Chapter Contents

CHAPTER 5

Perceiving Objects and Scenes

OPPOSITE PAGE This painting by Robert Indiana, titled *The Great Love,* provides examples of how different areas of a picture can be perceived as figure and ground. At first you may see the red areas, spelling the word "Love," standing out as the figure. It is also possible, however, to see small green areas as arrows on a red background, or the blue shapes in the center as three figures on a red background.
© 2010 Morgan Art Foundation Ltd./Artists Rights Society (ARS), New York. Carnegie Museum of Art, Pittsburgh/Gift of the Women's Committee.

VL The Virtual Lab icons direct you to specific animations and videos designed to help you visualize what you are reading about. The number beside each icon indicates the number of the clip you can access through your CD-ROM or your student website.

Some Questions We Will Consider:

- Why do some perceptual psychologists say "The whole differs from the sum of its parts"? (p. 104)
- How do we distinguish objects from their background? (p. 108)
- How do "rules of thumb" help us in arriving at a perception of the environment? (p. 109)
- Why are even the most sophisticated computers unable to match a person's ability to perceive objects? (p. 119)

Sitting in the upper deck in PNC Park in Pittsburgh, Roger looks out over the city (Figure 5.1). On the left, he sees a group of about 10 buildings and can tell one building from another, even though they overlap. Looking straight ahead, he sees a small building in front of a larger one, and has no trouble telling that they are two separate buildings. Looking down toward the river, he notices a horizontal yellow band above the right field bleachers. It is obvious to him that this is not part of the ballpark but is located across the river.

All of Roger's perceptions come naturally to him and require little effort. However, what Roger achieves so easily is actually the end result of complex processes. We can gain some perspective on the idea that perception is complex and potentially difficult, by returning to the "science project" that we described at the beginning of Chapter 1 (review page 4).

This project posed the problem of designing a machine that can locate, describe, and identify all objects in the environment and, in addition, can travel from one point to another, avoiding obstacles along the way. This problem has attracted the interest of computer scientists for more than half a century. When computers became available in the 1950s and '60s, it was predicted that devices with capacities approaching human vision would be available within 10 or 15 years. As it turned out, the task of designing a computer that could equal human vision was much more difficult

Bruce Goldstein

Figure 5.1 It is easy to tell that there are a number of different buildings on the left and that straight ahead there is a low rectangular building in front of a taller building. It is also possible to tell that the horizontal yellow band above the bleachers is across the river. These perceptions are easy for humans, but would be difficult for a computer vision system.

than the computer scientists imagined; even now, the problem has still not been solved (Sinha et al., 2006).

One way to illustrate the complexity of the science project is to consider recent attempts to solve it. Consider, for example, the vehicles that were designed to compete in the "Urban Challenge" race that occurred on November 3, 2007, in Victorville, California. This race, which was sponsored by the Defense Advanced Research Project Agency (DARPA), required that vehicles drive for 55 miles through a course that resembled city streets, with other moving vehicles, traffic signals, and signs. The vehicles had to accomplish this feat on their own, with no human involvement other than entering global positioning coordinates of the course's layout into the vehicle's guidance system. Vehicles had to stay on course and avoid unpredictable traffic without any human intervention, based only on the operation of onboard computer systems.

The winner of the race is shown in Figure 5.2. "Boss," from Carnegie Mellon University, succeeded in staying on course and avoiding other cars while maintaining an average speed of 14 miles per hour. The vehicle from Stanford came in second, and the one from Virginia Tech came in third. Teams from MIT, Cornell, and the University of Pennsylvania also successfully completed the course out of a total of 11 teams that qualified for the final race. VL 1

The feat of navigating through the environment, especially one that contains moving obstacles, is extremely impressive. However, even though these robotic vehicles can avoid obstacles along a defined pathway, they can't identify most of the objects they are avoiding. For example, even though "Boss" might be able to avoid an obstacle in the middle of the road, it can't tell whether the obstacle is "a pile of rocks" or "a bush."

Other computer-based machines have been designed specifically to recognize objects (as opposed to navigating a course). These machines can recognize some objects, but only after training on a limited set of objects. The machines can recognize faces, but only if the lighting is just right and the faces are viewed from a specific angle. The difficulty of computer face recognition is illustrated by the fact that systems designed to recognize faces at airport security checkpoints can accurately identify less than half of a group of specially selected faces (Sinha, 2002; also see Chella et al., 2000, and "If You Want to Know More," page 128, for more on computer perception).

Tartan Racing and Carnegie Mellon University

Figure 5.2 ▌ The "Boss" robotic vehicle on a test run on a track at Robot City in Pittsburgh. Notice that there is no human driver. Navigation is accomplished by onboard computers that receive information from numerous sensors on the vehicle, each of which has a specialized task. Sensors mounted on the back of the roof are laser range scanners that point down to detect lane markings. Sensors on the roof rack point down crossroads to detect and track vehicles when attempting to merge with traffic. The black sensors on the hood at the front of the vehicle are multiplane, long-range laser scanners used for tracking vehicles. The two white sensors on the corners of the front bumper are short-range laser scanners used to detect and track nearby vehicles. The four rectangles in the grill are radar sensors. The white sensors are short-range, used for detecting obstacles near the vehicle. The black sensors are long-range, for tracking vehicles when Boss is moving quickly or considering turning across traffic.

Why Is It So Difficult to Design a Perceiving Machine?

We will now describe a few of the difficulties involved in designing a perceiving machine. Remember that the point of these descriptions is that although they pose difficulties for computers, our human "perceiving machine" solves these problems easily.

The Stimulus on the Receptors Is Ambiguous

When you look at the page of this book, the image cast by the page on your retina is ambiguous. It may seem strange to say that, because it is obvious that the page is rectangular, but consider Figure 5.3, which shows how the page is imaged on your retina. Viewed from straight on, the rectangular page creates a rectangular image on the retina. However, other objects, such as the tilted rectangle or slanted trapezoid, can also create the same image.

The fact that a particular image on the retina (or a computer vision machine's sensors) can be created by many different objects is called the **inverse projection problem**. Another way to state this problem is as follows: If we know an object's shape, distance, and orientation, we can determine the shape of the object's image on the retina. However, a particular image on the retina can be created by an infinite number of objects.

The ambiguity of the image on the retina is also illustrated by Figure 5.4a, which appears to be a circle of rocks. However, looking at these rocks from another viewpoint

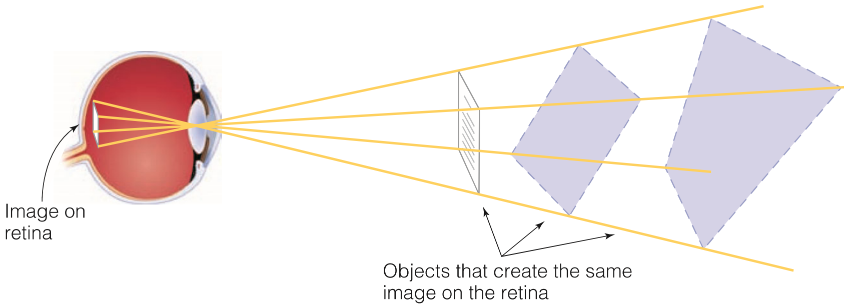


Figure 5.3 ▮ The principle behind the inverse projection problem. The page of the book that is near the eye creates a rectangular image on the retina. However, this image could also have been created by the tilted square, by the trapezoid and by many other stimuli. This is why we say that the image on the retina is ambiguous.

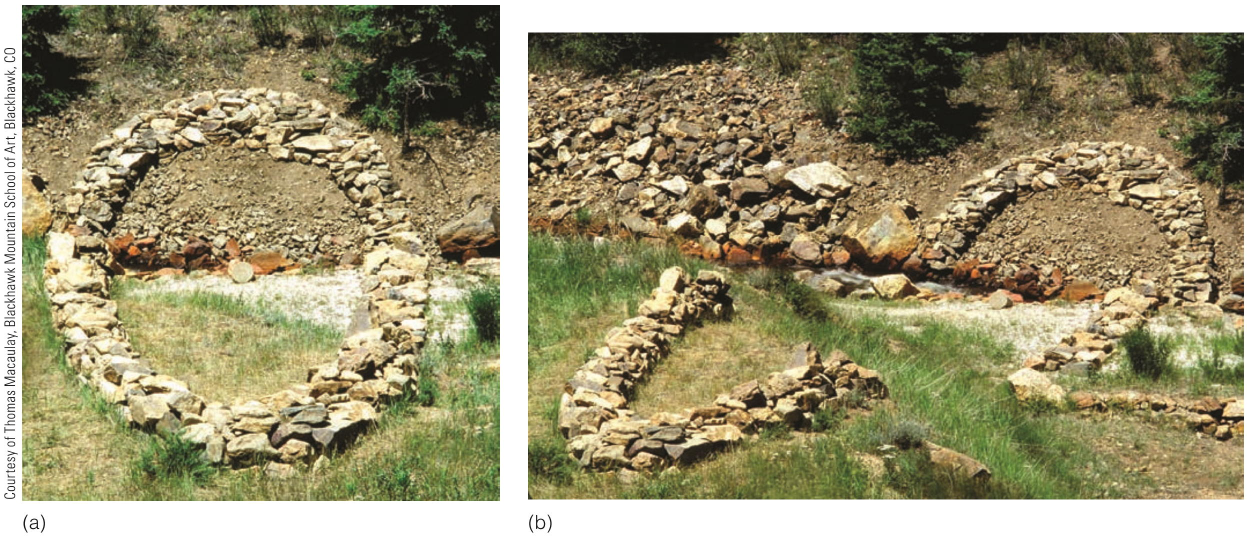


Figure 5.4 ▮ An environmental sculpture by Thomas Macaulay. (a) When viewed from exactly the right vantage point (the second-floor balcony of the Blackhawk Mountain School of Art, Black Hawk, Colorado), the stones appear to be arranged in a circle. (b) Viewing the stones from the ground floor reveals a truer indication of their configuration.

