CHAPTER 6

Christopher Small, Floating, 2014
Imagine that you’re moving quietly through a meadow, trying to get close enough to a bear cub to get a good picture. Suddenly you discover that Mother Bear is not happy about your photo session. She charges, and you run—back across the field, through a thicket of trees. A quick leap across a stream brings you to a slope that leads down to the road where your partner is waiting in a Jeep. You dive into the passenger side and roar off down the track to safety. As the bear lumbers off and your heartbeat returns to normal, your thoughts naturally turn to the acts of visual space perception that you just performed: You picked a path through the three-dimensional world that brought you to safety. You behaved as though you knew where the trees were. You acted as though you understood how far it was across the stream. All in all, you demonstrated a sophisticated grasp of the layout of the physical world around you.

Humans share this sophistication with a large part of the animal kingdom. Faced with the same bear, a rabbit or deer would have shown a similar grasp of the relevant issues in space perception (without the Jeep). The ability to perceive and interact with the structure of space is one of the fundamental goals of the visual system. It is also quite a formidable accomplishment, and in this chapter we’ll explore how we do it.

As a starting place, let’s assume that the external world exists. This is a philosophical position known as realism. It is not the only possibility. The positivists note that all you really have to go on is the evidence of your senses, so the world could be nothing more than an elaborate hallucination. For less philosophical elaborations of this viewpoint, we could consult the writings of Philip K. Dick and other science fiction authors. In this book, we’ll just assume that there is a real world out there to perceive.

The geometry of that real world is Euclidean (named in honor of the ancient Greek geometer Euclid, of the third century BCE). This means that parallel lines remain parallel as they are extended in space, that objects maintain the same size and shape as they move around in that space, that the internal angles of a triangle always add to 180 degrees, and so forth. All that stuff you learned in high school math classes is true in the real world.

Euclidean geometry is not the only geometry. Although the real world is Euclidean, the geometry of retinal images of that world is decidedly non-Euclidean. The geometry becomes non-Euclidean when the three-dimensional world is projected onto the curved, two-dimensional surface of the retina. Parallel lines in the world do not necessarily remain parallel in the retinal image, as Figure 6.1 illustrates. The angles of triangles don’t always add up to 180 degrees. The retinal area occupied by an object gets smaller as the object moves farther away from the eyeball. What all this means is that if we want to appreciate the Euclidean world, we have to reconstruct it from non-Euclidean input.

realism  A philosophical position arguing that there is a real world to sense.

positivism  A philosophical position arguing that all we really have to go on is the evidence of the senses, so the world might be nothing more than an elaborate hallucination.

Euclidean  Referring to the geometry of the world, so named in honor of Euclid, the ancient Greek geometer of the third century BCE. In Euclidean geometry, parallel lines remain parallel as they are extended in space, objects maintain the same size and shape as they move around in space, the internal angles of a triangle always add to 180 degrees, and so forth.
More precisely, generally our visual experience is a reconstruction of the world based on two non-Euclidean inputs: the two distinct retinal images. Close your left eye, stretch your left arm out in front of you, and hold up your left index finger. Then hold up your right index finger about 6 inches in front of your face, so that it appears to be positioned just to the left of the left index finger, as illustrated in Figure 6.2. Now, quickly open your left eye and close your right eye. If you positioned your fingers properly, your right finger will jump to the other side of your left finger. Although this demonstration is designed to exaggerate the different views of your two eyes, the point is a general one: the two retinal images always differ. They differ because the retinas are in slightly different places in your head. Just as you and the person standing next to you see somewhat different views of the world, so do your two eyes. Much of this chapter will be devoted to explaining how the visual system goes to quite elaborate lengths both to exploit and to reconcile these differences.

**FIGURE 6.1** The Euclidean geometry of the three-dimensional world turns into something quite different on the curved, two-dimensional retina. In the Euclidean world, the angles of a triangle add up to 180 degrees. In the non-Euclidean world of the retinal image, this need not be so.

**FIGURE 6.2** The two retinal images of a three-dimensional world are not the same. See the text for details.
Why have two eyes at all? Perhaps most fundamentally, having two eyes confers the same evolutionary advantage as having two lungs or two kidneys: you can lose one eye and still be able to see. A second advantage to doubling the number of eyes is that they enable you to see more of the world. This is especially true for animals like rabbits who have lateral eyes on the sides of their heads. A rabbit can actually see for 360 degrees around its head (Figure 6.3a). This explains why it is so hard to sneak up on a rabbit. Moreover, the rabbit can also see straight up above its head (Figure 6.3b) (Hughes, 1977).

Humans, with frontal eyes, still see more of the world with two eyes than with one. Our visual field is limited to about 190 degrees from left to right, 110 degrees of which is covered by both eyes (Figure 6.3c). The field is more restricted vertically: about 140 degrees, 60 degrees up to a limit defined by your eyebrows and 80 degrees down to your cheeks, as is shown in Web Essay 2.3: The Man Who Couldn’t Read. The exact size of your visual field will be limited by the specific anatomy of your cheeks and eyebrows. Overlapping, frontal, binocular visual fields give predator animals such as humans a better chance to spot small, fast-moving objects in front of them that might provide dinner. Prey animals, like rabbits, are often those with very wide visual fields allowing them to monitor the whole scene for predators.

With frontal eyes and overlapping visual fields, you get the advantage of two detectors looking at the same thing. For example, if two independent people each had a 50% chance of missing a target, the chance that both of them would miss it would be 50% × 50% = 25%. So the chance that at least one of the two would find that target is 100% − 25%, or 75%. This is known as probability summation. Something similar happens if the two eyes both look for the same hard-to-see target. In vision, this would be called binocular summation (Blake, Sloane, and Fox, 1981). Binocular summation may have provided the evolutionary pressure that first moved eyes toward the front of some birds’ and mammals’ faces. Under most circumstances, we do not get complete probability summation. The increase from a 50% chance to a 75% chance assumes two completely independent observers, but our two eyes are still embedded in one person. Nevertheless, we will do better at many tasks with two eyes than with just one (R. K. Jones and Lee, 1981).

Once the eyes moved to the front, though, evolution found an additional use for overlapping visual fields. Try this: Take the top off a pen and hold the top in one hand, the pen in the other, both about a foot in front of your face,
with your elbows bent. Now, close one eye and try to quickly put the cap on the pen. Repeat the same task with both eyes open. For most (but not all) people, this task is easier with two eyes than with one. This is a quick demonstration of the usefulness of **binocular disparity**—the differences between the two retinal images of the same scene. Disparity is the basis for a vivid perception of the three-dimensionality of the world that is not available with purely **monocular** (one-eyed) vision. The technical term for this binocular perception of depth is **binocular stereopsis**. The geometric and physiological bases for binocular stereopsis are the topic of a large portion of this chapter (see also I. P. Howard and Rogers, 2001).

When you decide you need a break from reading this chapter, take that break with one eye closed. You should be able to notice the loss of binocular stereopsis (and of part of your visual field), but a period of one-eyed visual experience will also make it clear that binocular stereopsis is not a necessary condition for depth perception or space perception. Rabbits do very well with very little binocular vision, and painters and movie directors manage to convey realistic impressions of depth on flat canvases and movie screens. On the other hand, binocular stereopsis does add a richness to perception of the three-dimensional world, as vividly described by Oliver Sacks (2006) in his article about “Stereo Sue,” a neuroscientist who regained stereopsis at the age of 48. We will talk about her later.

In this chapter, first we describe the set of **monocular depth cues** to three-dimensional space. After that, we turn to the more complicated topic of binocular stereopsis, a **binocular depth cue**. Finally, in the last section of the chapter, we consider how the various cues are combined to produce a unified perception of space.

**Monocular Cues to Three-Dimensional Space**

M. C. Escher (1898–1972) titled the drawing in **Figure 6.4 Relativity.** Escher was a master of the rules that govern our perception of space. Each bit of stairway, each landing, every person—all are drawn using cues that enable us to infer three dimensions from two. But when we try to follow those stairs, we find that Escher’s drawing cleverly fails to add up to a coherent representation of a place that could exist. Even when no one is trying to fool us, it is geometrically impossible (not to mention computationally infeasible) for the visual system to create a perfectly faithful reconstruction of Euclidean space, given the non-Euclidean input we receive through our eyes. The best we can do is to use depth cues to infer aspects of the three-dimensional world from our two-dimensional retinal images. On the basis of the retinal images and an implicit understanding of physics and geometry, we collect cues that provide hints about the likely structure of the space in front of us and the disposition of objects in that space.

**Figure 6.4** M. C. Escher, *Relativity*, 1953. Escher deploys monocular cues to depth in a way that is essentially correct at each location but that adds up to an impossible world.
Unless we’re stuck in an extremely impoverished perceptual environment (say, the Sahara during a sandstorm), every view of the world provides multiple depth cues. Usually the cues reinforce each other, combining to produce a convincing and reliable representation of the three-dimensional world. Occasionally, however, the cues are contradictory. Escher could fool us by deliberately manipulating depth cues and other routine visual inferences. He arranged sensible local cues into a globally implausible story. What cues does the visual system use to infer depth relations, and how do we use those cues to create a representation of the three-dimensional world? (See Web Activity 6.1: Monocular Depth Cues.)

**Occlusion**

Some of the cues to the layout of the three-dimensional world were introduced earlier in this book, because hints to the layout of space can also be hints about the structure of objects in that space. **Occlusion** is an example (see the section titled “Finding Edges” in Chapter 4). In Chapter 4, occlusion was a cue to the presence of an otherwise invisible edge. As a depth cue, occlusion gives information about the relative position of objects. Thus, in Figure 6.5 we are happy to infer a circle in front of a square in front of a triangle. Occlusion is present in almost every visual scene (we challenge you to find a situation in normal life where nothing blocks your view of anything else), and many researchers argue that it is the most reliable of all the depth cues. It is wrong only in the case of “accidental viewpoints” (remember those, from Chapter 4?). That is, the retinal image shown in Figure 6.5 could be produced by a circle and two oddly shaped puzzle pieces, as shown in Figure 6.6a. That scenario would require careful placement of the objects and the viewer. It is much more likely that Figure 6.5 would arise from a more generic view of a circle occluding a square occluding a triangle (Figure 6.6b).

We do not know from the occlusion cue alone whether the red square in Figure 6.5 is in front of a small green triangle, a larger but more distant triangle (maybe a green tree), or an even larger but even more distant green mountain. Occlusion is a **nonmetrical depth cue**; it just gives us the relative orderings of occluders and occludees. A **metrical depth cue** is one that does provide information about distance in the third dimension.

