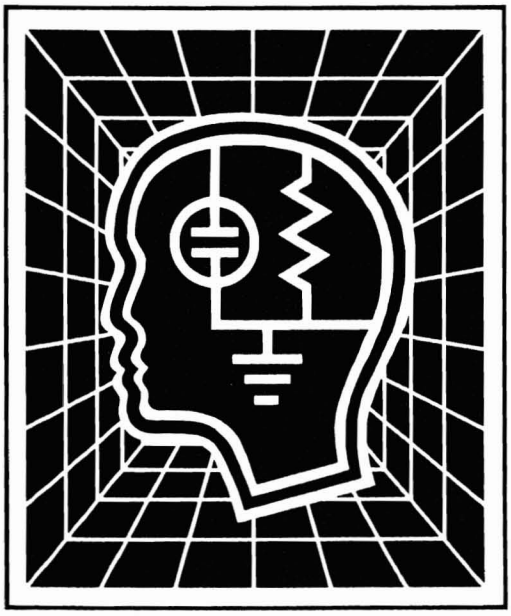
Part Three

Perception (Vision)



8. Vision
9. Computational Vision

Intelligence is a natural phenomenon. It developed in response to the requirement of living systems to predict changes in their environment—both as a result of their own actions, as well as those due to external agents and natural processes.

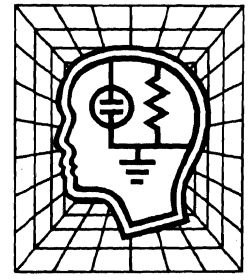
In this book, we discuss intelligence from two perspectives: cognition and (visual) perception. Cognition, covered in the preceding portion of the book, includes the general symbolic machinery which provides a basis for reasoning, planning and communication.

Visual perception, directly concerned with modeling the environment based on sensory information, is discussed in the following two chapters.

It might appear that cognition and perception are two different aspects of the same set of processes: cognition concerned with the nature of the reasoning mechanisms, and perception concerned with their application to modeling and understanding the external world. Unfortunately, things are not quite this simple. The propositional representations and techniques we previously discussed do not appear to be adequate to deal with the major problems of perception. On the other hand, the "iconic/isomorphic" representations that appear necessary for modeling sensor-derived data do not provide a basis for the reasoning techniques we currently understand. The story we tell in this book is far from complete.

8

Vision



Our purpose in this chapter is to explore the concept that visual perception is a form of intelligent behavior, and to examine the way in which organic vision evolved and functions. In particular we will address the following questions:

1. What is the relation between vision and intelligence?
2. Is vision a mechanical or a creative act?
3. What does it mean to "see" something; do all sighted organisms see the same world?
4. What types of visual systems has nature designed, and what universal principles underlie these designs?
5. How is the information from the eye coded into neural terms, into the language of the brain—what is the nature of the brain's description of the visual world, and how is it obtained?
6. How do context, expectations, and scene details blend together to create a perceived image—why do we see illusions; how do we perceive patterns?

THE NATURE OF ORGANIC VISION

Human vision is so effortless, we tend to forget, or possibly not even realize, that

there is a difficult problem to be solved. Most people assume that the eye furnishes the brain with a copy or model of the external world. This is not so. From a set of distorted two-dimensional images projected onto the retinas of our eyes, we must create a world. The eye is just a sensor; the visual cortex of the human brain is our primary organ of vision.

Any sensory organ is an information filter that extracts only part of the total information available to it. In addition, the sensory organ necessarily forms a *representation* or physical encoding of the received information that facilitates the answering of some questions about the environment, but makes it extremely difficult or impossible to answer others. For example, an examination of the encoded information produced by the human eye reveals that at any instant of time the eye has sensed only a small part of the electromagnetic spectrum, and has extracted an image of the scene from a particular viewpoint in space (parts of the scene will be occluded and not appear in the image). There has been no partitioning of the scene into meaningful elements, but the geometrical properties and relationships of objects have been retained in a somewhat accessible form.

At lower levels in the evolutionary scale, organic eyes act less like cameras, but rather more like a set of *goal-oriented* detectors. In the case of the frog's eye, a static scene results in very little information being recorded or transmitted. Only when the frog is looking at a moving object that might be something edible or an enemy, does the frog's eye transmit significant amounts of information to its brain.

Thus, an attempt to define vision on the basis of the structure of a particular type of receptor, i.e., the organic eye, will not address the important question of how the information acquired by the eye is transformed into an interpretation of the surrounding environment. Further, the quality (faithfulness), completeness, and even the encoding of the information provided by the eyes of different organisms vary considerably. It is more appropriate to consider vision to be the process of converting sensory information into knowledge of the shape, identity, or configuration of objects in the environment. This functional, rather than structural, definition concerns itself more directly with what we would intuitively say that vision is all about. We will find that:

- The main organ of vision is the component that does the interpretation, e.g., the brain in the human, rather than the human eye that does the sensing.
- Sensory organs other than the eye can be thought of as providing *visual* information to the interpretation organ. Examples of such sensory organs are the ear of the bat, the sense of touch of a blind person, and the heat detector of a pit viper.
- The memories of past visual experiences, and *wired-in* processing machinery may have a greater influence on how a scene is interpreted than the immediate information provided by the external sense organs.

In the following sections of this chapter, we will consider vision from both structural and functional viewpoints, and we will show that vision is a *creative* rather than a *mechanical* process.

THE EVOLUTION AND PHYSIOLOGY OF ORGANIC VISION

Our purpose in this section is to provide an understanding of the architecture of organic visual systems by examining the evolution and physiology of such systems. In particular, we would like to identify universal mechanisms devised by nature, that offer a solution to the problem of visual understanding of the world.

Seeing and the Evolution of Intelligence

When a camera, a human, an insect, and a frog look at the same scene they do not *see* the same image. In its simplest sense, we can define *seeing* as the physical recording of the pattern of light energy received from the world around us. *Perception*, the interpretation of what we see, is a much more complex process, and will be discussed in following sections.

Seeing, as defined above, consists of three operations: (1) the selective gathering-in of light emanating from the outside world, (2) the projection or focusing of this light on a light sensitive (photoreceptive) surface, and (3) the conversion of the light energy into a pattern of chemical change or electrical activity that is related in some specific way to the scene from which the light originated.

Most living organisms are in continual competition for the raw materials needed to sustain life, and thus they require knowledge of their surrounding environment. However, we cannot conclude that the more information an organism can gather, the better off it is, since the acquisition of information extracts a price in energy, organizational complexity, and the possibility of malfunction. Further, the nature of what is biologically useful information differs widely across the spectrum of living things. For example, the most important aspect of the light energy impinging on a plant, or on many one-celled animals, is the direction from which the light is coming. This detection task can be accomplished by comparing the amount of light energy received on the differently oriented external surfaces of the organism; there is no need to create an *image* of the surrounding environment.

Thus, in an evolutionary sense, the first simple eyes are light-sensitive cells on the body surfaces of organisms that respond only to light intensity, or to variations in intensity. As we proceed up the animal evolutionary scale, we find specially adapted light-sensitive cells appearing in various configurations on the skin of the organisms. Sometimes the cells appear in a randomly scattered arrangement, as in the case of the earthworm, but more commonly they form special arrangements, as in the lining of a depression or pit. The pit is more useful than a flat or convex surface arrangement of cells because the pit provides protection (especially if the opening of the pit can be narrowed in the presence of intense light or other dangerous conditions), more precise directional information, and is a better shadow detector (signaling the possible approach of a predator).

The evolution of the light-sensitive pit is thus justified as a non-imaging light detector with a simple function. However, as shown in Fig. 8-1, once light passes

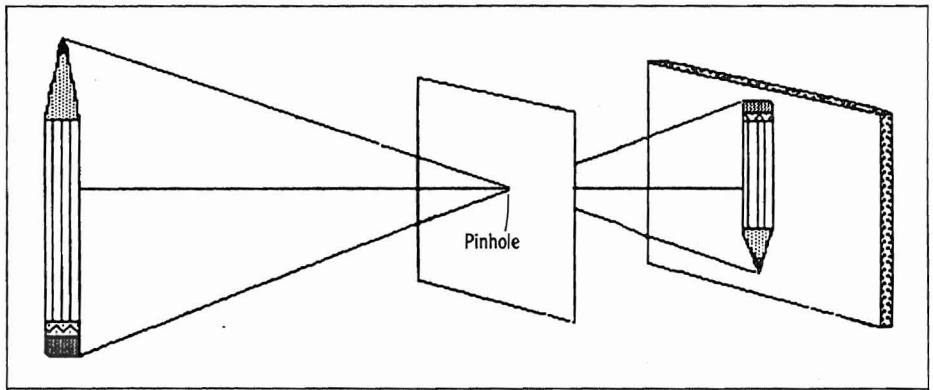


FIGURE 8-1 The Focusing of Light to Form an Image.

through a narrow opening, an image is formed. The existence of this more structured information about the environment could have provided a significant incentive for the incremental evolutionary development of a nervous system capable of interpreting and exploiting this information.[1] Thus, we see the possibility of a direct link between the evolution of vision and intelligence.

A simple organism generally cannot move very far or very fast, and is therefore primarily concerned with its local environment. The level of illumination and illumination gradient, touch, sound (vibration), smell, and taste (chemical analysis) provide all the information needed for adequate functioning. As the organism becomes physically more competent, the dimensions of the environment of its immediate concern expands. Things farther away become important, and sensors that provide such information, primarily vision in creatures that live on land and in the air, become essential. In particular, we note that vision at a distance is useful only in conjunction with a brain that is capable of planning some future course of action, as opposed to reflexive reaction to some local event. On the other hand, a highly competent brain would have little purpose in an organism with little information to process and no ability to use the results of such processing. Thus, it would appear that physical, perceptual, and intellectual competence are interdependent and must evolve as a coherent whole, rather than as independent entities.

[1]The inverted image, caused by a light-focusing eye or pit, could explain why the left brain hemisphere controls the right side of the vertebrate body, and vice versa. Assuming some primordial vertebrate had a single light-forming pit at its anterior end, the left side of this pit would receive signals from the visual field covering the right half of the organism's environment. Evolutionary pressures (as discussed in Chapter 2) would then cause the sensory processing machinery and motor control circuits for the right half of the body to develop in as close proximity as possible to the left light-sensitive portion of the pit. Subsequent evolutionary steps leading to the development of a more sophisticated brain apparently retained this initial structural plan.

Evolution and Physiology of the Organic Eye

We know that life first appeared on earth over 400 million years ago, since fossil records go back at least that far. Because of the similarities in the basic structural units of living things (e.g., all plant and animal forms are built using the same 20 amino acids linked into reasonably similar protein chains), it is likely that all life, as we now know it, had a common origin. Evolution, starting with the same raw material, has produced a small number of basic types of living things which have proved to be successful through the test of millions of years and billions of generations (see Box 2-1).

While there are probably on the order of 2 million distinct species,[2] there are probably less than a few hundred really distinct organizational plans on which these life forms are built. However, of all the varieties of life, only two basic types of imaging eyes have evolved and have come into widespread use. These two types are (1) the single lens, camera-like eye found mainly in the mollusks (especially the squid, cuttlefish, and octopus) and chordates (e.g., fish, amphibians, reptiles, birds, and mammals), and (2) the multilens compound eye found mainly in the arthropods (e.g., insects, lobsters, crabs, crayfish, spiders, centipedes).

The compound eye (Box 8-1) is functionally distinguished from the camera eye (Box 8-2) primarily with respect to achievable resolution, sensitivity, and geometric fidelity. Even though the compound eye does not produce a single coherent image, the light tubes associated with each of the *ommatidia* dissect the image of nearby objects into a mosaic that is similar to the mosaic produced by an image on the cells of the retina of the camera eye. Thus the nerve fibers leaving the compound eye can carry image information very similar to that of the nerve fibers of the camera eye. However, the single lens of the camera eye can form a sharp retinal image of objects located almost anywhere in its field of view.

