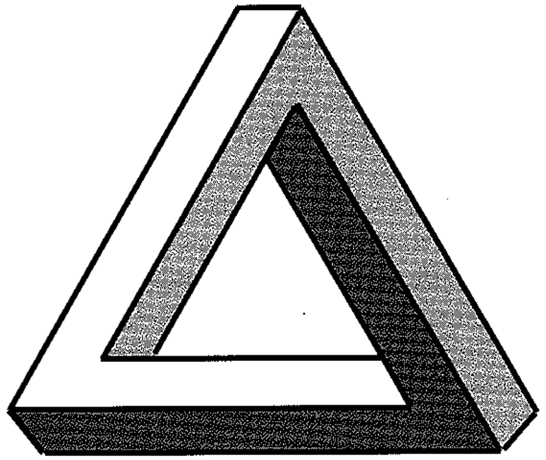
1

An Introduction to Vision Science



- 1.1 Visual Perception
  - 1.1.1 Defining Visual Perception
  - 1.1.2 The Evolutionary Utility of Vision
  - 1.1.3 Perception as a Constructive Act
    - Adaptation and Aftereffects
    - Reality and Illusion
    - Ambiguous Figures
  - 1.1.4 Perception as Modeling the Environment
    - Visual Completion
    - Impossible Objects
    - Predicting the Future
  - 1.1.5 Perception as Apprehension of Meaning
    - Classification
    - Attention and Consciousness
- 1.2 Optical Information
  - 1.2.1 The Behavior of Light
    - Illumination
    - Interaction with Surfaces
    - The Ambient Optic Array
  - 1.2.2 The Formation of Images
    - Optical Images
    - Projective Geometry
    - Perspective and Orthographic Projection
  - 1.2.3 Vision as an "Inverse" Problem
- 1.3 Visual Systems
  - 1.3.1 The Human Eye
    - Eye and Brain
    - Anatomy of the Eye
    - Physiological Optics
  - 1.3.2 The Retina
    - Neurons
    - Photoreceptors
    - Peculiarities of Retinal Design
    - Pathways to the Brain
  - 1.3.3 Visual Cortex
    - Localization of Function
    - Occipital Cortex
    - Parietal and Temporal Cortex
    - Mapping Visual Cortex
    - The Physiological Pathways Hypothesis

Most of us take completely for granted our ability to see the world around us. How we do it seems no great mystery: We just open our eyes and look! When we do, we perceive a complex array of meaningful objects located in three-dimensional space. For example, Figure 1.1.1 shows a typical scene on the Berkeley campus of the University of California: some students walking through Sather Gate, with trees and the distinctive Campanile bell tower in the background. We perceive all this so quickly and effortlessly that it is hard to imagine there being anything very complicated about it. Yet, when viewed critically as an ability that must be explained, visual perception is so incredibly complex that it seems almost a miracle that we can do it at all.

The rich fabric of visual experience that results from viewing natural scenes like the one in Figure 1.1.1 arises when the neural tissues at the back of the eyes are stimulated by a two-dimensional pattern of light that includes only bits and pieces of the objects being perceived. Most of the Campanile, for example, is hidden behind the trees, and parts of the trees are occluded by the towers of the gate. We don't perceive the Campanile as floating in the air or the trees as having tower-shaped holes cut in them where we cannot currently see them. Even objects that seem to be fully visible, such as the gate towers and the students, can be seen only in part because their far sides are occluded by their near sides. How, then, are we able so quickly and effortlessly to perceive the meaningful, coherent, three-dimensional scene that we obviously do experience from the incomplete, two-dimensional pattern of light that enters our eyes?

This is the fundamental question of vision, and the rest of this book is an extended inquiry into its answer from a scientific point of view. It is no accident that I began the book with a question, for the first step in any scientific enterprise is asking questions about things that are normally taken for granted. Many more questions will prove to be important in the course of our discussions. A few of them are listed here:

- Why do objects appear colored?
- How can we determine whether an object is large and distant or small and close?
- How do we perceive which regions in a visual image are parts of the same object?
- How do we know what the objects that we see are for?
- How can we tell whether we are moving relative to objects in the environment or they are moving relative to us?
- Do newborn babies see the world in the same way we do?
- Can people "see" without being *aware* of what they see?

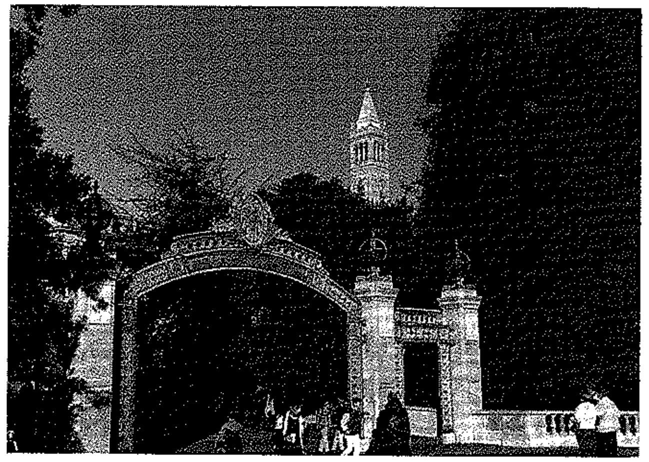


Figure 1.1.1 A real-world scene on the Berkeley campus. Viewers perceive students walking near Sather Gate with the Campanile bell tower behind a row of trees, even though none of these objects are visible in their entirety. Perception must somehow infer the bottom of the bell tower, the trees behind the gate towers, and the far sides of all these objects from the parts that are visible.

Posing such questions is just the first step of our journey, however, for we must then try to find the answers. The majority of this book will be devoted to describing how vision scientists do this and what they have discovered about seeing as a result. It turns out that different parts of the answers come from a variety of different disciplines—biology, psychology, computer science, neuropsychology, linguistics, and cognitive anthropology—all of which are part of the emerging field of **cognitive science**. The premise of cognitive science is that the problems of cognition will be solved more quickly and completely by attacking them from as many perspectives as possible.

The modern study of vision certainly fits this interdisciplinary mold. It is rapidly becoming a tightly integrated field at the intersection of many related disciplines, each of which provides different pieces of the jigsaw puzzle. This interdisciplinary field, which I will call **vision science**, is part of cognitive science. In this book, I try to convey a sense of the excitement that it is generating among the scientists who study vision and of the promise that it holds for reaching a new understanding about how we see.

In this initial chapter, I will set the stage for the rest of the book by providing an introductory framework for understanding vision in terms of three domains:

1. phenomena of visual perception,
2. the nature of optical information, and
3. the physiology of the visual nervous system.

The view presented in this book is that an understanding of all three domains and the relations among them is required to explain vision. In the first section of this chapter, we will consider the nature of visual perception itself from an evolutionary perspective, asking what it is for. We will define it, talk about some of its most salient properties, and examine its usefulness in coupling organisms to their environments for survival. Next, we will consider the nature of optical information, because all vision ultimately rests on the structure of light reflected into the eyes from surfaces in the environment. Finally, we will describe the physiology of the part of the nervous system that underlies our ability to see. The eyes are important, to be sure, but just as crucial are huge portions of the brain, much of which vision scientists are only beginning to understand. In each domain, the coverage in this introductory chapter will be rudimentary and incomplete. But it is important to realize from the very beginning that only by understanding all three domains and the relations among them can we achieve a full and satisfying scientific explanation of what it means to see. What we learn here forms the scaffold onto which we can fit the more detailed presentations in later chapters.

1.1 Visual Perception

Until now, I have been taking for granted that you know what I mean by "visual perception." I do so in large part because I assume that you are reading the words on this page using your own eyes and therefore know what visual experiences are like. Before we go any further, however, we ought to have an explicit definition.

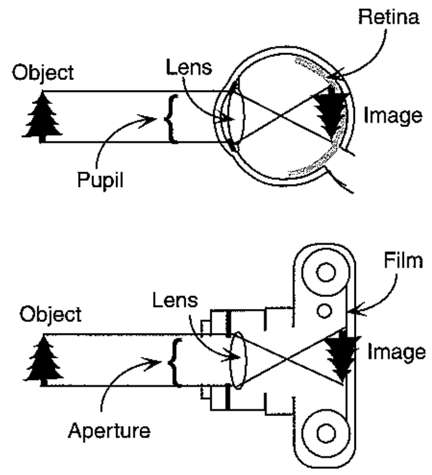


Figure 1.1.2 The eye-camera analogy. The eye is much like a camera in the nature of its optics: Both form an upside-down image by admitting light through a variable-sized opening and focusing it on a two-dimensional surface using a transparent lens.

1.1.1 Defining Visual Perception

In the context of this book, **visual perception** will be defined as the process of acquiring knowledge about environmental objects and events by extracting information from the light they emit or reflect. Several aspects of this definition are worth noting:

1. Visual perception concerns the *acquisition of knowledge.* This means that vision is fundamentally a cognitive activity (from the Latin *cognoscere*, meaning *to know* or *learn*), distinct from purely optical processes such as photographic ones. Certain physical similarities between cameras and eyes suggest that perception is analogous to taking a picture, as illustrated in Figure 1.1.2. There are indeed important similarities between eyes and cameras in terms of optical phenomena, as we will see in Section 1.2, but there are no similarities whatever in terms of *perceptual* phenomena. Cameras have no perceptual capabilities at all; that is, they do not *know* anything about the scenes they record. Photographic images merely contain information, whereas sighted people and animals acquire knowledge about their environments. It is this knowledge that enables perceivers to act appropriately in a given situation.

2. The knowledge achieved by visual perception concerns *objects and events in the environment.* Perception is not merely about an observer's subjective visual experiences, because we would not say that even highly detailed hallucinations or visual images would count as visual perception. We will, in fact, be very interested in the nature of people's subjective experience—particularly in Chapter 13 when we discuss visual awareness in detail—but it is part of visual perception only when it signifies something about the nature of external reality.

3. Visual knowledge about the environment is obtained by *extracting information.* This aspect of our definition implies a certain "metatheoretical" approach to understanding visual perception and cognition, one that is based on the concept of information and how it is processed. We will discuss this **information processing** approach more fully in Chapter 2, but for now suffice it to say that it is an approach that allows vision scientists to talk about how people see in the same terms as they talk about how computers might be programmed to see. Again, we will have more to say about the prospects for sighted computers in Chapter 13 when we discuss the problem of visual awareness.

4. The information that is processed in visual perception comes from the light that is *emitted or reflected by objects.* Optical information is the foundation of all vision. It results from the way in which physical surfaces interact with light in the environment. Because this restructuring of light determines what information about objects is available for vision in the first place, it is the appropriate starting point for any systematic analysis of vision (Gibson, 1950). As we will see in Section 1.2, most of the early problems in understanding vision arise from the difficulty of undoing what happens when light projects from a three-dimensional world onto the two-dimensional surfaces at the back of the eyes. The study of what information is contained in these projected images is therefore an important frontier of research in vision science, one that computational theorists are constantly exploring to find new sources of information that vision might employ.

1.1.2 The Evolutionary Utility of Vision

Now that we have considered what visual perception *is*, we should ask what it is *for*. Given its biological importance to a wide variety of animals, the answer must be that *vision evolved to aid in the survival and successful reproduction of organisms.* Desirable objects and situations—such as nourishing food, protective shelter, and desirable mates—must be sought out and approached. Dangerous objects and situations—such as precipitous drops, falling objects, and hungry or angry predators—must be avoided or fled from. Thus, to behave in an evolutionarily adaptive manner, we must somehow get information about what objects are present in the world around us, where they are located, and what opportunities they afford us. All of the senses—seeing, hearing, touching, tasting, and smelling—participate in this endeavor.

There are some creatures for which nonvisual senses play the dominant role—such as hearing in the navigation of bats—but for *homo sapiens*, as well as for many other species, vision is preeminent. The reason is that vision provides spatially accurate information from a distance. It gives a perceiver highly reliable information about the locations and properties of environmental objects while they are safely distant. Hearing and smell sometimes provide information from even greater distances, but they are seldom as accurate in identifying and locating objects, at least for humans. Touch and taste provide the most direct information about certain properties of objects because they operate only when the objects are actually in contact with our bodies, but they provide no information at all from farther distances.

Evolutionarily speaking, visual perception is useful only if it is reasonably accurate. If the information in light were insufficient to tell one object from another or to know where they are in space, vision never would have evolved to the exquisite level it has in humans. In fact, light is an enormously rich source of environmental information, and human vision exploits it to a high degree. Indeed, vision is useful precisely because it is so accurate. By and large, *what you see is what you get.* When this is true, we have what is called **veridical perception** (from the Latin *veridicus* meaning *to say truthfully*): perception that is consistent with the actual state of affairs in the environment. This is almost always the case with vision, and it is probably why we take vision so completely for granted. It seems like a perfectly clear window onto reality. But is it really?

In the remainder of this section, I will argue that perception is *not* a clear window onto reality, but an actively constructed, meaningful model of the environment that allows perceivers to predict what will happen in the future so that they can take appropriate action and thereby increase their chances of survival. In making this argument, we will touch on several of the most important phenomena of visual perception, ones to which we will return at various points later in this book.

1.1.3 Perception as a Constructive Act

The first issue that we must challenge is whether what you see is *necessarily* what you get: Is visual perception unerringly veridical? This question is important because the answer will tell us whether or not vision should be conceived as a "clear window onto reality."

Adaptation and Aftereffects. One kind of evidence that visual experience is not a clear window onto reality is provided by the fact that visual perception changes over time as it adapts to particular conditions. When you first enter a darkened movie theater on a bright afternoon, for instance, you cannot see much except the images on the screen. After just a few minutes, however, you can see the people seated near you, and after 20 minutes or so, you can see the whole theater surprisingly well. This increase in sensitivity to light is called **dark adaptation**. The theater walls and distant people were there all along; you just could not see them at first because your visual system was not sensitive enough.

Another everyday example of dark adaptation arises in gazing at stars. When you leave a brightly lit room to go outside on a cloudless night, the stars at first may seem disappointingly dim and few in number. After you have been outside for just a few minutes, however, they appear considerably brighter and far more numerous. And after 20–30 minutes, you see the heavens awash with thousands of stars that you could not see at first. The reason is not that the stars emit more light as you continue to gaze at them, but that your visual system has become more sensitive to the light that they do emit.

Adaptation is a very general phenomenon in visual perception. As we will see in many later chapters, visual experience becomes less intense[1] as a result of prolonged exposure to a wide variety of different kinds of stimulation: color, orientation, size, and motion, to name just a few. These changes in visual experience show that visual perception is not always a clear window onto reality because we have different visual experiences of the same physical environment at different stages of adaptation. What changes over time is our visual system, not the environment. Even so, one could sensibly argue that although some things may fail to be perceived because of adaptation, whatever *is* perceived is an accurate reflection of reality. This modified view can be shown to be incorrect, however, by another result of prolonged or very intense stimulation: the existence of visual aftereffects.

When someone takes a picture of you with a flash, you first experience a blinding blaze of light. This is a veridical perception, but it is followed by a prolonged experience of a dark spot where you saw the initial flash. This **afterimage** is superimposed on whatever else you look at for the next few minutes, altering your subsequent visual experiences so that you see something that is not there. Clearly, this is not veridical perception because the afterimage lasts long after the physical flash is gone.

Not all aftereffects make you see things that are not there; others cause you to misperceive properties of visible objects. Figure 1.1.3 shows an example called an **orientation aftereffect**. First, examine the two striped gratings on the right to convince yourself that they are vertical and identical to each other. Then look at the two tilted gratings on the left for about a minute by fixating on the bar between them and moving your gaze back and forth along it. Then look at the square between the two gratings on the right. The top grating now looks tilted to the left, and the bottom one looks tilted to the right. These errors in perception are further evidence that what you see results from an interaction between the external world and the present state of your visual nervous system.

Reality and Illusion. There are many other cases of systematically nonveridical perceptions, usually called **illusions**. One particularly striking example with which you may already be familiar is the **moon illusion**. You

---

[1] It may be confusing that during dark adaptation the visual system becomes *more* sensitive to light rather than less. This apparent difference from other forms of adaptation can be eliminated if you realize that during dark adaptation the visual system is, in a sense, becoming less sensitive to the *dark*.

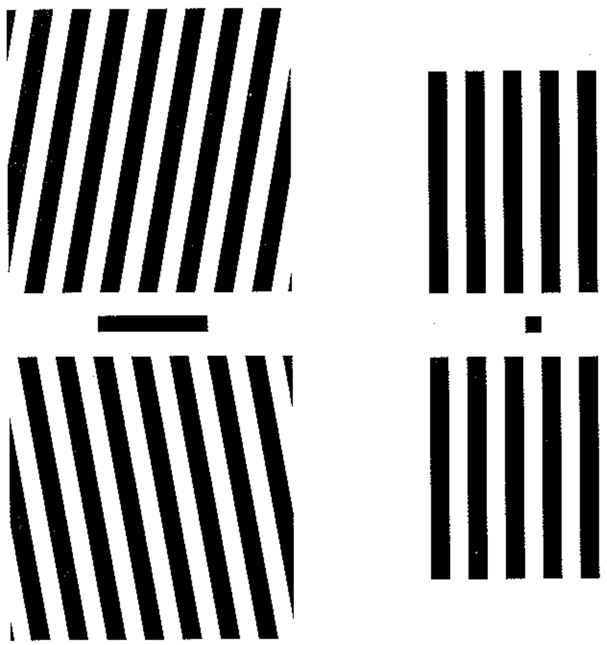


Figure 1.1.3 An orientation aftereffect. Run your eyes along the central bar between the gratings on the left for 30–60 seconds. Then look at the square between the two identical gratings on the right. The upper grating should now appear tilted to the left of vertical and the lower grating tilted to the right.

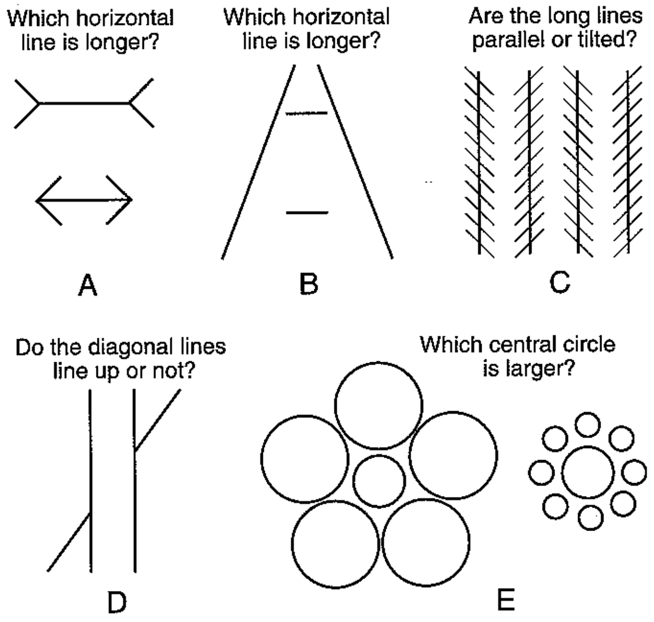


Figure 1.1.4 Visual illusions. Although they do not appear to be so, the two arrow shafts are the same length in A, the horizontal lines are identical in B, the long lines are vertical in C, the diagonal lines are collinear in D, and the middle circles are equal in size in E.

have probably noticed that the moon looks much larger when it is close to the horizon than it does when it is high in the night sky. Have you ever thought about why?

Many people think that it is due to refractive distortions introduced by the atmosphere. Others suppose that it is due to the shape of the moon's orbit. In fact, the optical size of the moon is entirely constant throughout its journey across the sky. You can demonstrate this by taking a series of photographs as the moon rises; the size of its photographic image will not change in the slightest. It is only our perception of the moon's size that changes. In this respect, it is indeed an illusion—a nonveridical perception—because its image in our eyes does not change size any more than it does in the photographs. In Chapter 7, we will discuss in detail why the moon illusion occurs (Kaufman & Rock, 1962; Rock & Kaufman, 1962). For right now, the important thing is just to realize that our perception of the apparent difference in the moon's size at different heights in the night sky is illusory.