**Size and Position Cues**

The image on the retina formed by an object out in the world gets smaller as the object gets farther away. (See Web Activity 3.1: Visual Angle for a review.) Moreover, your visual system knew this fact of **projective geometry** implicitly before you ever picked up this book and learned it explicitly. Projective occlusion  A cue to relative depth order in which, for example, one object obstructs the view of part of another object.

nonmetrical depth cue  A depth cue that provides information about the depth order (relative depth) but not depth magnitude (e.g., his nose is in front of his face).

metrical depth cue  A depth cue that provides quantitative information about distance in the third dimension.

**projective geometry**  For purposes of studying perception of the three-dimensional world, the geometry that describes the transformations that occur when the three-dimensional world is projected onto a two-dimensional surface. For example, parallel lines do not converge in the real world, but they do in the two-dimensional projection of that world.

**FIGURE 6.5** Occlusion makes it easy to infer relative position in depth.

**FIGURE 6.6** Figure 6.5 could be an “accidental” view of the pieces shown here in (a). It is much more likely, however, that it is a generic view of circle, square, and triangle, as shown in (b).
Geometry describes how the world is projected onto a surface. For example, a shadow is a projection of an object onto a surface. An implicit understanding of the rules of projective geometry can be said to undergird many of the depth cues described here. In this case, the visual system knows that, all else being equal, smaller things are farther away. Hence, the Plasticine balls in Figure 6.7 may appear to lie in different depth planes. We can call this depth cue relative size.

The impression of three-dimensionality in Figure 6.8 is more powerful than that in Figure 6.7 because we’ve added another cue. The critical difference is in the organization of the objects in the two figures. In Figure 6.8, the rabbits form an orderly texture gradient, with larger objects in one area and smaller objects in another. Because smaller is interpreted as farther away, this arrangement creates the perception of a ground plane receding into the distance.

In Figure 6.9, the rabbits are again arrayed in an orderly texture, but here we get less of a sense of depth. The difference between Figures 6.8 and 6.9 is that the former includes another depth cue that is not present in the latter: relative height. Imagine that you’re actually standing in a field of rabbits. Consider the rabbit at your feet (Figure 6.10). It will project its image in your lower visual field. The smaller image of a more distant rabbit will be projected higher in your visual field. Here, then, is another geometric regularity produced by projective geometry that the visual system can exploit: for objects on the ground plane, objects that are more distant will be higher in the visual field.

**FIGURE 6.7** This is a photograph of a collection of Plasticine balls that are resting on the same surface at the same distance from the camera. Nevertheless, the small ones appear to be farther away. Some portion of the visual system assumes that all of these items are the same. If one ball projects a smaller image on the retina and if we assume that the balls really are the same size, then the smaller one must be farther away. This is the cue of relative size.

**FIGURE 6.8** This rabbit texture gradient shows that the size cue is more effective when size changes systematically.

---

**relative size** A comparison of size between items without knowing the absolute size of either one.

**texture gradient** A depth cue based on the geometric fact that items of the same size form smaller images when they are farther away. An array of items that change in size smoothly across the image will appear to form a surface tilted in depth.

**relative height** As a depth cue, the observation that objects at different distances from the viewer on the ground plane will form images at different heights in the retinal image. Objects farther away will be seen as higher in the image.
Texture fields that provide an impression of three-dimensionality are really combinations of relative size and relative height cues. Remember the metaphor of “perceptual committees” in Chapter 4. Different modules in the visual system perform different tasks. The brain then combines the outputs to come up with a committee-like decision about the state of the world. In the

**FIGURE 6.9** Organized differently, this illustration of the same rabbits as those shown in Figure 6.8 does not produce the same sense of depth. A size cue is most effective when it is consistent with objects arranged on the ground, not on a wall.

**FIGURE 6.10** Relative height as a cue to depth. If we’re looking down at the ground plane, the image of a closer rabbit will lie above the image of a farther rabbit in the retinal image. Because the image is inverted, the closer rabbit lies lower in the visual field than the farther rabbit.
case of a texture field, multiple cues interact to produce a final perception. Figure 6.11 shows how this interaction can give rise to a size illusion. The rabbit at the upper left of the figure is actually the same physical size on the page as the rabbit at the lower right, but the one at the bottom looks smaller to most of us than the one at the top. Why? We infer, on the basis of relative height, that the rabbit at the bottom must be closer. If it is closer and it forms an image of the same size as the little rabbit at the top, it follows that the little rabbit at the bottom must be really little.

If we know what size something ought to be, that knowledge can be a depth cue in its own right. We infer that the woman in Figure 6.12a is holding her hand out at the end of an outstretched arm. Why do we make this guess? One alternative is that she’s holding her hand near her shoulder, as in Figure 6.12b. But if that were the case in Figure 6.12a, the hand would need to be a

![Figure 6.11](image1)

*Figure 6.11* The rabbit image at the top far left is the same size as the one at the bottom far right. If they don’t seem the same size, then you have been fooled by the depth cues.

![Figure 6.12](image2)

*Figure 6.12* The cue of familiar size. The hand in (a) looks closer than the one in (b) because we know how big hands should be relative to heads.
**SpaCe perCeption and binoCular viSion**

very big hand. Here, our knowledge of the normal relationship of hand size to head size makes all the difference. This is the depth cue of **familiar size**.

Recall that occlusion is a nonmetrical cue, providing only depth *order*. The relative size and relative height cues, especially taken together, provide some metrical information. This is illustrated in **Figure 6.13**, where the three balls appear to lie at measurable distances from each other. The blue ball seems closer to the red ball than to the green ball in depth, for example. Relative size and height do not tell us the *exact* distance to an object or between objects. These are **relative metrical depth cues**. Familiar size, however, could be an **absolute metrical depth cue**. If your visual system knew the actual size of an object and the visual angle of the object’s projection on the retina, it could (at least in theory) calculate the exact distance from object to eye. In practice, however, even if you know that your friend is 5 feet 10 inches tall, the visual system does not seem to know that fact with a precision that would let you know he’s standing exactly 12 feet away.

**Aerial Perspective**

In addition to its implicit knowledge of geometry and its learned knowledge of familiar size, the visual system “knows” about properties of the atmosphere. The triangles in **Figure 6.14a** give only a faint sense of depth, if any. They are just an array of identical shapes. Adding some grayscale information, however, as in **Figure 6.14b**, provides the impression of something like a mountain range receding in the distance. More specifically, the fainter mountains (triangles) may appear to be farther away than the darker ones. The depth cue at work here relies on an implicit understanding that light is scattered by the atmosphere, and that more light is scattered when we look through more atmosphere. Thus, objects farther away are subject to more scatter and appear

---

**Figure 6.13** The metrical cues of relative size and height can give the visual system more information than a nonmetrical cue like occlusion can. Not only does the red sphere in this image appear to be closest and the green sphere farthest away, but also the blue sphere is seen to be closer to the red sphere than to the green sphere.

**Figure 6.14** The triangles seem to recede into depth more in (b) than in (a).
haze or aerial perspective  A depth cue based on the implicit understanding that light is scattered by the atmosphere. More light is scattered when we look through more atmosphere. Thus, more distant objects are subject to more scatter and appear fainter, bluer, and less distinct.

linear perspective  A depth cue based on the fact that lines that are parallel in the three-dimensional world will appear to converge in a two-dimensional image.

fainter and less distinct. This cue is known as haze, or aerial perspective. Figure 6.15 shows a real-world example. Short wavelengths (blue) are scattered more than medium and long wavelengths. This is why the sky looks blue and why objects farther away look not only hazy, but bluish. The real image gives a stronger sense of depth. By now, it should be clear that this effect results from the contribution of other depth cues beyond the blue haze—for instance, occlusion and the known size of boats.

Linear Perspective

It would not be difficult to imagine the six lines shown in Figure 6.16 to be a sketch of the view out the windshield of a car moving down a road in a flat landscape. The depth cue in this case—linear perspective—is based on the rules that determine how lines in three-dimensional space are projected onto a two-dimensional image. The core piece of projective geometry in this case is that lines that are parallel in the three-dimensional world will appear to converge in the two-dimensional image, except when the parallel lines lie in a plane that is parallel to the plane of the two-dimensional image. Artists of the Italian Renaissance are said to have “discovered” linear perspective. That is not quite right. Your dog or cat knows about linear perspective. What Renaissance artists discovered was how to make the rules explicit, write them down, and turn linear perspective into a method for generating realistic depth in otherwise flat paintings. Filippo Di Ser Brunellesco (1377–1446) is typically given credit for bringing linear perspective into European art. Leon Battista Alberti (1404–1474) wrote the first book on the topic in 1435 (Alberti, 1970). Figure 6.17 shows an example of what could be done. This Architectural View is by Francesco di Giorgio Martini, a painter from the Italian city of Siena.

**FIGURE 6.16** Linear perspective. If we titled this image “Driving across Kansas,” you would understand, because the converging lines give the impression of parallel lines receding toward the horizon.
Painted around 1477, it is not the greatest work of Renaissance art, but it shows the basic rules of perspective at work. Parallel lines in the image plane, like the two front center pillars, would be parallel in the world. Look at all those other lines converging toward a vanishing point. Clearly, these are intended to show the parallel lines on the ground or on the sides of buildings running in depth. Of course, you can see occlusion, size, and texture cues at work, too.

*Pictorial Depth Cues and Pictures*

As befits a group of rules that allowed Martini to paint his picture with such good apparent depth, the depth cues discussed so far are known as *pictorial depth cues*. These are the cues produced by projection of the three-dimensional world onto the two-dimensional surface of the retina. A realistic picture or photograph is the result of projecting the three-dimensional world onto the two-dimensional surface of film or canvas. When that image is viewed from the correct position, the retinal image (in one eye, at least) formed by the two-dimensional picture will be the same as the retinal image that would have been formed by the three-dimensional world, and hence we see depth in the picture. In theory, this means that a picture should look correct from only one, precise viewing position. In fact, pictures look reasonable over quite a range of views. Were this not so, there would be only one good seat in the movie theater.

To correctly interpret the shapes of three-dimensional objects from two-dimensional pictures, people take the orientation of the flat surface of the image into account. This allows them to understand that the picture is, in fact, a picture and not the real thing; at the same time, they can calculate an accurate impression of the thing that is portrayed (Vishwanath, Girshik, and Banks, 2005). To illustrate this point, Marty Banks snapped the photo of one of the authors standing alongside a picture of himself that is shown in *Figure 6.18a*; here the “picture in the picture” appears reasonable. In *Figure 6.18b*, the same picture is shown stripped of its context. Now it appears quite distorted.