This ability to focus requires a sophisticated control mechanism that can identify something of interest in the prefocused image, and move or distort the lens to sharpen the boundary between the object of interest and the background. To take advantage of the sharp image, we need a very finely partitioned retina (each human retina has approximately 130 million light-sensitive cells, some with a diameter of one to two micrometers, which is on the order of a few wavelengths of visible light). We also need a computing capacity capable of dealing with this huge volume of data and making almost instantaneous decisions.

The less highly evolved nervous systems of the organisms employing the compound eye probably cannot use (and therefore do not need) the very high resolution required to provide precise shape information. The compound eye cannot be focused, and thus will produce a reasonable facsimile of a true image only for fairly close objects. The number of nerve cells carrying the mosaic information to the brain of the compound eye organism is typically a few thousand in contrast to

[2]Collections of living organisms similar in form and life history, and generally having the ability to interbreed.

the 1 million such fibers in the human optic nerve emanating from each eye.

By reasoning from structural properties of a sensing organ, it is possible to draw conclusions about what aspects of the visual information acquired by the organ are utilized. Because of the low quality of the image and lack of computing power to analyze it, the majority of the organisms possessing compound eyes probably do not rely primarily on vision in making final decisions about the identity of objects they must deal with; quite likely, such decisions are based on the chemical senses, which are highly developed in the arthropods.

Let us look at some other structural properties of organic image forming eyes, and see what they tell us about their owners. In addition to shape information, color is an important attribute for identifying or classifying objects. The mechanism by which organic eyes detect color differences is described in Appendix 8-1. The important point for our discussion is that a retinal cell that is sensitive to a particular color must of necessity ignore the light energy associated with other colors.

BOX 8-1 Insect (Compound) Eyes

Insects have compound eyes, each eye composed of many facets (ommatidia). Each facet has its own separate lens, a small collection of nerve cells, and a single exiting nerve fiber. Each facet points in a slightly different direction resulting in a form of *mosaic vision*—each facet sees a slightly different portion of the visual field. It is not yet known how distinct or complete an image can be formed in the insect brain, but the faceted eye is well suited for perceiving rapid movement.

Typical values for the number of facets in each eye are 4000 for the housefly, 9000 for the water beetle, 3900 for the queen honeybee, 6300 for the worker honeybee, and 13,000 for the drone. Seemingly, the drone needs a large number of facets to find the queen during the mating flight. It has also been observed that the size of the facets in the insect eye may vary significantly. For example, the dragonfly eye, which can discern moving objects several hundred feet away, shows a gradual reduction in facet size from top to bottom. This architecture, in which the upper part of the eye is used for distant vision, and the lower part for nearby vision (similar to bifocals), is designed to support its hunting pattern in which it first detects a distant flying insect, tracks, and finally captures it by scooping the prey up in a basket-shaped arrangement of its legs.

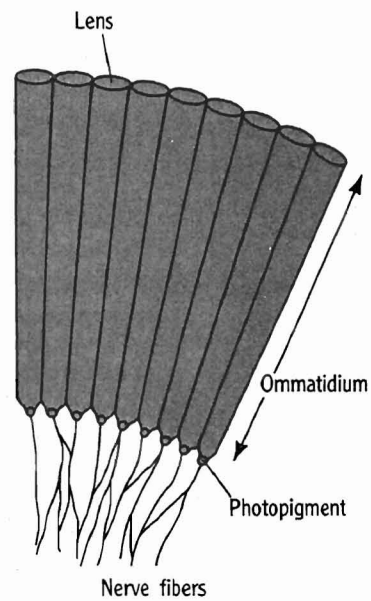


FIGURE 8-2 General Schematic of the Faceted Eye.

Figure 8-2 shows a general schematic of the faceted eye, indicating the interconnection of its nerve fibers. See also [Gregory 78, Hutchins 66].

Hence, the ability to see color means a loss of sensitivity in dim light, and the possibility of not detecting the movement of some object in the visual field because the moving object did not provide sufficient color contrast. The price that must be paid to see the world in color is too high for most organisms, and especially for those that use use their eyes primarily to detect, rather than identify, objects in their environment.

Most animals are color blind (e.g., the dog, cat, horse, cow, pig, sheep); of all the mammals, only man and some primates can see color.[3] Day-active birds, most reptiles, as well as all fishes that have been tested, can see color. Frogs and salamanders are color blind. Bees and a number of mollusks (squid and octopus) can see color.

The issue of sensitivity versus resolution occurs not only across species boundaries, but within the eyes of individual organisms. For example, in the human eye (as well as in many other vertebrates) there are two distinct visual systems based on two types of photoreceptor cells—*rod cells* and *cone cells*. The cone cells can extract color information, and are used for detailed vision; they are small in size, are most densely located near the center of each retina (especially in the area called the fovea), and in the foveal region they communicate with the brain through about as many ganglion cells as there are cone cells. The rod cells are two orders of magnitude more sensitive to light than the cone cells, their relative density is greatest in the peripheral regions of the retina, they are incapable of detecting color, and they work in groups which feed the brain through a much smaller number of shared ganglion cells. It appears that the main function of the rod cells is to detect anomalies (e.g., movement) in the visual field and then to allow the cone cells to do the detailed analysis via eye slewing and focusing. In very dim light, where the cone cells cannot function, the rod cells must also take over the responsibility for shape perception. Since the sensitive rod cells are most dense on the periphery of the retina, to identify an object at night, it is best not to look directly at it, but rather to look at the object out of the corner of your eye.

In regard to visual information, most insects and lower organisms are more concerned with detection than with classification. They have eyes that typically are designed for sensitivity and broad field of view, rather than precise resolution. These eyes either have large receptor cells, or rodlike networks feeding their brains through shared ganglion cells. The eyes of these insects do not focus nor are they independently movable; typically there are two compound eyes anchored to opposite sides of the head, where they monitor largely nonoverlapping fields of view.

Eye and Brain

How is the pattern of light energy projected onto the light-sensitive cells of the eye transformed into a model of the external world? Even for simple organisms, we know little of the structure of their neural machinery, and even less about the way the machinery actually functions. It would

[3] While almost all mammals possess some degree of hue discrimination, this faculty plays a small role in their behavior, typically being completely dominated by the intensity component of the received light.

appear that much of the visual processing in lower organisms is carried out in neural networks located adjacent to the photoreceptive cells. As we ascend the evolutionary scale, more of this processing is shifted to the brain (a process called *encephalization* which occurs in other senses and muscular control functions as well). The human retina, for example, is an extension of the embryonic brain tis-

BOX 8-2 The Human (Camera) Eye

The human (camera*) eye is a remarkable instrument with respect to both sensitivity and resolution. In clear air, a candle flame is just visible at a distance of ten miles; thus, 10^{-14} parts of the light produced by a single candle is sufficient to stimulate vision. The mechanical energy of a pea, falling from a height of one inch, would, if translated into luminous energy, be sufficient to give a faint impression of light to every person that ever lived [Pirenne 67]. Some of the parameters of the human visual system are:

- 120 million rod cells in each eye.
- 6 million cone cells in each eye.
- 2000 cone cells in each fovea in the region of maximum uniform density.
- 1 million nerve fibers in the optic nerve exiting each eye.
- Diameter of cone cells in fovea: 1 to 3 micrometers.
- 250 million receptor cells in the two eyes vs. 250,000 independent elements in a TV picture.
- Distance from effective center of lens to fovea: 17 mm.
- Interpupillary distance: 50 to 70 mm.
- Visual angle subtended by fovea: 20 minutes of arc for region uniform maximum cone density, 1 to 2 degrees for rod-free area, 5 degrees for a 50 percent drop in visual resolution (with the arm extended, the raised thumb subtends an angle of 2 to 2.5 degrees; one minute of arc corresponds to a retinal image of five micrometers).
- Angle with respect to visual axis of eye at which rod density is maximum: 15 to 20 degrees.
- Rod cells are on the order of 500 times more sensitive to light than cone cells.
- Visible portion of the electromagnetic spectrum: 0.4 to 0.7 micrometers.
- Wavelength of maximum rod sensitivity: 0.51 micrometers (green).
- Wavelength of maximum cone sensitivity: 0.56 micrometers (orange).
- Intensity range: 10^{16} (160 decibels).
- Minimum visual angle at which points can be separately resolved: 0.5 to 2 seconds of arc for alignment of lines (0.04 to 0.16 micrometer, less than 10% of the diameter of the smallest foveal cell); 10 to 60 seconds for dots (range of values is due to disagreement across reference sources). If the pupil is 3 mm in diameter and a 0.55-micrometer light is used, the image of a point will produce a central circle of 3.7 micrometers on the retina. This would mean that the illumination from two points 25 to 30 seconds apart would overlap.
- Object distance from eye for stereoscopic depth perception: 10 inches to 1500 feet (1500 feet corresponds to a retinal disparity of approximately 30 seconds of arc).
- Involuntary eye movements: 10 to 15 seconds of arc for tremor; slow drifts of up to 5 minutes of arc.

Figure 8-3 shows the anatomy and nervous organization of the human eye.

*Using a camera analogy to understand the operation of the human eye is an oversimplification in at least two important respects, (1) the measured performance of the eye is much better than its component specifications would permit if the eye really did behave as a camera, and (2) the eye has no shutter, and even though the scene information projected on the retina is in constant motion, our perception of the world is not blurred. The brain appears to extract information from the "optic flow" across the retina rather than by analyzing a static image.

sue, whereas the lens of the eye develops from embryonic skin tissue.

How then is the information from the eyes coded into neural terms—into the language of the brain—and then interpreted? When light strikes the retina, the decomposition (bleaching) of pigments in the rods and cones results in electrical activity, which is integrated in the bipolar and ganglion cells comprising the sixth and eighth levels of the ten-layer system of the retina (Fig. 8-3). As discussed in Chapter 2, the ganglion cells of the eye feed the brain with visual information coded into chains of electrical pulses. The rate of "firing" of the cells is proportional to the logarithm of the intensity of the original stimulation (Fechner's law). Other attributes of the illumination, such as color, are determined by which cells are firing. For example, as indicated in Appendix 8-1, the cone cells are differentially sensitive to the red, green, and blue components of the illumination because of differences in the chemical composition of their photosensitive pigments; these cells are intermixed in the fovea, and their relative excitation provides the brain with information about the color of the objects being viewed.

As depicted schematically in Fig. 8-4, the human retina is effectively divided vertically down the middle; the nerve fibers from the left half of each retina send information about the right half of the visual field to the *striate cortex* in the left occipital lobe of the brain. Similarly, the right half of each retina sends information about the left half of the visual field to the right striate cortex. The role played by the *lateral geniculate body* is not currently understood—it appears to simply relay the information it receives. (However, there is some evidence that it is functionally involved in the processing of color information.)

The nerve fibers from the eye, reaching the striate cortex, preserve the topology and much of the geometry of the imaged scene information; a portion of the striate cortex, called the *visual projection area*, is in approximate one-to-one spatial correspondence with the retina. Stimulation of nerve cells in this projection area by a weak electric current causes the subject to *see* elementary visual events, such as colored spots or flashes of light, in the expected location of the visual field. Lesions in the projection area lead to *blind spots* in the visual field consistent with the retina-to-cortex mapping, although some pattern vision is left intact. For example, contours of perceived objects are completed over blind spots.