There are many other illusions demonstrating that visual perception is less than entirely accurate. Some of these are illustrated in Figure 1.1.4. The two arrow shafts in A are actually equal in length; the horizontal lines in B are actually the same size; the long lines in C are actually vertical and parallel; the diagonal lines in D are actually collinear; and the two central circles in E are actually equal in size. In each case, our visual system is somehow fooled into making perceptual errors about seemingly obvious properties of simple line drawings. These illusions support the conclusion that perception is indeed fallible and therefore cannot be considered a clear window onto external reality. The reality that vision provides must therefore be, at least in part, a construction by the visual system that results from the way it processes information in light. As we shall see, the nature of this construction implies certain hidden assumptions, of which we have no conscious knowledge, and when these assumptions are untrue, illusions result. This topic will appear frequently in various forms throughout this book, particularly in Chapter 7.

It is easy to get so carried away by illusions that one starts to think of visual perception as grossly inaccurate and unreliable. This is a mistake. As we said earlier, vision is evolutionarily useful to the extent that it is accurate—or, rather, as accurate as it needs to be. Even illusory perceptions are quite accurate in most respects. For instance, there really *are* two short horizontal lines and two long oblique ones in Figure 1.1.4B, none of which touch each other. The only aspect that is inaccurately perceived is the single illusory property—the relative lengths of the horizontal lines—and the discrepancy between perception and reality is actually quite modest. Moreover, illusions such as these are not terribly obvious in everyday life; they occur most frequently in books about perception.

All things considered, then, it would be erroneous to believe that the relatively minor errors introduced by vision overshadow its evolutionary usefulness. Moreover, we will later consider the possibility that the perceptual errors produced by these illusions may actually be relatively harmless side effects of the same processes that produce veridical perception under ordinary circumstances (see Chapters 5 and 7). The important point for the present discussion is that the existence of illusions proves convincingly that perception is not just a simple registration of objective reality. There is a great deal more to it than that.

Once the lesson of illusions has been learned, it is easier to see that there is really no good reason why perception *should* be a clear window onto reality. The objects that we so effortlessly perceive are not the direct cause of our perceptions. Rather, perceptions are caused by the two-dimensional patterns of light that stimulate our eyes. (To demonstrate the truth of this assertion, just close your eyes. The objects are still present, but they no longer give rise to visual experiences.) To provide us with information about the three-dimensional environment, vision must therefore be an **interpretive process** that somehow transforms complex, moving, two-dimensional patterns of light at the back of the eyes into stable perceptions of three-dimensional objects in three-dimensional space. We must therefore conclude that the objects we perceive are actually interpretations based on the structure of images rather than direct registrations of physical reality.

Ambiguous Figures. Potent demonstrations of the interpretive nature of vision come from **ambiguous figures**: single images that can give rise to two or more distinct perceptions. Several compelling examples are shown in Figure 1.1.5. The vase/faces figure in part A can be perceived either as a white vase on a black background (A1) or as two black faces in silhouette against a white background (A2). The Necker cube in Figure 1.1.5B can be perceived as a cube in two different orientations relative to the viewer: with the observer looking down and to the right at the cube (B1) or looking up and to the left (B2). When the percept "reverses," the interpretation of the depth relations among the lines change; front edges become back ones, and back edges become front ones. A somewhat different kind of ambiguity is illustrated in Figure 1.1.5C. This drawing can be seen either as a duck facing left (C1) or as a rabbit facing right (C2). The interpretation of lines again shifts from one percept to the other, but this time the change is from one body part to another: The duck's bill becomes the rabbit's ears, and a bump on the back of the duck's head becomes the rabbit's nose.

There are two important things to notice about your perception of these ambiguous figures as you look at them. First, the interpretations are *mutually exclusive.* That

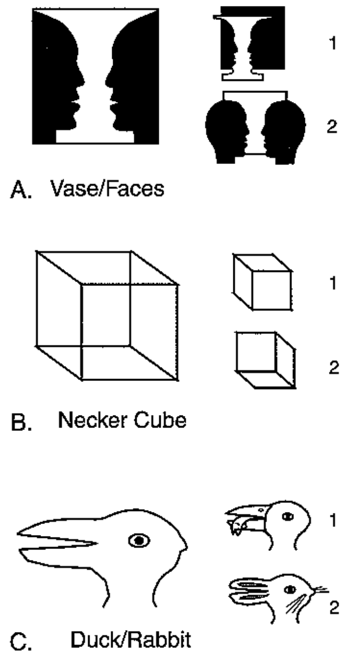


Figure 1.1.5 Ambiguous figures. Figure A can be seen either as a white vase against a black background or as a pair of black faces against a white background. Figure B can be seen as a cube viewed from above or below. Figure C can be seen as a duck (facing left) or a rabbit (facing right).

is, you perceive just one of them at a time: a duck *or* a rabbit, not both. This is consistent with the idea that perception involves the construction of an interpretive model because only one such model can be fit to the sensory data at one time. Second, once you have seen both interpretations, they are **multistable perceptions**, that is, dynamic perceptions in which the two possibilities alternate back and forth as you continue to look at them. This suggests that the two models compete with each other in some sense, with the winner eventually getting "tired out" so that the loser gains the advantage. These phenomena can be modeled in neural network theories that capture some of the biological properties of neural circuits, as we will see in Chapter 6.

1.1.4 Perception as Modeling the Environment

Ambiguous figures demonstrate the constructive nature of perception because they show that perceivers interpret visual stimulation and that more than one interpretation is sometimes possible. If perception were completely determined by the light stimulating the eye, there would be no ambiguous figures because each pattern of stimulation would map onto a unique percept. This position is obviously incorrect. Something more complex and creative is occurring in vision, going beyond the information strictly given in the light that stimulates our eyes (Bruner, 1973).

But *how* does vision go beyond the optical information, and *why*? The currently favored answer is that *the observer is constructing a model of what environmental situation might have produced the observed pattern of sensory stimulation.* The important and challenging idea here is that people's perceptions actually correspond to the models that their visual systems have constructed rather than (or in addition to) the sensory stimulation on which the models are based. That is why perceptions can be illusory and ambiguous despite the nonillusory and unambiguous status of the raw optical images on which they are based. Sometimes we construct the wrong model, and sometimes we construct two or more models that are equally plausible, given the available information.

The view that the purpose of the visual system is to construct models of the environment was initially set forth by the brilliant German scientist Hermann von Helmholtz in the latter half of the 1800s. He viewed perception as the process of inferring the most likely environmental situation given the pattern of visual stimulation (Helmholtz, 1867/1925). This view has been the dominant framework for understanding vision for more than a century, although it has been extended and elaborated by later theorists, such as Richard Gregory (1970), David Marr (1982), and Irvin Rock (1983), in ways that we will discuss throughout this book.

Care must be taken not to misunderstand the notion that visual perception is based on constructing models. Invoking the concept of models does *not* imply that perception is "pure fiction." If it were, it would not fulfill the evolutionary demand for accurate information about the environment. To satisfy this requirement, perceptual models must (a) be closely coupled to the information in the projected image of the world and (b) provide reasonably accurate interpretations of this information. Illusions show that our models are sometimes inaccurate, and ambiguous figures show that they are sometimes not unique, but both tend to occur only under unusual conditions such as in the books and laboratories of vision scientists. Everyday experience tells us that our perceptual models are usually both accurate and unique. Indeed, if the sensory information is rich and complex enough, it is nearly impossible to fool the visual system into interpreting the environment incorrectly (Gibson, 1966).

Visual Completion. Perhaps the clearest and most convincing evidence that visual perception involves the construction of environmental models comes from the fact that our perceptions include portions of surfaces that we cannot actually see. Look at the shapes depicted in Figure 1.1.6A. No doubt you perceive a collection of three simple geometrical figures: a square, a circle, and a long rectangle. Now consider carefully how this description relates to what is actually present in the image. The circle is partly occluded by the square, so its lower left portion is absent from the image, and only the ends of the rectangle are directly visible, the middle being hidden (or *occluded*) behind the square and circle. Nevertheless, you perceive the partial circle as complete and the two ends of the rectangle as parts of a single, continuous object. In case you doubt this, compare this perception with that of Figure 1.1.6B, in which exactly the same regions are present but not in a configuration that allows them to be completed.

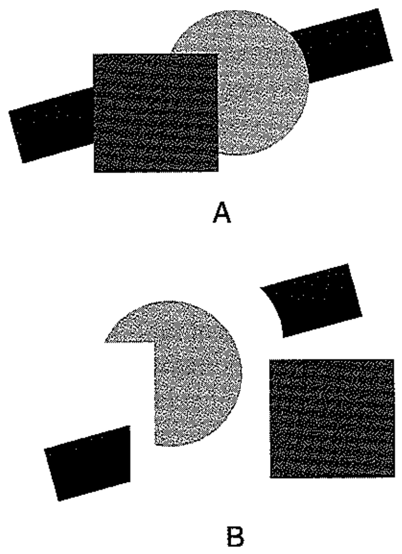


Figure 1.1.6 Visual completion behind partly occluding objects. Figure A is perceived as consisting of a square, a circle, and a rectangle even though the only visible regions are those shown separated in Figure B.

This perceptual filling in of parts of objects that are hidden from view is called **visual completion**. It happens automatically and effortlessly whenever you perceive the environment. Take a moment to look at your present surroundings and notice how much of what you "see" is actually based on completion of unseen or partly seen surfaces. Almost *nothing* is visible in its entirety, yet almost *everything* is perceived as whole and complete.

You may have noticed in considering the incompleteness of the sensory information about your present environment that visual perception also includes information about **self-occluded surfaces**: those surfaces of an object that are entirely hidden from view by its own visible surfaces. For example, only half of the cube that you perceive so clearly in Figure 1.1.7A is visible. Your perception somehow manages to include the three hidden surfaces that are occluded by the three visible ones. You would be more than a little surprised if you changed your viewpoint by walking to the other side and saw that the cube now appeared as in Figure 1.1.7B. Indeed, there are infinitely many possible physical situations that are consistent with Figure 1.1.7A, yet you automatically perceived just one: a whole cube.

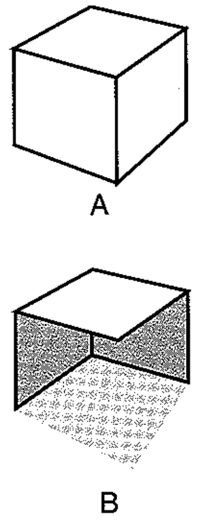


Figure 1.1.7 Visual completion due to self-occlusion. Figure A is invariably perceived as a solid cube, yet it is physically possible that its rear side looks like Figure B.

Completion presents an even more compelling case for the model-constructive view of visual perception than do illusions and ambiguous figures. It shows that what you perceive actually goes a good deal beyond what is directly available in the light reaching your eyes. You have very strong expectations about what self-occluded and partly occluded surfaces are like. These must be constructed from something more than the light entering your eyes, because the image itself contains no direct stimulation corresponding to these perceived, but unseen, parts of the world.

Impossible Objects. There is another phenomenon that offers an especially clear demonstration of the modeling aspect of visual perception. **Impossible objects** are two-dimensional line drawings that initially give the clear perception of coherent three-dimensional objects but are physically impossible. Figure 1.1.8 shows some famous examples. The "blivit" in Figure 1.1.8A looks sensible enough at first glance, but on closer inspection, it becomes clear that such an object cannot exist because the three round prongs on the left end do not match up with the two square ones on the right end. Similarly, the continuous three-dimensional triangle that we initially perceive in Figure 1.1.8B cannot exist because the surfaces of the locally interpretable sides do not match up properly (Penrose & Penrose, 1958).

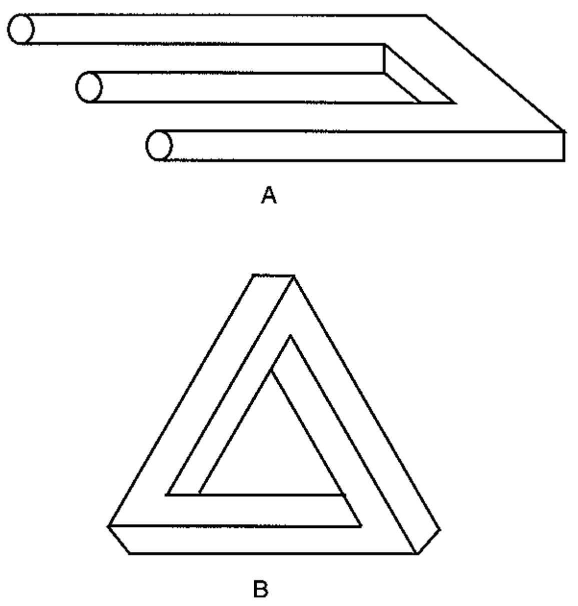


Figure 1.1.8 Impossible objects. Both the objects shown in this figure initially produce perceptions of coherent three-dimensional objects, but they are physically impossible. Such demonstrations support the idea that vision actively constructs environmental models rather than simply registering what is present.

One of the most interesting things about impossible objects is how clearly they show that our perceptions are internal constructions of a hypothesized external reality. If visual perception were merely an infallible reflection of the world, a physically impossible object simply could not be perceived. It would be as impossible perceptually as it is physically. Yet people readily perceive such objects when viewing properly constructed images. This fact suggests that perception must be performing an interpretation of visual information in terms of the three-dimensional (3-D) objects in the environment that might have given rise to the images registered by our eyes. Moreover, the kinds of errors that are evident in perceiving impossible objects seem to indicate that at least some visual processes work initially at a local level and only later fit the results into a global framework. The objects in Figure 1.1.8 actually make good sense locally; it is only in trying to put these local pieces together more globally that the inconsistencies become evident.

Predicting the Future. Supposing that the visual system does construct hypothetical models of reality rather than just sticking to information available in sensory stimulation, why might such a system have evolved? At some level, the answer must be that the models are more useful from an evolutionary standpoint than the images that gave rise to them, but the reason for this is not entirely clear. The usefulness of visual completion, for example, would seem to be that 3-D models representing hidden surfaces contain much more comprehensive information about the world than purely stimulus-based perceptions. The additional information in the constructed model is valuable because it helps the perceiving organism to predict the future. We have already considered one example in our discussion of Figure 1.1.7. Perceiving a whole three-dimensional cube provides the basis for expecting what we would see if we were to move so that new surfaces come into view. This is terribly important for creatures (like us) who are constantly on the move. A stable three-dimensional model frees us from having to reperceive everything from scratch as we move about in the world.

A perceptual model of the three-dimensional environment does not need to be modified much as we move around because the only thing that changes is our viewpoint relative to a largely stable landscape of objects and surfaces. In fact, the only time the model needs major modification is when model-based expectations are *disconfirmed* as unexpected surfaces come into view. Everyday experience tells us that this does not happen nearly as often as confirmation of our expectations. Thus, although constructing a three-dimensional model of the environment may initially seem like a poor evolutionary strategy, its short-term costs appear to be outweighed by its long-term benefits. It takes more time and effort to construct the complete model initially, but once it is done, it requires far less time and effort to maintain it. In the final analysis, the completed model is a remarkably economical solution to the problem of how to achieve stable and accurate knowledge of the environment.

The ability to predict the perceptual future is also evolutionarily crucial because we live in a world that includes moving objects and other mobile creatures. It is useful to know the current position of a moving object, but it is far more useful to know its direction and speed so that you can predict its future trajectory. This is particularly important when something is coming toward you, because you need to decide whether to approach, sidestep, flee, or ignore it. Without a perceptual model that somehow transcends momentary stimulus information, vision would not be able to guide our actions appropriately.

The view that the purpose of the brain is to compute dynamic, predictive models of the environment was set forth by British psychologist Kenneth Craik in 1943. He argued forcefully that organisms that can rapidly extrapolate the present situation into the future have an evolutionary advantage over otherwise identical organisms that cannot. An organism that can predict accurately is able to *plan* future actions, whereas one that cannot predict can only *react* once something has happened. There is an important caveat here, however: The process of extrapolation must work faster than the predicted event to be useful. Not surprisingly, then, most perceptual predictions are generated very quickly. Indeed, they are usually generated so quickly that we have no conscious experience of them unless they are violated. Even then, our conscious experience reflects the violation rather than the expectation itself.

1.1.5 Perception as Apprehension of Meaning

Our perceptual constructions of the external world go even further than completing unseen surfaces in a three-dimensional model, however. They include information about the meaning or functional significance of objects and situations. We perceive an object not just as having a particular shape and being in a particular location, but as a person, a dog, a house, or whatever. Being able to *classify* (or *recognize* or *identify*) objects as members of known categories allows us to respond to them in appropriate ways because it gives us access to vast amounts of information that we have stored from previous experiences with similar objects.

Classification. Perhaps the easiest way to appreciate the importance of classification is to imagine encountering some completely foreign object. You could perceive its physical characteristics, such as its color, texture, size, shape, and location, but you wouldn't know what it was or what you should do with it. Is it alive? Can it be eaten? Is it dangerous? Should you approach it? Should you avoid it? Such questions can seldom be answered directly from an object's physical characteristics, for they also depend on what kind of object it is. We embrace loved ones, flee angry dogs, walk around pillars, eat hamburgers, and sit in chairs. All this is so obvious that it scarcely seems worth mentioning, but without perceptually classifying things into known categories, it would be difficult to behave appropriately with the enormous variety of new objects that we encounter daily. We can simply walk around the pillar because past experience informs us that such objects do not generally move. But angry dogs can and do!

Classification is useful because objects within the same category share so many properties and behaviors. All chairs are not exactly alike, nor are all hamburgers, but one chair is a lot more like another than it is like any hamburger, and vice versa. Previous experience with members of a given category therefore allows us to predict with reasonable certainty what new members of that same class will do. As a consequence, we can deal with most new objects at the more abstract level of their category, even though we have never seen that particular object before.

Classifying objects as members of known categories seems simple, but it is actually quite an achievement. Consider the wide variety of dogs shown in Figure 1.1.9, for example. How can we recognize almost immediately that they are all dogs? Do dogs have some unique set of properties that enable us to perceive them as dogs? If so, what might they be? These are problems of object identification, one of the most difficult—and as yet unsolved—puzzles of visual perception. In Chapters 8 and 9, we will consider some current ideas about how this might happen.

Attention and Consciousness. It is an undeniable fact that the visible environment contains much more information than anyone can fully perceive. You must therefore be *selective* in what you attend to, and what you select will depend a great deal on your needs, goals, plans, and desires. Although there is certainly an important sense in which a hamburger is always a hamburger, how you react to one depends a great deal on whether you have just finished a two-day fast or a seven-course meal. After fasting, your attention would undoubtedly be drawn immediately to the hamburger; right after a big Thanksgiving dinner, you would probably ignore it, and if you did not, the sight of it might literally nauseate you.

This example demonstrates that perception is not an entirely **stimulus-driven process**; that is, perceptions

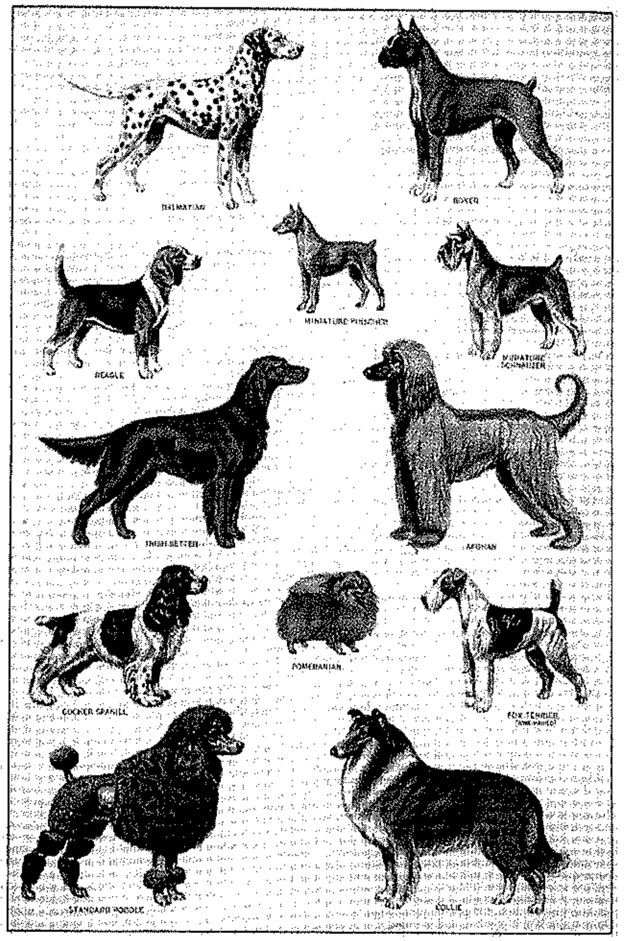


Figure 1.1.9 Many kinds of dogs. Visual perception goes beyond the physical description of objects to classify them into known categories. Despite the substantial physical differences in their appearance, all these animals are readily perceived as belonging to the category of dogs.

are not determined solely by the nature of the optical information present in sensory stimulation. Our perceptions are also influenced to some extent by **cognitive constraints**: higher-level goals, plans, and expectations. It would be strange indeed if this were not so, since the whole evolutionary purpose of perception, I have argued, is to make contact between the needs of the organism and the corresponding opportunities available in its environment. There are countless ways in which such higher-level cognitive constraints influence your perception, many of which involve the selective process of visual attention. As the hamburger example suggests, we look at different things in our surroundings depending on what we are trying to accomplish, and we may perceive them differently as a result. This point is perhaps so obvious that it goes without saying, but it is important nevertheless.