reveals that they aren't arranged in a circle after all (Figure 5.4b). Thus, just as a rectangular image on the retina can be created by trapezoid and other nonrectangular objects, a circular image on the retina can be created by objects that aren't circular. Although the example in Figure 5.4a leads human perceivers to the wrong conclusion about the rocks, this kind of confusion rarely occurs, because moving to another viewpoint reveals that the rocks aren't arranged in a circle.

These examples show that the information from a single view of an object can be ambiguous. Humans solve this problem by moving to different viewpoints, and by making use of knowledge they have gained from past experiences in perceiving objects.

Objects Can Be Hidden or Blurred

Sometimes objects are hidden or blurred. Can you find the pencil and eyeglasses in Figure 5.5? Although it might take a little searching, people can find the pencil in the foreground, and the glasses frame sticking out from behind the computer next to the scissors, even though only a small portion of these objects is visible. People also easily perceive the book, scissors, and paper as single objects, even though they are partially hidden by other objects.

This problem of hidden objects occurs any time one object obscures part of another object. This occurs frequently in the environment, but people easily understand that the part of an object that is covered continues to exist, and they are able to use their knowledge of the environment to determine what is likely to be present.

People are also able to recognize objects that are not in sharp focus, such as the faces in Figure 5.6. See how many of these people you can identify, and then consult the answers on page 130. Despite the degraded nature of these images, people can often identify most of them, whereas computers perform poorly on this task (Sinha, 2002).

Objects Look Different From Different Viewpoints

Another problem facing any perception machine is that objects are often viewed from different angles. This means that the images of objects are continually changing, de-

Bruce Goldstein

Figure 5.5 ▌ A portion of the mess on the author's desk. Can you locate the hidden pencil (easy) and the author's glasses (hard)?

Figure 5.6 ▌ Who are these people? See bottom of page 130 for the answers. *(From Sinha, P. (2002). Recognizing complex patterns.* Nature Neuroscience, 5, *1093–1097. Reprinted by permission from Macmillan Publishers Ltd. Copyright 2002.)*

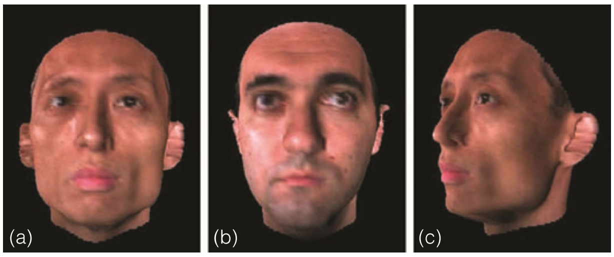


Figure 5.8 ▌ Which photographs are of the same person? *From Sinha, P. (2002). (Recognizing complex patterns.* Nature Neuroscience, 5, *1093–1097. Reprinted by permission from Macmillan Publishers Ltd. Copyright 2002.)*

pending on the angle from which they are viewed. Although humans continue to perceive the object in Figure 5.7 as the same chair viewed from different angles, this isn't so obvious to a computer. The ability to recognize an object seen from different viewpoints is called **viewpoint invariance.** People's ability to achieve viewpoint invariance enables them to identify the images in Figure 5.8a and c as being the same person, but a computer face recognition system would rate faces a and b as being more similar (Sinha, 2002).

The difficulties facing any perceiving machine illustrate that perception is more complex than it seems. But how do humans overcome these complexities? Early answers to this question were provided in the early 1900s by a group of psychologists who called themselves **Gestalt psychologists**—where *Gestalt,* roughly translated, means a whole configuration that cannot be described merely as the sum of its parts. We can appreciate the meaning of this definition by considering how Gestalt psychology began.

Bruce Goldstein

(a) (b) (c)

Figure 5.7 ▌ Your ability to recognize each of these views as being of the same chair is an example of viewpoint invariance.

The Gestalt Approach to Object Perception

We can understand the Gestalt approach by first considering an early attempt to explain perception that was proposed by Wilhelm Wundt, who established the first laboratory of scientific psychology at the University of Leipzig in 1879. Wundt's approach to psychology was called **structuralism**. One of the basic ideas behind structuralism was that perceptions are created by combining elements called **sensations**, just as each of the dots in the face in Figure 5.9 add together to create our perception of a face.

The idea that perception is the result of "adding up" sensations was disputed by the Gestalt psychologists, who offered, instead, the idea that *the whole differs from the sum of its parts.* This principle had its beginnings, according to a well-known story, in a train ride taken by psychologist Max Wertheimer in 1911 (Boring, 1942). Wertheimer got off the train to stretch his legs in Frankfurt and bought a toy stroboscope from a vendor who was selling toys on the train platform. The stroboscope, a mechanical device that created an illusion of movement by rapidly alternating two slightly different pictures, caused Wertheimer to wonder how the structuralist idea that experience is created from sensations could explain the illusion of movement he observed. We can understand why this question arose by looking at Figure 5.10a, which diagrams the principle behind the illusion of movement created by the stroboscope.

When two stimuli that are in slightly different positions are flashed one after another with the correct timing, movement is perceived between the two stimuli. This is an illusion called **apparent movement** because there is actually no movement in the display, just two stationary stimuli flashing on and off. How, wondered Wertheimer, can the movement that appears to occur between the two flashing stimuli be caused by sensations? After all, there is no stimulation in the space between the two stimuli, and therefore there are no sensations to provide an explanation for the movement. (A modern example of apparent movement is provided by electronic signs like the one in Figure 5.10b, which display moving advertisements or news headlines. The perception of movement in these displays is so compelling that it is difficult to imagine that they are made up of stationary lights flashing on and off.) VL 2

With his question about apparent movement as his inspiration, Wertheimer and two colleagues, Kurt Koffka and Ivo Kohler, set up a laboratory at the University of Frankfurt, called themselves Gestalt psychologists, and proceeded to do research and publish papers that posed serious problems for the structuralist idea that perceptions are created from sensations (Wertheimer, 1912). The following demonstration illustrates another phenomenon that is difficult to explain on the basis of sensations.

Figure 5.9 ▮ According to structuralism, a number of sensations (represented by the dots) add up to create our perception of the face.

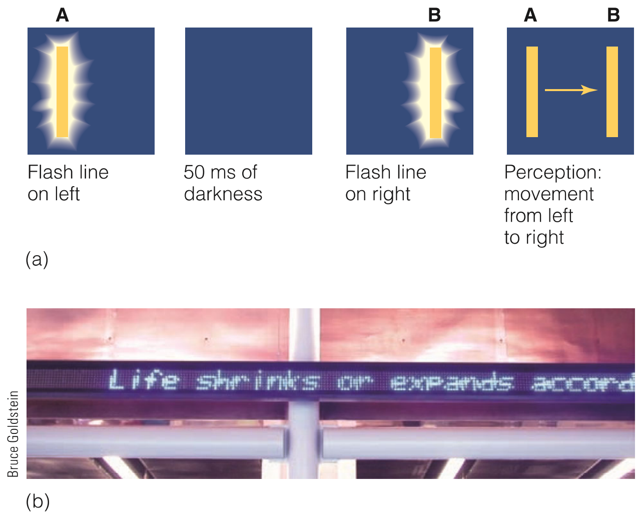
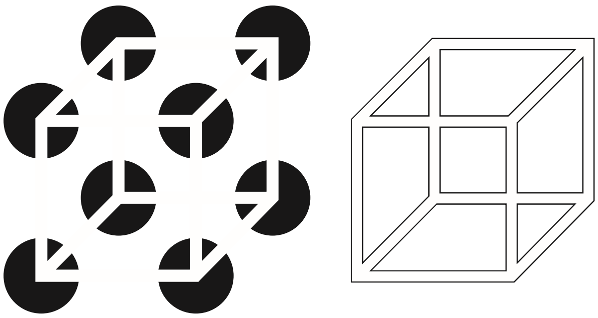


Figure 5.10 ▮ (a) Wertheimer's demonstration of apparent movement. (b) Moving electronic signs such as this one, in which the words are scrolling to the left, create the perception of movement by applying the principles of apparent movement studied by Wertheimer.