---

**FIGURE 6.17** *Architectural View* by Francesco di Giorgio Martini (1477), a very clear example of linear perspective. Parallel lines in the image plane, such as the pillars at the front center, remain parallel in the image. Parallel lines that would recede in depth in the three-dimensional world converge to a vanishing point in the two-dimensional image. (Staatliche Museen, Berlin.)
In Figure 6.18a, our visual system can compensate for the perceptual distortion because there is enough context to enable the viewer to attribute the distortion to the slant of the picture surface.

The technique known as **anamorphosis, or anamorphic projection**, illustrates that the ability to cope with distortion is limited. In anamorphic projection, the rules of linear perspective are pushed to an extreme. Now the projection of three dimensions into two dimensions creates a two-dimensional image that is recognizable only from an unusual vantage point (or sometimes with a curved mirror). The results are known as anamorphic art. As an example, there is an odd diagonal smear in the lower center of Hans Holbein’s sixteenth-century painting *The Ambassadors* (Figure 6.19a). If you could put your eye in exactly the right position to view the image, the smear would prove to be the skull shown in Figure 6.19b. Despite its successful recovery of the shapes in Figure 6.18a, the visual system cannot use knowledge about surface orientation to compensate for the distortion in Figure 6.19. In our own day, the sidewalk chalk
FIGURE 6.20  Modern-day anamorphic art. (a) In this photograph, artist Leon Keer creates what appears to be a large three-dimensional version of the classic video game, Pacman. But, as shown in (b), this is just a clever bit of anamorphic art. It is just a flat image that looks three-dimensional when viewed from the correct position.

Motion Cues

Beyond the pictorial depth cues, a number of additional sources of information are available to our visual system when we view real-world scene—cues that cannot be reproduced in a static two-dimensional picture. The first nonpictorial depth cue we will discuss is motion parallax. To appreciate its power (and to understand why photographs of the forest often don’t come out well), the best thing to do is go outside and lie under a tree. Gaze up into the branches and leaves with one eye covered and your head stationary. You will notice that leaves and branches form a relatively flat texture. You can see all the details, but you may have trouble deciding whether one little branch lies in front of or behind another. If you open the other eye, binocular stereopsis (introduced earlier and discussed in detail later in the chapter) will allow the branches and leaves to fill out a three-dimensional volume that was lacking before. Close the eye, and the volume collapses again. Now, move your head from side to side, and motion parallax will restore the sense of depth.

motion parallax  An important depth cue that is based on head movement. The geometric information obtained from an eye in two different positions at two different times is similar to the information from two eyes in different positions in the head at the same time.
How does motion provide a cue for depth? Suppose you’re sitting on a train, looking out the window at the countryside. At one instant you see the scene sketched in Figure 6.21a. A moment later, the scene has changed to the one in Figure 6.21b. Notice that as your train moved from left to right in the figure, all the objects shifted from right to left. But note that some things shifted more than others. The flower (F in the figure) moved almost all the way across your retinal image, the cow (C) moved a much shorter distance, and the tree (T) hardly changed positions at all. The term parallax refers to the geometric relationship revealed here: when you change your viewpoint while rolling down the tracks, objects closer to you shift position more than objects farther away. Of course, you don’t need to be on a train to experience motion parallax; just moving your head will do. The geometric information obtained from an eye in two different positions at two different times (motion parallax) is similar to the information from two eyes in different positions in the head at the same time (binocular stereopsis) (Durgin et al., 1995; Rogers and Collett, 1989).

Motion parallax provides relative metrical information about how far away objects are; and as the experiment with the tree branches proves, it can provide a compelling sense of depth in some situations in which other cues are not very effective. The downside of motion parallax is that it works only if the head moves (just moving the eyes back and forth won’t do, as you can easily prove to yourself). Now you know why a cat might bob its head back and forth as it plans a spectacular leap from the sofa to the table. Other motion signals produce information about depth. For example, objects get bigger and smaller as they get closer and farther away, so an object that is simply getting bigger on a screen can appear to be looming toward you. If everything is looming at once in a large field, this optic flow may make you feel like you

**FIGURE 6.21** Motion parallax. As you look out the window of a moving train, objects closer to you (like the flower in this illustration) shift position more quickly than do objects farther away (the tree) from one moment (a) to the next (b). This regularity can be exploited as a depth cue.
are moving toward the screen rather than like the objects on the screen are moving toward you. These topics are discussed in more detail in Chapter 8.

**Accommodation and Convergence**

Like a camera, the eyes need to be focused to see objects at different distances clearly. As we learned in Chapter 2, the human eye focuses via a process called **accommodation**, in which the lens gets fatter as we direct our gaze toward nearer objects. We also need to point our eyes differently to focus on objects at different distances. As the schematic eyeballs in Figure 6.22 move from the red dot to the blue dot, they rotate inward—a process called **convergence** (Figure 6.22a); refocusing on the red dot would require rotation outward, which is known as **divergence** (Figure 6.22b).

If we could monitor our state of accommodation and/or the extent to which our eyes were converged, we could use this information as a cue to the depth of the object we were trying to bring into focus: the more we have to converge and the more the lens has to bulge in order to focus on the object, the closer it is. In fact, we do use this information. When we focus on objects more than about 2–3 meters away, the lens is as thin as it can get and the eyes are diverged about as much as possible, so neither cue provides much useful information. But careful studies have shown that the visual system takes advantage of both cues for objects closer than this limit. Convergence is used more than accommodation (Fisher and Ciuffreda, 1988; Owens, 1987). Moreover, in principle these cues can tell us the exact distance to an object. Humans are not particularly precise about measuring the exact angles shown in Figure 6.22. Chameleons, on the other hand, do use the absolute metrical depth information from convergence to catch prey insects with their sticky tongues. Harkness (1977) showed this by fitting a chameleon with glasses that distorted the angle of convergence. The result was that the poor chameleon flicked out its tongue to the wrong distance and missed its intended dinner.

**accommodation** The process by which the eye changes its focus (in which the lens gets fatter as gaze is directed toward nearer objects).

**convergence** The ability of the two eyes to turn inward, often used in order to place the two images of a feature in the world on corresponding locations in the two retinal images (typically on the fovea of each eye). Convergence reduces the disparity of that feature to zero (or nearly zero).

**divergence** The ability of the two eyes to turn outward, often used in order to place the two images of a feature in the world on corresponding locations in the two retinal images (typically on the fovea of each eye). Divergence reduces the disparity of that feature to zero (or nearly zero).
Binocular Vision and Stereopsis

As defined earlier, the term *binocular disparity* refers to differences between the images falling on our two retinas, and *stereopsis* refers to the impression of three-dimensionality—of objects “popping out in depth”—that most humans get when they view real-world objects with both eyes. Like the accounts of other depth cues, the story of the route from binocular disparity to stereopsis is a story of the visual system exploiting the regularities of projective geometry to recover the three-dimensional world from its projections—this time, onto a pair of two-dimensional surfaces. We will illustrate the translation from disparity to stereopsis using the situation shown in Figure 6.23a, in which the viewer (call him Bob) is facing a scene that includes four colored crayons at different depths. Suppose that Bob is focusing his gaze on the red crayon, as shown in Figure 6.23b. The two lines in this figure trace the paths of the light rays that reflect off the red crayon and onto Bob’s retinas. (Similar experiences with crayon scenes are also demonstrated in Web Activity 6.2: Binocular Disparity.)

Because the visual system is designed so that the object of our gaze always falls on the fovea, the rays from the red crayon fall on the fovea in each of Bob’s eyes. Figure 6.24 shows the retinal image for the crayons in each eye. The red crayon is in the center of both images. We’ve added a dashed vertical

---

**FIGURE 6.23** This simple visual scene illustrates how geometric regularities are exploited by the visual system to achieve stereopsis from binocular disparity. (a) The viewer, Bob, is assumed to be fixing his gaze on the red crayon. (b) This top view traces the rays of light bouncing off the red crayon onto Bob’s retinas.

**FIGURE 6.24** The overlapping portions of the images falling on Bob’s left and right retinas. Because the retinal image is reversed, the blue and purple crayons on the right side of the scene in Figure 6.23 project to the left side of each retina, whereas the brown crayon on the left side of the scene projects to the right side of each retina. The size differences between the retinal images of the crayons in the two retinas are exaggerated in this figure compared with the differences we would observe if we saw this scene in the real world.
Now consider the retinal images of the blue crayon. As you saw in Chapter 2, the optics of the eye reverse left-right and up-down (see Figure 2.2). Thus, the blue crayon on the right side of the scene in Figure 6.23 falls on the left side in each of the two retinal images shown in Figure 6.24. In our imaginary scene, the blue crayon is placed so that the monocular retinal images of that crayon are formed at the same distance from the fovea in both eyes. We say that this crayon’s images fall on corresponding retinal points. The same can be said of the images of the red crayon, which fall on the two foveas.

In fact, any object lying on the Vieth-Müller circle—the imaginary circle that runs through the two eyeballs and the object on which Bob is fixated—should project to corresponding retinal points. This imaginary circle is drawn in gray in Figure 6.25. Objects that fall on corresponding retinal points are said to have zero binocular disparity. If the two eyes are looking at one spot (such as the red crayon), then there will be a surface of zero disparity running through that spot. That surface is known as the horopter. Any object placed on that imaginary surface in the world will form images on corresponding retinal points. As it happens, the horopter and the Vieth-Müller circle are not quite the same. If you are extremely fond of rather complicated geometry, you may want to pursue this topic in one of the following sources: I. P. Howard and Rogers, 1995 or 2001; or Tyler, 1991. Otherwise, the important point is that there is a surface of zero disparity whose position in the world depends on the current state of convergence of the eyes.

Objects that lie on the horopter are seen as single objects when viewed with both eyes. Objects significantly closer to or farther away from the surface of zero disparity form images on decidedly noncorresponding points in the two eyes, and we see two of each of those objects. This double vision is known as diplopia. Objects that are close to the horopter but not quite on it can still be seen as single objects. This region of space in front of and behind the horopter, within which binocular single vision is possible, is known as Panum’s fusional area (Panum, 1940). You can check this quite simply, by holding a red crayon (or pen) directly in front of you with your left hand, at a distance of about 20 cm, and keeping both eyes on it. Now hold a blue crayon (or pen) about 5 cm to either side of the red one with your right hand, and slowly move it nearer to your eyes and then farther away, while maintain careful fixation on the red corresponding retinal points. Two, monocular images of an object in the world are said to fall on corresponding points if those points are the same distance from the fovea in both eyes. The two foveas are also corresponding points.