One of the functions of attention is to bring visual information to consciousness. Certain properties of objects do not seem to be experienced consciously unless they are attended, yet unattended objects are often processed fully enough outside of consciousness to attract your attention. Everyday examples abound. You may initially not notice a stationary object in your visual periphery, but if it suddenly starts moving toward you, you look in its direction without knowing why, only then becoming consciously aware of its presence. While driving your car, you sometimes look over at the car next to you without knowing why, only to find that the driver has been looking at you. In both cases, visual processing must have taken place outside of consciousness, directing your attention to the interesting or important aspects of the environment: the moving object or the person looking at you. Once the object is attended, you become conscious of its detailed properties and are able to identify it and discern its meaning in the present situation.

Attending to an object visually usually means moving your eyes to fixate on it, but attention and visual fixation are not the same. You are probably familiar with the fact that you can be looking directly at something without attending to it in the slightest. Your thoughts may wander to some completely different topic, and once attention returns to the visual information, you may realize that you had no awareness of what was in your visual field during the diversion. Conversely, you can attend to an object without fixating on it. To demonstrate this, hold your hand out in front of you and fixate directly on your middle finger. Now, without moving your eyes, try attending to each of the other fingers in turn. It is not terribly easy, because you want to move your eyes at the same time as you shift your attention, but it clearly can be done.

Many high-level aspects of perception seem to be fully conscious. For example, when you look around the room trying to find your keys, you are certainly aware of the key-finding goal that directs your attention to various likely places in the room. Other aspects of perception are clearly *not* conscious, even in the same situation, such as knowing what makes an object "keylike" enough to direct your eyes at it during this visual search. In general, lower levels of perception do not seem to be accessible to, or modifiable by, conscious knowledge and expectations, whereas higher levels do.

Not much is yet known about the role of consciousness in perception. Indeed, we know surprisingly little even about the evolutionary advantage of conscious perception. There is a general belief that there must be one, but nobody has yet managed to give a good account of what it is. The basic question is what advantage there might be for a consciously perceiving organism over one that can perform all the same perceptual tasks but without having conscious visual experiences. The unconscious automaton can, by definition, engage in all of the same evolutionarily useful activities—successfully finding food, shelter, and mates while avoiding cliffs, predators, and falling objects—so it is unclear on what basis consciousness could be evolutionarily selected.

One possibility is that the problem is ill-posed. Perhaps the automaton actually could *not* perform all the tasks that the consciously perceiving organism could. Perhaps consciousness plays some crucial and as-yet-unspecified role in our perceptual abilities. We will return to these conjectures in Chapter 13 when we consider what is known about the relation between consciousness and perception.

1.2 Optical Information

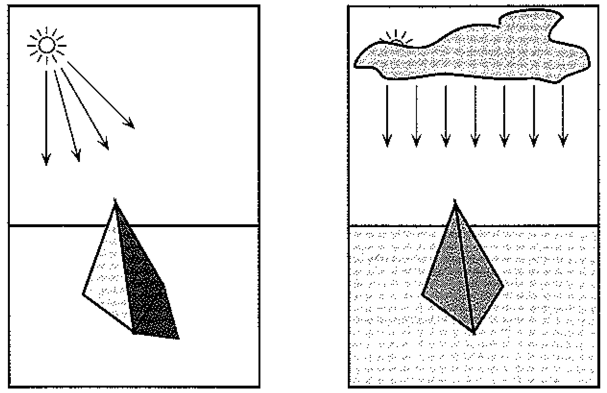
Our definition of visual perception implies that vision depends crucially on the interaction among three things: light, surfaces that reflect light, and the visual system of an observer that can detect light. Remove any one of these ingredients, and visual perception of the environment simply does not occur. It seems reasonable, therefore, to begin our study of vision by considering some basic facts about each of them. The present section will describe how light interacts with surfaces to produce the optical events that are the starting point of vision. The next section will describe the overall structure of the human visual system that processes information in these optical events. The remainder of the book discusses in detail how the visual system goes about extracting relevant information from light to produce useful perceptions of environmental scenes and events.

I argued in the preceding section that the evolutionary role of visual perception is to provide an organism with accurate information about its environment. For this to happen, the light that enters our eyes must somehow carry information about the environment. It need not carry *all* the information we ultimately get from looking at things, but it must carry enough that the rest can be inferred with reasonable accuracy. In this section, we will consider how light manages to carry information about the world of visible objects around us.

1.2.1 The Behavior of Light

The science concerned with the behavior of light is a branch of physics called **optics**. According to prevailing physical theory, light consists of minute packets of energy called **photons** that behave like waves in some respects and like particles in others. Throughout most of this book we will be concerned mainly with the particle behavior of photons, although the discussion of color vision in Chapter 3 will require consideration of its wavelike properties as well. Photons radiate outward from their source—a hot body such as the sun, a fire, or the filament of an incandescent light bulb—like infinitely tiny bullets that travel through air in perfectly straight lines at the enormous speed of 186,000 miles per second. When photons strike the surface of an object, we say that it is *illuminated.* The amount of visible light—that is, the number of photons—falling on a given surface per unit of time is called its **luminance**. The luminance of a light covaries to some degree with its perceived brightness, but the relation is far from simple, as we will discover in Chapter 3.

Illumination. Illumination refers to the lighting conditions in the environment. The simplest condition from an analytical standpoint is called **point-source illumination**. It refers to an idealized situation in which all the light illuminating a scene comes from a single, point-sized light source at a specific location. A single incandescent light bulb in an otherwise dark space would be a reasonable approximation, as would the sun on a cloudless day. Point-source illumination produces dark, well-defined shadows behind illuminated surfaces and strong shading effects on the illuminated surfaces themselves, as illustrated in Figure 1.2.1A. Both effects result from the fact that all of the direct (nonreflected) light is coming from a single location. In fact, one seldom encounters lighting conditions this simple in the real world. Point-



A. Point-Source Illumination B. Spatially Extended Illumination

Figure 1.2.1 Point-source illumination versus diffuse illumination. A single point source creates well-defined shadows (A), whereas spatially extended illumination does not (B).

source illumination is primarily of interest to vision theorists as a way to reduce the mathematical complexity of certain problems. It is used as a simplifying assumption, for example, in determining the shape of an object from the shading on its illuminated surfaces. If there are multiple point sources, such as a room with two or more incandescent lights, there are correspondingly two or more different shadows and shading patterns for each surface. Each additional light source thus complicates the optical structure of the environment.

In many real situations, light comes from **diffuse illumination**, in which light radiates from a relatively large region of space. To take an extreme example, the light of the sun on an overcast day is diffused almost uniformly over the entire sky, so nearly equal amounts of radiant light are coming from everywhere in the whole upper half of the visual environment. Under such conditions, both the shadow cast by illuminated surfaces and the shading on the surfaces themselves are much weaker and less well defined than under point-source illumination, as illustrated in Figure 1.2.1B. If you are a skier, you have probably noticed one of the effects of diffuse illumination under cloudy skies: It reduces the optical information that allows you to perceive the undulations in the snow surface (called "moguls") that cause a large proportion of falls. In fact, skiers have so much more trouble negotiating the slopes on cloudy days that they have invented an apt name for the situation: "flat light."

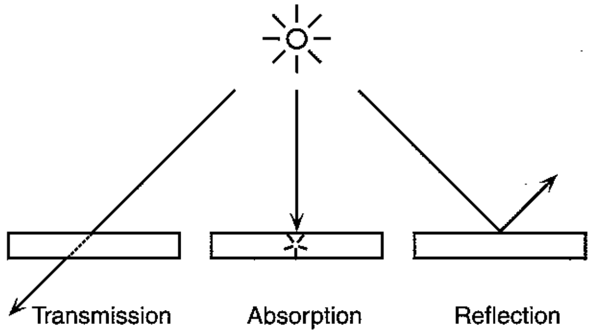


Figure 1.2.2 Interactions between light and surfaces. A photon can be transmitted through a surface, absorbed by it, or reflected off it.

Interaction with Surfaces. We said that photons travel in perfectly straight lines, but only until they strike the surface of an object. In almost every case, the surface produces a radical change in the behavior of the photons that strike it. It is these surface-induced changes in the behavior of photons that ultimately provide vision with information about the surfaces in the environment that produced them. The only surfaces that do not change the behavior of photons are completely transparent ones, and such surfaces would be literally invisible—if they existed. All real surfaces interact with light strongly enough under most conditions that they are visible to a vigilant observer.

When a photon strikes the surface of an object, one of three basic events takes place: It is either *transmitted* through the surface, *absorbed* by it, or *reflected* off it (see Figure 1.2.2). Transmitted light can either pass straight through the surface or be bent (*refracted*), as it is in Figure 1.2.3, which shows the lower part of a spoon in a glass of water as being displaced laterally. Of these photon-surface interactions, reflection is the most important for vision for two reasons. First, reflected light has been changed by its interaction with the surface, so it contains information about the surface. Second, reflected light is subsequently available to strike the receptive surface of an observer's eye, so it can transmit that information about the surface to the visual system. Nonreflected light is also important for understanding certain aspects of visual perception—absorbed light for color vision and transmitted light for perception of transparency—but in the present discussion, we will concentrate on the optical information in reflected light.

Figure 1.2.3 Refraction of light. When light is transmitted through an object, it can be bent (refracted), leading to erroneous perceptions, such as the misaligned appearance of the spoon handle in this glass of water.

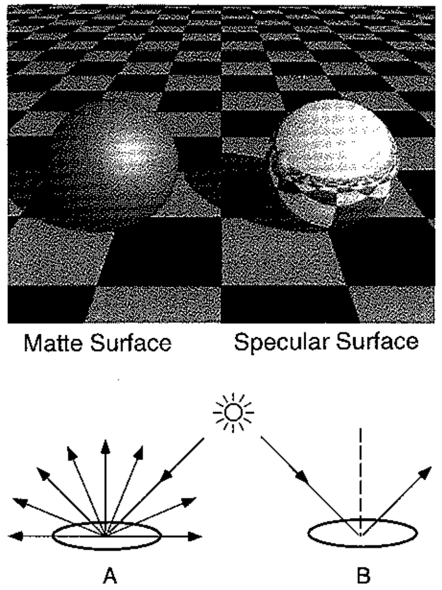


Figure 1.2.4 Matte versus specular surfaces. The two spheres in the picture differ only in how their surfaces reflect light. The matte sphere on the left reflects light equally in all directions, whereas the specular sphere on the right reflects it coherently, so that the angle of reflection equals the angle of incidence.

The change that a surface produces in a reflected photon is to alter its trajectory: The photon bounces off the surface in a direction that depends on both the direction from which it came and the microscopic structure of the surface. If the surface is highly polished, or **specular**, such as a mirror, the light is reflected in the single direction that is symmetric to the direction from which it came (see Figure 1.2.4B); the angle of incidence is equal to the angle of reflection. On the other hand, if the surface is dull or **matte**, such as a typical piece of paper, the light is scattered diffusely in many directions (see Figure 1.2.4A). Perfectly specular and perfectly matte surfaces are just the two idealized endpoints of a continuum, and all real surfaces fall somewhere in between. Figure 1.2.4 shows how different the same spherical shape looks with a highly specular surface compared to a highly matte surface. The reflectance properties of surfaces are actually even more complex than we have considered here because the degree of specularity can vary as a function of the angle of incident light, as is the case with semigloss surfaces. When viewed in a direction nearly parallel to the surface, they are much more specular than when viewed nearly perpendicular to it.

Let us now try to put together what we have said about light radiation from sources of illumination and light reflection by surfaces to come to a more complete understanding of the behavior of light in a real environment. All surfaces reflect some light except completely black ones (which absorb it all) and completely transparent ones (which transmit it all).[2] And most surfaces are more matte than specular. Together, these facts imply that some light bounces in almost every direction off almost every surface in the environment. As a result, light does not come just from the direction of radiant light sources, such as the sun and light bulbs; it also comes by reflection from virtually every surface in the environment. Surfaces thus act as **secondary light sources** by illuminating other surfaces with their reflected light. Moreover, all of these photons are bouncing from sur-

---

[2] In reality, there are neither totally black surfaces nor totally transparent ones. Even seemingly jet-black surfaces reflect a few percent of the photons that fall on them, and even seemingly crystal-clear glass or plastic reflects and absorbs a few percent of the incident photons.

face to surface at enormous speed, often being reflected off many surfaces before finally being absorbed.

Color Plate 1.1 illustrates the sometimes profound effect of different levels of light reflection in the appearance of a bathroom scene. Image A was generated by a sophisticated computer program that shows only the direct illumination entering the eye from the light source, before any light from the fixture has been reflected from surfaces into the eye. Image B shows the same scene with the addition of primary reflections from nonluminous surfaces. Image C then adds secondary reflections, and image D adds further reflections up to the fifth-order. Notice that nonluminous surfaces do not appear in the image until the primary reflections are included, that the reflection in the mirror does not appear until secondary reflections are included, and that specular highlights on the shower door are not apparent until higher-order reflections are modeled. Thus, the rich appearance of natural scenes depends importantly on the complex interactions of ambient light with the structure of the physical environment.

The net effect of all these reflections is that light is reverberating around the environment, filling it with light from virtually every direction. This fact is of paramount importance for vision, because it is this complex optical structure that enables vision to occur. More light comes from some directions than others, and that is why we are able to see surfaces of different colors in different directions. Equal amounts of light from all directions, called a **Ganzfeld**, just looks like an all-encompassing gray fog.

The Ambient Optic Array. The pioneering perceptual psychologist James J. Gibson conceptualized the optical information available in light in terms of what he called the **ambient optic array** (or **AOA**). The AOA refers to the light coming toward a given point of observation from all directions. It is called "ambient" because the observation point is literally surrounded by light converging on it from all directions. This means that if your eye were at this observation point, either light reflected from environmental surfaces or light emitted directly from the radiant source would be available from every direction. Vision is possible at that observation point because environmental surfaces structure the light in the AOA in complex, but lawful, ways. This lawfulness in the optical structure of the AOA provides the information that enables vision to occur.

When conceived in this way, vision can be likened to solving a puzzle. Surfaces in the world alter light by reflecting it in a way that forms the AOA at the current observation point with a particular complex structure. The visual system registers this structure and then tries, in effect, to reverse the process by determining the arrangement of surfaces that must exist in the environment to have structured the AOA in just that way. This all happens so quickly and effortlessly that we have no conscious knowledge about how it is done.

To appreciate more fully the nature of the AOA, let us consider a few examples. Figure 1.2.5A shows schematically the structure of the AOA from an observer's perspective in a room containing a stool and a window that looks onto a tree. Notice that light comes from *all directions* toward the point of observation and that the AOA defined at this point exists independently of an observer occupying it. Figure 1.2.5B illustrates the relationship between an observer's view of the world and the AOA: The eye samples a directional subset of the AOA. The shaded part of the AOA is not currently visible because the eye admits light only from the front. Figure 1.2.5C shows the resulting pattern of light that would be entering the observer's eye at this observation point. It depicts the momentary optical image that falls on the light-sensitive cells at the back of the eye. This is the starting point of vision, at least for that instant in time.

It is important to realize that *there is a differently structured AOA at every point in the environment.* Each one is unique, providing slightly different information about the environment. To illustrate this fact, consider what happens when this observer stands up. As he or she rises, the eye moves along a trajectory of different observation points, and visual stimulation is determined at each one by its unique AOA. Figure 1.2.5D depicts the final AOA of this event with solid lines, showing how certain parts of surfaces that were previously visible have become occluded and how other parts of surfaces that were previously occluded have now become visible. Figure 1.2.5E shows what optical image is entering the observer's eye at this moment so that you can see how it has changed.

The changes in optical information caused by the change in observation point highlight a distinction of great importance between the momentary static AOA and the temporally extended, dynamic AOA. Whereas the static AOA can be characterized by a pattern of light converging at the observation point, the dynamic AOA can be fully characterized only by the **optic flow** of light over time. Thus, the dynamic AOA provides the observer with information from an additional dimension that unfolds over time. This turns out to be enormously important for many perceptual phenomena, such as our ability to perceive the third dimension of the environment (depth or distance from the observer), to determine the shapes of moving objects, and to perceive our own trajectory through the environment as we move. We will consider each of these topics more fully in Chapters 5 and 10 when we discuss perception of depth and motion in detail.

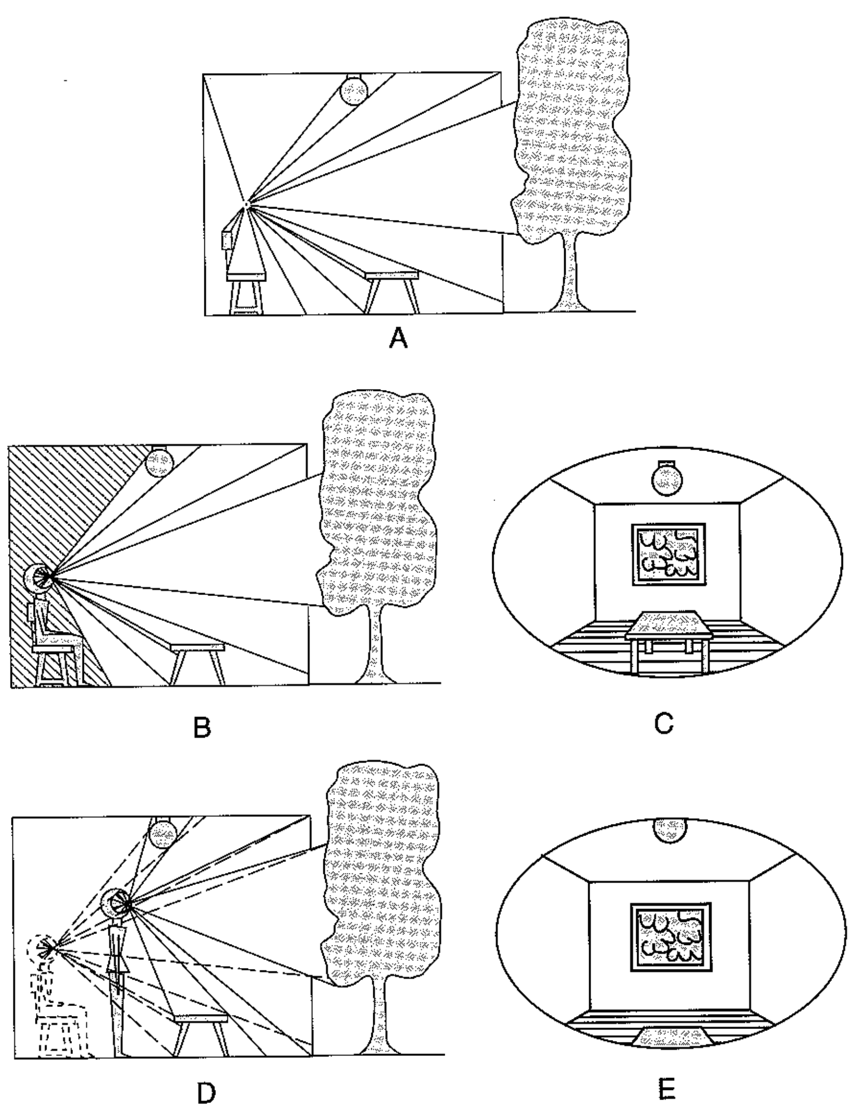


Figure 1.2.5 The ambient optic array (AOA). (A) All light converging at a given point defines the AOA at that position. (B) When an observer's eye is at that position, it samples the part of the AOA in front of the observer. (C) This sample is registered on the observer's retina as an image. (D) If the observer moves (for example, by standing), the AOA changes as illustrated. (E) This change in the AOA results in corresponding subtle changes in the observer's retinal image. (A, B, and D from Gibson, 1979.)