In the human, the region of the striate cortex immediately surrounding the visual projection area is called the *visual association area*. Electrical stimulation of cells in the association area give rise to complex recognizable visual hallucinations (images of known objects or even meaningful action sequences). Local lesions of this part of the occipital cortex neither reduce visual acuity nor lead to loss of any portion of the visual field; the essential symptom associated with such lesions is disturbance of the perception of complete visual complexes, the inability to combine individual impressions into complete patterns, and the inability to recognize complex objects or their pictorial representations. For example, some patients with visual association area lesions can describe individual parts of objects and can reproduce their outlines accu-

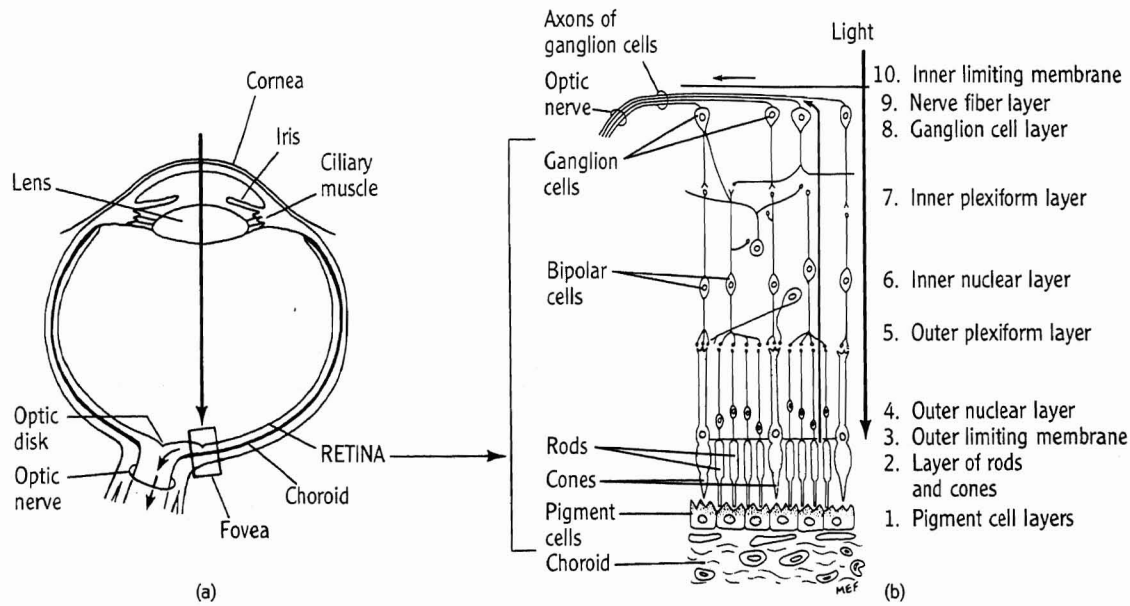
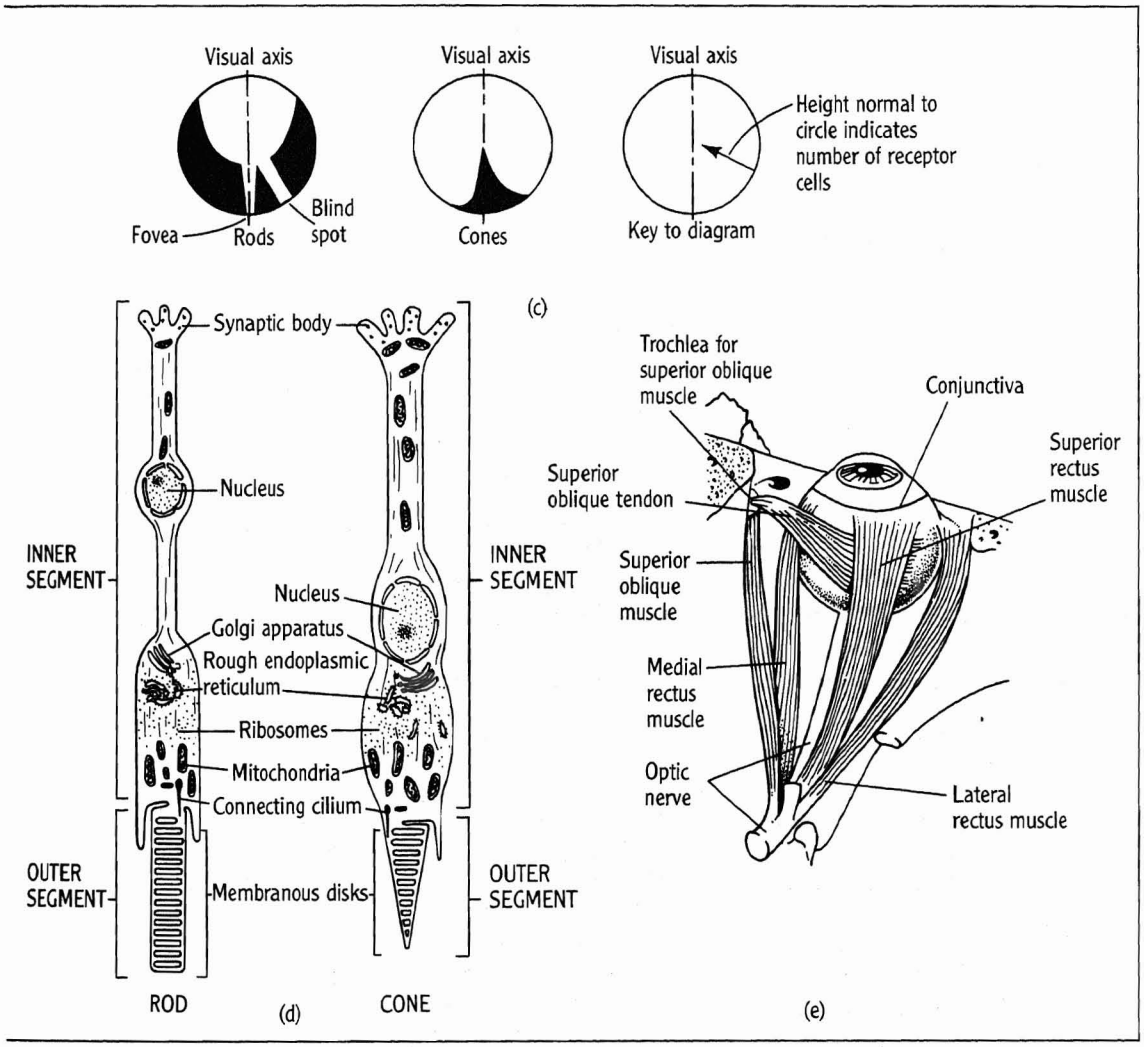


FIGURE 8-3
From Retina to Optic Nerve: the Conversion of Light Signals to Nerve Impulses.

(a) Anatomy of the human eye. The *optic disc* forms a blind spot where nerve fibers leave the eye to form the *optic nerve*. The *ciliary muscle* controls the focus of the *lens*. The pupil is the opening at the center of the *iris*. (b) Neural organization of the retina. The cellular arrangement in the retina appears to form ten layers, when viewed by light microscopy, numbered 1 through 10 from the innermost pigment layer to the outermost fiber layer. Light rays pass through the neural layer to reach the rod and cones. Nerve impulses are propagated in the opposite direction from the rods and cones to the optic nerve. [(a) and (b) from E. L. Weinreb. *Anatomy and Physiology*. Addison-Wesley, Reading, Mass., 1984, p. 246, with permission.] (c) Distribution of rod and cones in the human eye. (d) Diagrammatic representation of the photoreceptor cells showing the organelles as viewed by electron microscopy. (From E. L. Weinreb. *Anatomy and Physiology*. Addison-Wesley, Reading, Mass., 1984, p.247 with permission.) (e) The extrinsic muscles of the eye viewed from above. (From A. P. Spence and E. B. Mason. *Human Anatomy and Physiology*, 2nd edition. Benjamin Cummings, Menlo Park, Calif., 1983, p. 393 with permission.)

rately, yet are unable to recognize the objects as a whole; other patients are unable to see more than one object at a time in the visual field.

The integration of perception with the cognitive functions in the human is demonstrated in patients with lesions in the *parieto-occipital* regions of the brain (i.e., the regions physically located between the primary vision and speech centers). Such patients experience great difficulty in attempting to perform tasks involving spatial relationships. For example, even though they think they understand a task such as "Draw a triangle below a circle," they cannot appreciate



the difference between this task and that of drawing the circle below the triangle; they typically will draw the figures in the sequential order in which they are given in the instructions. The perception of embedded figures is affected by lesions almost anywhere in the cortex, although these effects are most pronounced when the lesions occur in the speech association areas (left temporal and parietal lobes). Injury to the parietal lobes can result in right-left reversals, copying problems, and in left visual field distortions. Right temporal lobe lesions appear to interfere with the understanding of complex pictorial material.

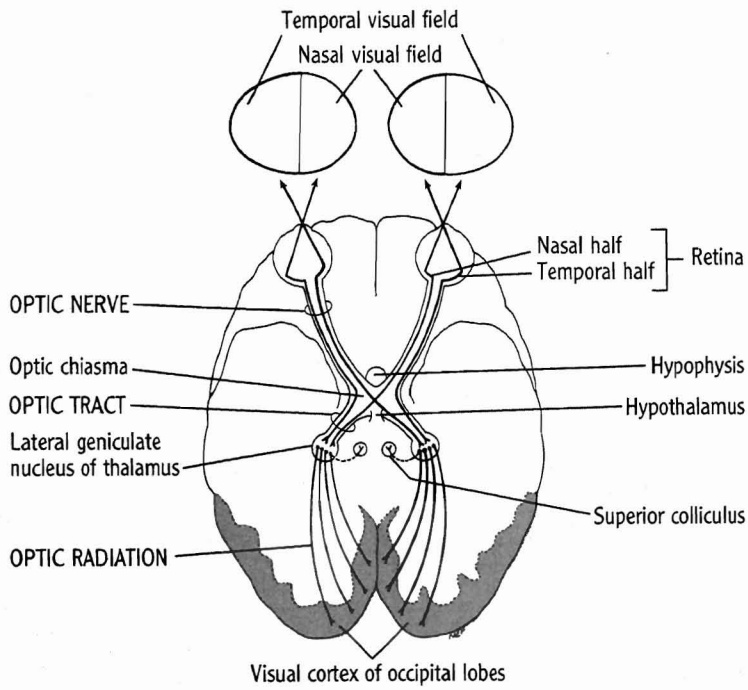


FIGURE 8-4 Nerve Pathways from the Retina to the Visual Cortex of the Brain.

Nerve fibers from the right side of each retina pass to the right side of the brain; nerve fibers from the left side of each retina pass to the left side of the brain. (From E. L. Weinreb. *Anatomy and Physiology*. Addison-Wesley, Reading, Mass., 1984, p. 254, with permission.)

Any defect of the human cortex leads to imperfect perception and reproduction of pictorial objects. On the other hand, the removal of the entire forebrain in fish or birds appears to have little effect on their visual discrimination. As in man, the destruction of the visual cortex in adult mammals leads to complete loss of pattern vision, but this is not true for young animals. For example, young cats in whom the striate cortex is completely destroyed show unimpaired performance on all visual tasks; obviously, some other portion of the cat's brain is capable of assuming the visual function.

In lower animals, subcortical centers play the major role in pattern vision. The *superior colliculus*, for instance, is the most important visual center in fish and amphibians. As we move up the evolutionary scale, the visual function is first concentrated almost completely in the striate cortex of the occipital lobe, as in the rat, and finally spreads out over the whole cortex of the brain when we reach the evolutionary level of the monkey.