Vieth-Müller circle The location of objects whose images fall on geometrically corresponding points in the two retinas. If life were simple, this circle would be the horopter, but life is not simple.

Horopter The location of objects whose images lie on corresponding points. The surface of zero disparity.

Diplopia Double vision. If visible in both eyes, stimuli falling outside of Panum’s fusional area will appear diplopic.

Panum’s fusional area The region of space, in front of and behind the horopter, within which binocular single vision is possible.
You should initially see the blue crayon/pen as single, when it is about the same distance from you as the red one, because it is within Panum’s fusional area. However, when it falls outside Panum’s area, it will appear double. Panum’s area provides a little room for small errors in eye alignment, while still maintaining single vision.

Armed with this terminology, let’s return to Bob and his crayons. Consider the retinal images of the brown crayon, lying just off the horopter. As Figure 6.24 and the view from above in Figure 6.26 show, rays of light bouncing off this crayon do not fall on corresponding retinal points: the crayon’s image is farther away from the fovea on the left retina than on the right retina. Relative to the horopter, this crayon forms retinal images with a nonzero binocular disparity. The purple crayon is even farther off the horopter (Figure 6.26b); it forms retinal images that are even more disparate (Figure 6.27).

The geometric regularity that the visual system uses to extract metrical depth information from binocular disparity should now be growing clear. The larger the disparity, the greater the distance in depth of the object from the horopter.
The direction in depth is given by the sign (that is, “crossed” or “uncrossed”) of the disparity, as illustrated in Figure 6.28. Suppose that Bob is looking at a red crayon with his eyes converged so that the red crayon falls on the fovea in each eye. A closer, blue crayon will form images on noncorresponding, disparate points. On the left retina, blue will lie to the left of red. Because the image is reversed, this means that, viewed from the left eye, blue is to the right of red. Viewed from the right eye, blue is to the left. Right in left, and left in right. This is known as crossed disparity (Figure 6.28a), and crossed disparity always means “in front of the horopter.” In Figure 6.28b, Bob is looking at the blue crayon, and the red is seen to the left with the left eye and to the right with the right. That’s uncrossed disparity, and uncrossed disparity always means “behind the horopter.”

Stereoscopes and Stereograms

Interestingly, although scientists had studied the geometry of binocular vision for millennia (the geometer Euclid was at it in the third century BCE), not until the nineteenth century was binocular disparity properly recognized as a depth cue. In the 1830s, Sir Charles Wheatstone invented a device called the stereoscope (Figure 6.29) that presented one image to one eye and a different image to the other eye, allowing the viewer to perceive depth. Wheatstone’s stereoscope was a device for simultaneously presenting one image to one eye and another image to the other eye. Stereoscopes can be used to present dichoptic stimuli for stereopsis and binocular rivalry.

**FIGURE 6.28** Crossed and uncrossed disparity. (a) Here Bob is foveating the red crayon. In the Bob’s-eye views, the closer, blue object is seen to the right in the left eye and to the left in the right eye. This situation is crossed disparity. (b) Here Bob has shifted his gaze and his horopter to the blue crayon. In the Bob’s-eye views, the farther, red object is seen to the left in the left eye and to the right in the right eye. This situation is uncrossed disparity.

**FIGURE 6.29** Wheatstone’s stereoscope. The viewer would bring her nose up the vertical rod at the center of the apparatus so that each eye was looking at the image reflected in one of the two mirrors. (From I. P. Howard and Rogers, 1995.)
different image to the other eye. The stereoscope proved that the visual system treats binocular disparity as a depth cue, regardless of whether the disparity is produced by actual or simulated images of a scene.

For the average citizen the stereoscope was not science; it was home entertainment. The Wheatstone stereoscope held two different images in two different places. In the 1850s, however, David Brewster and Oliver Wendell Holmes invented viewers (Figure 6.30) that held a card with a double image like that shown in Figure 6.30b. The double images were captured by cameras with two lenses separated by about 2 inches, the distance between the average human’s eyes. This arrangement allows stereo cameras to take a pair of pictures that mimic the images produced by the projective geometry of human binocular vision. Photographers traveled the world with these stereo cameras, capturing far-off scenes in a way that enabled a London schoolchild to see, for example, a vivid three-dimensional image of the British Army in Cape Town, South Africa. (For a guide to more of these historical images, as well as many other stereograms, see Web Essay 6.2: Stereo images on the Web.)

A stereoscope is very helpful, but you don’t actually need one to experience stereopsis. You can teach yourself a technique known as free fusion, which John Frisby (1980) called “the poor man’s stereoscope.” Figure 6.31 contains two almost identical pictures. If you cross your eyes hard, you should see four sets of squares. This is the phenomenon of double vision (diplopia), described in the previous section. Two of those sets are seen with the left eye and two with the right. The trick is to relax just a bit until you see just three sets of squares. The far-left set is seen only in the left eye; the far-right set, only in the right; but the middle set is the fusion of two sets, one seen by the left eye and another seen by the right eye. This fusion of the separate images seen by the two eyes makes stereopsis possible.

Achieving the perception of three, instead of four or two, sets of squares in Figure 6.31 is the first step toward free fusion. The second step is to bring the middle set into focus. As noted earlier, convergence and accommodation normally work in lockstep, so crossing your eyes automatically leads your ciliary muscles to make your lenses more spherical (unless you are presbyopic—see Chapter 2). Similar problems (in the opposite direction) will occur

**FIGURE 6.30** Stereopsis for the masses. (a) This Holmes stereoscope—among others—brought stereo photos into many mid-nineteenth-century homes. (b) A stereo photo of South African Light Horse, a scouting regiment of the British Army, on Adderly Street in Cape Town, South Africa, in 1900. If you can free-fuse (explained later in this section), you will be able to see this scene jump out in depth.
if you diverge your eyes. To see the middle set of squares clearly, you have to decouple accommodation and convergence. If you can manage it, then the image will come into focus and the three white squares will appear to lie at different depths in the middle set of squares. When you view them normally, notice that the white squares in the left and right panels look misaligned in opposite directions. Those are the monocular views. When you free-fuse, the opposite misalignments become the binocular disparity, and your visual system converts that disparity into a perception of depth.

The depth that you see depends on whether you converged or diverged your eyes. We described crossing, or converging, the eyes. It is also possible to free-fuse the images in Figure 6.31 by diverging your eyes. Divergence requires focusing on a point beyond the plane of the page so that the image of the left-hand set of squares falls on the left fovea and the image of the right-hand set falls on the right fovea. Because the images falling on the two retinas in the divergence method are reversed compared with the convergence situation, the disparities are reversed and the perceived depth will be reversed. If you converge, the top square will be the farthest back. If you diverge, it will appear closest to you. Either converging or diverging will produce a clear stereoscopic effect, so give it a try.

Before we go on, we should note that approximately 3–5% of the population lacks stereoscopic depth perception—a condition known as stereoblindness. Stereoblind individuals might be able to achieve the perception of three sets of squares in Figure 6.31, but the little white squares will not pop out in depth. Stereoblindness is usually a secondary effect of childhood visual disorders such as strabismus, in which the two eyes are misaligned. See the section “Recovering Stereo Vision” below. If you had such a visual disorder during childhood and/or you’ve been diagnosed with stereoblindness, we apologize but you just won’t perceive depth in the stereograms presented here and on the website. That said, many people who try and fail to see depth in stereograms have “normal” vision (wearing glasses doesn’t count as “abnormal” in this case). Those people just need practice, so don’t give up. Web Activity 6.3: Stereoscopes and Stereograms provides more stereograms for practice, and Web Essay 6.2: Stereo Images on the Web leads to another website with more tips for free-fusing.

Stereoblindness An inability to make use of binocular disparity as a depth cue. This term is typically used to describe individuals with vision in both eyes. Someone who has lost one (or both) eyes is not typically described as “stereoblind.”
Recovering Stereo Vision

As noted earlier, about 3–5% of the population are stereoblind, usually due to early childhood visual disorders. Can stereo vision be recovered later in life? Meet Stereo Sue and Stereo Bruce.

Susan Barry, a professor of neurobiology, had strabismus as an infant and had been stereoblind essentially all her life. Her book Fixing My Gaze provides a fascinating, informative, and beautifully written account of her acquisition of stereopsis following vision therapy. It describes her transformative journey from the many visual, social, and psychological challenges of a turned eye (a squint or strabismus) early in life, to the sudden enrichment of her perceptions of the world following successful unconventional visual therapy begun at 48 years of age. (An earlier article about her visual recovery was published in The New Yorker under the title of “Stereo Sue” by Oliver Sacks.)

Barry vividly recounts how acquiring stereoscopic vision led to a dramatic improvement of her perception of depth, or the appreciation of “the space between” objects. A particularly valuable insight is her argument for the inability of people with normal vision to appreciate the visual experience of being stereoblind. Naïvely one might think that this experience would be duplicated simply by closing one eye so all information about depth was conveyed by monocular cues. Not so, however, Barry argues: the monocular experience of a typically reared person who closes one eye has been informed by a lifetime of experience with stereoscopic vision and so is far different from that of a person who is stereoblind. As a result, Barry’s new stereoscopic vision brought much more to her life than just depth perception: Objects became clearer, motion perception more veridical, her ability to move around the world more confident. Even more poignant is her vivid description of the enhanced sense of touch she had developed over the years and its key role in informing her newly acquired sense of stereo vision. Barry did not simply “recover” stereopsis but rather had to relearn to see with stereo vision. Barry did not simply “recover” stereopsis but rather had to relearn to see with stereo vision. As blind or deaf individuals often describe, individuals deprived of a sense are not just “missing” a sense. Rather, they have developed an entirely different way of sensing the world. Upon sensory restitution, a fascinating but rather disturbing experience unfolds as the brain has to adapt to a new way to function.

Even more dramatic is the experience of Stereo Bruce (Bruce Bridgeman), a very perceptive vision scientist who had been stereo-deficient all his life. Remarkably, he recovered stereopsis after watching the 3D movie Hugo (Bridgeman, 2014). Whether this sort of immersive experience, with very large disparities along with many other depth cues, will be a generally effective treatment for abnormal stereopsis remains to be tested. However, these case studies, along with lab studies of perceptual learning that result in the recovery of stereopsis (Ding and Levi, 2011), call into question the notion that has been the received wisdom, that recovery of stereopsis can only occur during early childhood. The idea, dating back to the early twentieth century, has been that there is a “critical period” of development when the visual system is still plastic and capable of change. After that, it was thought, our basic visual capabilities are fixed. This led a number of practitioners to tell Susan Barry and her mother that “nothing could be done” about her vision (one suggested that she might need a psychiatrist). Since binocular neurons are present in the visual cortex of primates within the first week of life (see the “Development of Binocular Vision and Stereopsis” section below), Barry surmises that some of the innate wiring of her binocular connections remained intact, and that vision therapy taught her to move her eyes into position for stereo vision, “finally giving these neurons the information they were wired to receive” (Barry, 2009).
Random Dot Stereograms

For 100 years or so after the invention of the stereoscope, it was generally supposed that stereopsis occurred relatively late in the processing of visual stimuli. The idea was that the first step in free-fusing images such as those in Figure 6.32 would be to analyze the input as a face. We would then use the slight disparities between the left-eye and right-eye images of the nose, eyes, chin, and other objects and parts to enrich the sense that the nose sticks out in front of the face, that the eyes are slightly sunken, and so on.