1.2.2 The Formation of Images

If vision is to provide accurate information about the external world, then there must be a consistent relationship between the geometry of environmental surfaces and the light that enters the eye at a particular observation point. Indeed, there is. Figure 1.2.5 depicts the nature of this relationship by illustration. In this section, we will take a closer look at how optical images are formed on the back of the eye and how the laws of perspective projection describe this image formation process.

Optical Images. The situation involved in image formation is diagrammed in Figure 1.2.6. The external world has three spatial dimensions. Illumination bathes the objects in this three-dimensional space with light, and that light is reflected by surfaces into the observer's eye along straight lines (Figure 1.2.6A). Photons pass into the eye to form a two-dimensional, upside-down image on its back surface (Figure 1.2.6B and C). The object in the external world is often referred to as the **distal stimulus** (meaning distant from the observer), and its optical image at the back of the eye as the **proximal stimulus** (meaning close to the observer). As Figure 1.2.6A illustrates, the size of an object's image in the eye is usually specified by its **visual angle**: the number of degrees subtended by the image from its extremes to the focal point of the eye. It is important to understand that this angle measures the spatial dimensions of the proximal stimulus, not the distal stimulus. The same external object will subtend a smaller angle when it is farther away and a larger angle when it is closer to the observer's eye. This relationship between object size, object distance, and image size is important in understanding how we perceive the size and location of objects, as we will see in Chapter 5.

Perhaps the most important fact about the image formation process for understanding the problem of vision is that the image on the back of the eye has only *two* spatial dimensions rather than three; that is, the optical image on the back of the eye is like a slide projected onto a curved screen. This means that vital spatial information has been lost in going from the real, 3-D world to the eye's 2-D image of it. The dimension that is missing in the image is *depth*: the distance outward from the focal point of the observer's eye to the location of the environmental surface that reflected light into the eye. To perceive the world three dimensionally—which we obviously do—the dimension of depth must somehow be recovered from the information in the two spatial dimensions of the optical image.

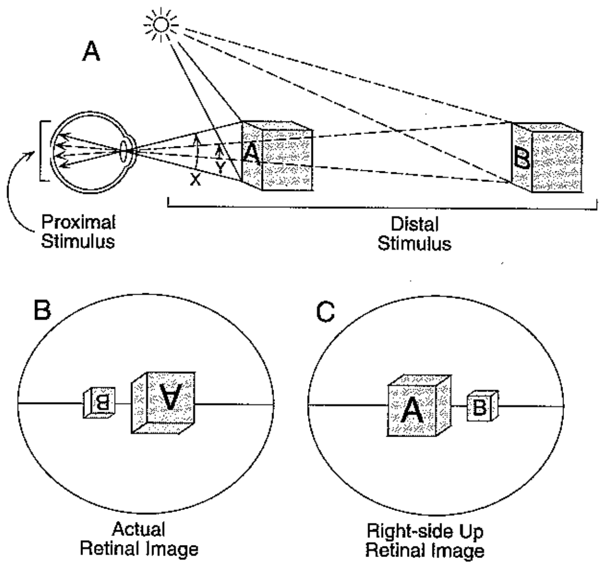


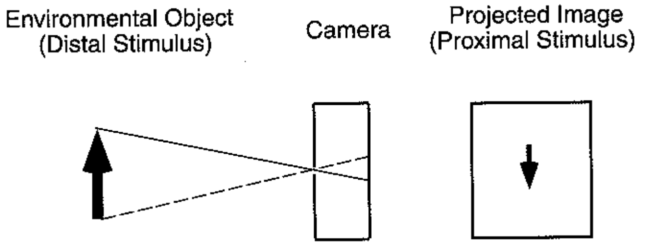
Figure 1.2.6 Image formation and the size-distance relation. Light is reflected by 3-D objects (the distal stimulus) into the eye, projecting a 2-D image (the proximal stimulus) onto the retina (part A). Because the distance to an object is not directly represented in its 2-D projection, environmental objects of the same size at different distances project images of different sizes, as illustrated in parts B and C.

Projective Geometry. The image formation process that maps the 3-D world to the 2-D image is highly lawful, and like most lawfulness, it can be analyzed mathematically. The most appropriate mathematics for the task is **projective geometry**: the study of how a higher-dimensional space is mapped onto a lower-dimensional one. In the case of static vision, the projective mapping of interest is from the 3-D space of the environment onto the 2-D space of the image plane. Projective geometry can therefore specify for a given 3-D scene of objects exactly where each point in the scene will project onto a given 2-D image plane and what properties of these images will be invariant over different projections. In dynamic vision, the projection of interest is from the 4-D structure of space-time onto the 3-D space of optic flow that unfolds over time on the 2-D surface at the back of the eye.

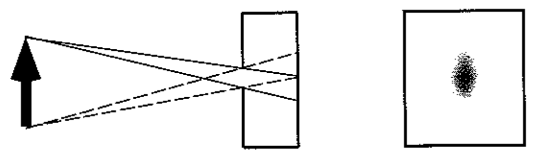
Projective geometry thus seems to be the ideal mathematical tool for understanding image formation. The problem is that projective geometry alone cannot model the full complexity of optical phenomena because it does not contain the appropriate structure for modeling reflection, absorption, or refraction of light. In a world filled with opaque surfaces, for example, only light reflected from the *closest* surface in a particular direction will enter the eye. Photons from all farther points will be either absorbed or reflected by the next closer surface, thus preventing them from reaching the eye. The complications that are introduced by the interaction of light with surfaces make projective geometry only partly useful for modeling the process of forming 2-D optical images from the 3-D world. Rather, one must understand in more detail how light from the 3-D world forms an image when it is projected onto a 2-D plane.

Perspective and Orthographic Projection. One way to form a 2-D image of the 3-D world is to place a pinhole at a given distance in front of an image plane or other 2-D surface inside an otherwise lightproof box. Such a device is called a **pinhole camera**. Because photons travel in straight lines, the light falling on each point of the image plane of the pinhole camera got there by being reflected (or emitted) from a particular point in environmental space. That point lies along the ray starting at its image-plane point and passing through the pinhole (see Figure 1.2.7A). This situation gives the basic geometry of **perspective projection** (or **polar projection**): the process of image formation in which the light converges toward a single focal point (or pole).

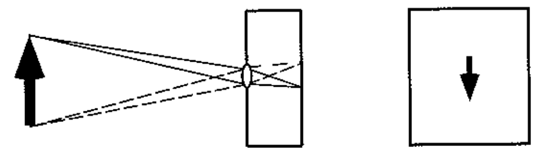
Good pinhole images are not as easy to create as the above description makes it seem. To get a crisp, clear image, the pinhole must be very small; about 0.4 mm in diameter is ideal. Because of this small aperture size, however, very little light falls on the image plane, so it must be observed under very dark conditions.[3] If the hole is made larger to let more light in, the image becomes blurred because all the light no longer goes through the single point of the pinhole but through many different points (see Figure 1.2.7B). This problem can be overcome by supplying the camera with a transparent convex **lens** at its opening to bend the incoming light inward to a point (called its **focal point**) some distance behind the lens (see Figure 1.2.7C). Thus, the lens provides a "virtual pinhole" at its focal point that makes the projected image at the back sharp and clear again, like the pinhole camera's image, only brighter because more light comes through the larger opening. As we will see in Section 1.3, the human eye contains such a lens whose job is to focus the image on the back surface of the eye.



A. Pinhole Aperture without Lens --> Sharp Image



B. Large Aperture without Lens --> Fuzzy Image



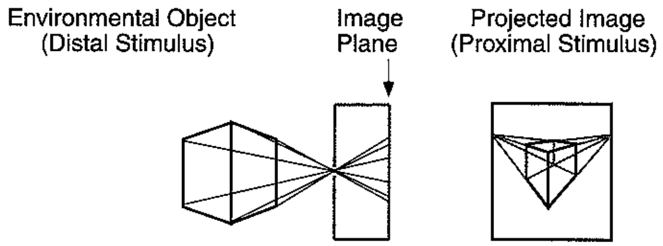
C. Large Aperture with Lens --> Sharp Image

Figure 1.2.7 The optics of pinhole cameras and lenses. (A) A pinhole camera with a small aperture produces sharp images without a lens. (B) A camera with a larger aperture but no lens produces fuzzy, out-of-focus images. (C) A camera with a large aperture and a lens can produce clear, well-focused images if the focal length of the lens is appropriate for the distance to the imaging surface.

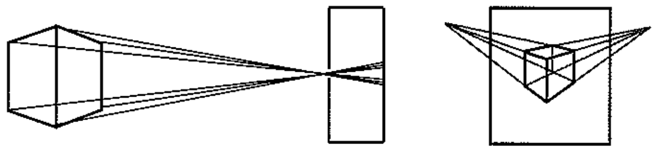
Assuming that the complicating effects of light-surface interaction can be incorporated into the model, the mathematics of perspective (or polar) projection are useful for modeling image formation by the human eye (see Figure 1.2.8A and B). Unfortunately, they are also

---

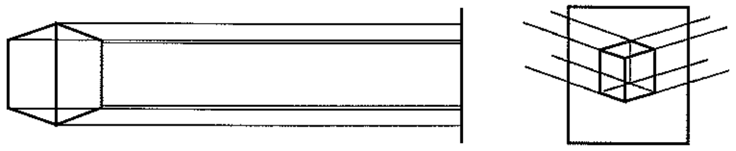
[3] A good way to do this is to make a light-tight box with a single pinhole on one side and a translucent flat surface behind it. When the back surface is viewed under darkened conditions—such as with a dark cloth covering one's head and the back of the box—you can observe the upside-down and backward image on the translucent surface.



A. Perspective Projection (Close)



B. Perspective Projection (Far)



C. Orthographic Projection

Figure 1.2.8 Perspective and orthographic projection. Perspective projection through a pinhole (A and B) produces convergence of parallel lines to a vanishing point, but close perspective (A) yields more convergence than far perspective (B). Orthographic projection (C) is based on parallel light rays and produces no convergence of parallel lines in the image.

rather complex, much more so than has been alluded to here. To simplify matters, visual theorists often employ **orthographic projection** (or **parallel projection**) instead of perspective projection to model the geometry of image formation (see Figure 1.2.8C). In this case, the image is conceptualized as being formed by light rays that travel parallel to each other and perpendicular to the image plane, rather than rays that converge at the pinhole.

The mathematical simplification that results from orthographic projection is that the depth dimension of the world—distances from image to objects—is simply ignored, whereas all spatial information in the plane perpendicular to the viewing direction is preserved without change. This means that when the distance from the image to the object is large relative to the depth of the object (Figure 1.2.8B), orthographic projection is a good approximation of perspective projection. Close up, however, the differences between orthographic and perspective projection become quite significant, as illustrated in Figure 1.2.8A.

One way of understanding the relationship between perspective and orthographic projection is to consider what happens when an object is moved farther and farther away from the pinhole of perspective projection (compare Figure 1.2.8A and Figure 1.2.8B). As this happens, the light rays projecting through the pinhole become more and more parallel so that, at an infinite distance, the light rays would be parallel, as they are in orthographic projection (Figure 1.2.8C). Thus, orthographic projection can be conceived as a limiting case of perspective projection, at which the distance between object and focal point is infinite. The important difference is that the perspective image of an object at an infinite distance is a single point, unlike its true orthographic projection which results in a spatially extended image.

1.2.3 Vision as an "Inverse" Problem

We have now described how light reflected from the 3-D world produces 2-D images at the back of the eye where vision begins. This process of image formation is completely determined by the laws of optics, so for any given scene with well-specified lighting conditions and a point of observation, we can determine with great accuracy what image would be produced. In fact, the field of **computer graphics** is concerned with exactly this problem: how to render images on a computer display screen that realistically depict scenes of objects by modeling the process of image formation. Many of the problems in this domain are now very well understood, as one can appreciate by examining some examples of state-of-the-art computer images that have been generated without recourse to any real optical processes whatsoever. The images in Color Plate 1.1, for example, were rendered by a ray-tracing algorithm that simulates image formation from an internal model of the surfaces in the room and the behavior of the light that illuminates them. In effect, the program simulates the optical events of photon emission, reflection, transmission, and absorption to construct an image of a "virtual" environment that does not exist in the physical world. Such programs allow the effects of different orders of light reflection to be illustrated (e.g., in Color Plate 1.1A–D) because the program can be stopped after each cycle of simulated reflection to see what the image looks like. This is not possible with real optical image formation.

The early stages of visual perception can be viewed as trying to solve what is often called the **inverse problem**: how to get from optical images of scenes back to knowledge of the objects that gave rise to them. From this perspective, the most obvious solution is for vision to try to *invert* the process of image formation by undoing the optical transformations that happen during image formation.

Unfortunately, there is no easy way to do this. The difficulty is that the mathematical relation between the environment and its projective image is not symmetrical. The projection from environment to image goes from three dimensions to two and so is a well-defined function: Each point in the environment maps into a unique point in the image. The inverse mapping from image to environment goes from two dimensions to three, and this is not a well-defined function: Each point in the image could map into an infinite number of points in the environment. Therefore, logic dictates that for every 2-D image on the back of our eyes, there are infinitely many distinct 3-D environments that could have given rise to it.

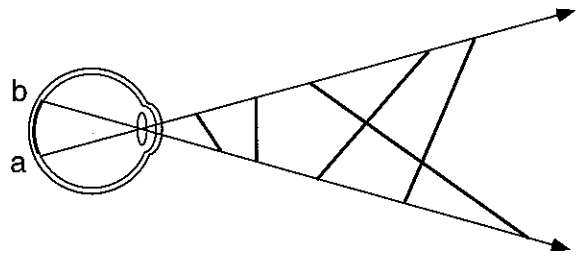


Figure 1.2.9 An illustration of inverse projection. A single line segment on the retina can be the projection of an infinite variety of lines in the environment.

Figure 1.2.9 illustrates the indeterminacy of inverse projection by showing that a single line segment in an optical image could have resulted from the projection of an infinite number of lines in the environment. The reason is that the inverse problem is *underspecified* (or *underconstrained* or *underdetermined*) by the sensory data in the image. There is no easy way around this problem, and that is why visual perception is so complex. In fact, were it not for the fact that our brains manage to come up with the correct solution most of the time, it would be tempting to conclude that 3-D visual perception is simply impossible!

We know that 3-D perception *is* possible precisely because the human visual system manages to do it with such remarkable accuracy under most circumstances. How does it solve this seemingly insoluble problem? Different theorists have taken different approaches, as we shall see in Chapter 5, but the dominant one is to assume that 3-D perception results from the visual system making a lot of highly plausible assumptions about the nature of the environment and the conditions under which it is viewed. These assumptions constrain the inverse problem enough to make it solvable most of the time. If the assumptions are true, the resulting solution will be veridical. Vision is thus a **heuristic process** in which inferences are made about the most likely environmental condition that could have produced a given image. The process is *heuristic* because it makes use of inferential rules of thumb—based on the additional assumptions—that are not always valid and so will sometimes lead to erroneous conclusions, as in the case of perceptual illusions. Under most everyday circumstances, however, the assumptions *are* true, and so normal visual perception is highly veridical. We will encounter these additional assumptions throughout the book, particularly in Chapter 5 when we discuss depth perception in detail. Then we will be able to see in what sense they allow the visual system to go beyond the information given in the optical image so that the seemingly impossible inverse problem can be solved.

1.3 Visual Systems

We know that the inverse problem can be solved because the human visual system solves it—maybe not all the time and maybe not with perfect accuracy, but enough of the time and with sufficient accuracy to provide us with excellent information about the environment. Much of the rest of this book is concerned with what is currently known about how the visual system accomplishes this feat. To begin, we will now take a quick look at the overall structure of the part of the nervous system that is known to be involved in processing visual information. The description that we give here will be brief and superficial in many respects. This is intentional. Its purpose is merely to provide a scaffolding of background knowledge about the biological structure of the visual system so that it will be available for later discussions that complete the picture. Once we have mastered some of the basic facts about the "hardware" of this system—its **anatomy** or physical structure—we can begin to ask better informed questions about what it does—its **physiology**, or biological function.

1.3.1 The Human Eye

Although it has been known since antiquity that eyes are the sensory organs of vision, an accurate understanding of how they work is a relatively recent achievement. The Greek philosopher Plato (427–347 B.C.) believed that an "inner fire" gave rise to rays that emanated from the eye toward perceived objects. Epicurus (341–270 B.C.) rejected this emanation theory, believing instead that tiny replicas of objects were somehow transmitted rapidly into the mind through the eyes. Galen (A.D. 130–200) later elaborated these ideas with physiological details, proposing that after the rays emanated from the eye, they interacted with the object and then returned to the eye. In the lens of the eye, he believed, these rays interacted with a "visual spirit" that flowed from the brain to the eye and back, bringing with it the replicas of perceived objects.

The modern era of physiological optics did not really begin until the brilliant Arabic philosopher Alhazen (A.D. 965–1040) hit upon the idea that the eye is like a pinhole camera, as we discussed in Section 1.2.2. The important insight that he achieved was that vision occurs when light from external sources is reflected from surfaces of objects and enters the eye. Even so, an accurate understanding of the optics of the eye required the invention and understanding of lenses. The noted astronomer Johannes Kepler (1571–1630) finally put these elements together into a reasonable approximation of the modern theory of physiological optics, discussed below.

Eye and Brain. Although Galen had many important facts about vision wrong, he was right in believing that both eyes and brain are essential (see Figure 1.3.1). We now know that optical information from the eyes is transmitted to the primary visual cortex in the occipital lobe at the back of the head, as shown in Figure 1.3.1. This information is then sent to many other visual centers located in the posterior temporal and parietal cortex, as illustrated in Figure 1.3.2. Some estimates put the percentage of the cortex involved with visual function at more than 50% in the macaque monkey (DeYoe & Van Essen, 1988; Van Essen et al., 1990), although it is probably slightly lower in humans. The complete visual system thus includes much of the brain as well as the eyes, and the whole eye-brain system must function properly for the organism to extract reliable information about the environment from the ambient optic array. The eyes must collect and register information contained in light, and the brain must then process that information in ways that make it useful for the organism.

The fact that both eye and brain are required for vision means that a person who has normal eyes but damage to visual parts of the brain might be as "blind" as a person who has a normal brain but eyes that fail to work. Indeed, both sorts of blindness exist. Damage to the eyes from accidents or disease sometimes prevents them from doing their job of registering optical information and/or sending it on to the visual areas of the brain. Such conditions cause the sort of blindness most people know about: lack of sight because no information from light gets into the system. Although far less common, there are also people who cannot see—or at least do not have visual experiences—yet have eyes that work quite normally. This kind of blindness—called **blindsight**—results from damage to certain critical parts of the visual cortex due to disease, surgery, or stroke (see Section 13.2.2). Damage to other parts of visual cortex does not result in blindness but can cause debilitating selective deficits in perception. Some patients can see well enough to describe faces accurately but cannot identify even members of their own family by sight (see Section 9.2.5). Others can describe and draw simple everyday objects but cannot name them or use them properly. It has even been reported that some people see stationary scenes perfectly well but cannot perceive motion, experiencing instead a series of frozen snapshots (see Section 10.1.4). We will discuss these fascinating problems in more detail at appropriate places throughout the book, particularly at the end when we consider conscious experiences of seeing.