How do the neural circuits of the brain produce its perceptions of the world? There is very little we can say about the relationship between brain architecture and the performance of high-level functions, but we know a little about how some of the more elementary neural processing is accomplished. In particular, there are a few processing tricks that nature has discovered that are so important that we find them employed in almost all eyes (or visual systems) of all species. One of these is *lateral inhibition* (or *center-surround inhibition*) for detecting when something different or unusual has occurred in the visual field; this technique is discussed in Box 8-3. At a less detailed level, we also have some insight into how

BOX 8-3 Lateral Inhibition and Adaptation—The Enhancement of Contract in Space and Time

Most of the biologically important information about the surrounding environment provided by our senses remains essentially constant from one instant to the next. It would be inefficient for our sensory systems to keep telling us things we already know, and indeed, this does not happen: animal nervous tissue is designed so that its response diminishes, and even stops, with repeated stimulation. This process, called adaptation, can be dramatically demonstrated by using a special apparatus to cause an image to be projected onto a fixed location on the retina (normally, even a static scene would move around on the retina due to the constant movement of the eye). When such "stabilized" images are produced, they quickly disappear from conscious perception. Sometimes, coherent meaningful segments of the visual field will reappear, only to fade out again. This proves that adaptation is occurring not only in the retinal tissue, but also at higher levels in the brain.

Most sensory tissue (retina of the eye, cochlea of the ear, pressure-sensitive nerves of the skin), and even portions of the brain (cerebellar and cerebral cortex), is organized so that stimulation of any given location produces inhibition in the surrounding nerve fibers. It is shown (next chapter) that the effect of this structural organization of nervous tissue, called *lateral inhibition*, is to (mathematically) differentiate the signals being processed. In the case of visual information, such (spatial) differentiation causes gradual changes in the contrast between an object and its background to become more abrupt, thus enhancing the ability of the visual system to detect objects of interest in the visual field. For example, if you align two sheets of paper so that only a narrow slit is visible along the length of the intensity step wedge shown in Fig. 8-5, and count the number of regions which appear to have different intensities, this number will be less than that obtained when the full step wedge is visible. Lateral inhibition is an area effect, and peering through the narrow slit prevents it from operating; local contrast is insufficient to allow us to see all the intensity boundaries when lateral inhibition is suppressed.

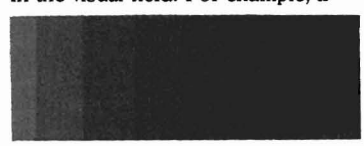


FIGURE 8-5
Intensity Step Wedge Used to Demonstrate Lateral Inhibition.

In human perception, the contrast enhancing effect of lateral inhibition produces what are called *Mach bands*. If a sharp shadow is produced on a flat surface, a thin bright band will appear to parallel the shadow line on the illuminated side, and a corresponding dark band on the occluded side. These bands are not physically present, but are subjective phenomena—essentially "overshoot" and "undershoot" caused by our neural circuits mathematically differentiating the step discontinuity in illumination.

BOX 8-4 Feature Detection and the Frog's Eye

In Appendix 8-2 we note that neural circuits in the visual cortex of the human brain appear to detect generic image features, such as oriented line segments. In contrast, lower organisms tend to search directly for goal-specific features; the corresponding computation is often carried out by neural networks in the sensing organ. The frog's eye provides a good example.

In 1953, Barlow [Barlow 53] found that one particular type of ganglion cell in the frog's retina was excited when a black disk, subtending a degree or so of arc, was moved rapidly to and fro within the receptor field. This caused a vigorous discharge that could be maintained as long as the movement continued. Barlow suggested that these retinal neurons were "bug detectors" and that the frog's feeding responses might partially originate in the retina.

A classic work on the physiological basis of information extraction from a visual image was by Lettvin et al [Lettvin 59]. They found that the frog eye uses four different types of neural structure to extract patterns of information from the visual signal: (1) edge detectors that respond strongly to the border between light and dark regions; (2) moving contrast detectors that respond when an edge moves; (3) dimming detectors that respond when the overall illumination is lowered; and (4) convex edge detectors that react when a small dark, roughly circular object moves in the field of vision. The response of this detector increases as the object moves steadily closer to the frog.

The convex edge detector provides the visual information used by the frog to detect and catch flies. Note that this detector requires motion: a dead fly is of no interest to the frog.

Similar experiments were subsequently carried out with higher animals. However, by the 1970s it was realized that in higher organisms, visual perception is considerably more than a collection of specialized neural feature detectors.

stereoscopic vision works (Appendix 8-2) and how, at least in some animals, features or attributes of perceived objects are computed (Box 8-4).

THE PSYCHOLOGY OF VISION

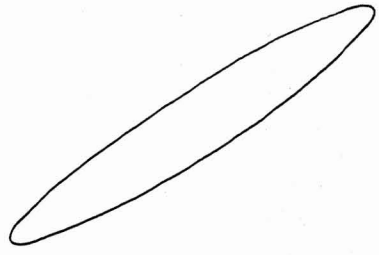
Our purpose in this section is to explore the nature of the algorithmic techniques employed by organic visual systems through an examination of their successes and failures in interpreting both natural and contrived images.

Perceiving the Visual World: Recognizing Patterns

Humans and a few other organisms live in a world of shape and color. We perceive ourselves as moving through a stationary environment, rather than ourselves being stationary and the surrounding environment as moving. We can recognize and actually perceive common objects as having an expected shape, even though we view them from different distances, orientations, and under unknown lighting conditions. Thus, if we look obliquely at a circle drawn on a flat surface, we see the expected circular shape, even though the image projected onto our retina is an ellipse, (Fig. 8-6). We can adjust to more than ten orders of magnitude of light intensity variation without conscious awareness, and are not bothered by shadows or partial occlusions. It is obvious that perception is not the inevitable result of a set of stimulus patterns, but

rather a best interpretation of sensory data based on the past experience of both the organism and its ancestors. While the senses do not directly give us a faithful model of the world, they do, however, provide evidence for checking hypotheses about the nature of our surrounding environment. Perhaps the most concise way of summing up our visual capability is that, except in the case of physical injury, it appears to operate flawlessly, spontaneously, and without surprises (or indeed, we are both shocked and surprised).

(a)



(b)

FIGURE 8-6
Shape Constancy—an Ellipse Seen as a Circle.

(a) Photograph of a wheel (photo courtesy of O. Firschein).
(b) Actual shape of wheel as it appears in the photograph.

Of the full range of perceptual skills needed to completely model the world, we will limit our discussion in this section to the problem of recognizing patterns, and of necessity, our discussion will be descriptive rather than explanatory.

At the highest levels of performance, shape recognition involves the ability to ignore variations in size, brightness, position, and orientation. It is known, for example, that if a person (or animal) learns to identify a shape using one part of his retina, he is able to identify the same shape when it is presented to other parts of the retina, or even the other eye. On the other hand, when the brightness of an object and its background are inverted, even humans are sometimes unable to recognize the object. Many species (e.g., human, rat) can recognize a shape from its outline, while others (e.g., octopus) have great difficulty in recognizing the outline if they have been trained to recognize the filled-in shape.

When a rat or octopus is trained to discriminate between a horizontally elongated rectangle, and a square with sides equal to the height of the rectangle, and is shown the rectangle rotated by 90 degrees, it treats the rotated rectangle as if it were the square. After being trained with a square and a triangle, an octopus responds to a diamond (45-degree rotation of the square) as if it were the triangle. Humans have great difficulty recognizing faces presented upside-down.

The inability of some higher organisms to deal with such apparently simple variations as a 90- or 180-degree rotation, seems, at first, to be rather strange. After all, there are almost trivial mechanical procedures that could undo such a variation. One possibility is that while the human visual system can generally decompose (partition) the visual field into meaningful subunits (see below), most other organisms may not have this ability. If a visual system cannot extract and manipulate portions of an image, then the *normalization* operations needed for robust pattern vision become almost hopelessly complex, and impossible to implement.

We know that even simple organisms have the ability to recognize patterns. (The development of vision in infants is discussed in Box 8-5.) For example, bees and ants utilize visual landmarks in finding their way back to the nest. The wasp *Philanthus* locates the entrance to its nest by the arrangement of visual markings and objects around it; bees can recognize their hive by colored marks at the entrance, and can determine if the colors have the proper spatial arrangement when approaching the hive from an arbitrary direction. However, bees cannot distinguish among geometrical shapes (triangle, square, circle, ellipse), but appear in such cases to be able to respond only to some gross measure of the degree to which the figure is branchy (vs. solid) or divided up into parts (Fig. 8-7).

In discussing pattern vision, we have tended to talk about it as if it were a single integrated function in any given organism. This is an oversimplification—biologically important visual tasks are often handled by special mechanisms. For example, in the human, visual pattern perception is not performed uniformly for all tasks. It appears that the human brain has distinct procedures for processing visual information about faces—a task of great biological importance.

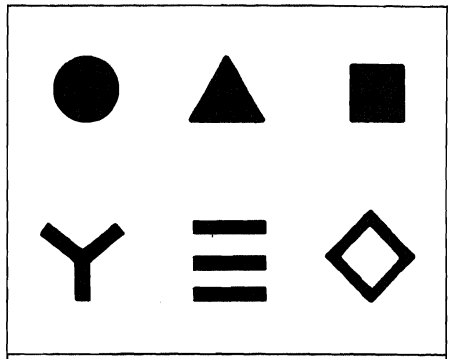


FIGURE 8-7
Perception of Shape by the Bee.

For the bee, figures of the upper row are indistinguishable from one another. The same applies to the lower row. But the bee readily distinguishes figures of the lower row from those of the upper row. (After V. B. Wigglesworth. *The Principles of Insect Physiology*, p. 207.)

In lower organisms, the ability to recognize hereditary prey, or a mate, appears to be carried out with a sophistication orders of magnitude beyond that exhibited for abstract patterns. For example, the *Pepsis* wasps hunt tarantulas as food for their larvae (their own food consists of flower nectar). Only the female hunts, and she is a specialist—each species of wasp hunts only one species of tarantula. She searches for her prey in desertlike terrain, even locating and entering the tarantula's underground tunnels

or burrows. In order to achieve her purpose, she must force the tarantula (a very large and fierce hunter itself) over on its back and sting it in the soft membrane between the basal joints. The almost always successful wasp now digs a sloping foot-long tunnel of just the right size to hold the paralyzed tarantula, drags the tarantula into this burrow, cements her eggs to its abdomen, and then seals the

BOX 8-5 Development of Vision in Infants

To study infant perception, the experimenter exposes the child to visual stimulation and notes the reaction such as eye movements, eye fixations, sucking, head turning, and reaching. Although the design of experiments using infant subjects has improved in recent years, there is still ambiguity in the results concerning inborn versus acquired capabilities and this leads to controversy among the various theories of early perception. Some of the difficulty lies in the fact that important parts of the visual system are not fully developed at birth. This includes the retina, the lateral geniculate nucleus, and the visual cortex. Many of the changes in early development therefore reflect maturation of the neural system, particularly the visual cortex. The other problem in carrying out experiments is that the infant cannot be kept in a controlled environment, and it is therefore difficult to isolate characteristics that are under investigation. A good review of the field is given in Banks and Salapatek [Banks 83] and in Flavell [Flavell 85]; some of the highlights are given below.

Newborn infants. Newborn acuity is very poor (about 20/600), as is contrast sensitivity, both improving considerably during the first 6 months. Infants prefer to fixate some patterns over others, and repetitive patterns over random ones, but there is no good theory as to why one is preferred over another. Faces are preferred over nonsense patterns of equal contour density, and familiar faces are preferred over unfamiliar ones [Salapatek 77]. Newborns do not scan a figure very extensively; their gaze gets captured by a single feature or part of a figure.

2–3 months. By 3 months of age, the scanning limitations are overcome. Infants can distinguish patterns on the basis of their shape and form. By 2 months of age infants can make some color discrimination, and by 3 months color vision is quite good, and there is some improvement in color discrimination after that. Three month old subjects can perceive touching objects as two objects rather than one. There is some suggestion that they react to "looming objects," objects that appear to move toward them suddenly [Yonas 81].