Bela Julesz, a Hungarian radar engineer who spent most of his career at Bell Labs in New Jersey, thought the conventional wisdom might be backward. He theorized that stereopsis might be used to discover objects and surfaces in the world. Why would this be useful? Julesz thought that stereopsis might help reveal camouflaged objects. A mouse might be the same color as its background, but out in the open it would be in front of the background. A cat that could use stereopsis to break the mouse’s camouflage would be a more successful hunter. (Cats do have stereopsis, by the way [Blake, 1988; R. Fox and Blake, 1971].) To prove his point, Julesz (1964, 1971) made use of random dot stereograms (RDSs). An example is shown in Figure 6.33. If you can free-fuse these images, you will see a pair of squares, one sticking out like a bump, the other looking like a hole in the texture (which one is the bump and which is the hole depends, again, on whether you converge or diverge your eyes).
The important point about RDSs is that we cannot see the squares in either of the component images. We cannot see the squares using any monocular depth cues. These are shapes that are defined by binocular disparity alone. Julesz called such stimuli Cyclopean, after the one-eyed Cyclops of Homer’s Odyssey. Wheatstone showed with his stereoscope that binocular disparity is a necessary condition for stereopsis. Julesz demonstrated with the RDS that disparity is sufficient for stereopsis. To understand how RDSs are made, visit Web Activity 6.3: Stereoscopes and Stereograms, which also explains “magic eye” and “wallpaper” stereograms, known more technically as “autostereograms.”

Stereo Movies, TV, and Video Games
Whereas the Brewster stereoscope provided home entertainment in Victorian times, in our day 3D movies, TV, and video games are the new stereoscope. Actually, 3D movies have been around for more than 60 years. In general, 3D movies make use of stereoscopic photography (which records the images as seen from two slightly different perspectives) and 3D glasses (which separate the images to the two eyes, so that disparity provides additional cues to depth). Some of the earliest 3D movies, were viewed with “anaglyphic” glasses, with a red lens for one eye and a blue or green filter for the other eye (Figure 6.34). When the images on the screen were filtered through the differently colored filters, one eye saw one set of images, and the other eye saw the other set. You’ve may have seen the anaglyphic method used in comic books. Later, filmmakers used polarization to separate the images to the two eyes, and polarizing glasses were used to view the films. A recent advance on the anaglyphic method is the use of interference filters to present specific wavelengths of red, green, and blue to the right eye, and different wavelengths of red, green, and blue to the left eye. Very expensive goggles that filter out the specific wavelengths enable the filmgoer to see a three-dimensional image and enable the movie theater to charge a few extra dollars per ticket.

Many modern 3D movies, TV, and video games use special glasses that contain liquid crystal display technology (known as LCD shutter glasses) and
enable the presentation of images to each eye that are synchronized with the images on the movie, TV, or computer screen. In this method, alternate frames of the movie are presented to different eyes. All of the systems described here have the drawback that they require special glasses or goggles to see in three dimensions. However, certain clever methods do not require glasses. For example, in a method known as lenticular printing, the images are digitally split and interleaved (left, right, left, right, ...) at a fixed spacing. A special array of many tiny lenses placed on the screen ensures that the left eye sees only the “left” image and the right eye sees only the “right” image. This setup permits three-dimensional viewing without glasses or other paraphernalia. The development for the consumer market of virtual reality (VR) technology is making 3D games and movies more widely available. VR technology enables the user to be immersed in a virtual 3D world. Its commercial potential was evident when, in 2014, Facebook bought Oculus, the VR company that developed the pioneering Oculus Rift system.

Using Binocular Stereopsis

The applications of stereopsis go well beyond entertainment. The military has known for a long time that you can get more information out of aerial surveillance if your view of the ground is stereoscopic. However, if you’ve ever looked out the window from thousands of feet in the air, you may have noticed the ground looks rather flat. This is yet more geometry. Binocular stereopsis can provide useful information about metric depth only for distances up to 40 meters (Palmisano et al., 2010). With eyes a few centimeters apart, you don’t get adequate disparity from more distant targets. What you need are eyes separated by hundreds of feet. This can be done if you have a plane and a special camera. Figure 6.35 reprints a figure from a 1951 issue of Popular Mechanics in which Colonel George W. Goddard showed the public how images taken from two vantage points produced stereo images during the Korean War. We can also use stereopsis to have a better look inside the body. Figure 6.36 shows a stereo pair of mammograms, X-rays of the breast used to detect breast cancer. Actually, in order to create another free-fusion demonstration, we’ve printed the image of the left breast twice, to the left and the to right of the image of the right breast in the middle. If you free-fuse the images so that you see four breast images, one of the two center images will
have the correct disparities (depending on whether you diverge or converge your eyes to free-fuse). The correct one will show the white wire—a marker for the surgeon—on top of the breast. You can see that the breast tissue is characterized by a network of intersecting structures. This is like our earlier example of looking at the little branches of a tree while lying on your back with just one eye open. It can be very hard to tell which line-like structures actually intersect and which ones lie at different depths. This turns out to be important when reading a mammogram. A starburst structure might be a sign of cancer, but not if it is an accidental viewpoint (see Chapter 4) of structures at different depths in the breast that just happen to form a suspicious pattern in a two-dimensional projection. Stereopsis can disambiguate this situation. If you can free-fuse these images, you will see the texture in three dimensions, and you will be able to determine how different structures relate to each other in depth. Stereoscopic displays are beginning to be used in radiology (Held and Hui, 2011), and they can reduce the error rate in these important tasks (Getty, D’Orsi, and Pickett, 2008).

Is binocular stereopsis useful in everyday life? In people with normal binocular vision, visually guided hand movements are significantly impaired when viewing is restricted to one eye (Fielder and Mosely, 1996), likely owing to the fact that binocular depth thresholds are about a factor of 10 better than monocular thresholds (McKee and Taylor, 2010). These results are mirrored in patients with amblyopia (“lazy eye”) for whom many observed visuomotor deficits are due to impaired stereopsis, and in particular impaired visual feedback control of movements, rather than visual acuity loss (Grant and Moseley, 2011). Loss of stereopsis may also result in unstable gait, especially reduced accuracy when a change of terrain (e.g., steps) occurs, and difficulties for children in playing some sports.

Further Discussion of stereo sensation can be found in Chapter 10 (sound localization; pages 291–295) and Chapter 14 (binaural rivalry in olfaction; page 443).

Stereoscopic Correspondence

If you successfully free-fused the random dot patterns in Figure 6.33, you solved a truly daunting problem. Even if you didn’t, if you have normal binocular vision, you are solving the correspondence problem all the time. The correspondence problem is the problem of figuring out which bit of the image in the left eye should be matched with which bit in the right eye. Figures 6.37 and 6.38 use an extremely simple situation to illustrate why correspondence is so tricky. There are, of course, just three dots in Figure 6.37. Figure 6.38a traces the paths of the rays of light from the printed circles on the page to the
images on the viewer’s retinas. The retinal images of the circles are labeled to make it clear which image on the left retina corresponds to which image on the right retina, but your visual system has no such labels. All it knows about is the retinal images, as shown in Figure 6.38b. Figure 6.38c shows another possible geometric interpretation of the situation: if the left retinal image of circle 2 is matched to the right retinal image of circle 1, and the left retinal image of circle 3 is matched to the right retinal image of circle 2, you will perceive four circles, with the inner pair of circles perceived as floating in front of the outer pair. (In fact, you may be able to experience this for yourself if you can cross your eyes correctly.)

With only three elements in the visual scene, it isn’t hard to imagine how the visual system might achieve the proper correspondence: first match the two circles whose images fall on the foveas; then match the two images to the left of the foveas with each other; then match the two images to the right of the foveas. Before random dot stereograms, a similar logic seemed reasonable for more complex scenes too. Go back to the face in Figure 6.32. Our visual systems could solve the correspondence problem by first finding the parts of the two faces and then matching nose to nose, mouth to mouth, and so forth. The RDS in Figure 6.33, however, contains thousands of identical black and white dots falling on each retina. How can we be sure that the dot in the center of the fovea of one eye corresponds to the dot in the center of the other eye? Even if we knew that, could we really match each dot in the right eye with just one dot in the left eye? If there were a little dirt on the page, would the whole process collapse? How in the world does our visual system succeed in making the proper matches?

Matching thousands of left-eye dots to thousands of right-eye dots in Figure 6.33 would require a lot of work for any computational system. However, the problem is simpler if we look at a blurred version of the stereogram. Blurring leaves only the low-spatial-frequency information. Figure 6.39 shows the low spatial frequencies of the stereogram from Figure 6.33. Now, rather than thousands of dots, we have just a few large blobs. Now you could imagine a process that, for example, matched the black blob in the upper left corner of the left image with the very similar blob in the right image. Crude matches

**FIGURE 6.38** Interpreting the visual information from the three circles in Figure 6.37. It would require careful placement, but four dots in the world could produce three dots in each eye as in (c).
In addition to starting with low-spatial-frequency information, David Marr and Tomaso Poggio (1979) suggested two more heuristics for solving the correspondence problem. They called these the uniqueness and continuity constraints. The uniqueness constraint acknowledges that a feature in the world is represented exactly once in each retinal image. Working in the opposite direction, the visual system knows that each monocular image feature (e.g., a nose or a dot) should be paired with exactly one feature in the other monocular image. Notice that Figure 6.38c would not violate uniqueness. Each dot in the world would be represented exactly once in each retinal image. The odd thing is that two dots in the real world could be represented by the same dot in the retinal image. The continuity constraint holds that, except at the edges of objects, neighboring points in the world lie at similar distances from the viewer. Accordingly, disparity should change smoothly at most places in the image. (These constraints are difficult to illustrate on a static page, but Web Activity 6.4: Stereoscopic Correspondence provides dynamic explanations.) With those constraints, the correspondence problem is not entirely solved but it is made much more tractable. There are not so many possible solutions.