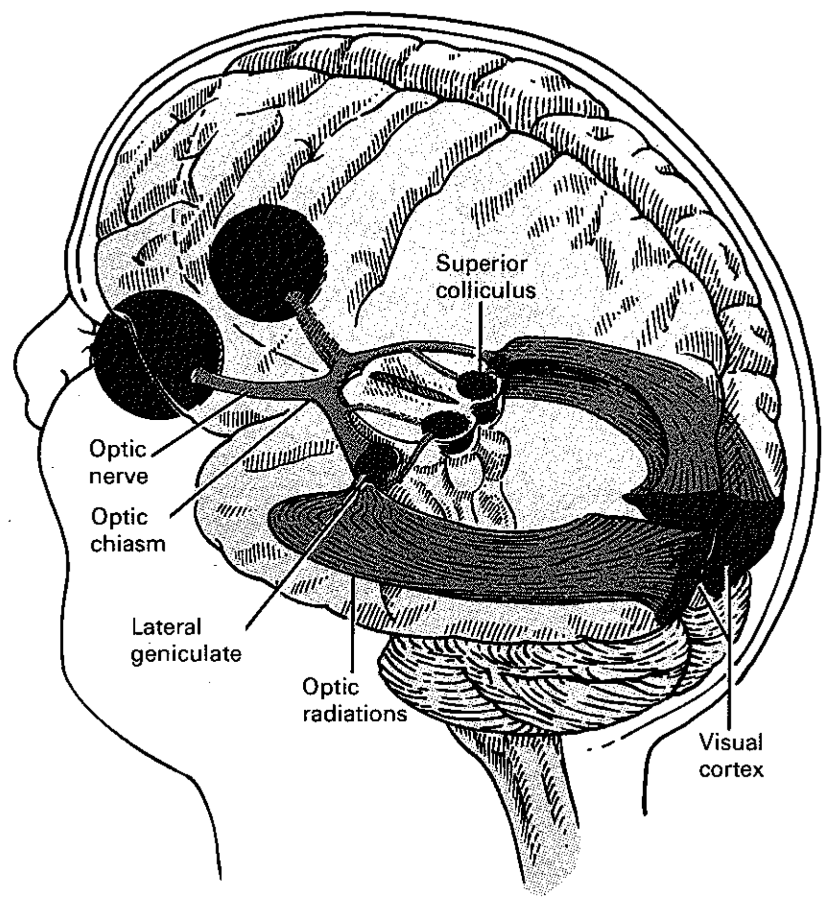


Figure 1.3.1 The human visual system. Visual processing begins in the eyes and is relayed to the brain by the optic nerve. The primary visual pathway then goes from the lateral geniculate nucleus to occipital cortex via the optic radiations. From there, visual information travels to other parts of the brain. A secondary pathway goes from the optic nerve to the superior colliculus and then to other brain centers. (From Rosenzweig & Leiman, 1982.)

Anatomy of the Eye. There are some obvious anatomical facts about the eyes that almost everyone knows. Humans have two eyes, which are approximately spherical in shape except for a bulge at the front. Located at about the horizontal midline of the head, they sit in nearly hemispherical holes in the skull, called the **eye sockets**, that hold them securely in position yet allow

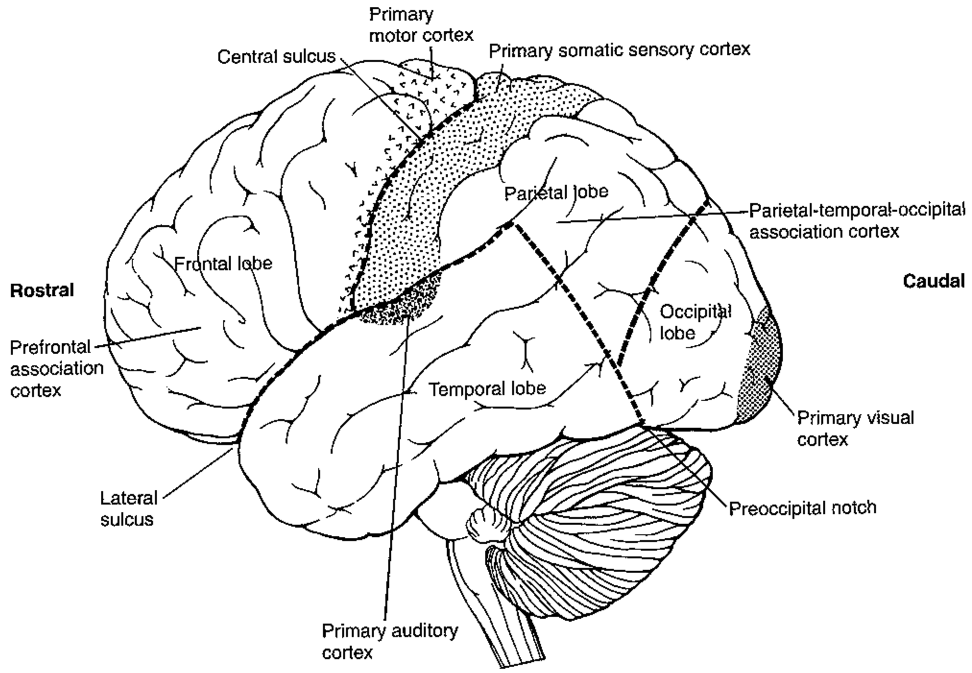


Figure 1.3.2 Visual areas of the human cortex. From primary visual cortex in the occipital lobe, visual information separates into two major pathways: a lower (ventral) one that goes to the inferior regions of the temporal lobe and an upper (dorsal) one that goes to the parietal lobe. Both pathways eventually make their way to the frontal lobe. (From Kandel, Schwartz, & Jessell, 1991.)

their direction to be changed by rotation. Each eye is moved by the coordinated use of six small, strong muscles, called the **extraocular muscles**, which are controlled by specific areas in the brain (see Section 11.1.2). Eye movements are necessary for scanning different regions of the visual field without having to turn the entire head and for focusing on objects at different distances. Eyelids and eyelashes protect the eyes, and tears keep them moist and clean.

Few of these simple facts are universally true of nonhuman eyes, however. Some species, such as pigeons and owls, cannot move their eyes in sockets but must move their whole heads. This limits the rate at which new views of the world can be registered, simply because heads are so much heavier and harder to turn than eyes. Eye position also differs in important ways across species. Although human eyes are both positioned at the front of our heads, many animals have their eyes located very much nearer the sides. Frontal placement provides two visual fields with a large area of overlap, as shown in Figure 1.3.3 but a correspondingly smaller total view of the environment. The benefit of overlapping visual fields is that binocular ("two-eyed") vision is important for precise depth perception, as we will discover in Section 5.3. Being able to gauge the distance to an object precisely is evolutionarily advantageous for predators, who need this information for an effective attack, and so the eyes of hunters tend to be placed frontally with overlapping fields. In contrast, panoramic visual fields are advantageous for the hunted to monitor as much of the world as possible for danger, and so prey tend to have laterally placed eyes. Other animals have frontally or laterally placed eyes for a variety of reasons, all of which reflect the same tradeoff between accuracy of depth perception and coverage of the visual world.

Physiological Optics. Because they constitute the "front end" of the visual system, the eyes have two important optical functions in common with cameras: to gather light reflected from surfaces in the world and to focus it in a clear image on the back of the eye. If insufficient light is admitted, the image will be dim and ineffective for vision. If the image is not clearly focused, fine-grained optical information will be irrevocably lost, and spatial perception will suffer.

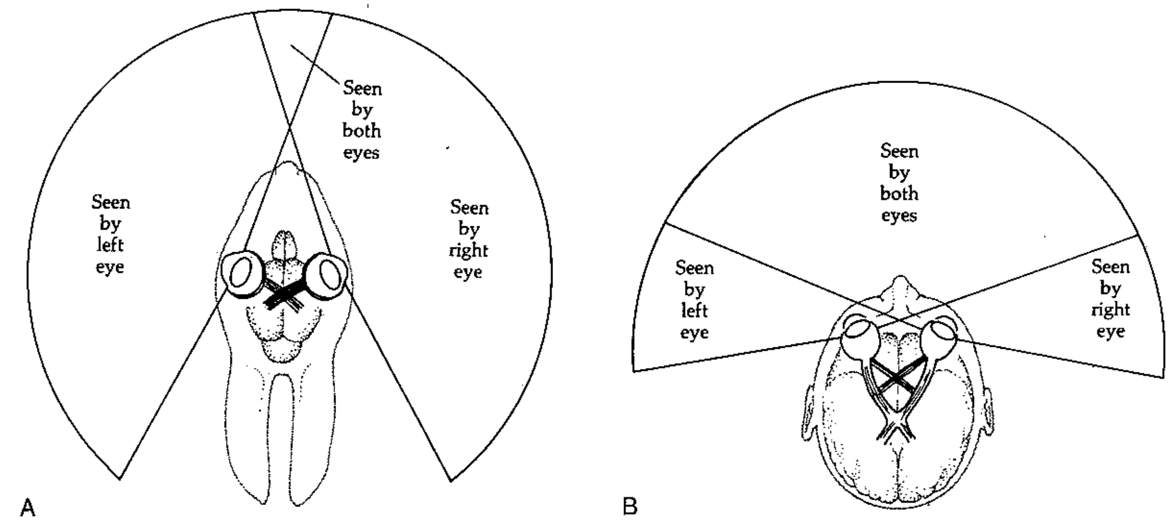


Figure 1.3.3 Overlap of monocular visual fields. In rabbits (A), lateral placement of the eyes produces a wide field of view but little binocular overlap. In people (B), frontal placement of the eyes produces a narrower field of view but a large area of binocular vision. (From Sekuler & Blake, 1985.)

There are many parts of the eye that accomplish different optical functions (see Figure 1.3.4). To find out how the eye is constructed and how it registers light, let's follow a photon through its various structures. First, light enters the **cornea**, a transparent bulge on the front of the eye behind which is a cavity filled with a clear liquid, called the **aqueous humor**. Next it passes through the **pupil**, a variably sized opening in the opaque **iris**, which gives the eye its external color. Just behind the iris, light passes through the *lens*, whose shape is controlled by **ciliary muscles** attached to its edge. The photon then travels through the clear **vitreous humor** that fills the central chamber of the eye. Finally, it reaches its destination, striking the **retina**, the curved surface at the back of the eye. The retina is densely covered with over 100 million light-sensitive **photoreceptors**, which convert light into neural activity. This information about the light striking the retina is then sent to the visual centers in the brain.

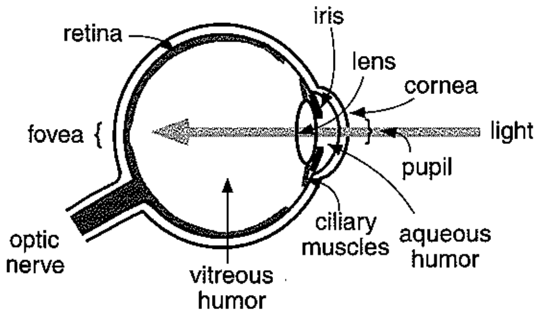


Figure 1.3.4 A cross section of the human eye. Light enters the eye through the cornea, aqueous humor, lens, and vitreous humor before striking the light-sensitive receptors of the retina, where light is converted into electrochemical signals that are carried to the brain via the optic nerve.

Each of the eye's components performs a critical role in the eye's sensory capabilities. The amount of light striking the retina is controlled by the iris and pupil. When illumination is low, the pupil dilates so that more light strikes the retina. When illumination is high, the pupil constricts so that less light strikes it. Interestingly, pupil size also changes in response to psychological factors. For instance, positive emotional reactions dilate the pupil, as shown by the fact that most men's pupils dilate when viewing pictures of nude females and most women's pupils dilate when viewing pictures of nude males (Hess & Polt, 1960). Pupil size also reflects mental effort, dilating when concentration is intense (Hess & Polt, 1964; Kahneman & Beatty, 1967). All of this occurs without our knowledge, since we have no conscious control over the mechanisms that alter the size of our pupils.

Light entering the eye is useful only if it is focused on the retina in a reasonably sharp image. Because the

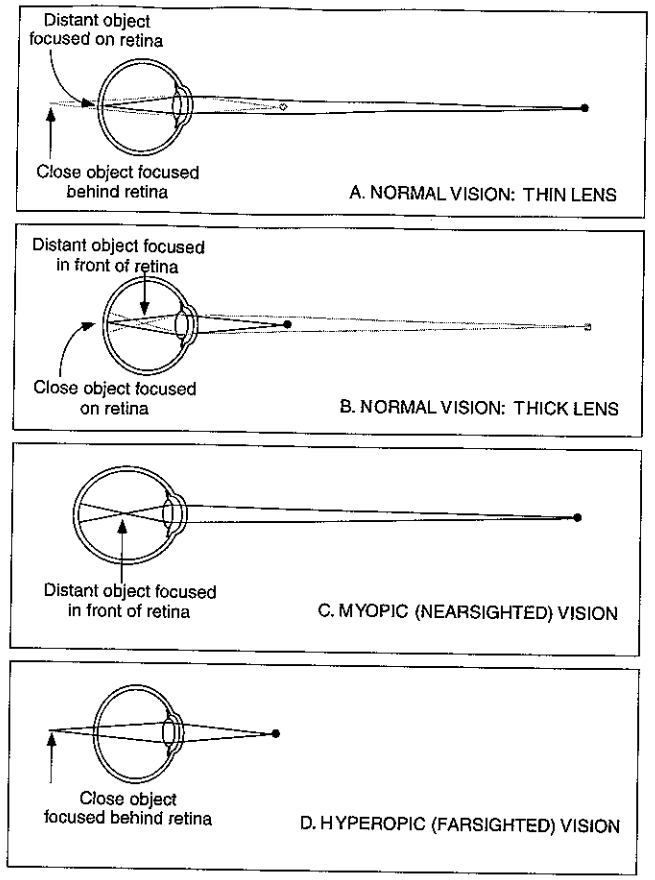


Figure 1.3.5 Focusing light by the lens. In a normal eye (A), a thin lens focuses light from distant objects, but not from close objects, on the retina. In a normal eye (B), a thick lens focuses light from close objects, but not from far objects, on the retina. In an uncorrected myopic (nearsighted) eye (C), light from distant objects is focused in front of the retina. In an uncorrected hyperopic (farsighted) eye (D), light from near objects is focused behind the retina.

pupil is much bigger than a pinhole, the light must be bent inward toward the center of the eye to focus it on the retina (see Figure 1.3.5). The curvature of the cornea does most of the job, and the lens does the rest. The lens is particularly important, however, because of its *variable focusing ability*. The lens's optical properties can be altered by changing its shape, a process called **accommodation**. To bring distant objects into focus on the retina, the lens must be thin (see Figure 1.3.5A). This is accomplished by relaxing the ciliary muscles. To focus on nearby objects, the lens must be thick (see Figure 1.3.5B). This is accomplished by contracting the ciliary muscles. People who are nearsighted (or **myopic**) have excellent near vision but cannot focus distant objects properly because their lens is too thick for the depth of their eye (see Figure 1.3.5C). People who are farsighted (or **hyperopic**) can see well far away but cannot focus on nearby objects properly because their lens is too thin for the depth of their eye (see Figure 1.3.5D). As people get older, the lens gradually loses its elasticity, so it cannot become thick enough for near vision—a condition known as **presbyopia**. At about 40–50 years of age, many people who have never worn glasses before begin to need them for reading and other near work, and many people who already wore glasses must switch to bifocals.

Most of the light that comes though the pupil gets absorbed by photoreceptors in the retina and causes neurochemical responses in them. Some gets absorbed by the lens, vitreous humor, blood vessels, and various nonreceptor cells in the retina. This is to be expected, because none of these structures are perfectly clear, and some of them (like blood vessels and the pigment epithelium) are quite dark, absorbing a great deal of light. But the majority of light entering the eye makes it through to the retina.

1.3.2 The Retina

After the optics of the eye have done their job, the next critical function of the eye is to convert light into neural activity so that the brain can process the optical information. To understand how this occurs, we must briefly explain the basic building blocks of the brain and how they work.

Neurons. The main functional component of the brain is generally believed to be the **neuron**: a specialized type of cell that integrates the (input) activity of other neurons that are connected to it and propagates that integrated (output) activity to other neurons. This process of integration and transmission is accomplished by a complex series of biochemical events within the neuron. The parts of a neuron are illustrated in Figure 1.3.6, and their function can be described as follows:

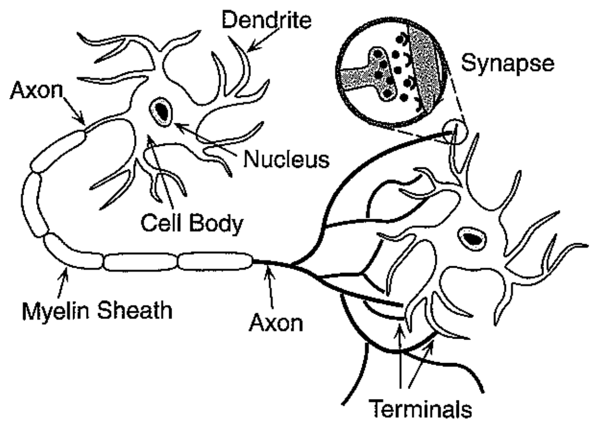


Figure 1.3.6 A typical neuron. A neuron is a cell that consists of a cell body that integrates graded electrical signals from its dendrites and transmits the result via discrete action potentials. These spikes travel along an axon, which is encased in a myelinated sheath, to terminals, where neurotransmitters are released at synapses to stimulate the dendrites of other neurons.

1. The **dendrites** are thin protrusions from the cell body that collect chemical signals from other neurons and convert them into electrical activity along the thin membrane that encloses the cell. This electrical activity is a **graded potential**: an electrical difference between the inside and outside of the dendrite whose value can vary continuously within a range, depending on how strongly the dendrite has been stimulated by other neurons.

2. The **cell body** contains the nucleus and cellular machinery. The membrane around the cell body integrates the electrical signals arriving from all the dendrites, again coded in terms of a graded potential, and converts it into a series of all-or-none electrical potentials (called **action potentials**, **nerve impulses**, or simply **spikes**) that are propagated along the axon.

3. The **axon** is a long, thin projection of the neuron along which action potentials are propagated to other neurons, often over a considerable distance. Most of the axon of most neurons is covered by a **myelin sheath**, which speeds the conduction of action potentials. The strength of the integrated signal that the axon transmits is encoded primarily in its **firing rate**: the number of electrical impulses it generates in a given amount of time (e.g., spikes per second).

4. The **terminals** are the branching ends of the axon at which the electrical activity of the axon is converted back into a chemical signal by which it can stimulate another neuron. This is accomplished by releasing a **neurotransmitter** into the small gap between the terminal and the dendrite of the next neuron. Neurotransmitters are chemical substances that are capable of exciting the dendrites of other neurons. The signal strength transmitted at the terminal is determined by the amount of neurotransmitter released.

5. The **synapse** is the small gap that exists between the terminals of one neuron and the dendrites of another. The neurotransmitter that is released into the synapse rapidly crosses the gap and affects the next neuron's dendrite by occupying specialized sites on its membrane. This is where the chemical signal from one neuron is converted to an electrical signal in the next, as described in item 1 above.

Neurons thus receive input from some neurons and send their output to other neurons. But before any of this neural activity can occur, something must convert energy in the environment into the form needed by these neurons. In the visual system, this function is carried out by *photoreceptors*: specialized retinal cells that are stimulated by light energy through a complex process that we will describe shortly. Once the optical information is coded into neural responses, some initial processing is accomplished within the retina itself by several other types of neurons, including the **horizontal**, **bipolar**, **amacrine**, and **ganglion cells**, all of which integrate responses from many nearby cells (see Figure 1.3.7). We will consider the function of these other neurons in the retina in later chapters (e.g., Chapters 3 and 4). The axons of the ganglion cells carry information out of the eye through the **optic nerve** and on to the visual centers of the brain.