DEMONSTRATION

Making Illusory Contours Vanish

Consider the picture in Figure 5.11. If you see this as a cube like the one in Figure 5.11b floating in space in front of black circles, you probably perceive faint **illusory contours** that represent the edges of the cube (Bradley & Petry, 1977). These contours are called illusory because they aren't actually present in the physical stimulus. You can prove this to yourself by (1) placing your finger over the two black circles at the bottom or (2) imagining that the black circles are holes and that you are looking at the cube through these holes. Covering the circles or seeing the cube through the holes causes the illusory contours to either vanish or become more difficult to *see.* ▮



(a) (b)

Figure 5.11 ▌ (a) This can be seen as a cube floating in front of eight discs or as a cube seen through eight holes. In the first case, the edges of the cube appear as illusory contours. (b) The cube without the black circles. *(Based on "Organizational Determinants of Subjective Contour: The Subjective Necker Cube," by D. R. Bradley and H. M. Petry, 1977,* American Journal of Psychology, 90*, 253–262. American Psychological Association.)*

When you made the contours vanish by placing your finger over the black circles, you showed that the contour was illusory and that our perception of one part of the display (the contours) is affected by the presence of another part (the black circles). The structuralists would have a hard time explaining illusory contours because there is no actual contour, so there can't be any sensations where the contour is perceived. **VL 3, 4**

Additional displays that are difficult to explain in terms of sensations are bistable figures, like the cube in Figure 5.11b, which switch back and forth as they are viewed, and illusions, in which perceptions of one part of a display are affected by another part. (See Virtual Labs 5–7.) Making the contours vanish by imagining that you are looking through black holes poses a similar problem for the structuralists. It is difficult to explain a perception that is present one moment and gone the next in terms of sensations, especially since the stimulus on your retina never changes. **VL 5–7**

Having rejected the idea that perception is built up of sensations, the Gestalt psychologists proposed a number of principles, which they called laws of perceptual organization.

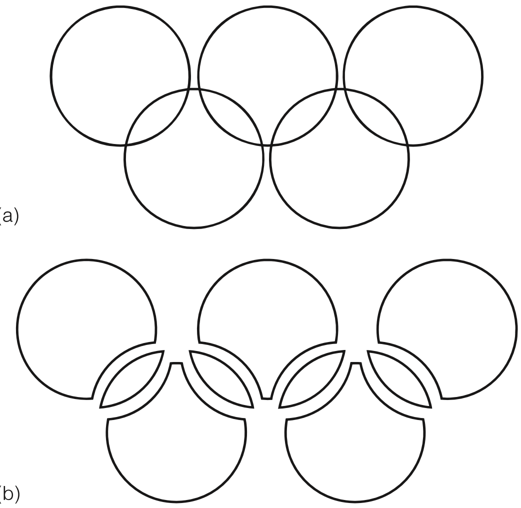
The Gestalt Laws of Perceptual Organization

Perceptual organization involves the grouping of elements in an image to create larger objects. For example, some of the dark areas in Figure 5.12 become grouped to form a Dalmation and others are seen as shadows in the background. Here are six of the laws of organization that the Gestalt psychologists proposed to explain how perceptual grouping such as this occurs. **VL 8**

Pragnanz *Pragnanz,* roughly translated from the German, means "good figure." The law of pragnanz, also called the law of good figure or the law of simplicity, is the central law of Gestalt psychology: *Every stimulus pattern is seen in such a way that the resulting structure is as simple as possible.* The familiar Olympic symbol in Figure 5.13a is an example of the law of simplicity at work. We see this display as five circles and not as a larger number of more complicated shapes such as the ones in Figure 5.13b. **VL 9**

Similarity Most people perceive Figure 5.14a as either horizontal rows of circles, vertical columns of circles, or both. But when we change the color of some of the columns, as in Figure 5.14b, most people perceive vertical columns of circles. This perception illustrates the law of similarity: *Similar things appear to be grouped together.* This law causes circles of the same color to be grouped together. Grouping

Image not available due to copyright restrictions



(a) (b)

Figure 5.13 ▌ (a) This is usually perceived as five circles, not as the nine shapes in (b).

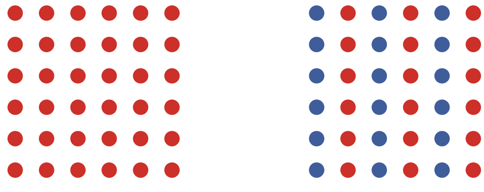


Figure 5.14 ▮ (a) Perceived as horizontal rows or vertical columns or both. (b) Perceived as vertical columns.

can also occur because of similarity of shape, size, or orientation (Figure 5.15). VL 10

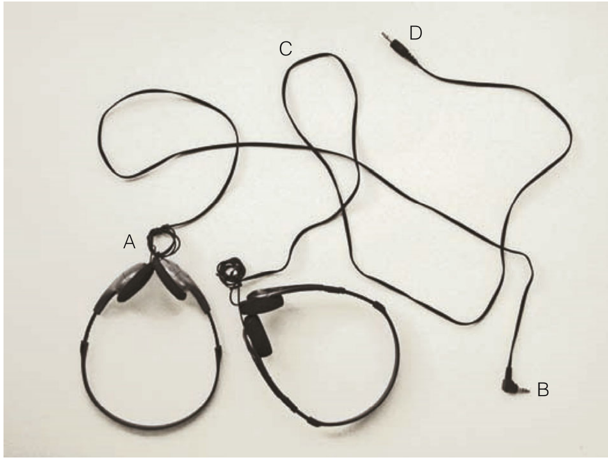
Grouping also occurs for auditory stimuli. For example, notes that have similar pitches and that follow each other closely in time can become perceptually grouped to form a melody. We will consider this and other auditory grouping effects when we describe organizational processes in hearing in Chapter 12.

Good Continuation In Figure 5.16 we see the wire starting at A as flowing smoothly to B. It does not go to C or D because those paths would involve making sharp turns and would violate the law of good continuation: *Points that, when connected, result in straight or smoothly curving lines are seen as belonging together, and the lines tend to be seen in such a way as to follow the smoothest path.* Another effect of good continuation is shown in the Celtic knot pattern in Figure 5.17. In this case, good continuation assures that we see a continuous interweaved pattern that does not appear to be broken into little pieces every time one strand overlaps another strand. Good continuation also helped us to perceive the smoothly curving circles in Figure 5.13a. VL 11, 12

Proximity (Nearness) Our perception of Figure 5.18a as two pairs of circles illustrates the law of proximity, or nearness: *Things that are near each other appear to be grouped together.* VL 13

Donald Miralie/Getty Images

Figure 5.15 ▮ What are they looking at? Whatever it is, Tiger Woods and Phil Mickelson have become perceptually linked because of the similar orientations of their arms, golf clubs, and bodies.



Bruce Goldstein

Figure 5.16 ▮ Good continuation helps us perceive two separate wires, even though they overlap.

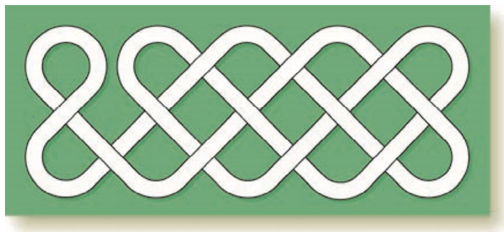


Figure 5.17 ▮ Because of good continuation, we perceive this pattern as continuous interwoven strands.

Common Region Figure 5.18b illustrates the principle of common region: *Elements that are within the same region of space appear to be grouped together.* Even though the circles inside the ovals are farther apart than the circles that are next to each other in neighboring ovals, we see the circles inside the ovals as belonging together. This occurs because each oval is seen as a separate region of space (Palmer, 1992; Palmer & Rock, 1994). Notice that in this example common region overpowers proximity. Because the circles are in different regions, they do not group with each other, as they did in Figure 5.18a, but with circles in the same region.

Uniform Connectedness The principle of uniform connectedness states: *A connected region of visual properties, such as lightness, color, texture, or motion, is perceived as a single unit.* For example, in Figure 5.18c, the connected circles are perceived as grouped together, just as they were when they were in the same region in Figure 5.18b.

Synchrony The principle of synchrony states: *Visual events that occur at the same time are perceived as belonging together.* For example, the lights in Figure 5.18d that blink together are seen as belonging together.