4–6 months. "Biological motion," caused by luminous spots placed on the hip, arm joints, and leg joints of a person running in place in a darkened room is preferred to random motion of the spots [Fox 82]. Biological motion right side up is preferred to upside down biological motion.

A "visual cliff" consists of a horizontal sheet of glass resting just above a patterned surface on one side and spanning a deep depression on the other side. Prelocomotive babies (younger than 7 months) when slowly lowered toward the glass put out their hands just prior to touchdown on the shallow side but not on the deep side [Svedja 79], showing that they perceive the depth.

Spelke [Spelke 82] showed that infants of 4 months perceive as one continuous object a reciprocating rod whose center is hidden by a block. This holds even if the two visible parts of the partially occluded object differ from one another in size, shape, color, texture, and alignment.

In the postinfancy period, the child develops the ability to attend selectively to wanted information in the sensed input, while tuning out or disregarding unwanted information [Flavell 85].

tunnel with soil and sand. The wasp grub feeds on the tarantula, and on the order of a year later emerges from the burrow as an adult wasp. The searching, fighting, and building activities described above appear to require perceptual abilities equal to almost any capability of the human visual system—except that such perceptual behavior in the wasp is not general; for most other visual tasks its behavior has the various limitations mentioned earlier.

Perceptual Organization

Everything we see, we see for the first time. While parts of a scene may correspond to objects we have some previous acquaintance with, we almost never see the same objects in the same configuration under the same lighting conditions from the same perspective in space. Unless we can decompose or partition a scene into coherent and independently recognizable entities, the complexity of natural scenes would seem to render human-type vision impossible.

How can we partition a scene into independent components without already knowing what might be present? If we were only searching for a few well-known objects, we might attempt to exhaustively determine if each of the objects were present at each possible location in the visual field. However, there are probably thousands of objects that can appear in an almost infinite variety of configurations and orientations that we can recognize; exhaustive matching against stored models is not a reasonable explanation of human perception.

It is largely agreed that there must be a set of generic criteria, applied independently of scene content, that underlies the procedures discovered by nature for partitioning the visual field. Discontinuities in scene properties (e.g., distance, material composition, motion) are the most likely clues as to where partitions should be inserted. A significant portion of the work in computational vision (next chapter) is devoted to the partitioning or perceptual organization problem, the critical issue being that of relating image intensity variations to physical discontinuities in the scene.

Psychologists have also attempted to discover the *laws* underlying the partitioning decisions made by the human visual system. One of the earliest and intuitively most acceptable collections of such laws was proposed by Wertheimer in 1923 and elaborated by Koffka in 1935. These *gestalt laws* include:

- **The Law of Proximity.** Stimulus elements that are close together tend to be perceived as a group (Fig. 8-8a).
- **The Law of Similarity.** Similar stimuli tend to be grouped; this tendency can even dominate grouping due to proximity (Fig. 8-8b).
- **The Law of Closure.** Stimuli tend to be grouped into complete figures (Fig. 8-8c).
- **The Law of Good Continuation.** Stimuli tend to be grouped so as to minimize change or discontinuity (Fig. 8-8d).
- **The Law of Symmetry.** Regions bounded by symmetrical boarders tend to be perceived as coherent figures (Fig. 8-8e).
- **The Law of Simplicity.** Ambiguous stimuli tend to be resolved in favor of the simplest alternative. For example, if

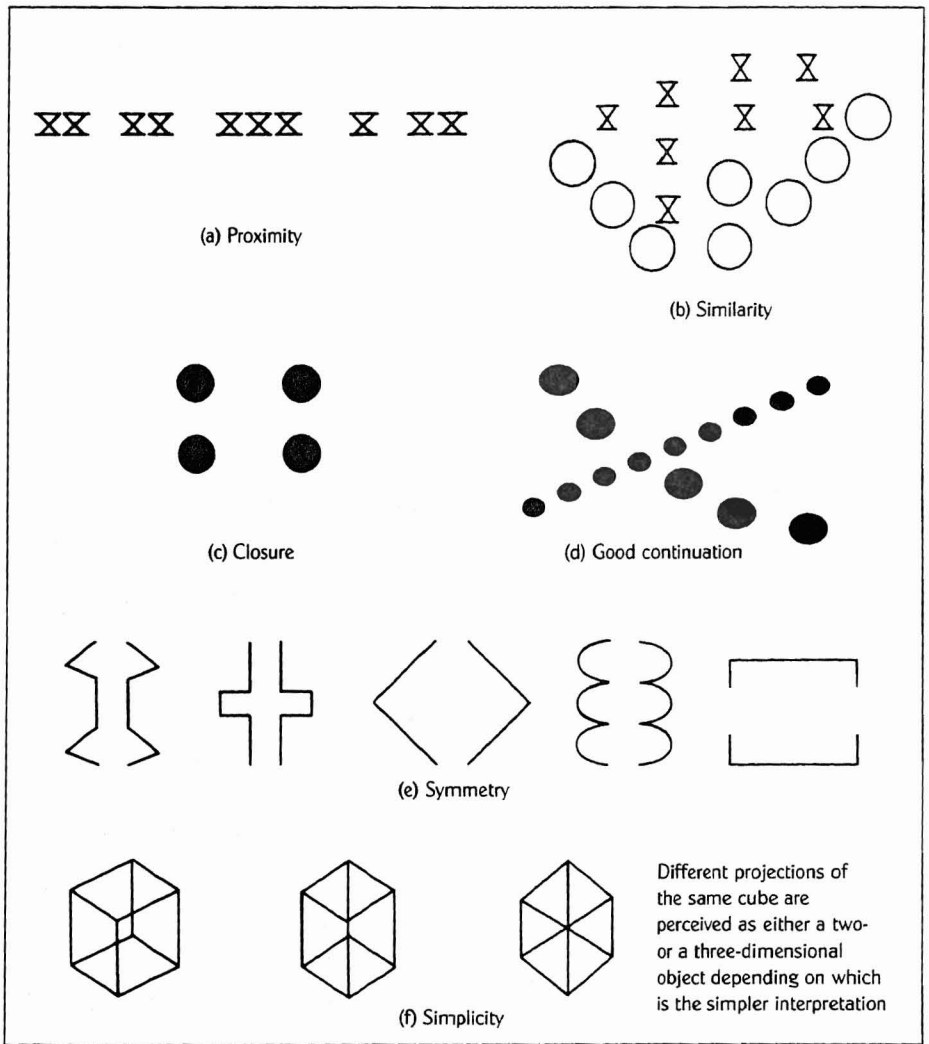


FIGURE 8-8
The Gestalt Laws of Perceptual Organization.

(a) Proximity. (b) Similarity. (c) Closure. (d) Good continuation. (e) Symmetry. (f) Simplicity.

fewer different angles and line lengths are required to describe a figure as three-dimensional, the observer will select this alternative over the two-dimensional interpretation (Fig. 8-8f).

- **The Law of Common Fate.** If a group of dots were moving with uniform velocity through a field of similar though stationary dots, the moving dots would be perceived as a coherent group.

None of these laws are as simple as they first appear. For example, proximity grouping seems to be based on measurements in perceived space (as opposed to proximity measured by retinal distance) and is influenced by prior experience as demonstrated in Fig. 8-9.

A major problem with the above set of gestalt laws is that there is no explanation as to the purpose they serve, or how the given criteria contribute toward achieving the intended purpose. It is possible to argue that all perceptual decisions are implied explanations of how sensed data relates to scene content. As an explanation, any partitioning decision must satisfy criterion for believability—i.e., completeness (explaining "all" the data), stability (consistency of explanation), and limited complexity (economy of explanation). This alternative viewpoint does not conflict with the gestalt laws, but rather provides a broader basis for understanding them. Additional ideas about the nature of perceptual organization, such as the existence of a primitive perceptual vocabulary and a "preattentive visual system" are discussed in Chapter 9.

Visual Illusions

To understand how something works, we often have to stress it, take it apart, or even break it. How can we discover the nature of the algorithms employed by organic (especially human) visual systems? Examining neurologic structure in an attempt to deduce function is a hopeless task for anything other than the simplest types of mechanical or reflex mechanisms. Introspection is unreliable, available only in the human, and even here, language is not always suitable for describing perception or intent. Further, the operations of many (if not most) visual functions are not accessible to introspection.

Observation of performance suffers from two defects—it is not always clear that the organism and the experimenter have the same task in mind (or that the

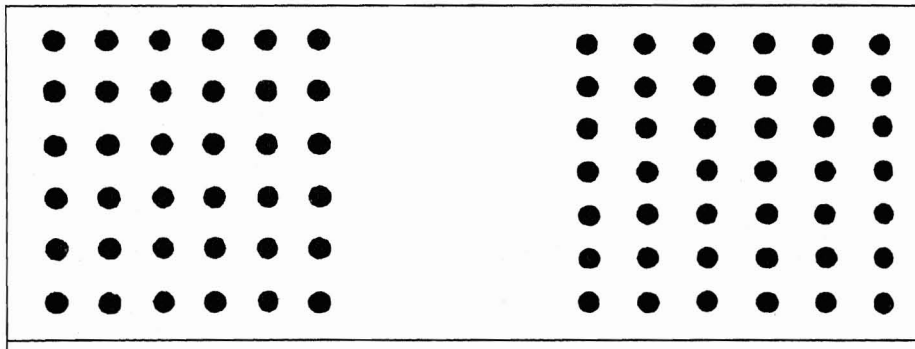


FIGURE 8-9 Proximity Grouping Can Be Altered by Recent Experience.

When first viewed, the grouping of the dots on the left is ambiguous. After looking at the array on the right for one or two minutes, however, the array on the left appears to be grouped into rows.

organism is seriously interested in performing the task). Furthermore, it takes a great deal of ingenuity to design a task in which performance will fully explain the underlying algorithms. Visual illusions have provided much of the information on which theories of the functioning of organic visual systems are based. What are visual illusions, and why are they so fascinating?

In a sense, everything we perceive is an illusion since we can never exactly recreate objective reality through our senses. The term "illusion" is reserved for those situations in which our perceptions differ markedly from what we know corresponds to the actual physical situation. Further, there must be no reason to believe that processing involved in the perception of an illusion is in any way unusual or unique; generally the causative factor should be some circumstance associated with the scene, or the context under which it is viewed. Illusions are fascinating because we expect our sense of vision to be infallible (seeing is believing). We are not surprised when our muscles fail to do exactly what we want, and we know that we can misinterpret the direction of a sound, or can confuse a very hot object with a cold one via our sense of touch, but we almost never question our visual decisions—in fact, we routinely trust our lives to them. How is it possible then, that we are so readily misled by visual illusions, even when we know what the true situation is?

The most remarkable aspect of human vision is not that it is subject to failure, but rather how accurate it is in spite of its limited and distorted inputs. The human eye is far from a perfect instrument; it distorts the image that it projects onto the retina (even under the best conditions) because of its finite aperture, out-of-round lens and cornea, different index of refraction at different wavelengths, and imperfect focusing machinery. It is obvious that there are a number of relatively independent problems that the visual system must deal with, and it is not unreasonable to assume that illusions are due to a failure of one or more of the mechanisms set up to deal with these problems. It is very unlikely that a single mechanism underlies all illusions. In particular, the visual system is able to compensate for the various distortions introduced by the limitations of the imaging system of the eye.

- It can compensate for the change in appearance of objects due to the projective transformation inherent in even a perfect camera-type imaging system.
- It can resolve the ambiguity resulting from the projection of the three-dimensional world onto a two-dimensional retina.
- It can provide true information about surface reflectance and color under a wide variety of illumination conditions.
- It can provide a stable frame of reference and an unblurred image, even when the eye is constantly in motion.