Photoreceptors. There are two distinct classes of photoreceptor cells in the retina: **rods** and **cones**. Their names were chosen to describe their shapes, as shown in the scanning electron micrograph in Figure 1.3.8. Rods are typically longer and have untapered (rodlike) ends, whereas cones are shorter, thicker, and have tapered (conelike) ends. Rods are more numerous (about 120 million), extremely sensitive to light, and located everywhere in the retina except at its very center (see Figure 1.3.9). They are used exclusively for vision at very low light levels (called **scotopic conditions**): at night, at twilight, or in dimly lighted rooms. Cones are less abundant ("only" 8 million), much less sensitive to light, and heavily concentrated in the center of the retina, although some are found scattered throughout the periphery (see Figure 1.3.9). They are responsible for

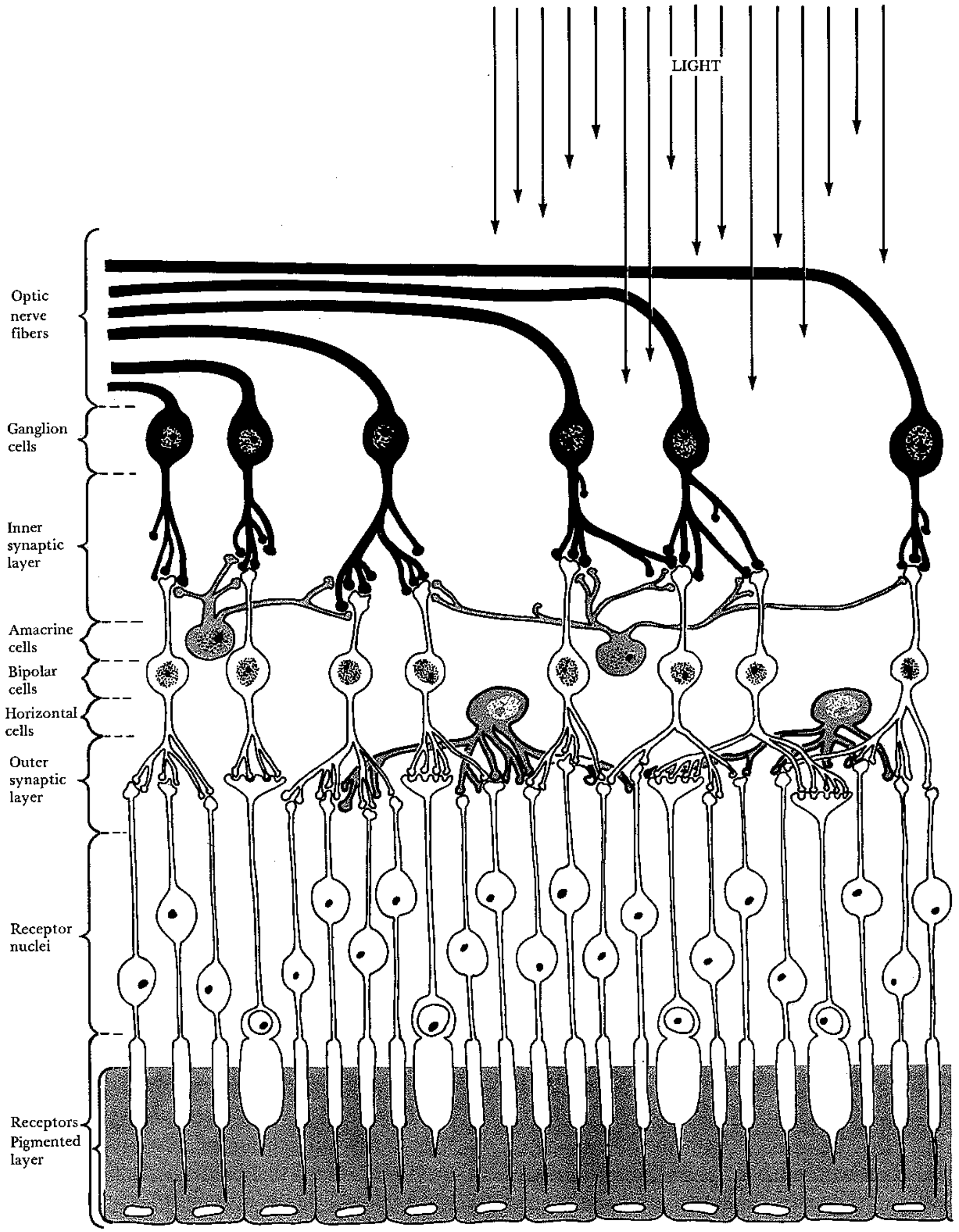


Figure 1.3.7 The human retina. The retina consists of five major types of neurons: receptors (the rods and three kinds of cones, shown near the bottom), bipolar cells, ganglion cells, horizontal cells, and amacrine cells. (From Lindsay & Norman, 1977.)

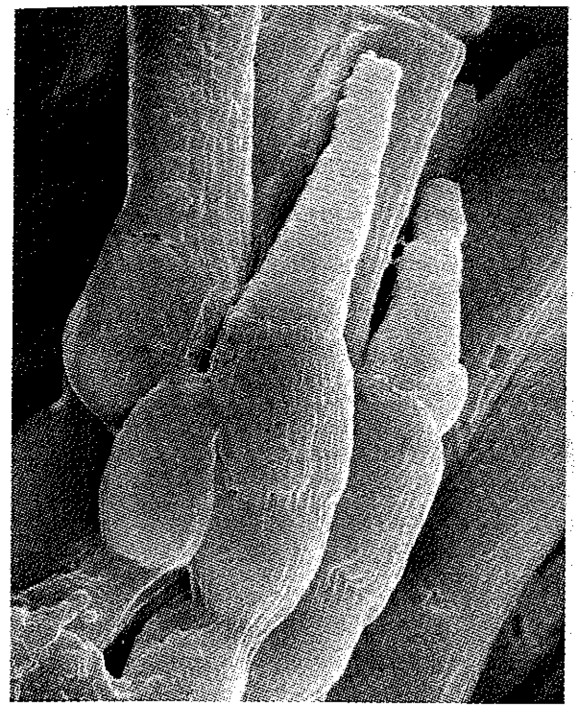


Figure 1.3.8 Scanning electron micrograph of rods and cones. The outer segments of rods have an untapered cylindrical shape, and those of cones have a tapered conical shape. (From Lewis, Zeevi, & Werblin, 1969.)

our visual experiences under most normal lighting conditions (called **photopic conditions**) and for all our experiences of color. There is a small region, called the **fovea**, right at the center of the retina that contains nothing but densely packed cones (see Figure 1.3.9). The visual angle covered by the fovea is only about 2 degrees, the size of your thumbnail held at arm's length. It is more important than its small size would suggest, however, because it is here that both color and spatial vision are most acute.

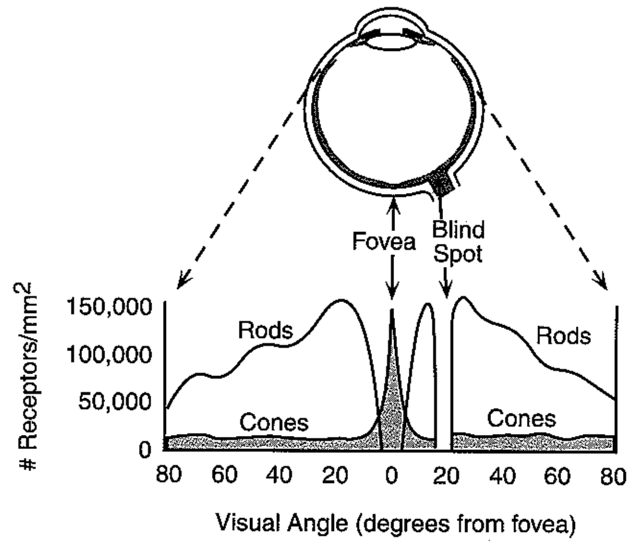


Figure 1.3.9 Distribution of rods (solid curve) and cones (shaded region) in the human retina. Notice that the fovea is populated almost exclusively by cones and that rods are much more plentiful than cones in the periphery.

The obvious question at this point is *how* photoreceptors manage to change the electromagnetic energy of photons into neural activity. It is a complex and truly ingenious process that is now reasonably well understood. Both rods and cones have two basic parts, as shown in Figure 1.3.10: the **inner segment** which contains the nucleus and other cellular machinery, and the **outer segment** which contains billions of light-sensitive pigment molecules. These pigment molecules are embedded in the membranes of hundreds of disks stacked like pancakes perpendicular to the long axis of the outer segment.

The pigment in rods is called **rhodopsin**, and a great deal is now known about how this photosensitive molecule converts light into electrochemical energy. When a photon strikes a rhodopsin molecule and is absorbed by it, the molecule changes its shape in a way that alters the flow of electric current in and around the pigment molecule. We will not be concerned with the details of this complex biochemical reaction except to note that its result is to produce electrical changes in the outer membrane of the receptor. These changes are then propagated down the outer membrane to the synaptic region of the receptor, where chemical transmitters affect the next neuron.

The electrical changes that result from many photons being absorbed within the same receptor are integrated in the response of its outer membrane. The resulting overall change in the electrical potential between inside and outside of the cell is *graded,* in that it is continuous (rather than discrete), unlike that in most other parts of the nervous system. In quantitative terms, the graded response of a photoreceptor is a *logarithmic function* of the number of photons absorbed. This means that the same overall increase in output will require very few addi-

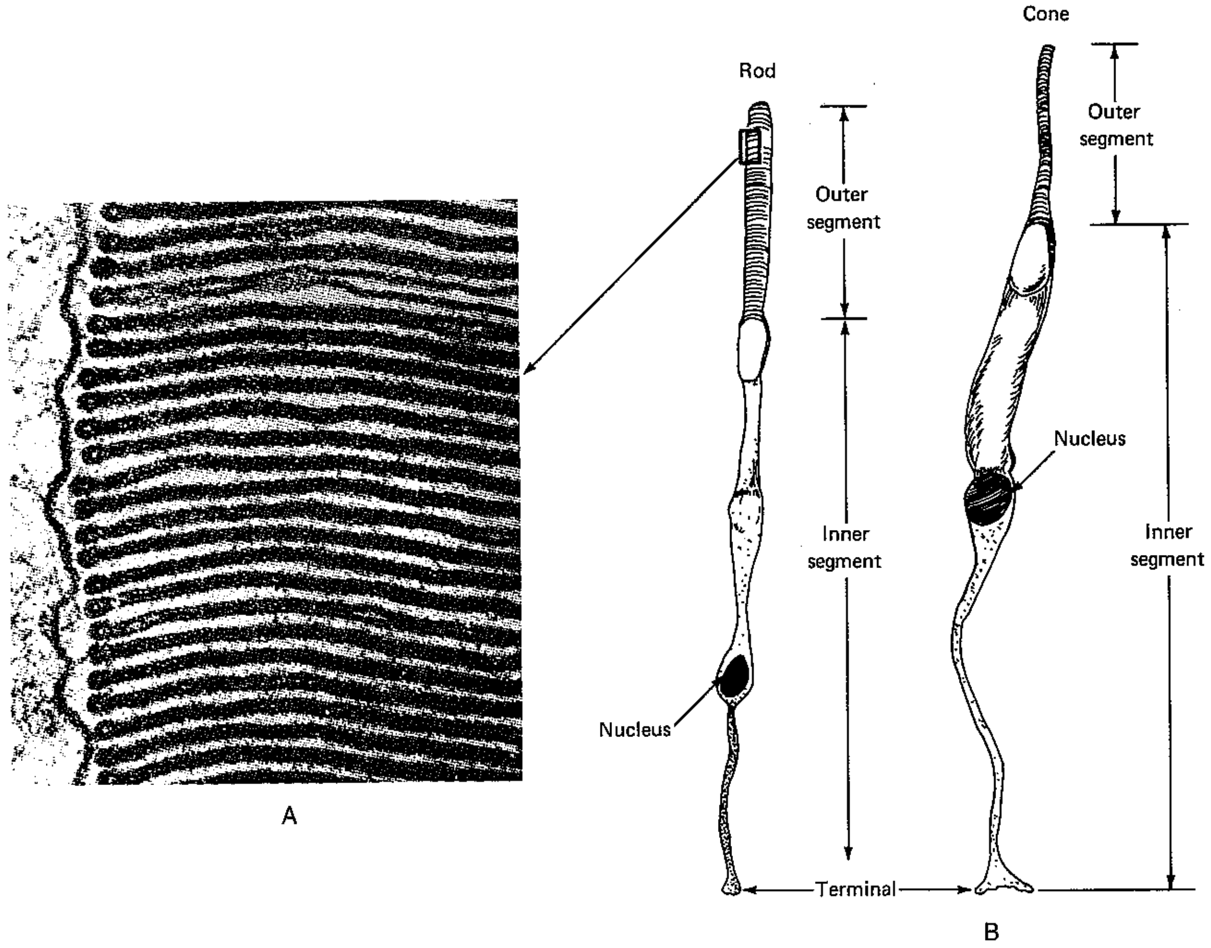


Figure 1.3.10 The anatomy of rods and cones. The outer segment contains thin disks with pigment molecules embedded in them. Absorption of light by these molecules results in chemical changes that cause electrical changes in the inner segment. These electrical signals are transmitted to neurons via the terminal. (From Allen, 1967.)

tional photons at low light levels but lots of additional photons at high light levels (see Figure 1.3.11).

This complex chain of events in the outer segment is called **pigment bleaching** because the change in molecular shape brought about by light also causes the molecule to change color. Before a rhodopsin molecule is bleached by absorbing light, it looks deep purple; afterward, it is almost transparent. Because it becomes transparent when it absorbs a photon, it is very unlikely to absorb another photon after it is bleached. Once bleached, the pigment molecules are restored to their prior unbleached state by the action of enzymes in the **pigment epithelium** behind the retina.

These changes in pigment bleaching can actually be measured in the eyes of living humans by a process called **retinal densitometry** (Rushton & Campbell, 1954). A beam of light with known intensity is focused on the retina. Part of the light is absorbed by the receptors (and pigment epithelium), and the rest is reflected back out the eye. The intensity of the reflected beam can then be measured. By making this measurement at different times after a bright flash of light that bleaches most of the pigment molecules, the rate at which pigment regeneration takes place can be determined. Full regeneration takes about 30 minutes in rods; it takes only about 6 minutes in cones.

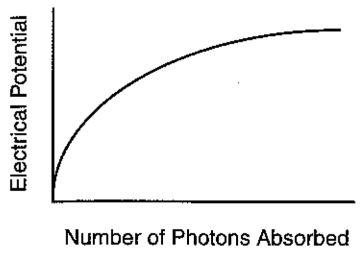


Figure 1.3.11 The logarithmic relation between absorption of photons by receptors and their output. When few photons are absorbed, small changes in input produce large changes in output. When many photons are absorbed, large changes in input produce small changes in output.

These differences in regeneration time in the rods and cones result in the uneven course of dark adaption alluded to in Section 1.1.3. The precise development of dark adaptation in visual experience is measured objectively by finding the observer's **absolute threshold**—the dimmest light that can be seen—at various times after an observer is placed in a dark room. The results from such an experiment are shown in Figure 1.3.12. The intensity of the dimmest visible light is plotted as a function of how long the observer has been in darkened conditions. This graph reflects dark adaptation in the fact that the threshold curve decreases over time: The longer the observer has been in the dark, the dimmer is the spot that the subject can just barely see. Notice, however, that this decrease in threshold is not smooth, having two distinct parts: one from 0 to about 8 or 10 minutes and the other from 10 to about 40 minutes. Many further experiments have shown that these two parts reflect the different time courses of dark adaptation in rods and cones, as described above.

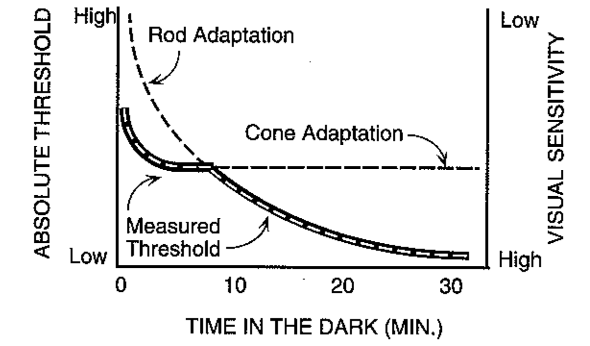


Figure 1.3.12 The dark adaptation function. Visual threshold (left scale) and sensitivity (right scale) are plotted as a function of time in the dark. The function results from two different components. Cones adapt rapidly but reach only a moderate level of sensitivity. Rods adapt more slowly but reach a high level of sensitivity (low threshold).

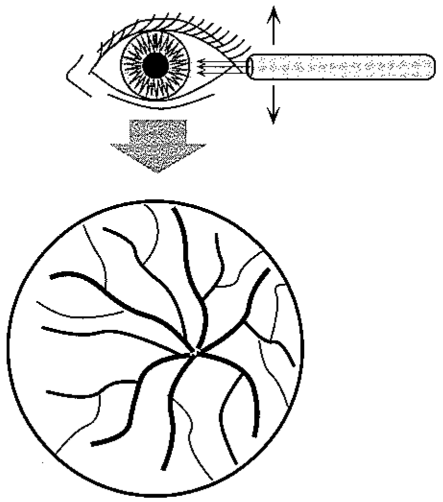


Figure 1.3.13 Seeing your retinal blood vessels. To see the blood vessels on your retina, carefully place a small penlight against the outer corner of your eye and gently shake it up and down. The weblike lines that you then see are caused by the shadows of your retinal blood vessels.

Peculiarities of Retinal Design. Although the human eye is an excellent organ for detecting light, certain aspects of its design seem peculiar enough to warrant special mention. For example, one would logically expect the receptor cells to be the *first* layer of the retina that incoming light encounters, but they are actually the *last* (see Figure 1.3.7). Not only that, but both rods and cones appear to be pointing *backward*; the light-sensitive outer segment is the most distant part of the receptor cell from incoming light. The reason for this unusual arrangement is probably that the enzymes that are needed for pigment regeneration are in the pigment epithelium, which is opaque. Because the receptor disks must be adjacent to this vital biochemical resource, they must also be at the back of the retina. Luckily for our visual abilities, the retinal cells and axons in front of the receptors are fairly transparent, so the optical quality of the image does not suffer as much as one might imagine.

In addition to the nearly transparent retinal cells that lie in the path of incoming light, however, there are also many blood vessels that nourish the retina. Because

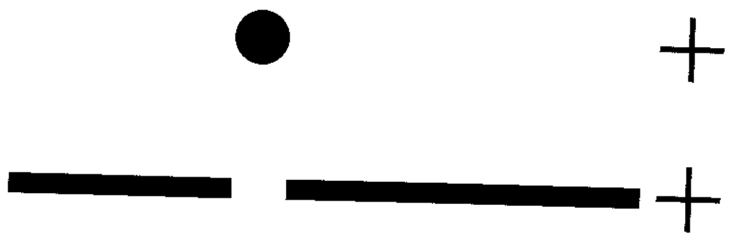


Figure 1.3.14 Spotting your blind spot. With your right eye closed, fixate the upper cross with your left eye. Starting with the book close to your face, move it slowly away. At a distance of about 9 inches, the spot on the left will disappear. Doing the same thing with the lower display will cause the gap in the line to fill in so that it appears complete and uninterrupted.

these are dark, one would expect them to be clearly visible all the time. The reason we do not normally see them is that the brain adapts completely to their presence and fills in the part of the image over which the blood vessels cast their shadows. In fact, you *can* see your retinal blood vessels under the proper circumstances. When you get an eye examination, for example, you can see them quite vividly as the ophthalmologist shines a light into your eye at unusual angles. You can also see them for yourself using the following procedure: Look at a uniform surface (a plain wall or blank piece of paper will do) and shine a pen flashlight directly into the white of your eye against the outside corner, as illustrated in Figure 1.3.13. When you gently shake the penlight up and down, you will be able to see the blood vessels clearly because their shadows are now moving over different receptors, ones to which the brain has not adapted.

Another curiosity in the anatomical design of the eye exists where the axons of the ganglion cells leave the eye at the optic nerve. This region is called the **optic disk** (also known as the **blind spot**) and it contains no receptor cells at all. However, we do not experience blindness there, except under very special circumstances. There are two reasons for this: First, the blind spots are positioned so that receptors in one eye register what is missed in the blind spot of the other eye. Second, the visual system fills in this region with appropriate sensory qualities, just as it does the shadows of the blood vessels. We do not yet know how this is accomplished, but it works quite effectively.