Common Fate The law of common fate states: *Things that are moving in the same direction appear to be grouped together.* Thus, when you see a flock of hundreds of birds all flying

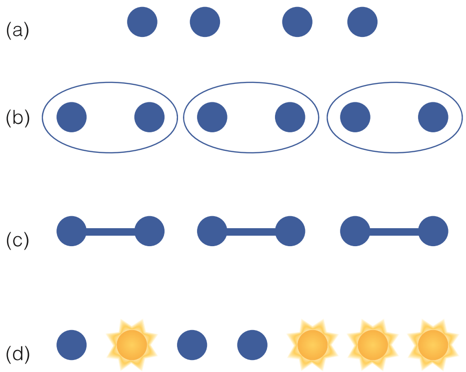


Figure 5.18 ▌ Grouping by (a) proximity; (b) common region; (c) connectedness; and (d) synchrony. Synchrony occurs when the yellow lights blink on and off together.

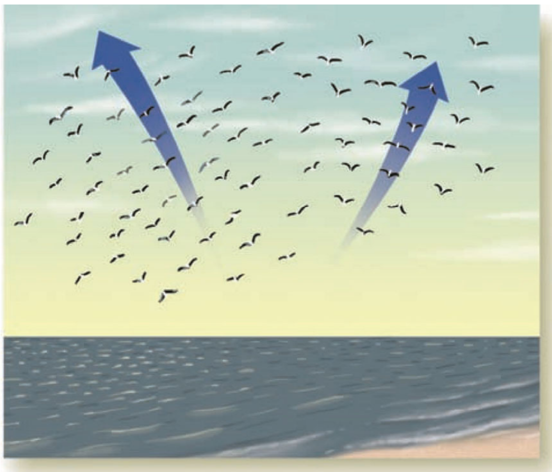


Figure 5.19 ▌ A flock of birds that are moving in the same direction are seen as grouped together. When a portion of the flock changes direction, their movement creates a new group. This illustrates the law of common fate.

together, you tend to see the flock as a unit, and if some birds start flying in another direction, this creates a new unit (Figure 5.19). Notice that common fate is like synchrony in that both principles are dynamic, but synchrony can occur without movement, and the elements don't have to change in the same direction as they do in common fate. VL 14

Meaningfulness or Familiarity According to the law of familiarity, *things that form patterns that are familiar or meaningful are likely to become grouped together* (Helson, 1933; Hochberg, 1971). You can appreciate how meaningfulness influences perceptual organization by doing the following demonstration.

DEMONSTRATION

Finding Faces in a Landscape

Consider the picture in Figure 5.20. At first glance this scene appears to contain mainly trees, rocks, and water. But on closer inspection you can see some faces in the trees in the background, and if you look more closely, you can see that a

Figure 5.20 ▌ *The Forest Has Eyes* by Bev Doolittle (1984). Can you find 13 hidden faces in this picture? E-mail the author at bruceg@email.arizona.edu for the solution.

number of faces are formed by various groups of rocks. See if you can find all 13 faces hidden in this picture.

Some people find it difficult to perceive the faces at first, but then suddenly they succeed. The change in perception from "rocks in a stream" or "trees in a forest" to "faces" is a change in the perceptual organization of the rocks and the trees. The two shapes that you at first perceive as two separate rocks in the stream become perceptually grouped together when they become the left and right eyes of a face. In fact, once you perceive a particular grouping of rocks as a face, it is often difficult *not* to perceive them in this way—they have become permanently organized into a face. This is similar to the process we observed for the Dalmatian. Once we see the Dalmatian, it is difficult not to perceive it.

Perceptual Segregation: How Objects Are Separated From the Background

The Gestalt psychologists were also interested in explaining **perceptual segregation**, the perceptual separation of one object from another, as Roger did when he perceived each of the buildings in Figure 5.1 as separate from one another. The question of what causes perceptual segregation is often referred to as the problem of **figure–ground segregation**. When we see a separate object, it is usually seen as a **figure** that stands out from its background, which is called the **ground**. For example, you would probably see a book or papers on your desk as figure and the surface of your desk as ground. The Gestalt psychologists were interested in determining the properties of the figure and the ground and what causes us to perceive one area as figure and the other as ground.

What Are the Properties of Figure and Ground? One way the Gestalt psychologists studied the properties of figure and ground was by considering patterns like the one in Figure 5.21, which was introduced by Danish psychologist Edgar Rubin in 1915. This pattern is an example of **reversible figure–ground** because it can be perceived alternately either as two blue faces looking at each other, in front of a white background, or as a white vase on a blue background. Some of the properties of the figure and ground are:

- The figure is more "thinglike" and more memorable than the ground. Thus, when you see the vase as figure, it appears as an object that can be remembered later. However, when you see the same white area as ground, it does not appear to be an object and is therefore not particularly memorable. VL 15
- The figure is seen as being in front of the ground. Thus, when the vase is seen as figure, it appears to be in front of the dark background (Figure 5.22a), and when the faces are seen as figure, they are on top of the light background (Figure 5.22b).

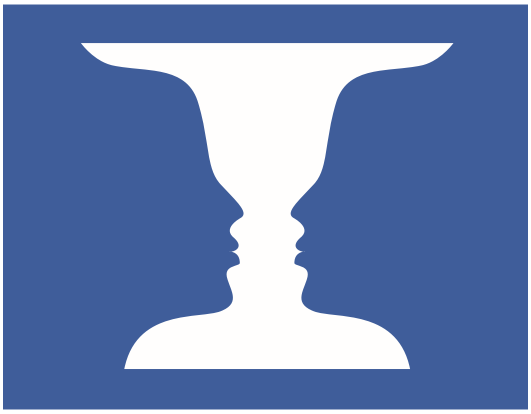


Figure 5.21 A version of Rubin's reversible face–vase figure.

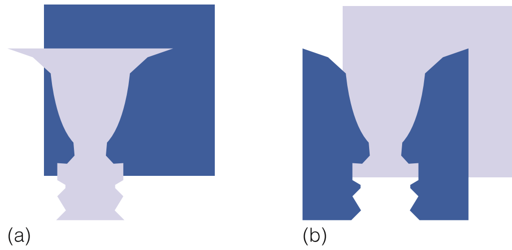


Figure 5.22 (a) When the vase is perceived as figure, it is seen in front of a homogeneous dark background. (b) When the faces are seen as figure, they are seen in front of a homogeneous light background.

- The ground is seen as unformed material and seems to extend behind the figure.
- The contour separating the figure from the ground appears to belong to the figure. This property of figure, which is called **border ownership**, means that, although figure and ground share a contour, the border is associated with the figure. Figure 5.23 illustrates border ownership for another display that can be perceived in two ways. If you perceive the display in Figure 5.23a as a light gray square (the figure) sitting on a dark background (the ground), then the border belongs to the gray square, as indicated by the dot in Figure 5.23b. But if you perceive the display as a black rectangle with a hole in it (the figure) through which you are viewing a gray surface (the ground), the border would be on the black rectangle, as shown in Figure 5.23c.

What Factors Determine Which Area Is Figure? What factors determine whether an area is perceived as figure or ground? Shaun Vecera and coworkers (2002) used the phenomenological method (see page 13) to show that regions in the lower part of a display are more likely to be perceived as figure than regions in the upper

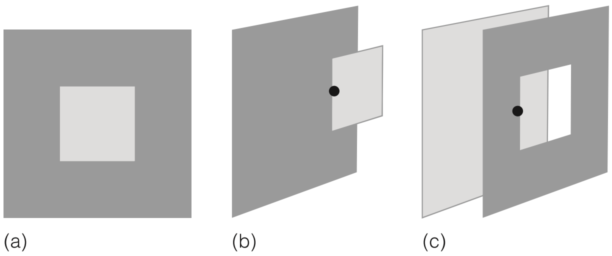


Figure 5.23 ▌ (a) This display can be perceived in two ways. (b) When it is perceived as a small square sitting on top of a dark background, the border belongs to the small square, as indicated by the dot. (c) When it is perceived as a large dark square with a hole in it, the border belongs to the dark square.

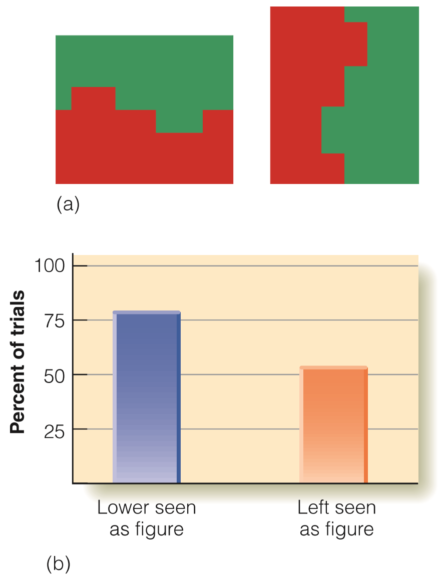


Figure 5.24 ▌ (a) Stimuli from Vecerra et al. (2002). (b) Percentage of trials on which lower or left areas were seen as figure.

part. They flashed stimuli like the ones in Figure 5.24a for 150 milliseconds (ms) and asked observers to indicate which area they saw as figure, the red area or the green area. The results, shown in Figure 5.24b, indicate that for the upper–lower displays, observers were more likely to perceive the lower area as figure, but for the left–right displays, they showed only a small preference for the left region. From this result, Vecera concluded that there is no left–right preference for determining figure, but there is a definite preference for seeing objects lower in the display as figure. The conclusion from this experiment is that the lower region of a display tends to be seen as figure.

