: An introduction to neural and electronic networks, Ed. Zornetzer, S.F., Davis, J. L. and Lau, C. 43-72.

Van Hoesen, G. W. (1993). The modern concept of association cortex. Current Opinion in Neurobiology. 3: 150-154.

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

Weyand, T. G. and J. G. Malpeli (1989). Responses of neurons in primary visual cortex are influenced by eye position. Abstracts of the Society of Neuroscience. 15: 1016.

Whitehead, S. D. and D. H. Ballard (1990). Active perception and reinforcement learning. Neural Computation. 2: 409-419.

Whitehead, S. D., and D. Ballard (1991). Connectionist designs on planning. In: Neural Information Processing Systems 3, Ed. by R. P. Lippmann,

J. E. Moody, and D. Touretsky, San Mateo, CA: Morgan Kaufmann pp. 357-370.

Wise, R. A. (1982). Neuroleptics and operant behavior: The anhedonia hypothesis. Behavioral and Brain Sciences. 5: 39-87.

Wurtz, R. H., & Mohler, C. W. (1976). Enhancement of visual response in monkey striate cortex and frontal eye fields. Journal of Neurophysiology. 39: 766-772.

Young, M. P. (1993). The organization of neural systems in the primate cerebral cortex. Proceedings of the Royal Society: Biological Sciences. 252: 13-18

Yuille, A. L. and S. Ullman (1990). Computational theories of low-level vision. In: Visual Cognition and Action. Ed. by D. N. Osherson, S. M. Kosslyn, and J. M. Hollerbach. Cambridge, Mass.: MIT Press.

Zijang, J. H., & Nakayama, K. (1992). Surfaces vs. features in visual search. Nature. 359: 231-233.

Zipser, D. and Andersen, R. A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. Nature, 331, 679-684.