To spot your blind spot, hold this book near your face, close your right eye, and fixate on the upper-right cross in Figure 1.3.14 with your left eye. Now move the book slowly further away while continuing to focus on the upper-right cross. When the circle on the left is in your blind spot, it will disappear. Notice how the visual system fills in this area with the background whiteness of the page. This is a curious fact, since it is not obvious how the visual system can "see" the white page, which is *not* there, and fail to "see" the dot, which *is*.

Higher brain processes, probably in visual cortex, seem to fill in the part of the visual field corresponding to the blind spot with appropriate information, which is then experienced consciously. You can convince yourself of this using the lower-right cross and the broken line next to it. Again, close your right eye and fixate the cross with your left eye as you move the book further away. When the gap in the line is in your blind spot, the gap will disappear. It is filled in with an illusory line that completes the broken one. As we said before, what you see isn't always what you get!

Yet another peculiarity of retinal physiology concerns the encoding of information about light intensity. Since the task of the eye is to convert light into neural responses, the natural expectation is that receptors would not respond in darkness and would increase their synaptic output as the intensity of light increases. But exactly the opposite actually occurs in the receptors of vertebrates (Toyoda, Nosaki & Tomita, 1969). The response to a flash of light is a *decrease* in synaptic activity at the receptors' output. Although this result greatly surprised visual scientists, it probably should not have. The important fact is only that neural activity should preserve the information present in light intensities over the visual field, and this task can be accomplished equally well by either positive or negative correlations between neural activity and light intensity. As it turns out, receptors in invertebrate eyes work in the opposite—and more intuitively obvious—way, increasing their output for increases in light intensity. And the seeming anomaly of decreased activity in vertebrate receptors is rectified by the next synapse in which increased light produces increased release of neurotransmitter.

Pathways to the Brain. The axons of the ganglion cells leave the eye via the optic nerve which leads to the **optic chiasm**, named for its resemblance to the Greek letter χ ("chi," pronounced "kye" as in "sky"). Here the fibers from the nasal side of the fovea in each eye cross over to the opposite side of the brain while the others remain on the same side (see Figure 1.3.15). The result is that the mapping from external visual fields to the cortex is completely crossed: All of the information from the left half of the visual field goes to the right half of the brain, while all the information from the right visual field goes to the left half of the brain.

From the optic chiasm, there are two separate pathways into the brain on each side. The smaller one (only a few percent) goes to the **superior colliculus**, a nucleus in the brain stem. This visual center seems to process primarily information about where things are in the world and to be involved in the control of eye movements. The larger pathway goes first to the **lateral geniculate nucleus** (or **LGN**) of the thalamus and then to the **occipital cortex** (or **primary visual cortex**). We will have much more to say about the extensive processing that goes on in the visual cortex in Chapter 4, but for now we will merely describe the gross anatomical and physiological organization of this complex structure.

1.3.3 Visual Cortex

Two facts about the human cortex are obvious from inspecting it with the naked eye. The first is that its surface is highly convoluted or folded. This is because the cortex is actually a layered sheet of neurons, and the convolutions result from trying to fit a large sheet of cortex into an small inflexible skull. A second obvious macroscopic feature of the cortex is that it is divided into two halves, or **cerebral hemispheres**, that are approximately symmetrical. Otherwise, the cortex looks pretty much homogeneous, and it is completely unclear from looking at it how it might work. Indeed, its mechanisms are so obscure that early scientists failed to realize that the brain was the organ of mental functioning. Aristotle believed that the seat of mental capacities was in the heart and that the brain was essentially a heat sink to cool the blood!

Localization of Function. One of the first questions to be raised about the brain, once its true importance was realized, was whether or not its functions are *localized*: Are different mental faculties located in different anatomical regions, or are all functions spread throughout the entire brain? Early pseudoscientific support for the localization hypothesis came from **phrenology**, the study of the shape of people's skulls. Phrenologists claimed that the size of the lumps, bumps, and bulges on a person's skull indicated the size and development of the brain structures underneath. Their approach was to find correspondences between skull measurements and assessments of mental attributes, such as "ambition," "calculation," and "spirituality." The result was a collection of bizarre phrenological maps, as shown in Figure 1.3.16, charting the position of such functions. These supposed correlations were generally unfounded, however, and real scientific support for the concept of localization of function did not appear until the effects of brain damage were studied systematically during the late nineteenth century.

The scientific basis of localization of function was established when physicians began to perform postmortem analyses of the brains of patients who had acquired specific mental disabilities during their lifetimes from strokes or head injuries. They found that certain types of deficits were very strongly correlated with damage to certain regions. Among the findings was that visual dysfunction characteristically occurred when there was damage to the posterior parts of the brain, mainly in the occipital lobe (see Figure 1.3.2). Particularly important was a study by a young Japanese physician of visual impairments in soldiers resulting from gunshot wounds to the head in the Russo-Japanese War of 1904–1905 (Glickstein, 1988). As a result of this and other neuropsychological studies, it is now well established that the occipital lobe is the primary cortical receiving area for visual information and that there are other cortical areas that are similarly specialized for the other sensory modalities: audition, taste, touch, and smell.

Still, there has been considerable controversy as to whether or not there is further localization of function.

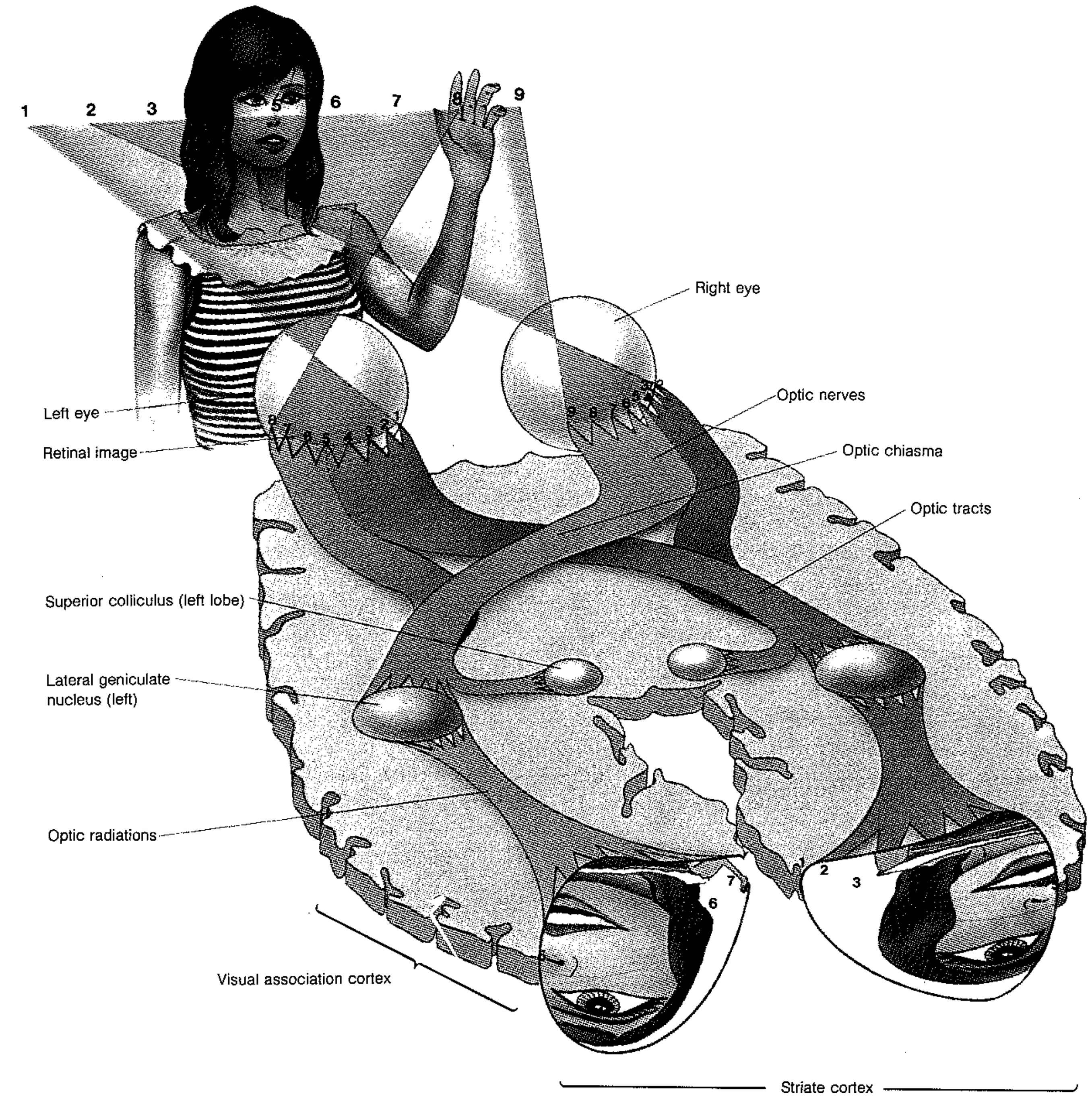


Figure 1.3.15 Neural pathways from the eye to visual cortex. Information from the inner (nasal) portion of each retina crosses over to the opposite side of the brain so that each side of the brain receives input only from the opposite half of the visual field. Note that the representation of the central portion of the visual field receives disproportionate representation in visual cortex. (From Frisby, 1979.)

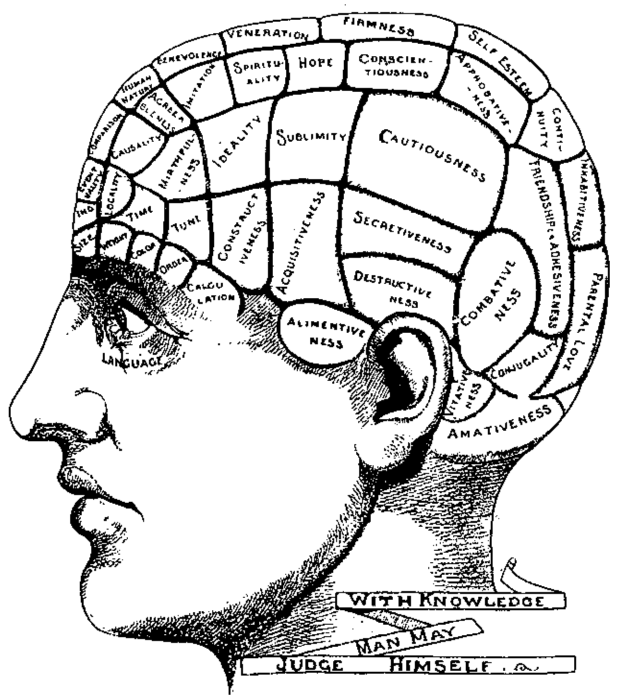


Figure 1.3.16 A phrenological map. Phrenologists believed that they could determine the location of brain functions by measuring characteristics of the skull and correlating them with psychological abilities. (From Kolb & Whishaw, 1996.)

In a well-known series of studies, the physiological psychologist Karl Lashley and his colleagues tried to localize memories—visual and otherwise—for the information that rats learned from running mazes (Lashley, 1929, 1931, 1950). They removed various amounts of cortex at various different locations and never found any specific locus that seemed to correspond to those memories. The only variable that seemed to make a difference was the total amount of cortex that was excised: The more brain tissue that was removed, the worse the animal performed, a result that Lashley christened the "law of mass action." Because of this finding, he questioned the notion that memories are localized at all, instead proposing that they are distributed more or less equally over the whole cortex.

We still do not know the extent to which function is precisely localized, but the evidence increasingly supports the view that it is. In visual cortex of monkeys, for example, over 30 different visual areas have been identified, with more being discovered each year (Van Essen & DeYoe, 1995). The best known and understood of these visual centers lie in the **occipital**, **parietal**, and **temporal lobes** of the cortex, the locations of which are indicated in Figure 1.3.2.

Occipital Cortex. The most complete anatomical and physiological data about visual cortex come from old-world monkeys, such as the macaque. Behavioral analyses have shown that their visual abilities are strikingly similar to those of humans, making them a good animal model for the human visual system (De Valois & De Valois, 1988). Since the cellular exploration of visual cortex began more than two decades ago with the pioneering studies of Hubel and Wiesel (1959, 1962), a great deal has been learned about both the anatomy and the physiology of various mammalian visual systems. The structure and function of dozens of distinct visual areas are being explored, and their interrelationships are being determined by a variety of techniques. Although it would be a gross overstatement to say that we understand how visual cortex works, we are at least beginning to get some glimmerings of what the assorted pieces might be and how they might fit together.

The first steps in cortical processing of visual information take place in the **striate cortex**. This part of the occipital lobe receives its input from the LGN on the same side of the brain, so the visual input of striate cortex, like that of LGN, is completely crossed: The left visual field projects to the right striate cortex, and the right visual field projects to the left striate cortex (see Figure 1.3.15). Both sides are activated by the thin central vertical strip, measuring about 1 degree of visual angle in width, that separates the two sides of the visual field. The cells that are sensitive to this strip in one side of the brain are connected to the corresponding cells on the other side of the brain through the **corpus callosum**, the large fiber tract that allows communication between the two cerebral hemispheres.

The mapping from retina to striate cortex is *topographical* in that nearby regions on the retina project to nearby regions in striate cortex. This transformation preserves *qualitative* spatial relations but distorts *quantitative* ones, much as an image on a rubber sheet can be distorted when it is stretched without being torn. Figure 1.3.15 depicts the approximate distortion produced by this topological transformation: The central area of the visual field, which falls on or near the fovea, receives proportionally much greater representation in the cortex than the periphery does. This is called the **cortical magnification factor**. It reflects the fact that we have more detailed spatial information about objects in the central region of the retina than about those in peripheral regions.[4] It does *not* imply that perception of environmental space is somehow distorted so that objects in the center of the visual field appear to be bigger than those in the periphery. You can convince yourself of this simply by holding your two hands out at arm's length and focusing on one of them. If perception of space were distorted by the cortical magnification factor, the hand that you focus on would seem much larger than the other hand, but this does not happen. Instead, you merely perceive the fine details of the focused hand more clearly, such as its lines, veins, and fingernails.

This is but a small part of what is known about the internal structure of striate cortex. We will save the rest of the story for Chapter 4, in which the physiology of this area of cortex plays a central role in our understanding of image-based processing. Its output projects to many other parts of the visual cortex, including other areas in the occipital lobe as well as parts of the parietal and temporal lobes. Different areas are involved in different sorts of visual processing. One of the important features that marks such areas as "visual" is that, like striate cortex, they are organized topographically with respect to retinal locations. Indeed, visual cortex is a veritable patchwork quilt of small maps that code different aspects of retinal stimulation, including brightness, color, motion, depth, texture, and form.

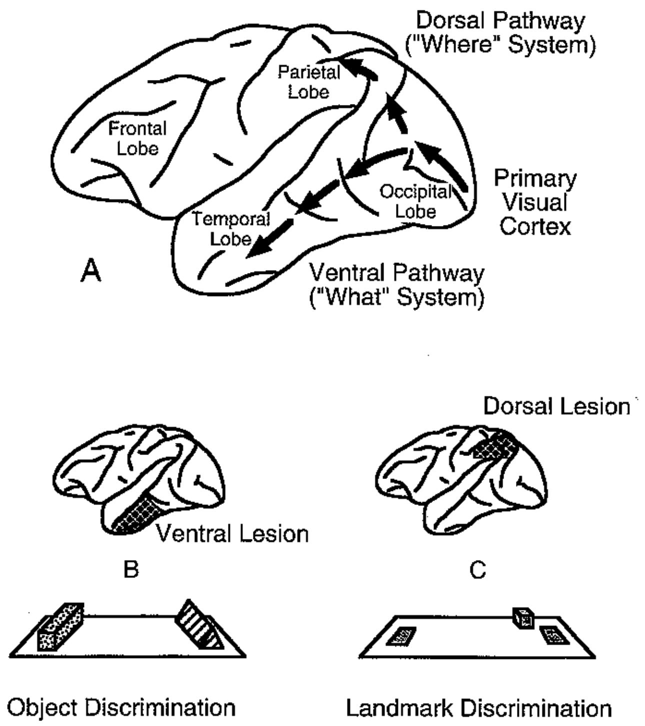


Figure 1.3.17 Two visual pathways in monkey cortex. The lower (ventral) pathway goes from occipital cortex to the temporal lobe and is believed to be specialized for object recognition (the "what" system). The upper (dorsal) pathway goes from occipital cortex to parietal cortex and is believed to be specialized for object location (the "where" system). Parts B and C illustrate experiments that support this division. (See text for details.) (After Goldstein, 1989.)

Parietal and Temporal Cortex. Part of the "big picture" about how cortical functioning is organized has come from lesion studies by physiologists Mortimer Mishkin, Leslie Ungerleider, and their colleagues. They have reported convincing evidence of a pronounced difference between the function of the visual areas in the temporal versus parietal lobes of the monkey's cortex (Figure 1.3.17A). The inferior temporal centers in the lower (ventral) system seem to be involved in *identifying objects*, whereas the parietal centers in the upper (dorsal) system seem to be involved in *locating objects*. These two pathways are often called the **"what" system** and the **"where" system**, respectively. The evidence for this claim comes from experiments in which monkeys were required to perform two different kinds of tasks after one or the other area of cortex had been surgically removed (Ungerleider & Mishkin, 1982; Mishkin, Ungerleider, & Macko, 1983).

One task was an *object discrimination* problem, as illustrated in Figure 1.3.17B. After being familiarized with a particular object, the monkey has to select the familiar object over a novel one to receive a food reward. This is an easy task for a normal monkey. It turns out also to be easy for a monkey with a portion of its parietal lobe removed. However, it is extremely difficult for the monkeys that are missing their inferior temporal cortex. The second task, called a *landmark discrimination* problem, is illustrated in Figure 1.3.17C. It also required the monkey to make a choice between two objects, but this time the two objects were identical in shape, differing only in their spatial proximity to a third, landmark object. This is also an easy task for normal monkeys, but this time it is the parietally lesioned monkeys that have trouble and the temporally lesioned ones that are unaffected.

The results of this and several other experiments support Ungerleider and Mishkin's (1982) hypothesis that the temporal pathway processes the shape information required to identify objects and that the parietal pathway processes the location information required to determine where objects are. More recently, other investigators have suggested that the parietal pathway is more accurately described as subserving spatially guided motor behavior (the "how" system), such as reaching and grasping (Goodale, 1995; Milner & Goodale, 1995). In any case, it seems almost inevitable that these two different kinds of information must get together somewhere in the brain so that the "what-where" connection can be made, but it is not yet known where this happens. One likely candidate is the frontal lobes, since they receive projections from both the parietal and temporal areas. The information processing might take a more circuitous route, however, and go through several intermediate cortical regions before linking up in some as-yet-unidentified place.

The distinction between these two visual pathways appears to be important in humans as well as monkeys. People who have damage in certain areas of their temporal cortex exhibit **visual agnosia**: a deficit in identifying certain kinds of objects by sight. One form of visual agnosia is specific to faces. The patient cannot recognize anybody by sight—even a spouse, parent, or child—but can immediately do so by hearing them speak. This disability is not due to the lack of visual experiences, for such patients are able to describe the faces they see quite precisely, including the presence of freckles, glasses, and so on. They just cannot tell whose face it is. This is easy to understand in terms of a breakdown in the ventral "what" pathway that leads to the temporal lobe.

Some patients with brain damage in the parietal lobe suffer from a syndrome known as **unilateral neglect**. Neglect is a complex pattern of symptoms that we will discuss more fully in Chapter 11; one of its main features is the apparent inability to attend to objects in the half of the visual field opposite to their brain damage. A person with a lesion in the right parietal lobe, for example, will eat only the food on the right side of the plate and draw only the right half of a picture he or she is asked to copy. It is as though objects in the left visual field were not there at all. Although many features of neglect are not yet well understood, it seems consistent with the possibility that some aspect of the "where" system is damaged.

These lesion results are particularly useful for understanding the overall function of gross regions of the brain, but they do not address more specific issues concerning the precise neural processing that underlies complex abilities like these. Doing so requires more detailed physiological studies using a variety of other techniques, including anatomical tracing of neural projections from one area to another and recording from individual cells in specific brain areas to find out what kinds of information they process. We do not yet understand this level of functioning in detail for areas outside primary visual cortex, but we will now briefly consider some of what is known at the present time.