Figure 5.25 illustrates four other factors that help determine which area will be seen as figure. In Figure 5.25a (symmetry), the symmetrical red areas on the left are seen as figure, as are the symmetrical yellow areas on the right.

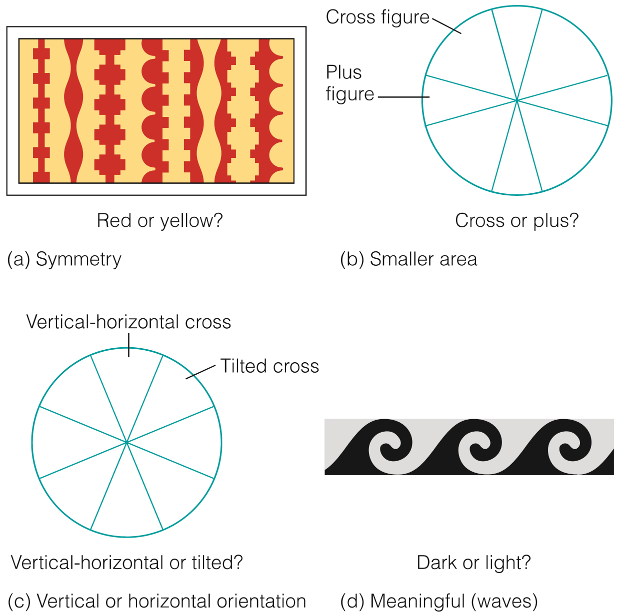


Figure 5.25 ▌ Examples of how (a) symmetry, (b) size, (c) orientation, and (d) meaning contribute to perceiving an area as figure.

In Figure 5.25b (smaller area), the smaller plus-shaped area is more likely to be seen as figure. In Figure 5.25c (vertical or horizontal areas), the vertical–horizontal cross tends to be seen as figure. In Figure 5.25d (meaningfulness), the fact that the dark areas look like waves increases the chances that this area will be seen as figure. VL 16

The Gestalt "Laws" as Heuristics

Although the Gestalt psychologists called their principles "laws of perceptual organization," most perceptual psychologists call them the Gestalt "principles" or "heuristics." The reason for rejecting the term *laws* is that the rules of perceptual organization and segregation proposed by the Gestalt psychologists don't make strong enough predictions to qualify as laws. Instead, the Gestalt principles are more accurately described as **heuristics**—rules of thumb that provide a best-guess solution to a problem. We can understand what heuristics are by comparing them to another way of solving a problem, called algorithms.

An **algorithm** is a procedure that is *guaranteed* to solve a problem. An example of an algorithm is the procedures we learn for addition, subtraction, and long division. If we apply these procedures correctly, we get the right answer every time. In contrast, a heuristic may not result in a correct solution every time. For example, suppose that you want to find a cat that is hiding somewhere in the house. An algorithm for doing this would be to systematically search every room in the house (being careful not to let the cat sneak past you!). If you do this, you will eventually find the cat, although it may take a while. A heuristic for finding the cat would be to first look in the places where the cat likes to hide. So you check under the bed and in the hall closet. This may not always lead to finding the cat, but if it does, it has the advantage of usually being faster than the algorithm.

We say the Gestalt principles are heuristics because, like heuristics, they are best-guess rules that work most of the time, but not necessarily all of the time. For example, consider the following situation in which the Gestalt laws might cause an incorrect perception: As you are hiking in the woods, you stop cold in your tracks because not too far ahead, you see what appears to be an animal lurking behind a tree (Figure 5.26a). The Gestalt laws of organization play a role in creating this perception. You see the two shapes to the left and right of the tree as a single object because of the Gestalt law of similarity (because both shapes are the same color, it is likely that they are part of the same object). Also, good continuation links these two parts into one because the line along the top of the object extends smoothly from one side of the tree to the other. Finally, the image resembles animals you've seen before. For all of these reasons, it is not surprising that you perceive the two objects as part of one animal.

Because you fear that the animal might be dangerous, you take a different path. As your detour takes you around the tree, you notice that the dark shapes aren't an animal after all, but are two oddly shaped tree stumps (Figure 5.26b). So in this case, the Gestalt laws have misled you.

The fact that heuristics are usually faster than algorithms helps explain why the perceptual system is designed to operate in a way that sometimes produces errors. Consider, for example, what the algorithm would be for determining what the shape in Figure 5.26a really is. It would involve walking around the tree, so you can see it from different angles and perhaps taking a closer look at the objects behind the tree. Although this may result in an accurate perception, it is slow and potentially risky (what if the shape actually *is* a dangerous animal?).

The advantage of our Gestalt-based rules of thumb is that they are fast, and correct most of the time. The reason they work most of the time is that they reflect properties of the environment. For example, in everyday life, objects that are partially hidden often "come out the other side" (good continuation), and objects often have similar large areas of the same color (similarity). We will return to the idea that perception depends on what we know about properties of the environment later in the chapter.

Although the Gestalt approach dates back to the early 1900s, it is still considered an important way to think about perception. Modern researchers have done experiments like Vecera's (Figure 5.24) to study some of the principles of perceptual organization and segregation proposed by the Gestalt psychologists, and they have also considered issues in addition to organization and segregation. We will now describe a more recent approach to object perception called *recognition by components* that is designed to explain how we recognize objects.

Recognition-by-Components Theory

How do we recognize objects in the environment based on the image on the retina? **Recognition-by-components (RBC) theory**, which was proposed by Irving Biederman (1987), answers this question by proposing that our recognition of objects is based on features called **geons**, a term that stands for "geometric ions," because just as ions are basic units of molecules (see page 29), these geons are basic units of objects. Figure 5.27a shows a number of geons, which are shapes such as cylinders, rectangular solids, and pyramids. Biederman proposed 36 different geons and suggested that this number of geons is enough to enable us to mentally represent a large proportion of the objects that we can easily recognize. Figure 5.27b shows a few objects that have been constructed from geons.

To understand geons, we need to introduce the concept of **non-accidental properties (NAPs)**. NAPs are properties of *edges* in the retinal image that correspond to the

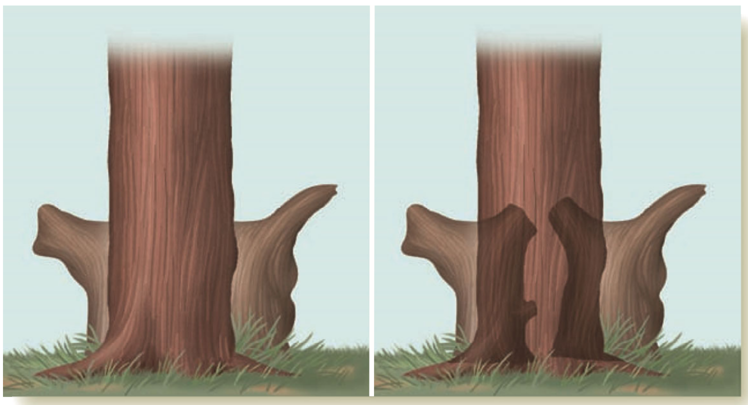

(a) (b)

Figure 5.26 (a) What lurks behind the tree? (b) It is two strangely shaped tree stumps, not an animal!