**The Physiological Basis of Stereopsis and Depth Perception**

Now that we know something about the theoretical basis of stereopsis, we can ask how it is implemented by the human brain. The most fundamental requirement is that input from the two eyes must converge onto the same cell. As noted in Chapter 3, this convergence does not happen until the primary visual cortex, where most neurons can be influenced by input from both the left and right eyes—i.e., they are binocular (Hubel and Wiesel, 1962). A binocular neuron has two receptive fields, one in each eye. In binocular primary visual cortex neurons, the receptive fields in the two eyes are generally very similar, sharing nearly identical orientation and spatial-frequency tuning, as well as the same preferred speed and direction of motion (Hubel and Wiesel, 1973). Thus, these cells are well suited to the task of matching images in the two eyes.

Many binocular neurons respond best when the retinal images are on corresponding points in the two retinas, thereby providing a neural basis for the horopter. However, many other binocular neurons respond best when similar images occupy slightly different positions on the retinas of the two eyes (Barlow, Blakemore, and Pettigrew, 1967; Pettigrew, Nikara, and Bishop, 1968). In other words, these neurons are tuned to a particular binocular disparity (diagramed in Figure 6.40).

**uniqueness constraint** In stereopsis, the observation that a feature in the world is represented exactly once in each retinal image. This constraint simplifies the correspondence problem.

**continuity constraint** In stereopsis, the observation that, except at the edges of objects, neighboring points in the world lie at similar distances from the viewer. This is one of several constraints that have been proposed as helpful in solving the correspondence problem.
Recall the distinction, from earlier in this chapter, between metrical and nonmetrical depth cues. Stereopsis can be used both metrically and nonmetrically. Nonmetrical stereopsis might just tell you that a feature lies in front of or behind the plane of fixation. Gian Poggio and his colleagues (Poggio and Talbot, 1981) found disparity-tuned neurons of this sort in V2 (which stands for “visual area two”) and some higher cortical areas. Some neurons responded positively to disparities near zero—that is, to images falling on corresponding retinal points. Other neurons were broadly tuned to a range of crossed (near) or uncrossed (far) disparities. On the other hand, stereopsis can also be used in a very precise, metrical manner. Indeed, stereopsis is a “hyperacuity” like Vernier acuity (see Web Essay 3.1: Hyperacuity) with thresholds smaller than the size of a cone. Both of these forms of stereopsis have their uses, and functional magnetic resonance imaging (fMRI) data suggest that the dorsal where pathway (see Chapter 4) is most interested in metrical stereopsis, while the ventral what pathway makes do with more categorical, near-versus-far information (Preston et al., 2008).

The neural bases of other depth cues have also been investigated. For example, when we discussed motion parallax earlier, we suggested moving your head back and forth while looking into the branches of a tree in order to create a more vivid impression of the depth relationships among the branches and twigs. To exploit that cue properly, you need to know how your head is moving (see Chapter 12) and how items in the visual field are moving (see Chapter 8). Nadler, Angelaki, and DeAngelis (2008) looked for the neural substrate of parallax in the middle temporal area (area MT) of the brain of macaque monkeys. As we will see in Chapter 8, this area is very important in the perception of motion. Nadler et al. set up an apparatus where the monkey

![Fig. 6.40](https://example.com/fig6.40.png)  
**FIGURE 6.40** In these simplified diagrams of receptive fields for two binocular-disparity–tuned neurons in primary visual cortex, the red neuron “sees” stimuli falling on the red receptive fields, and the blue neuron responds to stimuli falling on the blue receptive fields (these receptive fields overlap on the right retina). (a) The overall picture, showing the fixation point in relation to the two retinas. (b) The red neuron responds best to a stimulus closer to and slightly to the right of the fixation point. (c) The blue neuron responds best if its preferred stimulus is behind and slightly to the left of fixation.
was moved from side to side while items on the screen also moved. If the monkey was integrating signals about its head movement with the motion signals, then the objects on the screen should have been seen in depth. Otherwise, they would have been seen as just moving in the plane of the screen. It turns out that cells in area MT can signal the sign of depth (near or far) based on this motion parallax signal alone.

Other visual areas also contribute to the complex business of inferring the three-dimensional world from two-dimensional retinal images. Anzai and DeAngelis (2010) suggest that early visual areas, particularly V2, are involved in computing depth order (who’s in front?), based on the contour completion and border ownership process we discussed in Chapter 4 (see Figures 4.5 and 4.13). Intermediate visual areas such as V4 (meaning “visual area four”) encode depth intervals, based on relative disparities, and higher cortical areas such as inferotemporal cortex are involved in the representation of complex three-dimensional shapes.

FURTHER DISCUSSION of phase as it relates to vision can be found in Chapter 3 on pages 62–64. Sensitivity to phase is also important for temporal coding of sound frequency (Chapter 9, page 278).

Combining Depth Cues

If the chapters of this book were novels, this chapter could be said to have the same plot as the discussion of object recognition in Chapter 4, but with different characters. In Chapter 4 we talked about a set of cues that enable us to group local features together into possible objects and then to recognize those objects. We described the process as a sort of committee effort in which different sources of information all contribute their opinions and where we see the committee decision without necessarily knowing how that decision was reached. In this chapter we’ve covered multiple sources of depth information and they, too, need to be combined. Any or all of these cues might be available to the visual system when we’re viewing any visual scene. None of the cues are foolproof, and none work in every possible situation. For example, relative height produces inconsistent or misleading information if we can’t see the point at which an object touches the ground. All we really have is a collection of guesses about possible depth relations between different objects in our visual field.

By carefully combining and weighting these guesses, the visual system generally arrives at a coherent, and more or less accurate, representation of three-dimensional space. Helmholtz, writing in the nineteenth century (and translated into English in the twentieth), called this automatic cue combination process “unconscious inference” (Helmholtz, 1924). In recent years, a number of vision researchers have been attempting to put this sort of argument on the more rigorous mathematical footing of the Bayesian approach that we mentioned in Chapter 4 and that is the subject of a more detailed essay on the website, Web Essay 4.4: Bayesian Analysis.

The Bayesian Approach Revisited

Recall that the basic insight of Reverend Thomas Bayes was that prior knowledge could influence estimates of the probability of a current observation. Let’s apply this idea to a concrete, depth perception example. Suppose our visual system is confronted with the retinal image shown in Figure 6.41. There are infinite possible ways to produce this retinal image. Actually, this is a bit of a

Bayesian approach A way of formalizing the idea that our perception is a combination of the current stimulus and our knowledge about the conditions of the world—what is and is not likely to occur. The Bayesian approach is stated mathematically as Bayes’ theorem—\( P(A|O) = P(A) \times P(O|A) / P(O) \)—which enables us to calculate the probability \( P \) that the world is in a particular state \( A \) given a particular observation \( O \).
problem for the use of Bayes’ theorem. We don’t really know the prior probabilities (M. Jones and Love, 2011). Still, we can acknowledge that limitation, and we can still make good use of the basic insight that some hypotheses are more likely than others and that these prior probabilities can shape our interpretation of the world. Three hypotheses about our pennies are shown in Figure 6.42. Maybe the two pennies are the same size, but the one on the left is slightly farther away than the one on the right (Figure 6.42a). Maybe the penny on the right is much smaller, but also much closer, than the penny on the left (Figure 6.42b). Maybe the two pennies are equidistant, but the penny on the left is smaller and has had a bite taken out of it (Figure 6.42c). If you don’t see how the set of possibilities could be infinite, remember that size and distance can vary continuously over a large range. In Figure 6.42b, the big penny could be on the moon (but it would have to be a really big penny).

How does the visual system decide what we’re actually seeing? Which interpretation seems most likely? That is the core of the Bayesian approach (except that it’s all automatic; our conscious selves do not get to make the decision). In our experience, all pennies are the same size. This cue of familiar size is one source of prior knowledge in this case. This makes the prior probability of the hypotheses shown in Figure 6.42a higher than the prior probabilities of the other two hypotheses. Furthermore, for the scene in Figure 6.42c to produce the retinal image in Figure 6.41, we would have to be seeing the scene from one of those unusual and unlikely “accidental viewpoints.” It is much more likely that the points of contact between the images of the two pennies reflect occlusion. If we were to plug all these probabilities into the math of Bayes’ equation, we would find that, given the image in Figure 6.41, the most likely answer is the scene depicted in Figure 6.42a.

![Figure 6.41](image)

**Figure 6.41** Retinal image of a simple visual scene.

![Figure 6.42](image)

**Figure 6.42** Three of the infinite number of scenes that could generate the retinal image in Figure 6.41. Which of these is the most probable?
In thinking about combining depth cues, our choice of the metaphor of a committee is not arbitrary. We could have talked about an election, but that would have implied something like “one cue one vote.” On a committee, you might have one member who is stronger than the others and wins all the arguments. You might give more weight to the committee member who comes prepared with the best information. The committee might defer to one member on one topic and another member on a different topic. Something like this last option is described by Held, Cooper, and Banks (2012). Binocular-disparity information can be very precise, but that is only true near the plane of fixation (remember Panum’s fusional area?). Blur can be quite a good cue too, but it is actually better away from the plane of fixation. When Held et al. made stimuli that had disparity cues, blur cues, or both, they found that disparity drove responses where disparity was more reliable and that blur drove responses in parts of the three-dimensional world where it was more reliable. This is different from just letting every cue have its say, and it makes us realize that the visual system must be estimating how reliable each depth cue might be.

Illusions and the Construction of Space

If our visual perception of the world is our best guess about the causes of visual input, then interesting things should happen when a guess is wrong. In some sense, as with the pennies we just discussed, a guess is wrong whenever we look at a two-dimensional picture and see it as three-dimensional. As noted, however, we are not really fooled into thinking that the picture is three-dimensional. It would be more accurate to say that we make a plausible guess about the three-dimensional world that is being represented in the two-dimensional picture.

What about a situation like that shown in Figure 6.43? One of the five pairs of horizontal lines (and only one) shows two lines of the same length. Can you pick the correct pair (without a ruler)? In fact, it is the second from the left. Odds are you picked the third or fourth pair, even though the bottom line in both of those images is physically longer than the top line. This is known as the Ponzo illusion, named after Mario Ponzo, who described the effect in 1913. What causes this illusion? For many years, a popular family of theories has held that the illusion is a guess gone wrong—a situation in which we overinterpret the depth cues in a two-dimensional image. The basic idea is illustrated in Figure 6.44. Maybe the two tilted lines that induce the illusion in each image of Figure 6.43 are being interpreted by the visual system as linear-perspective cues like the train tracks in Figure 6.44. If so, then objects that were the same size in the two-dimensional image would represent objects of different sizes in the three-dimensional world.