Mapping Visual Cortex. Much of visual cortex in humans and closely related primates is hidden within the folds of the cortex. Figure 1.3.18 shows an anatomically correct depiction of the location and arrangement of some of the principle areas of visual cortex (areas V1 through V5) in the brain of macaque monkeys. Notice that these areas are part of the highly convoluted sheet of cortical neurons, much of which is not visible on the exterior surface of the brain. These are only a few of the visual areas, however, many of which lie quite far from primary visual cortex.

An anatomically distorted map of the currently known visual areas in the macaque monkey's cortex is shown in Figure 1.3.19. It is a strange view of a brain because the convoluted surface of the cortex has been "unfolded" so that areas hidden within the cortical folds (called *sulci*) can be seen in approximately correct spatial relations to visible areas. A side view of a normal macaque brain is given in the inset so that you can see

---

[4] This topological transformation actually takes place in a series of less radical stages. There is some magnification of the central regions in going from receptor to ganglion cells, more going from ganglion to LGN cells, and still more going from LGN to striate cells.

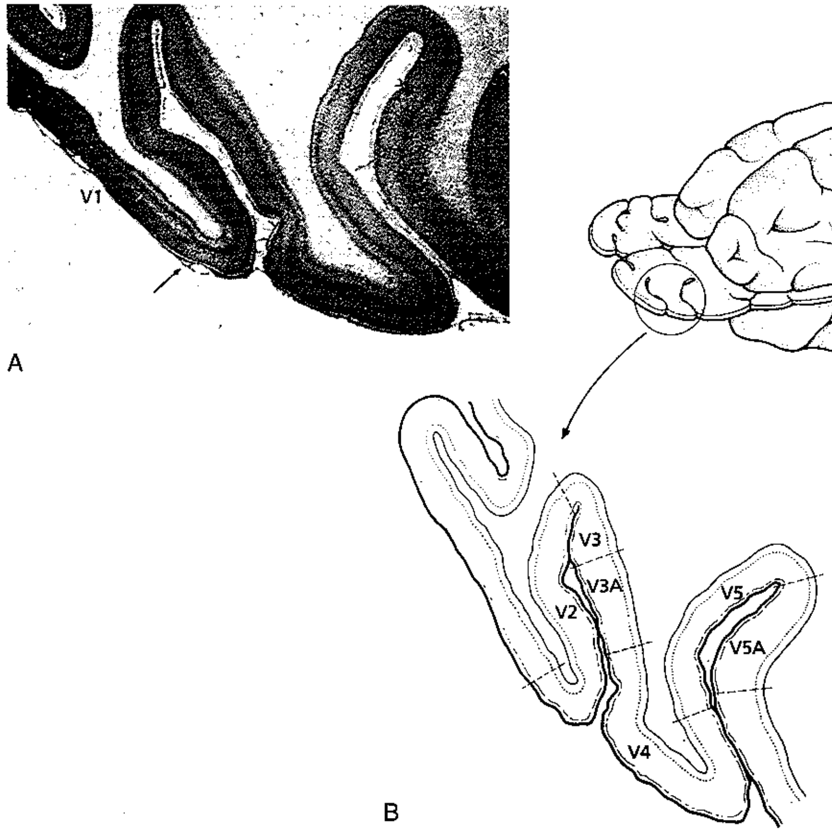


Figure 1.3.18 The location of primary visual cortex in macaque monkeys. The anatomical positions of striate cortex (area V1) and several prestriate areas (V2 through V5) are shown in a (displaced) horizontal slice through the brain. The cellular structure of these areas is shown in the inset (A) with the transition between V1 and prestriate areas indicated by an arrow. (From Zeki, 1993.)

how the visible areas are arranged in an intact (properly "folded") brain. It shows that visual processing takes up almost the entire back half of the cortex. Notice that in the flattened map, the border between areas V1 and V2 has been cut so that the sizes of the different areas do not become too distorted.

The first cortical stage of visual processing—called *striate cortex*, **primary visual cortex**, or area **V1**—is the largest and is located at the very back of the occipital lobe. It receives the majority of ascending projections from the LGN and is responsible for the first few operations of visual processing. We know a great deal about the precise anatomy and physiology of this area, perhaps more than for any other area in the brain. In Chapter 4, when we consider spatial processing in detail, we will explore the architecture of this part of visual cortex. For now, we will just describe its gross anatomical connections to other areas of visual cortex.

It was originally thought that there might be a strict serial ordering of visual processing, each area projecting to the next in a linear sequence. That simple hypothesis was quickly laid to rest as researchers found more and more anatomical connections among visual areas. It is now abundantly clear that a great deal of processing takes place in parallel across different areas, each region projecting fibers to several other areas but by no means to all of them. A schematic diagram of some of the currently known direct connections is given in Figure 1.3.20. Although simplified for clarity, it indicates the interconnections between some of the best known and most studied visual areas. The connections are generally bidirectional; that is, if area X projects to area Y, then Y projects back to X as well. It turns out that the projections in the two different directions are not completely symmetrical, however, in that they originate and terminate in different layers of the cortex.

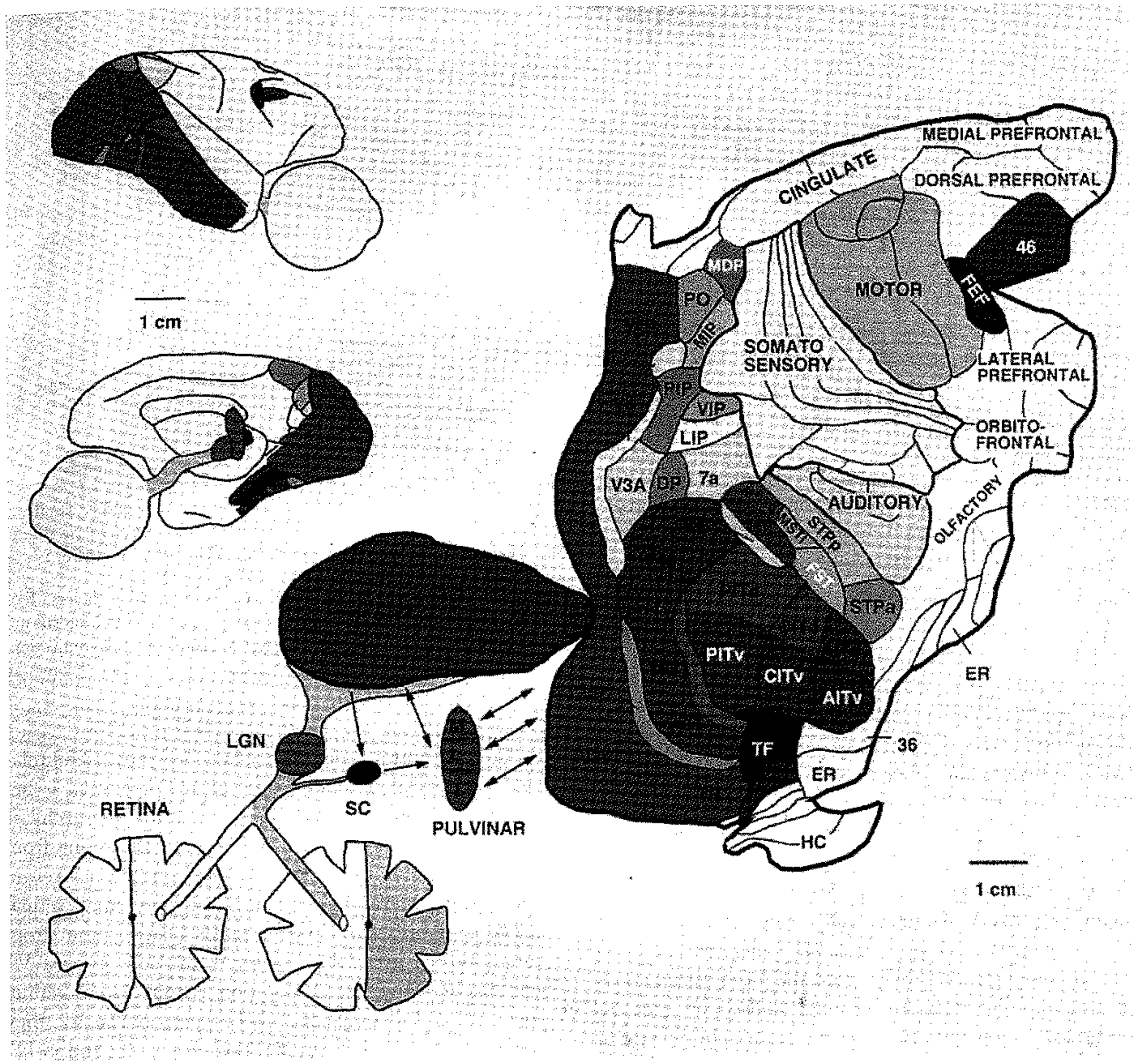


Figure 1.3.19 A flattened map of visual areas in monkey cortex. The principle areas of cortex currently known to be involved in vision are shown in three views of a macaque brain. The top left diagram shows a lateral (side) view of a normal brain (facing right), and the one below it shows a medial (central) view (facing left). The large diagram to the right and below shows a more detailed view of the visual system after the cortex has been flattened. Note that the border between V1 and V2 has been cut to minimize size distortions. (From Van Essen, Anderson, & Felleman, 1992.)

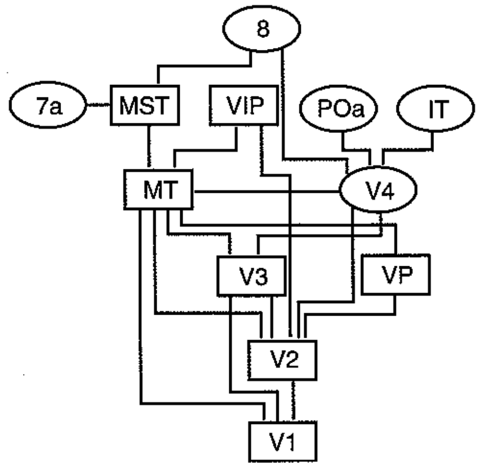
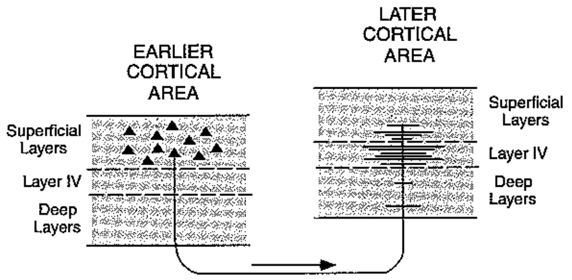
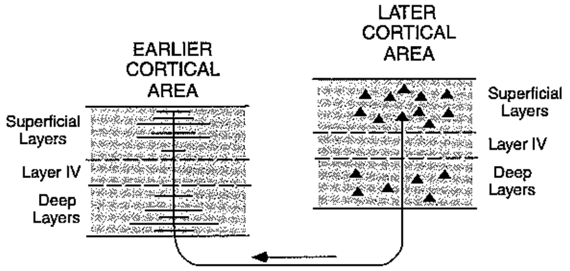


Figure 1.3.20 Interconnections between cortical areas. This diagram summarizes just a few of the known connections between visual areas of monkey cortex. (From Van Essen & Maunsell, 1983.)

The cerebral cortex has a **laminar structure**, in that it is constructed in layers. Visual cortex has six major anatomically defined layers, with several more sublayers being defined by physiological evidence. Of these, the fourth seems to be the input layer for "forward" or "ascending" projections from lower parts of the nervous system. This is certainly true in area V1, where the ascending fibers from the LGN are known to synapse mainly in layer 4. For other cortical areas, it is somewhat less clear which projections are "forward" and which are "backward" or "descending." In one direction, the projections systematically originate in the superficial layers of cortex and terminate primarily in layer 4 (Figure 1.3.21A). These are called *forward projections* by analogy with those from the LGN. Projections in the opposite direction—called *feedback* or *backward connections*—typically originate and terminate outside of layer 4 (Figure 1.3.21B). This distinction between forward and backward connections has been used to define the hierarchy diagrammed in Figure 1.3.20. Each area is located at a level in the hierarchy just above the highest area from which it receives forward projections. As you can see, the many different visual centers have complex interconnections, few of which are yet well understood. In later chapters, we will discuss what is known about the function (or physiology) of some of the better understood of these cortical areas.



A. FORWARD CONNECTIONS



B. BACKWARD CONNECTIONS

Figure 1.3.21 Interconnections between cortical layers. Forward connections originate in upper layers of cortex and terminate in central layers, primarily in layer 4. Feedback connections originate and terminate outside of this central region. (From Van Essen & Maunsell, 1983.)

Do these purely anatomical facts bear any useful relation to cortical function? One clear correlation is between the level of a cell in the anatomical hierarchy of Figure 1.3.20 and size of the region of the retina from which it receives information. Cells from lower levels in the cortical hierarchy receive input from small retinal areas, and cells from higher levels receive input from larger retinal areas. For example, the cells in area V1 can be activated by stimulation within foveal retinal areas 0.1–0.5 degrees of visual angle wide; in V2, they are typically 0.5–1.0°; in V4, they are 1–4°; and in IT, they are often 25° or more (Desimone, Moran, & Spitzer, 1988). The increase in receptive field size from lower to higher visual areas in cortex is thus 100-fold or more.

The Physiological Pathways Hypothesis. A further possible relation between anatomical structure and physiological function has begun to emerge during the last decade or so. The hypothesis is that there are separate neural **pathways** for processing information about different visual properties such as color, shape, depth, and motion. This idea arose from studies suggesting that different areas of cortex were specialized for processing different properties (e.g., Zeki, 1978, 1980). It later became increasingly apparent that this specialization had roots earlier in the visual system. Livingstone and Hubel (1987, 1988) summarized much of this anatomical, physiological, and perceptual evidence and proposed that these four types of information are processed in different neural pathways from the retina onward. They traced these differences from two classes of retinal ganglion cells (one for color and form, the other for depth and motion) to the LGN and from there to different regions of V1, V2, and beyond.

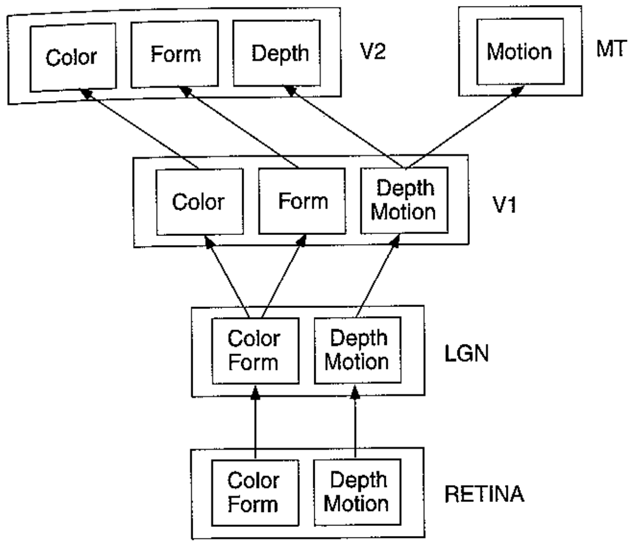


Figure 1.3.22 Schematic diagram of the visual pathways hypothesis. Some theorists believe that color, shape, motion, and depth are processed independently in the visual system. This diagram summarizes a simplified form of the theory.

They report evidence from single cell recordings that color, form, motion, and stereoscopic depth information are processed in distinct subregions of V1 and V2, as indicated schematically in Figure 1.3.22. These areas then project to distinct higher-level areas of cortex: movement and stereoscopic depth information to area **V5** (also called **MT**, Medial Temporal cortex), color to area **V4**, and form through several intermediate centers (including V4) to area **IT** (InferoTemporal cortex). From these areas, the form and color pathways may project to the ventral "what" system for object identification and the depth and motion pathways to the dorsal "where" system for object localization. We will review this theory in Chapter 4 when we describe the physiology of visual cortex in greater detail.

The nature of visual processing in higher level areas of cortex is much less clear than in area V1. For example, some cells in cortical area IT have been found to be strongly activated by the sight of a monkey's hand and other cells to a monkey's face (Gross, Rocha-Miranda, & Bender, 1972). The nature of the spatial processing that occurs between V1 and IT remains mysterious, however. The motion analysis in area MT provides output to area **MST** (Medial Superior Temporal cortex) and several other parietal areas. But again, very little is known about what specific processing occurs in these later centers.

As vague as this story is for monkey cortex—and parts of it are fairly controversial—the understanding of visual areas in human cortex is even less clear. Researchers in human neuropsychology have begun to discover some interesting correlations between locations of strokes and tumors and the visual deficits that they produce, but the evidence is complex and often difficult to evaluate. The use of brain-imaging techniques such as PET and fMRI (see Section 2.2.3) is just beginning to provide useful information about localization of visual function in humans, but their promise is enormous. We will mention some of the clinically significant findings at various points later in this book when we discuss the relevant subject matter. Some of the most interesting neurological phenomena will arise in the last chapter when we discuss the complex but fascinating topic of visual awareness.

Suggestions for Further Reading

The Nature of Vision

There are many excellent books about the phenomena of visual perception. The ones that I find most readable and broadly compatible with the views presented in this book are the following:

Gregory, R. L. (1970). *The intelligent eye.* New York: McGraw-Hill.

Rock, I. (1984). *Perception.* New York: Scientific American Books

Optical Information

The classic discussions of the relations among the world, optical information, and visual perception are found in the three principal works of James J. Gibson:

Gibson, J. J. (1950). *The visual world.* Boston: Houghton Mifflin.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Visual Systems

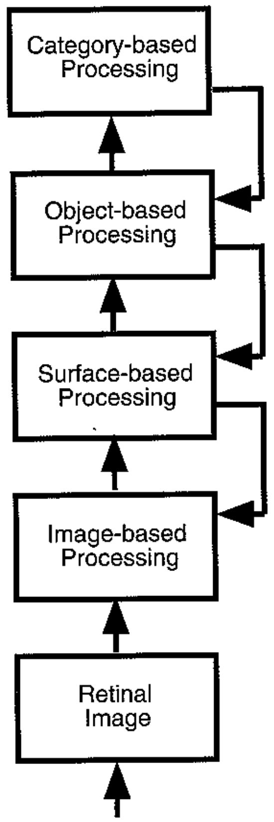
There are many good treatments of the physiology of vision. Among the most informative and readable are the books by Nobel laureate David Hubel and by Semir Zeki.

Hubel, D. (1995). *Eye, brain, and vision.* New York: Scientific American Books.

Zeki, S. (1993). *A vision of the brain.* Oxford: Blackwell.

2

Theoretical Approaches to Vision



- 2.1 Classical Theories of Vision
  - 2.1.1 Structuralism
  - 2.1.2 Gestaltism
    - Holism
    - Psychophysiological Isomorphism
  - 2.1.3 Ecological Optics
    - Analyzing Stimulus Structure
    - Direct Perception
  - 2.1.4 Constructivism
    - Unconscious Inference
    - Heuristic Interpretation
- 2.2 A Brief History of Information Processing
  - 2.2.1 Computer Vision
    - The Invention of Computers
    - Blocks World
    - Computational Approaches to Ecological Optics
    - Connectionism and Neural Networks
  - 2.2.2 Information Processing Psychology
  - 2.2.3 Biological Information Processing
    - Early Developments
    - Single-Cell Recording
    - Autoradiography
    - Brain Imaging Techniques
- 2.3 Information Processing Theory
  - 2.3.1 The Computer Metaphor
  - 2.3.2 Three Levels of Information Processing
    - The Computational Level
    - The Algorithmic Level
    - The Implementational Level
  - 2.3.3 Three Assumptions of Information Processing
    - Informational Description
    - Recursive Decomposition
    - Physical Embodiment
  - 2.3.4 Representation
  - 2.3.5 Processes
    - Implicit versus Explicit Information
    - Processing as Inference
    - Hidden Assumptions
    - Heuristic Processes
    - Hidden Assumptions versus Ecological Validity
    - Top-Down versus Bottom-Up Processes
- 2.4 Four Stages of Visual Perception
  - 2.4.1 The Retinal Image