Under normal circumstances, while any one of these compensating functions might fail, there appears to be enough redundancy to allow a very clever integrating system to detect and correct the error so that the final perception is faithful to the physical situation. Illusions are produced when we are presented with an impoverished visual environment that

eliminates the normal redundancy, and overloads or deliberately misinforms a single functional system.

Thus, the Ponzo illusion (Fig. 8-10) appears to be due to the fact that we are interpreting what is really a two-dimensional picture as if it were a three-dimensional object. This implies that we have built-in machinery for automatically compensating for the shrinking of the image of an object with increasing distance: something that is perceived (via other processing channels in the visual system) to be located further away, but projects onto the retina with the same length as something nearby, is judged to be larger.

In the Necker cube (Fig. 8-11), two configurations—given face in front, given face in back—are reasonable interpretations of the imaged data, and there are no cues to cause one interpretation to dominate. Therefore, the perceptual system appears to formulate and offer us in turn these alternative hypotheses. This is an indication of a reasoning process carried

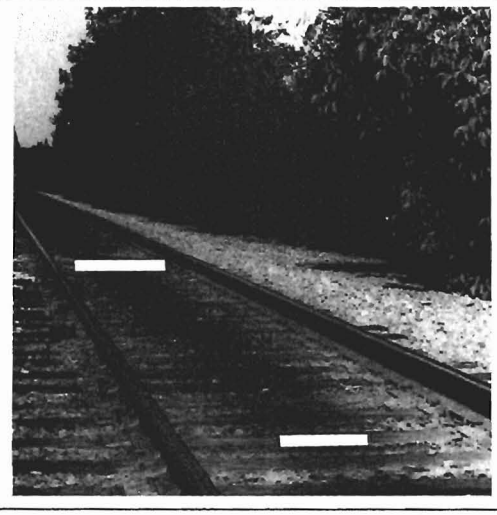


FIGURE 8-10
The Ponzo Illusion: Apparent Depth Alters Size Perception.

A feature that is near the narrowing end of the exterior lines gets expanded and looks longer than it would if placed below, where the lines widen. The perspective effect induced by the converging lines causes our visual system to make size corrections for the three-dimensional phenomenon of change in size with distance. (Photo courtesy of O. Firschein.)

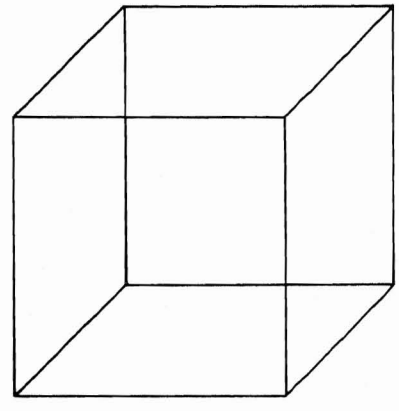


FIGURE 8-11
Perceptual Ambiguity (Multistable Perception).

This figure alternates in depth: a face of the cube sometimes appears as the *front*, and sometimes as the *back* face. We can think of these ways of seeing the figure as the result of alternative perceptual *hypotheses*. The visual system entertains alternative hypotheses, and will settle for one solution only when there are no obvious alternatives.

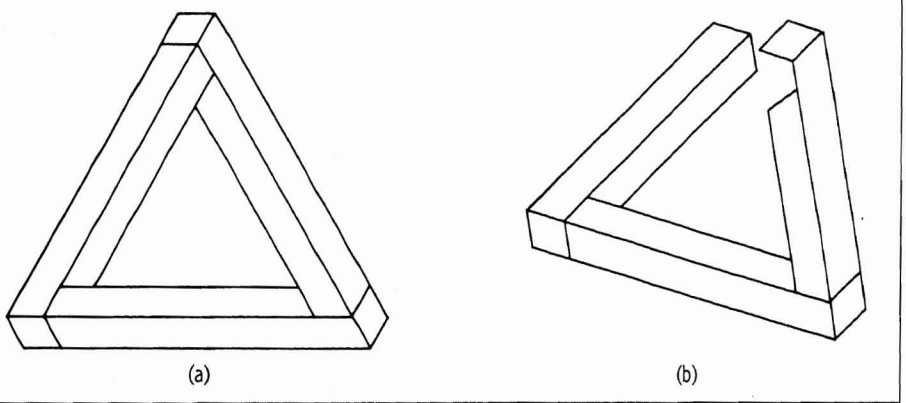


FIGURE 8-12
The "Impossible" Triangle.

(a) A seemingly impossible three-dimensional object. (b) The same three-dimensional "triangle" from a slightly different view, showing that it is possible after all.

out by the perceptual system at a level below our conscious awareness.

In the *impossible* triangle (Fig. 8-12) we have two separate phenomena. First, when certain visual cues are present, we assume we are viewing a coherent object in three-dimensional space. Once we have made this assumption, our visual system assumes that the object we are viewing is in *general position*; that is, a slight change in our viewing position should leave our basic perception of the object unchanged. This second assumption appears to be hard-wired into our processing—even when we know the assumption is invalid for a given situation, we cannot avoid invoking it. There are undoubtedly other such hard-wired assumptions that nature has decided are good bets to make in general, even if there are occasional exceptions to their validity.

Finally, in the case of the subjective contour (Fig. 8-13), we use circumstantial evidence (e.g., gaps in the black circles) to deduce the presence of an occluding object. Once this decision has been made, other "channels" in the visual system alter our perceived interpretation of intensities, distance, etc., to make the complete interpretation a consistent one. This again shows the existence of a system capable of deductive reasoning and consistency maintenance operating below the level of our conscious awareness. Vision is not the simple task our introspection tells us about.

Visual Thinking, Visual Memory, and Cultural Factors

We briefly discuss three phenomena related to vision; (1) visual thinking, the use of images to aid the reasoning process;

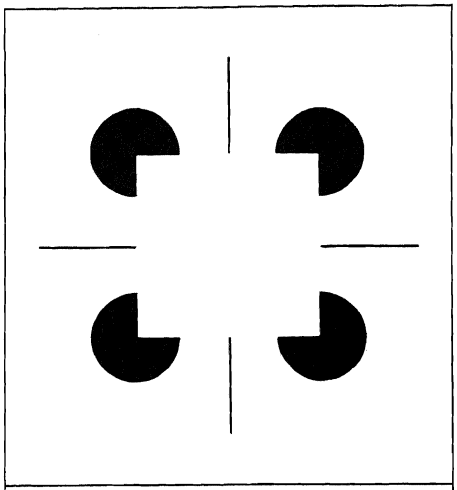


FIGURE 8-13 Subjective Contours.

The white square that appears to be occluding the black circles is an illusion—it is actually the same intensity as the background.

(2) visual memory, the use of visual images for remembering; and (3) pictorial perception and culture, the effect of our culture in teaching us how to understand pictures.

Visual Thinking. Some people believe that all thinking is basically perceptual in nature, and that the ancient dichotomy between seeing and thinking, between perceiving and reasoning, is false and misleading. In fact, some feel that visual thinking is a skill that can be learned, and that it improves with practice (see [Arnheim 69]). In Chapter 3 we described the visual thinking used by the physicist Richard Feynman, and by Friedrich Kekule, the chemist who discovered the structure of the benzene ring in a dream.

Visual Memory. It has been known since the time of the Greeks that a list of objects can be effectively memorized if they are set in the context of a vivid visual scene. The more vivid the scene, the better the objects are remembered. Typically, a reference set of images is used that is easy to remember, and that has a natural order. One such reference set using the numbers from one to ten is presented in Box 8-6. Another approach uses a mental traverse through a place known to the user, for example, a walk through one's house. Objects to be remembered are vividly associated with the reference images. Recall then consists of summoning up the reference images and remembering the object associated with each. The implication here is that we employ distinct mechanisms (and representations) for both symbolic and for iconic information, and that our storage and recall ability for iconic information is significantly better than that available for symbolic information.

Pictorial Perception and Culture. A picture is a pattern of lines and shaded areas on a flat surface that depicts some aspect of the real world. The ability to recognize objects in pictures is so common in most cultures that it is often taken for granted that such recognition is universal in man. Experiments described by Deregowski [Deregowski 74] show that people of one culture perceive a picture differently from people of another, and that the perception of pictures calls for some form of learning.

Conventions for depicting spatial arrangements of three-dimensional objects in a flat plane can give rise to difficulties

in perception. These conventions give the observer depth cues that tell him the objects are not all the same distance from him. Inability to interpret such cues is bound to lead to misunderstanding of the picture as a whole. For example, a typical cue is given when the larger of two known objects is drawn considerably smaller to indicate that it is farther away. Another cue is overlap, in which portions of nearer objects overlap and obscure portions of objects that are farther away. A third cue is perspective, the convergence of lines known to be parallel to suggest distance. In experiments carried out in many parts of Africa, it was found that both children and adults found it difficult to perceive depth in such pictorial material.

Some cultures use pictures that depict the essential characteristics of an object even if these characteristics cannot be seen from a single viewpoint. In such *split drawings*, an elephant appears in a top view with its four legs spread out, two on each side (Fig. 8-14). This split type of drawing to represent three-dimensional objects appears, and has been developed to a high artistic level, in various cultures. It is used by children in all cultures, even in those cultures where the style is considered manifestly wrong by adults.

BOX 8-6 A Memory Technique based on Images

The task of memorizing a list of thirty digits printed on a piece of paper, after a few seconds of inspection, would probably be impossible for all but a very small number of people. On the other hand, an aerial picture of the Golden Gate Bridge could easily be memorized so that at some future time it could be distinguished from a variety of other scenes. It seems clear that the means by which we try to remember the information required for these two visual tasks is considerably different. In the case of the numerals, our memorization is primarily based on assigning a specific name, the name of the numeral, to each depicted object. In the case of the natural scene, our memory is primarily that of a picture. This box indicates how names of objects can be remembered by connecting them to a set of reference images.

A set of ten reference images for use in memorization and recall are contained in the following rhyme. The images are easy to remember because the name of each image rhymes with its corresponding number in the sequence:

One is a bun
Two is a shoe
Three is a tree
Four is a door
Five is a hive
Six is a stick
Seven is heaven
Eight is a gate
Nine is wine
Ten is a hen.

To remember a shopping list of (1) eggs, (2) milk, (3) meat, and (4) apples, we would make vivid associations with the first four reference images. The more vivid the association, the stronger the retention of the item will be: (1) the egg has been crushed by the bun and sticky egg yolk is flowing out of the bun; (2) milk has been poured into a shoe and is running out of the eyelets of the shoe; (3) meat is hanging from the branches of a tree. It smells bad and the flies are buzzing around it; and (4) apples have been thrown at the door. They have left a trail of apple slime on the door, and there is apple mush in front of the door.

This memory technique not only enhances our ability to memorize a list of items, but it also permits us direct access to the *n*th item on the list. For example, to remember the third term, "three is a tree," we simply recall the image associated with the tree.