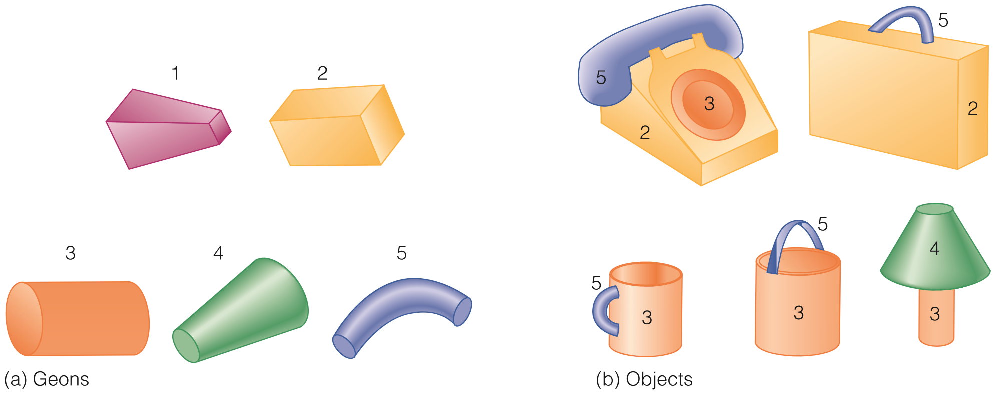


Figure 5.27 ▮ (a) Some geons. (b) Some objects created from these geons. The numbers on the objects indicate which geons are present. Note that recognizable objects can be formed by combining just two or three geons. Also note that the relations between the geons matter, as illustrated by the cup and the pail. *(Reprinted from "Recognition-by-Components: A Theory of Human Image Understanding," by I. Biederman, 1985,* Computer Vision, Graphics and Image Processing, 32, *29–73. Copyright © 1985, with permission from Elsevier.)*

properties of *edges* in the three-dimensional environment. The following demonstration illustrates this characteristic of NAPs.

DEMONSTRATION

Non-Accidental Properties

Close one eye and look at a coin, such as a quarter, straight on, so your line of sight is perpendicular to the quarter, as shown in Figure 5.28a. When you do this, the edge of the quarter creates a curved image on the retina. Now tilt the quarter, as in Figure 5.28b. The edge of this tilted quarter still creates an image of a curved edge on the retina. Now tilt the quarter so you are viewing it edge-on, as in Figure 5.28c. When viewed in this way, the edge of the quarter creates an image of a straight edge on the retina. ▮

In this demonstration, the property of *curvature* is called a *non-accidental property,* because the only time it doesn't occur is when you view the quarter edge-on. Because this edge-on viewpoint occurs only rarely, it is called an **accidental viewpoint.** Thus, the vast majority of your views of circular objects result in a curved image on the retina. According to RBC, the image of a curved edge on the retina indicates the presence of a curved edge in the environment.

RBC proposes that a key property of geons is that each type of geon has a unique set of NAPs. For example, consider the rectangular-solid geon in Figure 5.29a. The NAP for this geon is three parallel straight edges. You can demonstrate the fact that these edges are NAPs by viewing a rectangular solid (such as a book) from different angles, as shown in Figure 5.30. When you do this, you will notice that most

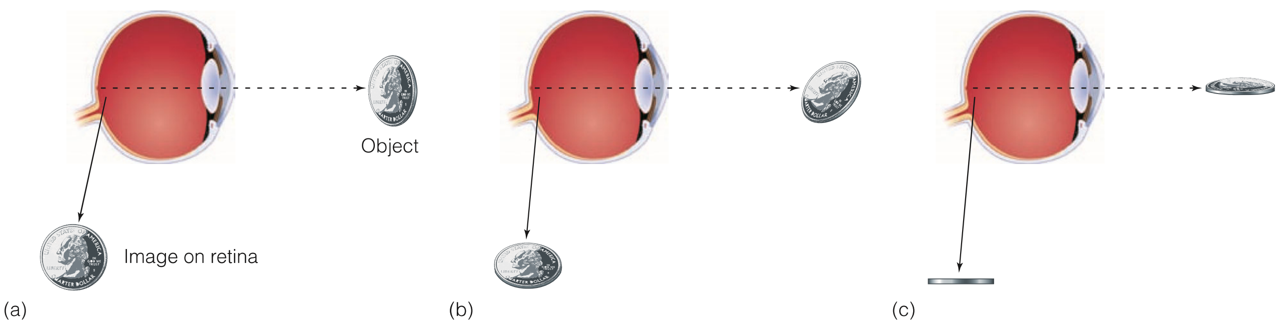
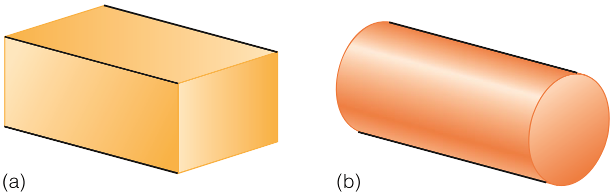


Figure 5.28 ▮ What happens to a quarter's image on the retina as it is tilted. Most views, such as (a) and (b), create a curved image on the retina. The rare accidental viewpoint shown in (c) creates an image of a straight line on the retina.



(a) (b)

Figure 5.29 ▌ (a) Rectangular-solid geon. The highlighted three parallel edges are the non-accidental property for this geon. (b) Cylindrical geon. The highlighted two parallel edges are the non-accidental property of this geon.

Bruce Goldstein

(a) (b) (c)

Figure 5.30 ▌ This book's non-accidental property (NAP) of three parallel edges are seen even when the book is viewed from different angles, as in (a) and (b). When viewed from an accidental viewpoint, as in (c), this NAP is not perceived.

of the time you can see three parallel straight edges, as in Figures 5.30a and b. Figure 5.30c shows what happens when you view the book from an accidental viewpoint. The three parallel edges are not visible from this viewpoint, just as the quarter's curvature was not visible when it was viewed from an accidental viewpoint.

The NAP for the cylinder geon in Figure 5.29b is two parallel straight edges, which you see as you view a cylindrical object such as a pencil or pen from different angles. Like the rectangular geon, the cylindrical geon has an accidental viewpoint from which the NAP is not visible (what is the accidental viewpoint for the cylinder?).

The fact that each geon has a unique set of NAPs results in a property of geons called **discriminability**—each geon can be discriminated from other geons. The fact that NAPs are visible from most viewpoints results in another property of geons, *viewpoint invariance* (see page 103)—the geon can be identified when viewed from most viewpoints.

The main principle of recognition-by-components theory is that if we can perceive an object's geons, we can identify the object (also see Biederman & Cooper, 1991; Biederman, 1995). The ability to identify an object if we can identify its geons is called the **principle of componential recovery**. This principle is what is behind our ability to identify objects in the natural environment even when parts of the objects are hidden by other objects. Figure 5.31a shows a situation in which componential recovery can't occur because the visual noise is arranged so that the object's geons cannot be identified. Luckily, parts of objects are rarely obscured in this way in the natural environment, so, as we see in Figure 5.31b, we can usually identify geons and, therefore, are able to identify the object.

Another illustration of the fact that our ability to identify objects depends on our ability to identify the object's geons is shown by the tea kettle in Figure 5.32a. When we view it from the unusual perspective shown in Figure 5.32b, we can't identify some of its basic geons, and it is therefore more difficult to identify in Figure 5.32b than in Figure 5.32a.

RBC theory also states that we can recognize objects based on a relatively small number of geons. Biederman (1987) did an experiment to demonstrate this, by briefly presenting line drawings of objects with all of their geons and with some geons missing. For example, the airplane in Figure 5.33a, which has a total of 9 geons, is shown with only 3 of its geons in Figure 5.33b. Biederman found that

(a) (b)

Figure 5.31 ▌ (a) It is difficult to identify the object behind the mask, because its geons have been obscured. (b) Now that it is possible to identify geons, the object can be identified as a flashlight. *(Reprinted from "Recognition-by-Components: A Theory of Human Image Understanding," by I. Biederman, 1985,* Computer Vision, Graphics and Image Processing, 32, *29–73. Copyright © 1985, with permission from Elsevier.)*

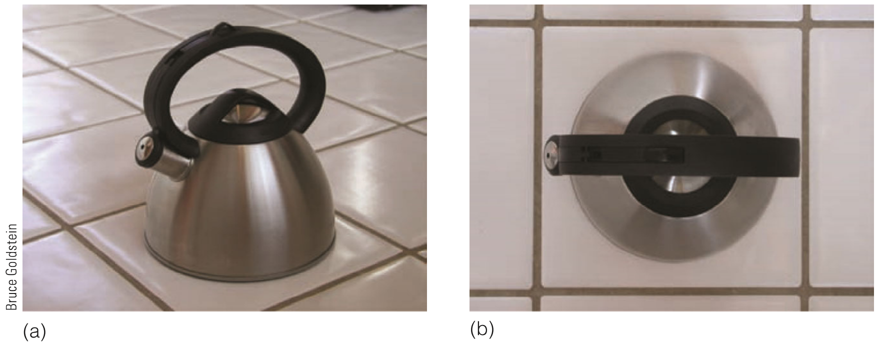


Figure 5.32 ▮ (a) A familiar object. (b) The same object seen from a viewpoint that obscures most of its geons. This makes it harder to recognize the object.