Such accounts are very compelling and exist for a wide range of visual illusions (Gillam, 1980). (See Web Essay 6.3: The Moon Illusion.) They are consistent with the view that the job of the visual system is to use available cues to make an intelligent guess about the world (Gregory, 1966, 1970). Just
because an answer is plausible, however, doesn’t mean it’s right. In Figure 6.45a, line B looks longer than line A. That makes sense if we’re interpreting these lines as lines lying at different distances on the wall of the colonnade. As in the Ponzo illusion, if line B is farther away than A, then the same image size implies a larger size in the real world. But what about lines C and D? Surely D would be interpreted as farther away than C. Does it look convincingly larger? Figure 6.45b reveals what you probably guessed—that all five of the lines in Figure 6.45a are the same length.

Prinzmetal, Shimamura, and Mikolinski (2001) use a demonstration like this as part of their argument that the Ponzo illusion is not really a by-product of depth cues. They argued that it reflects a more general aspect of the visual system’s response to tilted lines and is related to illusions like the Zollner and Hering illusions illustrated in Figure 6.46 (which we will not try to explain;
see Prinzmetal and Beck, 2001). The point is debatable. After all, line E looks very big down there at the apparent end of the colonnade in Figure 6.45a. Who is right? It could be that both positions hold a piece of the truth. Perhaps the visual system’s response to tilted lines is related to the role of those lines in creating an impression of depth. Going back to Figure 6.44, that would mean that the Ponzo illusion was not based on some version of the railroad track story but that the processes that give rise to a three-dimensional interpretation of Figure 6.44 also give rise to illusions like those in Figures 6.43, 6.45, and 6.46.

**Binocular Rivalry and Suppression**

The preceding sections demonstrated that objects in the world often project images on our two retinas that do not overlap (that is, the images fall on non-corresponding retinal points) and that the visual system is physiologically prepared to deal with these discrepancies via disparity-tuned neurons in striate cortex and beyond. But what happens when completely different stimuli are presented to the two eyes? You can answer this question for yourself by fixating on a small object across the room, such as a clock, and moving your hand up so that your fingers occlude the object in the right eye (making sure the left eye still has an unobstructed view). It would seem a bad idea to fuse the fingers and the clock into a single perception of something that does not exist in the world. Accordingly, the visual system chooses instead to suppress one image and perceive the other. In the present situation, you probably see the clock as though you were looking through a hole in your hand. (For an even more compelling perception of a hole in your hand, click on “Another Demonstration” in Web Activity 6.2: Binocular Disparity.)

How does the visual system “decide” what to see? The more interesting of the two stimuli is likely to be dominant. Interesting in this case has several meanings. The most important factor is which stimulus is more salient to the early stages of cortical visual processing. High contrast is more salient than low contrast, bright is better than dim, moving objects are more interesting than stationary ones, and so forth (Fahle, 1982). The meaning of the stimulus also has an effect (Yu and Blake, 1992), as does what you’re attending to (Ooi and He, 1999).

The competition between the two eyes for control of visual perception, known as **binocular rivalry**, is never completely won by either eye (Alais and Blake, 2005; Wheatstone, 1852) or either stimulus (Blake and Logothetis, 2002). If you stare at the combination of the clock and hand long enough, your fingers will eventually conquer the visual territory, only to surrender it back to the clock a moment later. The battle is easier to see if the two combatants are more closely matched. If you free-fuse the two panels of Figure 6.47, your visual system will not actually combine the perpendicular stripes in the two center squares. Instead, you will see a battle between the vertically and horizontally...
striped patches, with regions of dominance growing and shrinking over time, as illustrated in Figure 6.48.

Binocular rivalry might seem an odd situation that would arise only in a vision lab or a perception course, but a moment’s reflection should convince you that the stimuli for rivalry are actually very common. Recall that Panum’s fusional area is the region of space surrounding the horopter within which images on the two retinas can be fused. As noted earlier, objects located too far off the horopter—outside Panum’s area—are subject to diplopia. In fact, the proportion of the three-dimensional visual world that falls inside Panum’s area is actually fairly small. Think about two pairs of corresponding retinal points: A_L and A_R, and B_L and B_R (the L and R stand for “left” and “right,” respectively). Now imagine that one object falls on two noncorresponding retinal points: A_L and B_R. It follows that something else must be forming an image on A_R. There will be noncorresponding images on the corresponding points A_L and A_R (and, similarly, on B_L and B_R). Those are the conditions for rivalry.

The fact that the two eyes are seeing different images brings us to an ancient problem: the problem of binocular single vision. Why, when we see one elephant in one eye and one elephant in the other eye, don’t we perceive two elephants? Given what we’ve learned in this chapter, we’re now in a position to answer this question. If the elephant is within Panum’s area, we fuse its two images into a single stereoscopic perception. If it is outside Panum’s area, we normally suppress one of the copies. Why don’t we see the rivalry? In part, because we aren’t looking. Our attention and eyes are typically directed toward the foveated object or toward objects falling on roughly corresponding points in each eye. Moreover, acuity is so bad in the periphery of the visual fields that, even when objects are vying for binocular precedence, the rivalry is quite indistinct.

The classic demonstrations of rivalry pit a stimulus in one eye against a stimulus in the other eye. More recently, it has become clear that rivalry is part of a larger effort by the visual system to come up with the most likely version of the world, given the current retinal images (sounds Bayesian again,

**Figure 6.47** Binocular rivalry. If you free-fuse these two images, you will be able to watch the blue verticals and orange horizontals engage in the perceptual battle known as binocular rivalry.

**Figure 6.48** If blue vertical bars are shown to one eye while orange horizontal bars are shown to the other, the two stimuli will battle for dominance.
Binocular rivalry is not just a fight between the two eyes. If you can free-fuse these pairs of images, you will see a typical example of rivalry. (a) Different images in each eye struggle for dominance. The chimp and the text struggle for dominance in panel (b) too, but this is pattern rivalry: the two eyes are actually cooperating to put together coherent views of the chimp or the text. (From Kovacs et al., 1996.)

Rivalry is a very useful tool for probing one of the more vexing problems in the neuroscience of perception: what parts of the visual system give rise to the conscious experience of seeing something? For these purposes, the great feature of rivalry is that it dissociates the stimulus on the retina from the stimulus that you see. Now, imagine that you record from single cells somewhere in the monkey visual system (Logothetis and Schall, 1989) or you use fMRI to look at the working human brain. If you train the monkeys or ask humans to monitor their perception (“Do you see vertical or horizontal?”), you can ask whether the neural signal in a specific part of the visual pathway follows the physical stimulus or the perceived stimulus. It is a complex and evolving story, but one that clearly shows that conscious visual awareness is not something that happens in one discrete step in a chain of visual processing (Blake and Wilson, 2011).

Development of Binocular Vision and Stereopsis

What is binocular vision like in infants? Babies are born with two eyes, but are they born with stereopsis? If not, how does binocular function develop? In a wonderful conversation about visual development, Davida Teller and Tony Movshon (1986) recalled a lecture by a disillusioned developmental psychologist, John McKee, who argued that the field could be summed up by three laws:

1. As children get older, they get better at things.
2. Whatever it is, girls do it before boys.
3. Everything develops along with everything else.

To these “laws,” Teller added a summary statement: “Things start out badly, then they get better; then, after a long time, they get worse again” (Teller and Movshon, 1986).

As it turns out, research over the last 25 years or so has shown that visual development provides support for the first and second “laws,” but not for a strict form of the third. The development of binocular vision and stereopsis provides one of the strongest violations of that third law.

Most visual functions indeed start off badly (but not as badly as we used to think) and then improve steadily until they reach adult levels; however, the development of stereopsis is surprising, in that infants are essentially blind to disparity until about 4 months of age. At that point, stereopsis appears quite suddenly—almost out of the blue. Of course, measuring stereopsis (or anything else) in infants is no easy task, but developmental psychologists have been very inventive in designing methods for assessing development.
Despite differences in techniques and procedures, most investigators agree about the onset of stereopsis. Figure 6.50 summarizes the results of several studies, showing the age at which stereopsis can first be detected. These studies (and others like them) looked for evidence indicating that infants could reliably detect a large binocular disparity (typically on the order of 30–60 arc minutes). The agreement among the studies is remarkable. Infants are essentially stereoblind before 3 months, with most infants showing a sudden onset of stereopsis between 3 and 5 months.

Stereopsis is not an all-or-none phenomenon. Just as an individual’s acuity is a measure of his ability to resolve spatial detail, **stereoacuity** is a measure of the smallest binocular disparity that can generate a sensation of depth. Once an infant develops stereopsis, stereoacuity increases rapidly to near adult levels (Figure 6.51). Birch and Petrig (1996) found that stereoacuity rose from essentially nothing before 4 months to near adult levels by 6 months! This time course is very different from the time course in the development of simple acuity. Though coarsely present at birth, basic acuity takes years to reach adult levels. The same difference between basic acuity and stereoacuity is seen in monkeys (O’Dell and Boothe, 1997), but the overall rate of development is faster. Interestingly, not only stereoacuity but several other visual functions develop at a rate approximately four times faster than in humans, as if one monkey week were the equivalent of one human month. In keeping with this rule of one human month being equal to one monkey week, stereopsis can be detected in monkeys within the first 3–5 weeks of life, compared with the 3- to 5-month window of onset observed in humans.

**Figure 6.50** The onset of stereopsis. This figure shows the percentage of infants demonstrating stereopsis for the first time as a function of their age. In three separate studies (the three different colors), almost all infants showed stereopsis for the first time between 3 and 5 months. (After Birch, 1993.)

**Figure 6.51** The development of stereoacuity. Stereoacuity develops to adult levels within the first 6–7 months of life. Data points below the dashed line indicate unmeasurable stereoacuity. (After Birch and Petrig, 1996.)

**stereoacuity** A measure of the smallest binocular disparity that can generate a sensation of depth.
How, then, do we explain the sudden emergence of stereopsis in humans at about 4 months? Although a newborn infant makes convergence eye movements to track a target as it approaches her nose, accurate and consistent convergence probably does not occur until 3–4 months of age. But we can’t conclude that inaccurate convergence prevents stereopsis from developing earlier than 4 months, because convergence need not be very accurate in order to detect large disparities. Moreover, several studies used repeating gratings, which would be fused at some disparity even if convergence were inaccurate. Before 4 months, babies don’t respond to these stereoscopic stimuli either.