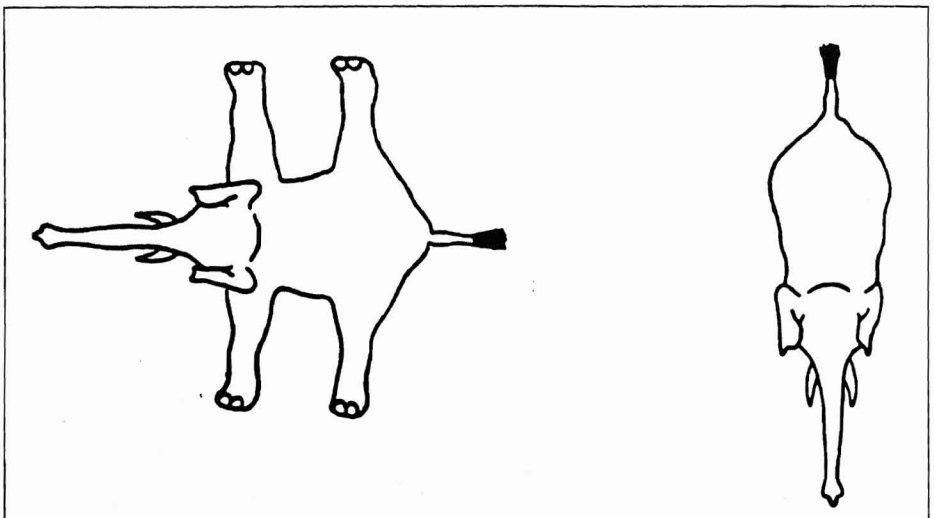


FIGURE 8-14
Conventions for the Two-Dimensional Representation of Three-Dimensional Objects are Culture-Specific.

Split-elephant drawing (left) was generally preferred by African children and adults to the top view perspective drawing (right). (From J. B. Deregowski, in: *Image, Object, and Illusion.* W. H. Freeman. San Francisco, Calif., 1974, with permission.)

DISCUSSION

What distinguishes living from inanimate objects? In one sense, inanimate objects respond to the current state of the universe; that is, their behavior can be simply described by reference to the *current* values of some set of physical variables. The behavior of living entities, on the other hand, appears to be most readily described in terms of the *future* values of these variables. All living organisms attempt to model and predict how their environment will change with time—their actions are based on these predictions; certainly the model of reality that the human invokes allows him to peer forward and backward in time. Metaphorically speaking, nervous tissue is a time machine that somehow has managed to free itself from moving in lock step with the clock that drives the inanimate universe.

Since the universe is too complex and interrelated for any practical model to completely capture the details of even a local environment, living organisms must continually compare current and predicted values to physical reality and adjust the relevant model parameters. The information to accomplish this task is provided by the senses. Different species have different models that impose different information needs, and thus their sensors measure different attributes of their

environment—in a sense, they live in different worlds.

No finite organism can completely model the infinite universe, but even more to the point, the senses can only provide a subset of the needed information; the organism must correct the measured values and guess at the needed missing ones. In most organisms these guesses are made automatically by algorithms embedded in their neural circuitry, and are the best bet the organism can make based on the past experience of its species. Even good bets occasionally fail, so it is likely that all organisms experience illusions. Indeed, even the best guesses can only be an approximation to reality—perception is a creative process.

In spite of the apparent diversity of organic life, nature has returned again and again to just a few solutions to the problems of perception. If we ignore minor differences in design, only two basic types of eyes have found widespread use, and the underlying neural components are almost identical in all animal life. We even find strong similarities in the circuits which do the initial processing of sensed data (e.g., the use of lateral inhibition). It is only in the later stages of neural processing that significant structural and functional differences can be found.

It would appear from neurological studies that most of the human brain is involved in visual perception, and we have earlier presented arguments to support the view that intelligence evolved to support the perceptual process. When a person says "I see" after solving a difficult mathematical or conceptual problem, he is voicing a piece of wisdom that we are just beginning to appreciate, that his perceptual machinery, diverted from its nominal tasks, probably played a substantial role in producing the solution.

Appendixes

8-1

Color Vision and Light

At any instant of time, the light incident on a surface, or received by the human eye, can be characterized by its intensity (energy) and color (frequency). The spectral peak of ambient light incident on the surface of the earth changes throughout the course of the day (see Fig. 8-15). During most of the day[4] the sun provides the main source of illumination and the ambient peak hovers around 550 nanometers. However, during the twilight period, when the sun's rays in their longer path through the earth's atmosphere are

[4]During moonlit evenings, the spectral peak is similar to that of broad daylight since the moon reflects the sun's light rather uniformly over the visible spectrum. The other main source of illumination at night is the airglow that results from oxygen activation in the earth's atmosphere; this source has a sharp "green" spectral peak at 558 nm. Starlight provides a faint source of illumination with a spectral peak shifted somewhat toward the red end of the spectrum.

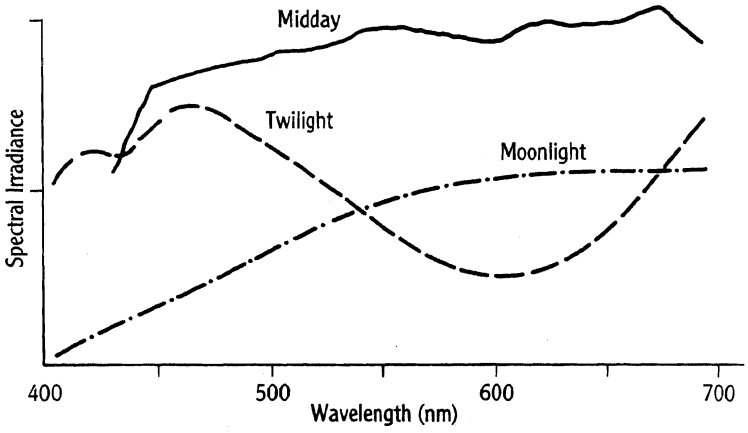


FIGURE 8-15
Spectral Irradiance During Midday, Twilight, and Moonlight.

(After J. N. Lythgoe. *The Ecology of Vision.* Clarendon Press, Oxford, 1979, p. 95.)

more scattered by small airborne particles, the ambient light peak shifts toward the blue end of the spectrum (the Purkinje shift). For day-active animals, especially those that need to move about at twilight, the sensitive rod system, with a spectral peak shifted toward the blue frequencies and a high-resolution cone system tuned to operate in direct sunlight, is an excellent adaptation to environmental conditions (Fig. 8-16).

There are many considerations relevant to the evolutionary appearance of color vision in a species. For example, many mammals are nocturnal, or evolved from nocturnal ancestors, and such species would be unlikely to find a use for color vision which, as we previously noted, requires high light levels. However, it has been argued [Lythgoe 79] that a visual system with two sets of detectors, one tuned to that of the spectral peak of the ambient light, and the other set of detectors with their spectral peak offset somewhat, can provide a more effective means for detecting a bright target against a dark background than a singly tuned set of detectors. Thus, even though not evolved for this purpose, such a two-spectral response system would provide the sensory information required for the subsequent evolution of color vision.

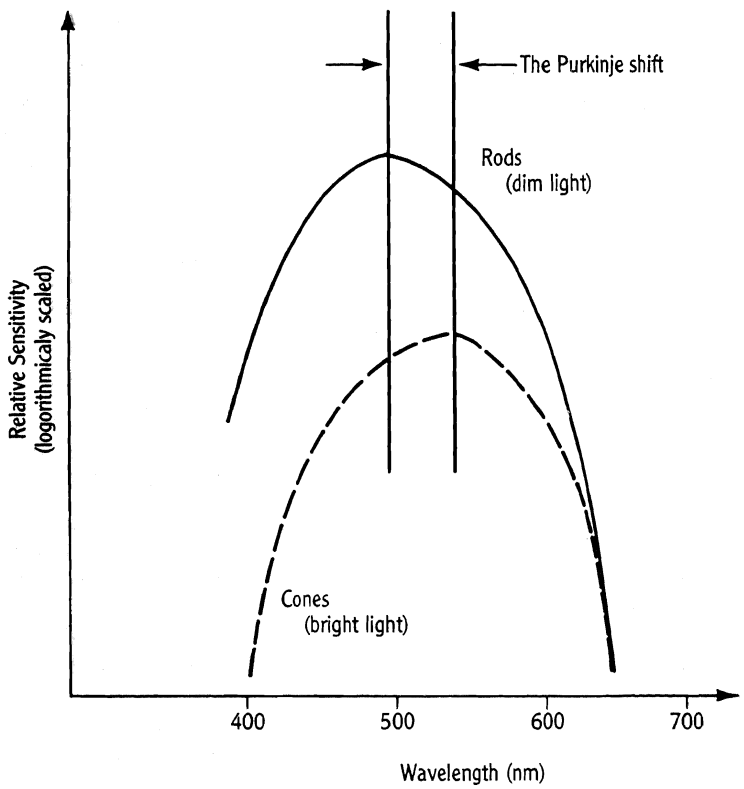


FIGURE 8-16
Spectral Sensitivity of the Human Eye.

Only the nominal shape of the sensitivity curve is shown here. Precise values can be found in G. Wald. *Science* 101; 653-658, 1945.

APPENDIX 8-1

Given the appropriate environmental conditions, color enhances an organism's ability to identify visible objects, determine their physical properties (e.g., ripe compared to immature fruit), and can play an important role in visual communication. For example, sexual displays and body markings are typically based on color cues in organisms with color vision.

How does the eye measure the spectral attributes of light energy impinging on it, how is this information represented internally, and how is it transformed into the subjective impression we call color?

The Young-Helmholtz theory, formulated in the early part of the nineteenth century, asserts that there are three color-sensitive types of receptors in the eye which correspond respectively to red, green, and blue, and that all color perception is the result of the relative strength of the signals received from these three receptor systems. This theory, while possibly valid for simple patches of light, is not sufficient to explain human color perception in complex natural scenes. Edwin Land [Land 59] has demonstrated that our final perception of color at any point in a scene is dependent on colors perceived in other parts of the scene, and that in complex scenes we can perceive colors that cannot be exactly reproduced by a simple mixture of the three primary colors (e.g., highly saturated brown)—in fact, he demonstrated that a mixture of, say, red and white light could induce the human visual system to perceive a realistically colored scene. Other possible problems with the until recently dominant Young-Helmholtz theory includes the fact that the "blue" (short wavelength) pigmented cones are not spatially distributed in the same manner as the red and green cones—they are not as common and are actually absent from the central fovea.[5] The above results imply that any account of color vision based strictly on local stimulation of retinal nerve cells is doomed to fail—color perception involves processes occurring at higher levels in the brain. Nevertheless, we can still profit by understanding the local sensory apparatus the eye employs to detect and encode the spectral attributes of incident light.

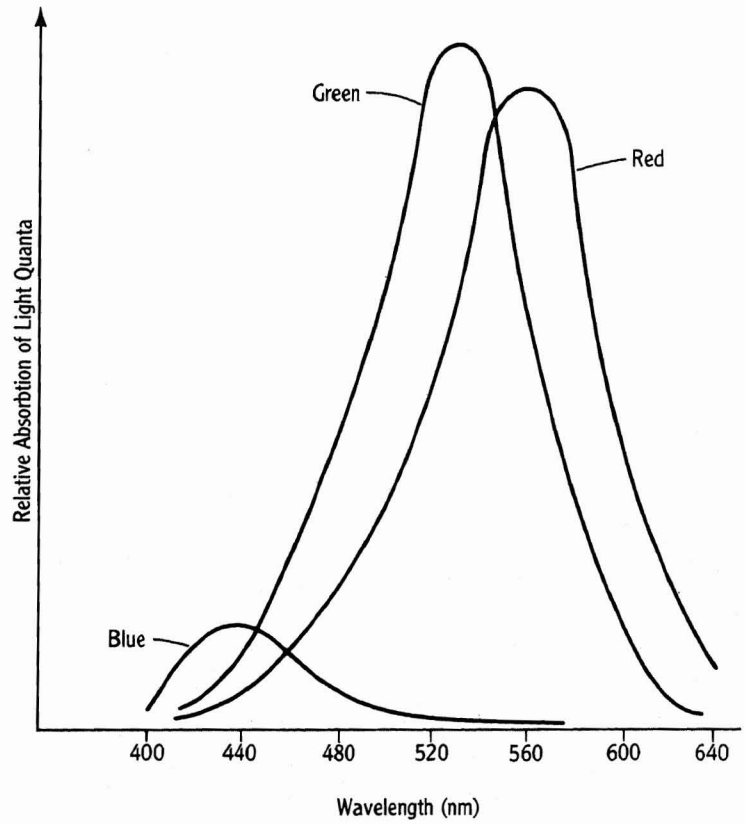


FIGURE 8-17
Absorption of Light Energy by Three Populations of Cones in the Human Eye.