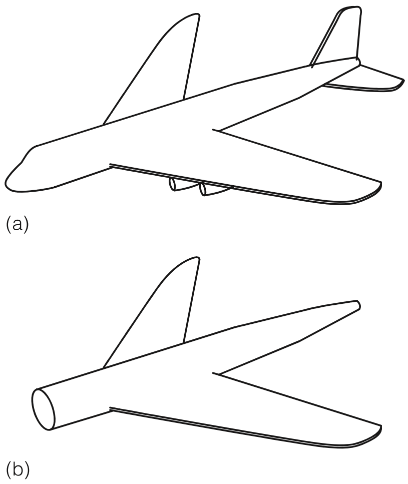


Figure 5.33 ▮ An airplane, as represented (a) by 9 geons and (b) by 3 geons. *(Reprinted from "Recognition-by-Components: A Theory of Human Image Understanding," by I. Biederman, 1985, Computer Vision,* Graphics and Image Processing, 32, *29–73. Copyright © 1985, with permission from Elsevier.)*

9-geon objects such as the airplane were recognized correctly about 78 percent of the time based on 3 geons and 96 percent of the time based on 6 geons. Objects with 6 geons were recognized correctly 92 percent of the time even when they were missing half their geons.

RBC theory explains many observations about shape-based object perception, but the idea that our perception of a complex object begins with the perception of features like geons is one that some students find difficult to accept. For example, one of my students who, having read the first four chapters of the book was apparently convinced that perception is a complex process, wrote, in reaction to reading about RBC theory, that "our vision is far too complex to be determined by a few geons."

This student's concern can be addressed in a few ways. First, there are factors in addition to geons that help us identify objects. For example, we might distinguish between two birds with the same shape on the basis of the texture of their feathers or markings on their wings. Similarly, there are some things in the environment, such as clouds, that are difficult to create using geons (although even clouds are sometime arranged so that geons are visible, leading us to see "objects" in the sky).

The fact that there are things that RBC can't explain is not surprising because the theory was not meant to explain everything about object perception. For example, although edges play an important role in RBC, the theory is not concerned with the rapid processes that enable us to perceive these edges. It also doesn't deal with the processes involved in grouping objects (which the Gestalt approach does) or with how we learn to recognize different types of objects.

RBC does, however, provide explanations for some important phenomena, such as view invariance and the minimum information needed to identify objects. Some of the most elegant scientific theories are simple and provide partial explanations, leaving other theories to complete the picture. This is the case for RBC.

TEST YOURSELF 5.1

1. What are some of the problems that make object perception difficult for computers but not for humans?
2. What is structuralism, and why did the Gestalt psychologists propose an alternative to this way of looking at perception?
3. How did the Gestalt psychologists explain perceptual organization?
4. How did the Gestalt psychologists describe figure–ground segregation?
5. What properties of a stimulus tend to favor perceiving an area as "figure"? Be sure you understand Vecera's experiment that showed that the lower region of a display tends to be perceived as figure.
6. How does RBC theory explain how we recognize objects? What are the properties of geons, and how do these properties enable us to identify objects from different viewpoints and identify objects that are partially hidden?

Perceiving Scenes and Objects in Scenes

So far we have been focusing on individual objects. But we rarely see objects in isolation. Just as we usually see actors in a play on a stage, we usually see objects within a scene (Epstein, 2005). A **scene** is a view of a real-world environment that contains (1) background elements and (2) multiple objects that are organized in a meaningful way relative to each other and the background (Epstein, 2005; Henderson & Hollingworth, 1999).

One way of distinguishing between objects and scenes is that objects are compact and are *acted upon,* whereas scenes are extended in space and are *acted within.* For example, if we are walking down the street and mail a letter, we would be *acting upon* the mailbox (an object) and *acting within* the street (the scene).

Perceiving the Gist of a Scene

Perceiving scenes presents a paradox. On one hand, scenes are often large and complex. However, despite this size and complexity, you can identify most scenes after viewing them for only a fraction of a second. This general description of the type of scene is called the **gist of a scene**. An example of your ability to rapidly perceive the gist of a scene is the way you can rapidly flip from one TV channel to another, yet still grasp the meaning of each picture as it flashes by—a car chase, quiz contestants, or an outdoor scene with mountains—even though you may be seeing each picture for a second or less. When you do this, you are perceiving the gist of each scene (Oliva & Torralba, 2006).

Research has shown that it is possible to perceive the gist of a scene within a fraction of a second. Mary Potter (1976) showed observers a target picture and then asked them to indicate whether they saw that picture as they viewed a sequence of 16 rapidly presented pictures. Her observers could do this with almost 100-percent accuracy even when the pictures were flashed for only 250 ms (milliseconds; 1/4 second). Even when the target picture was only specified by a description, such as "girl clapping," observers achieved an accuracy of almost 90 percent (Figure 5.34).

Another approach to determining how rapidly people can perceive scenes was used by Li Fei-Fei and coworkers (2007), who presented pictures of scenes for times ranging from 27 ms to 500 ms and asked observers to write a description of what they saw. This method of determining the observer's response is a nice example of the phenomenological method, described on page 13. Fei-Fei used a procedure called *masking* to be sure the observers saw the pictures for exactly the desired duration. VL 17

METHOD | Using a Mask to Achieve Brief Stimulus Presentations

To present a stimulus, such as a picture, for just 27 ms, we need to do more than just flash the picture for 27 ms, because the perception of any stimulus persists for about 250 ms after the stimulus is extinguished—a phenomenon called **persistence of vision**. Thus, a picture that is presented for 27 ms will be *perceived* as lasting about 275 ms. To eliminate the persistence of vision it is therefore necessary to flash a **masking stimulus**, usually a pattern of randomly oriented lines, immediately after presentation of the picture. This stops the persistence of vision and limits the time that the picture is perceived.

Typical results of Fei-Fei's experiment are shown in Figure 5.35. At brief durations, observers saw only light and dark areas of the pictures. By 67 ms they could identify some large objects (a person, a table), and when the duration was increased to 500 ms they were able to identify smaller objects and details (the boy, the laptop). For another picture, of an ornate 1800s living room, observers were able to identify the picture as a room in a house at 67 ms and to identify details, such as chairs and portraits, at 500 ms. Thus, the overall gist of the scene is perceived first, followed by perception of details and smaller objects within the scene.

What enables observers to perceive the gist of a scene so rapidly? Aude Oliva and Antonio Torralba (2001, 2006) propose that observers use information called **global image**

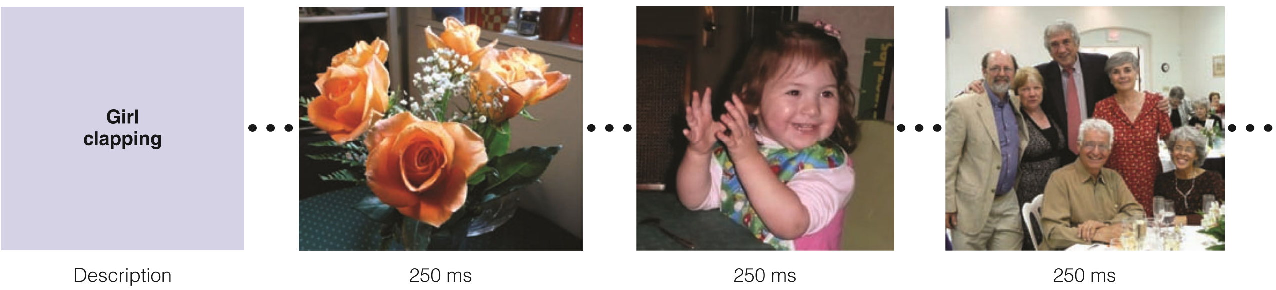


Figure 5.34 ▌ Procedure for Potter's (1976) experiment. She first presented either a target photograph or, as shown here, a description, and then rapidly presented 16 pictures for 250 ms each. The observer's task was to indicate whether the target picture had been presented. In this example, only 3 of the 16 pictures are shown, with the target picture being the second one presented. On other trials, the target picture is not included in the series of 16 pictures.

