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?

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.

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?

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 disciplinesbiology, psychology, computer science, neuropsychology, linguistics, and cognitive anthropologyall 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,

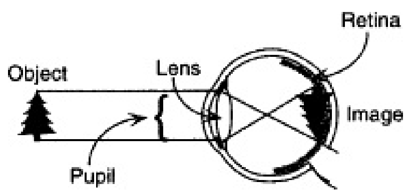
2. the nature of optical information, and

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.



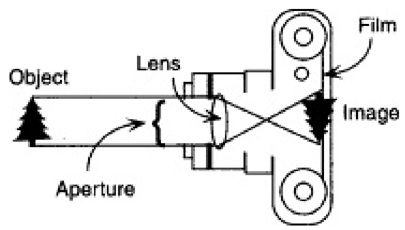


Figure 1.1.2
The eye-camera analogy. The eye is much like a camera in the nature of its optics: Both form an upside-down image by admitting light through a variable-sized opening and focusing it on a two-dimensional surface using a transparent lens.

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 experienceparticularly in Chapter 13 when we discuss visual awareness in detailbut 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 situationssuch as nourishing food, protective shelter, and desirable matesmust be sought out and approached. Dangerous objects and situationssuch as precipitous drops, falling objects, and hungry or angry predatorsmust 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 sensesseeing, hearing, touching, tasting, and smellingparticipate in this endeavor.

There are some creatures for which nonvisual senses play the dominant rolesuch as hearing in the navigation of batsbut 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*.

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?

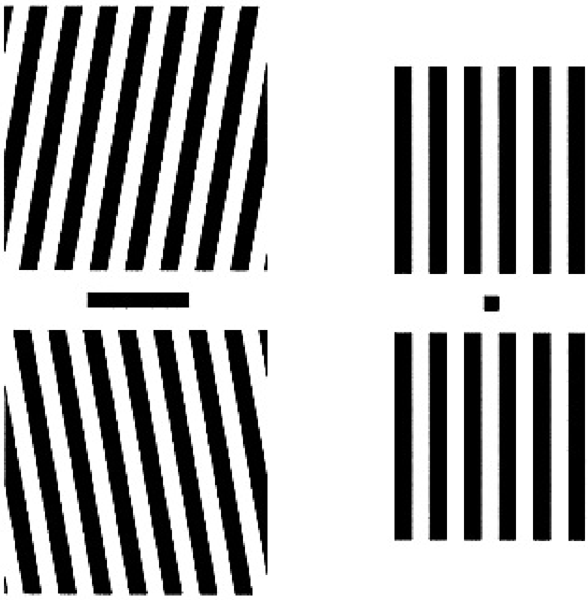


Figure 1.1.3
An orientation aftereffect. Run your eyes along the central bar between the gratings on the left for 30-60 seconds. Then look at the square between the two identical gratings on the right. The upper grating should now appear tilted to the left of vertical and the lower grating tilted to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

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 illusiona nonveridical perceptionbecause 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.

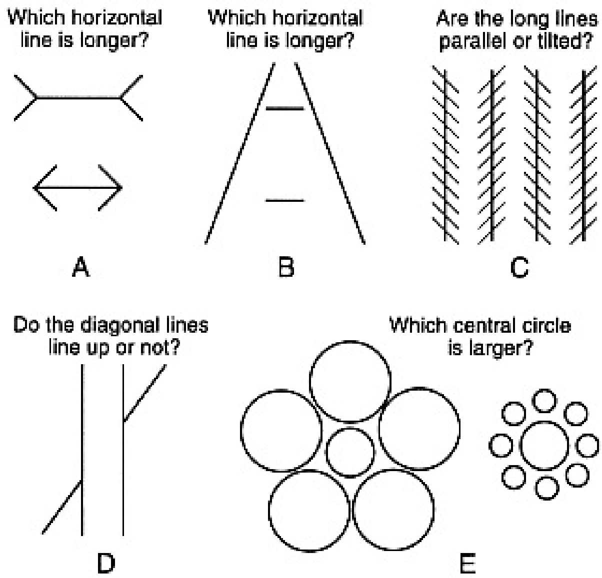


Figure 1.1.4
Visual illusions. Although they do not appear to be so, the two arrow shafts are the same length in A, the horizontal lines are identical in B, the long lines are vertical in C, the diagonal lines are collinear in D, and the middle circles are equal in size in E.

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 accurateor, 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 propertythe relative lengths of the horizontal linesand 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.

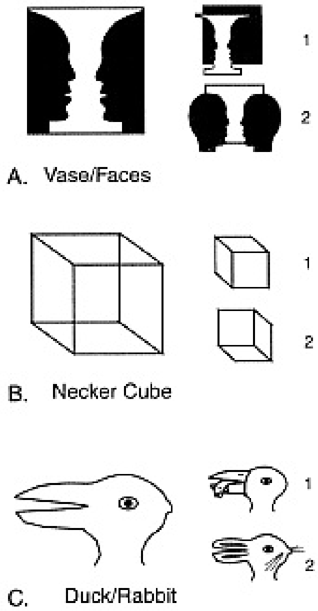


Figure 1.1.5
C. Duck/Rabbit
Ambiguous figures. Figure A can be seen either as a white vase against a black background or as a pair of black faces against a white background. Figure B can be seen as a cube viewed from above or below. Figure C can be seen as a duck (facing left) or a rabbit (facing right).

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 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 yon 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