Only the nominal shape of the absorption spectra is shown here. More precise values can be found in G. Wald. *Science* 145; 1007-1017, 1964.

[5]The blue cones seem to be distinguished from the red and green cones in many other important respects, e.g., they do not contribute to the perception of boundaries between differently colored regions, and make little, if any, contribution to the total luminance signal.

The photosensitive part of the eye is a mosaic of rod and cone receptor cells as described in the main text and Box 8-2. Only the cone cells are directly involved in color vision; there are approximately 6 million such cells distributed over each retina, but they are most densely concentrated in the fovea of the eye where there are no rod cells. The rod-free region of each eye (2 degrees in diameter) contains approximately 50,000 cones.

The retinal image consists of a pattern of light energy. This image is transformed into a pattern of nerve activity by the presence of photosensitive pigments in the rods and cones that absorb part of the incident light energy. The rod pigment, rhodopsin, has been successfully extracted and studied. While not as well understood, all the cones appear to be anatomically alike (although their connections differ), and are distributed into at least three populations with distinct spectral responses (see Fig. 8-17). Some recent theories of color vision suggest the presence of receptors with a fourth spectral response, possibly implying an indirect role in color perception for the rod cells.

8-2

Stereo Depth Perception and the Structure of the Human Visual Cortex

With the exception of about two percent of the population, the normal human visual system can convert the overlapping flat images projected onto the retinas of its two eyes into a three-dimensional model of the surrounding environment. We see the world in depth, a luxury shared with most other primates and many predators, e.g., predatory birds. In contrast, many two-eyed animals, such as the rabbit and the pigeon, have panoramic rather than stereo vision: their eyes are placed primarily to look in different directions, rather than to provide the overlapping coverage needed for binocular depth perception.

If we can match corresponding points or objects in the two retinal images, then simple geometric triangulation can be employed to compute the distance (depth) to these objects.[6] The machinery that the human visual system employs to perform the stereo function, while not completely understood, appears to be organized as described below.

Each retinal ganglion cell has a receptive field (the patch of retinal receptor cells supplying the ganglion cell) that consists of an excitatory center and an inhibitory surround. Thus, each such ganglion cell responds best to a roughly circular spot of light of a particular size in a particular part of the visual field. The path from the receptor cells in the retina to the cells in the visual cortex is indicated schematically in Fig. 8-18.

The first of the two major transformations performed by the visual cortex is the integration of information from the retinal ganglion cells so that the cortical cells respond to specifically oriented line segments rather than to spots of light. Depending on the particular cell, its maximum response will be triggered by a moving bright line on a dark background, or the reverse, or it may be a moving boundary between light and dark regions. The orientation of the line, as well as its speed and direction of motion, are also important; note that head and eye movement will cause even a static object in the scene to move across the retina. There appears to be a hierarchy of cell types, with simpler ones feeding the more complex cells. Neurons in the visual cortex with orientation specificity vary in their complexity. "Simple" cells appear to obtain their inputs from a line of retinal cells, and the far more numerous "complex" cells behave as though they receive their input from a number of simple cells, all with the same receptive field orientation, but differing slightly in the exact location of their fields.

The second major transformation performed by the visual cortex is to combine inputs from the two eyes; aside from seeing things in depth, we see a single world, even though the two eyes provide slightly different views of this world. The cells in the visual cortex receiving direct input from the retinas (through the lateral geniculate "relay stations") are all simple "monocular"

[6]Computer stereo techniques are discussed in Chapter 9, Box 9-5.

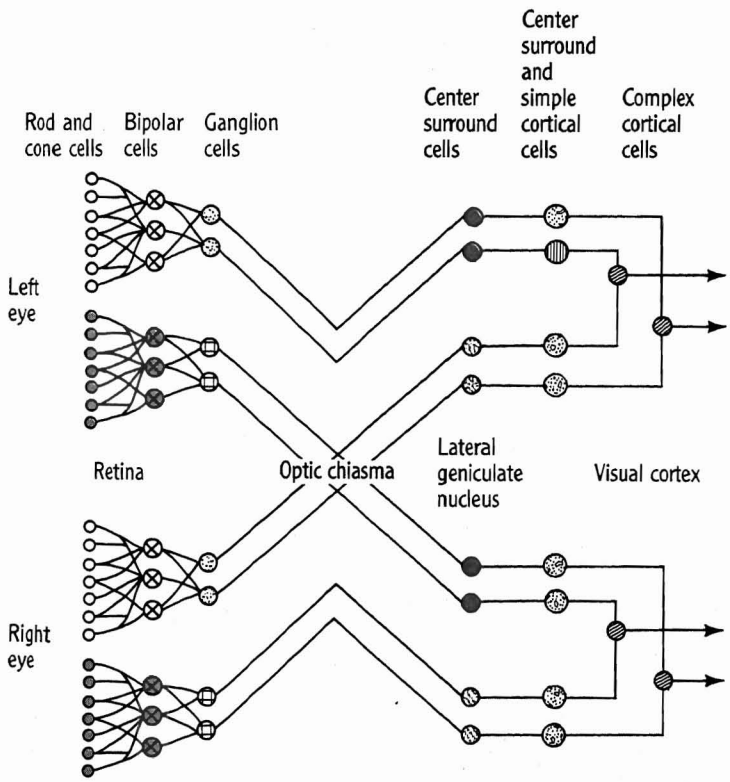


FIGURE 8-18
Schematic Diagram of the Path from Receptor Cells to the Visual Cortex in the Human Stereo System.

Approximately half of the complex cells are binocular and the other half monocular; almost all of the center-surround and simple cells are monocular.

cells that receive stimulation from exactly one of the two eyes, but not both. About half of the complex cells are monocular, and the rest are binocular, i.e., they can be influenced independently by both eyes. The left and right receptive field inputs to a binocular complex cell are generally identical in all respects, except that the stimulation ability of one eye typically dominates the other; all degrees of dominance can be found.

The highly specific stimulus pattern requirements for the firing of "complex" binocular neurons could provide a means for identifying the parts of the left and right images corresponding to the same features. Because the number of identical features in any local region of the image is likely to be small, similar features lying in roughly corresponding regions on each retina can be assumed to correspond to the same real-world object. Since the right and left visual fields depict objects at a variety of different depths in the world, these fields cannot be coherently superimposed, and there is evidence [Pettigrew 79] that "disparity-specific" complex binocular neurons (for a range of disparities) provide local depth information.

The visual cortex is subdivided into roughly parallel columns of tissue, (swirled as in Fig. 8-19, rather than planar, as shown in the simplified schematic drawing of Fig. 8-20), approximately normal to the surface of the cortex. Each column is partitioned into 50 micrometer-thick slabs containing neurons with like receptive field orientation; adjacent slabs have 10 degree shifts in their line orientation. Slabs are arranged into coherent blocks with each block containing a right eye dominant column, and a left eye

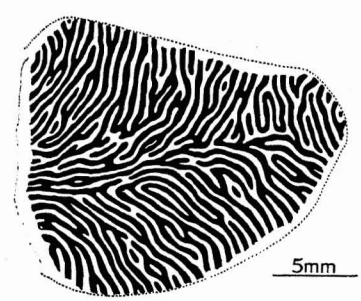


FIGURE 8-19
Section of Monkey Brain Showing Ocular-Dominance Columns.

(From D. H. Hubel and T. N. Wiesel. *Proceedings of the Royal Society* 198; 35, 1977. Reprinted with permission.)

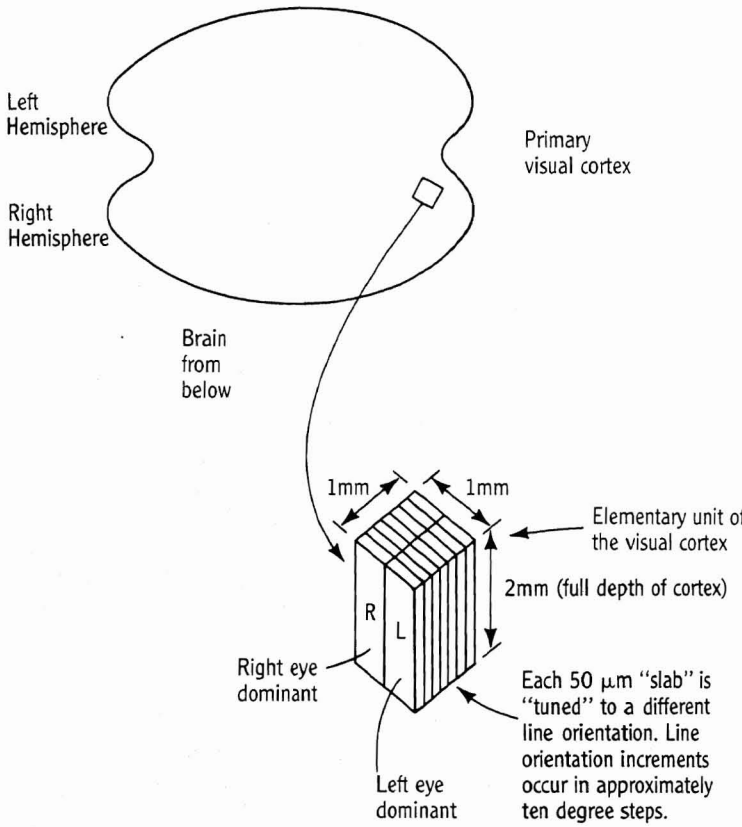


FIGURE 8-20
Schematic Depiction of the Elementary Unit of the Visual Cortex.

dominant column. Blocks near the center of gaze have small receptive fields; blocks corresponding to receptive fields of increasing eccentricity (further out in the visual field) have large receptive fields.

As we show in the following chapter on computational vision, binocular stereo is not the only method for obtaining depth from two-dimensional images. A single imaging sensor, e.g., a single eye, can recover depth information by viewing a sequence of images, and even a single image offers many depth cues. For example, objects at different depths produce retinal images that move at different speeds when the head is moved, a phenomenon known as *head movement parallax*. Other monocular depth cues include:

- The muscular tension needed to bring different objects into focus, *accommodation*
- Occlusion of more distant objects by near ones
- "Aerial perspective" in which distant objects tend to be hazy and assume a bluish tint
- More distant objects which generally appear higher up in the visual field
- "Linear perspective," i.e., convergence of parallel receding lines
- Shading and texture gradients, as shown in Fig. 8-21, which encode depth information.

Nevertheless, the speed, reliability, and accuracy of binocular stereo cannot be matched by the above monocular approaches. One additional advantage of binocular stereo over monocular vision occurs in recognizing patterns. Even when each individual eye fails to see a camouflaged object, a binocular stereo system can still fuse local cues into a clearly visible depth image, [Julesz 74].

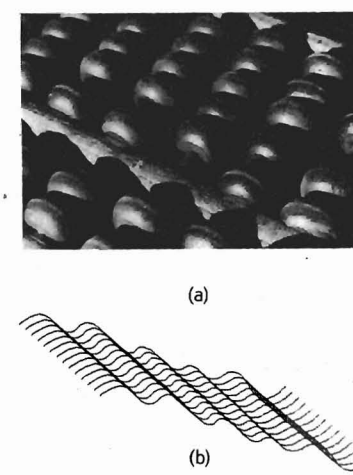


FIGURE 8-21
Obtaining Depth From Shading and Texture Cues.

(a) Depth from shading (photo by O. Firschein); (b) Depth from texture.