An alternative view is that the failure of stereopsis to develop prior to 4 months might mean that some part of the visual system is immature. Disparity-sensitive neurons in the primary visual cortex (V1) are one plausible candidate for that immature part, but recent anatomical and physiological data suggest that we need to look beyond V1 for an explanation of why infants do not exhibit stereopsis.

Yuzo Chino and his colleagues made the most detailed quantitative study of the binocular responses of V1 neurons of infant monkeys (Chino et al., 1997). They presented a pair of drifting sine wave gratings dichoptically (one to each eye). The sine waves were identical in spatial frequency, orientation, contrast, and velocity, and these values were chosen to maximize the cell’s response. A phase difference between the two monocular gratings will also create binocular disparity. Some cells in the visual cortex are sensitive to this phase shift (Ohzawa and Freeman, 1986a, 1986b). When Chino et al. varied the relative spatial phase of the two drifting sine waves (Figure 6.52), the response of a binocular neuron waxed and waned (Figure 6.52b). The dashed lines in Figure 6.52b show the levels of response for the two eyes, stimulated alone. Notice that for some interocular phase differences, the binocular response was considerably higher than the response through either eye alone. At other interocular phase differences, the binocular response was lower than the response through either eye alone. This sinusoidal binocular phase tuning is the hallmark of this type of binocular neuron in the visual cortex. Using this sensitive method, Chino et al. (1997) found that within the first week of life—well before the onset of stereopsis—infant monkeys had practically the same proportion of phase disparity–sensitive neurons that adults have in primary visual cortex.

In addition, the ocular dominance properties of these infant monkeys were essentially identical to those of adults. Other investigators have also found that ocular dominance columns in the input layers of V1 are essentially adultlike at birth (Horton and Hocking, 1996).

What do these studies tell us about the development of stereopsis? The results suggest that the neural apparatus in V1 of newborns is capable of combining signals from the two eyes and that it is sensitive to interocular disparities. So why are newborns blind to disparity? One possibility is that the extraction of relative disparity, which is needed for stereoacuity, takes place beyond V1, possibly in V2. At this time we do not know much about how V2 neurons in newborns respond to disparity, but emerging evidence suggests that other receptive-field properties mature later in V2 than in V1 (Zhang et al., 2005; Zheng et al., 2007).

Another possibility is that the problem is in V1. Although V1 cells of newborn monkeys are adultlike in their response to interocular phase dispar-
ity, these neurons remain immature in several important ways. They do not have adult sensitivity to monocular spatial frequency or direction of motion. Moreover, they are much less responsive overall than are adult neurons (that is, their peak firing rates are considerably lower). In addition, these neurons display more interocular suppression than adult neurons do. Thus, it is also possible that because of the immaturity in V1 neurons, the signals they send to the next stage of processing are too weak or confused to support stereopsis.

**Abnormal Visual Experience Can Disrupt Binocular Vision**

The presence of all this binocular hardware, even if it is immature, strongly suggests that extensive binocular visual experience is not necessary for binocular connections to form in V1. These connections are present at birth or very shortly thereafter, so we don’t need to “learn” or develop binocular vision. However, the normal development of adult binocular vision and stereopsis does require visual experience. In Chapter 3 we learned about Hubel and Wiesel’s work on the **critical period**, the period during early visual development when normal binocular visual stimulation is required for normal cortical development. During this period the visual cortex is highly susceptible to any

**FIGURE 6.52** Interocular phase difference gratings are used to study disparity tuning. (a) The sinusoidal grating drifts differently in one eye than in the other, producing a sinusoidal change in spatial phase disparity. (b) The response of a binocular neuron. Notice that the binocular response varies with phase, and the monocular responses do not. (After Chino et al., 1997.)
disorder that alters normal binocular visual experience. In cats and monkeys this critical period is approximately the first 3–4 months of life.

We cannot study a child’s visual cortex the way we might study a monkey’s or a cat’s. How, then, can we estimate the critical period in humans? Some humans are born with two eyes that do not point at the same spot in the world. This not uncommon disorder is known as **strabismus**, as we learned in Chapter 3. The incidence is about 3%. In **esotropia**, one eye is pointed too far toward the nose (“cross-eyed”). In **exotropia**, the deviating eye is pointed too far to the side. There are various ways to treat strabismus. For example, it is possible to surgically correct the position of the eyes. For the present discussion, however, the important point is that it is possible to test adults who had misaligned eyes at different times and for different durations during childhood.

Recall from Chapter 3 that exposure to lines tilted to one side of vertical will make vertical lines appear tilted to the other side. This is known as the **tilt aftereffect**. One characteristic of the tilt aftereffect is that it shows interocular transfer (transfer of the effect from one eye to the other). If we show the adapting lines to one eye, we can measure an aftereffect through the other eye. This result is generally taken to show that the cells responsible for the effect are binocular: they receive input from both eyes (for some details, see Wolfe and Held, 1981). Individuals who exhibited strabismus during the first 18 months of life do not show normal interocular transfer (Banks, Aslin, and Letson, 1975; Hohmann and Creutzfeldt, 1975). This result provides an indirect estimate of the period during which binocular connections in humans are susceptible to abnormal input.

Let’s explore in a bit more detail why strabismus disrupts binocular vision. We will use left esotropia as an example. In left esotropia (Figure 6.53), the left eye is turned in. As a consequence, although the object of fixation (the yellow brick in this case) lands on the fovea of the right eye, in the left eye it lands on a region in the “nasal” retina. (The nasal retina is the half of the retina closer to the nose.) This means that the images of the yellow brick are in noncorresponding points in the two eyes. What will the patient see? If an adult becomes esotropic (perhaps because of an injury), she will experience *diplopia* (double vision), seeing two bricks instead of one. However, people who exhibit strabismus early in life often experience no such problem. Why? Notice that *something* is present at the fovea of the left eye. In Figure 6.53, this something is a purple pentagon. Thus, in strabismus, normally corresponding points in the two eyes receive conflicting information (this situation is known—not unreasonably—as “confusion”). To eliminate diplopia and confusion, the brain **suppresses** one of the two images. It is simply not consciously perceived. In esotropia, the most common pattern is suppression of the input from the eye that is turned in. So, in the example in Figure 6.53, the patient would most likely suppress visual input from the left eye.

Binocular rivalry is a form of suppression, so some suppression is an important part of normal visual experience. Unfortunately, early-onset strabismus can have other, more serious effects on the developing visual nervous system, and on visual performance. For example, strabismus greatly reduces the number of binocular neurons in the visual cortex (Wiesel, 1982). Cells that

---

**strabismus** A misalignment of the two eyes such that a single object in space is imaged on the fovea of one eye and on a nonfoveal area of the other (turned) eye.

**esotropia** Strabismus in which one eye deviates inward.

**exotropia** Strabismus in which one eye deviates outward.

**tilt aftereffect** The perceptual illusion of tilt, produced by adaptation to a pattern of a given orientation.

**suppression** In vision, the inhibition of an unwanted image. Suppression occurs frequently in people with strabismus.

---

**FIGURE 6.53** Left esotropia. The patient wants to fixate on the yellow brick, but the left eye is turned too far toward the nose; as a result, the left fovea is pointing at a different location (here the purple pentagon), while the image of the yellow brick falls to the right of the left fovea.
would normally be driven by both eyes are dominated by only one. You would be correct if you suspected that this situation would disrupt stereopsis. Birch and her colleagues (e.g., Stager and Birch, 1986) followed the development of stereopsis in normal and esotropic infants. Figure 6.54 illustrates the percentage of infants who showed a measurable ability to perceive stereoscopic depth. The red line shows that by about 6 months of age, almost all normal infants demonstrate stereopsis. In contrast, the esotropes (all of whom were diagnosed with esotropia by 6 months of age) initially demonstrated a normal pattern of stereopsis development. After 4 months, however, very few esotropic infants demonstrated stereopsis.

This result has an interesting parallel in cortical physiology. Chino and his colleagues (Kumagami et al., 2000) made otherwise normal monkeys strabismic. They found that a brief period of experimental strabismus shortly after the age of onset of stereopsis produced a greater loss of disparity sensitivity and more binocular suppression in V1 neurons than did an earlier episode of strabismus. These physiological and perceptual deficits appear to be permanent and have important implications for the surgical treatment of infantile esotropia. Almost all surgeons agree that treatment should be early, but there has been a lot of debate about just how early. These results suggest that the treatment should take place before the age at which stereopsis normally develops, in order to minimize the damage done by esotropia.

Summary

1. Reconstructing a three-dimensional world from two non-Euclidean, curved, two-dimensional retinal images is one basic problem faced by the brain.
2. A number of monocular cues provide information about three-dimensional space. These include occlusion, various size and position cues, aerial perspective, linear perspective, motion cues, accommodation, and convergence.
3. Having two eyes is an advantage for a number of reasons, some of which have to do with depth perception. It is important to remember, however, that it is possible to reconstruct the three-dimensional world from a single two-dimensional image. Two eyes have other advantages over just one: expanding the visual field, permitting binocular summation, and providing redundancy if one eye is damaged.
4. Having two laterally separated eyes connected to a single brain also provides us with important information about depth through the geometry of the small differences between the images in each eye. These differences, known as binocular disparities, give rise to stereoscopic depth perception.
5. Random dot stereograms show that we don’t need to know what we’re seeing before we see it in stereoscopic depth. Binocular disparity alone can support shape perception.
6. Stereopsis has been exploited to add, literally, depth to entertainment—from nineteenth-century photos to twenty-first-century movies. It has also served to enhance the perception of information in military and medical settings.
7. The difficulty of matching an image element in one eye with the correct element in the other eye is known as the correspondence problem. The brain uses several strategies to solve the problem. For example, it reduces the initial complexity of the problem by matching large “blobs” in the low-spatial-frequency information before trying to match every high-frequency detail.
8. Single neurons in the primary visual cortex and beyond have receptive fields that cover a region in three-dimensional space, not just the two-dimensional image plane. Some neurons seem to be concerned with a crude in-front/behind judgment. Other neurons are concerned with more precise, metrical depth perception.

9. When the stimuli on corresponding loci in the two eyes are different, we experience a continual perceptual competition between the two eyes known as binocular rivalry. Rivalry is part of the effort to make the best guess about the current state of the world based on the current state of the input.

10. All of the various monocular and binocular depth cues are combined (unconsciously) according to what prior knowledge tells us about the probability of the current event. Making the wrong guess about the cause of visual input can lead to illusions. Bayes’ theorem is the basis of one type of formal understanding of the rules of combination.

11. Stereopsis emerges suddenly at about 4 months of age in humans, and it can be disrupted through abnormal visual experience during a critical period early in life.