Image not available due to copyright restrictions

features, which can be perceived rapidly and are associated with specific types of scenes. Some of the global image features proposed by Oliva and Torralba are:

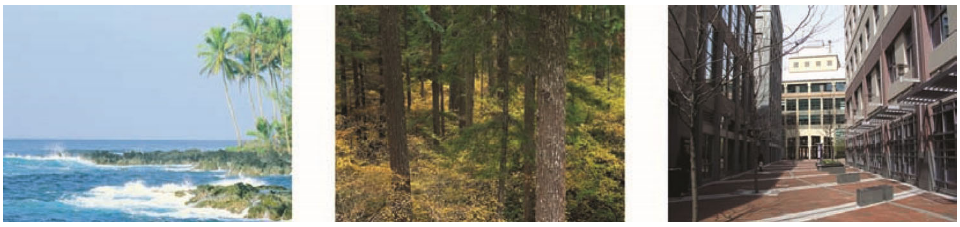
- *Degree of naturalness.* Natural scenes, such as the beach and forest in Figure 5.36, have textured zones and undulating contours. Man-made scenes, such as the street, are dominated by straight lines and horizontals and verticals.
- *Degree of openness.* Open scenes, such as the beach, often have a visible horizon line and contain few objects. The street scene is also open, although not as much as the beach. The forest is an example of a scene with a low degree of openness.
- *Degree of roughness.* Smooth scenes (low roughness) like the beach contain fewer small elements. Scenes with high roughness like the forest contain many small elements and are more complex.
- *Degree of expansion.* The convergence of parallel lines, like what you see when you look down railroad tracks that appear to vanish in the distance, or in the street scene in Figure 5.36, indicates a high degree of expansion. This feature is especially dependent on the observer's viewpoint. For example, in the street scene, looking directly at the side of a building would result in low expansion.
- *Color.* Some scenes have characteristic colors, like the beach scene (blue) and the forest (green and brown). (Goffaux et al., 2005)

Global image features are *holistic* and *rapidly perceived.* They are properties of the scene as a whole and do not depend on time-consuming processes such as perceiving small details, recognizing individual objects, or separating one object from another. Another property of global image features is that they contain information that results in perception of a scene's structure and spatial layout. For example, the degree of openness and the degree of expansion refer directly to characteristics of a scene's layout, and naturalness also provides layout information that comes from knowing whether a scene is "from nature" or contains "human-made structures."

Global image properties not only help explain how we can perceive the gist of scenes based on features that can be seen in brief exposures, but also illustrate the following general property of perception: Our past experiences in perceiving properties of the environment plays a role in determining our perceptions. We learn, for example, that blue is associated with open sky, that landscapes are often green and smooth, and that verticals and horizontals are associated with buildings. Characteristics of the environment such as this, which occur frequently, are called **regularities in the environment**. We will now describe these regularities in more detail.

Regularities in the Environment: Information for Perceiving

Although observers make use of regularities in the environment to help them perceive, they are often unaware of the specific information they are using. This aspect of perception is similar to what occurs when we use language. Even though people easily string words together to create sentences in conversations, they may not know the rules of grammar that specify how these words are being combined.



Courtesy of Aude Oliva

Figure 5.36 Three scenes that have different global image features. See text for description.

Similarly, we easily use our knowledge of regularities in the environment to help us perceive, even though we may not be able to identify the specific information we are using. We can distinguish two types of regularities, *physical regularities* and *semantic regularities.*

Physical Regularities **Physical regularities** are regularly occurring physical properties of the environment. For example, there are more vertical and horizontal orientations in the environment than oblique (angled) orientations. This occurs in human-made environment (for example, buildings contain lots of horizontals and verticals) and also in natural environments (trees and plants are more likely to be vertical or horizontal than slanted) (Coppola et al., 1998). It is, therefore, no coincidence that people can perceive horizontals and verticals more easily than other orientations, an effect called the **oblique effect** (Appelle, 1972; Campbell et al., 1966; Orban et al., 1984).

Why should being exposed to more verticals and horizontals make it easier to see them? One answer to this question is that *experience-dependent plasticity,* introduced in Chapter 4 (see page 80), causes the visual system to have more neurons that respond best to these orientations. The fact that the visual system has a greater proportion of neurons that respond to verticals and horizontals has been demonstrated in experiments that have recorded from large numbers of neurons in the visual cortex of the monkey (R. L. Devalois et al., 1982; also see Furmanski & Engel, 2000, for evidence that the visual cortex in humans responds better to verticals and horizontals than to other orientations).

Another physical characteristic of the environment is that when one object partially covers another, the contour of the partially covered object "comes out the other side." If this sounds familiar, it is because it is an example of the Gestalt law of good continuation, which we introduced on page 106 and discussed in conjunction with our "creature" behind the tree on page 110 (Figure 5.26). Other Gestalt laws (or "heuristics") reflect regularities in the environment as well.

Consider, for example, the idea of uniform connectedness. Objects are often defined by areas of the same color or texture, so when an area of the image on the retina has the property of uniform connectedness, it is likely that this area arises from a single environmental shape (Palmer & Rock, 1994). Thus, uniformly connected regions are regularities in the environment, and the perceptual system is designed to interpret these regions so that the environment will be perceived correctly. The Gestalt heuristics are therefore based on the kinds of things that occur so often that we take them for granted. Another physical regularity is illustrated by the following demonstration.

DEMONSTRATION

Shape From Shading

What do you perceive in Figure 5.37a? Do some of the discs look as though they are sticking out, like parts of three-dimensional spheres, and others appear to be indentations? If you do see the discs in this way, notice that the ones that appear to be sticking out are arranged in a square. After observing this, turn the page over so the small dot is on the bottom. Does this change your perception? ▮

Figures 5.37b and c show that if we assume that light is coming from above (which is usually the case in the environment), then patterns like the circles that are light on the top would be created by an object that bulges out (Figure 5.37b), but a pattern like the circles that are light on the bottom would be created by an indentation in a surface (Figure 5.37c). The assumption that light is coming from above has been called the **light-from-above heuristic** (Kleffner & Ramachandran, 1992). Apparently, people make the light-from-above assumption because most light in our environment comes from above. This includes the sun, as well as most artificial light sources.

Another example of the light-from-above heuristic at work is provided by the two pictures in Figure 5.38. Figure 5.38a shows indentations created by people walking in the sand. But when we turn this picture upside down, as

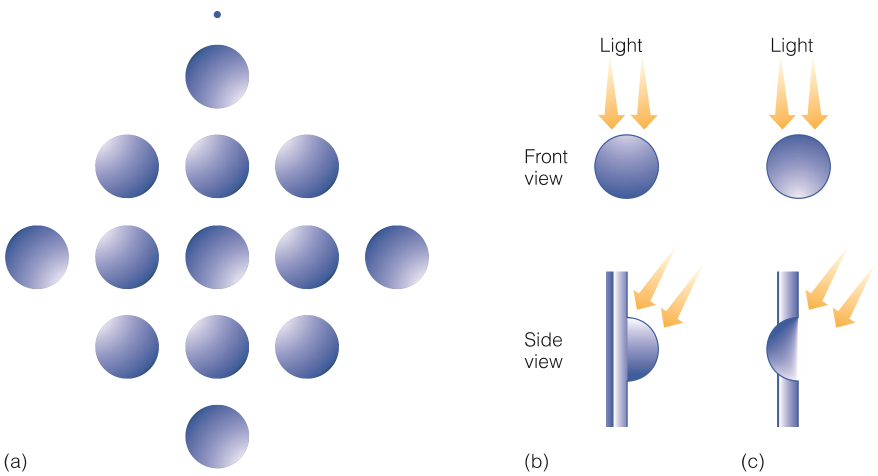


Figure 5.37 ▮ (a) Some of these discs are perceived as jutting out, and some are perceived as indentations. Why? Light coming from above would illuminate (b) the top of a shape that is jutting out and (c) the bottom of an indentation.

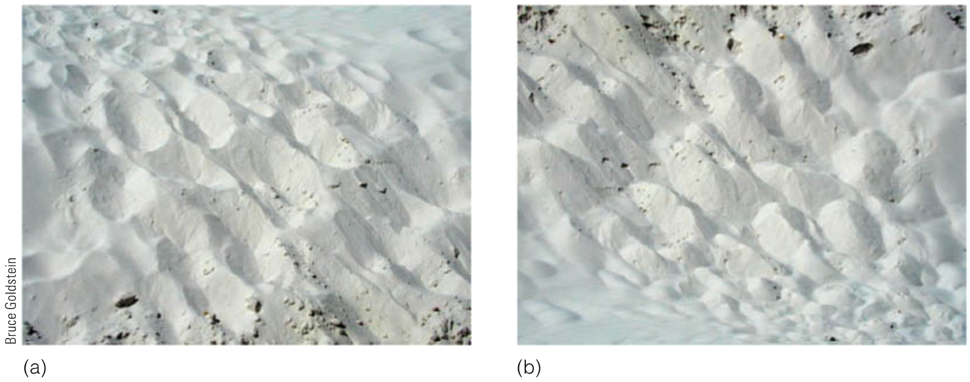


Figure 5.38 ▮ Why does (a) look like indentations in the sand and (b) look like mounds of sand? See text for explanation.

shown in Figure 5.38b, then the indentations in the sand become rounded mounds. VL 18–20

It is clear from these examples of physical regularities in the environment that one of the reasons humans are able to perceive and recognize objects and scenes so much better than computer-guided robots is that our system is customized to respond to the physical characteristics of our environment. But this customization goes beyond physical characteristics. It also occurs because we have learned about what types of objects typically occur in specific types of scenes.

Semantic Regularities In language, *semantics* refers to the meanings of words or sentences. Applied to perceiving scenes, semantics refers to the meaning of a scene. This meaning is often related to the function of a scene—what happens within it. For example, food preparation, cooking, and perhaps eating occur in a kitchen; waiting around, buying tickets, checking luggage, and going through security checkpoints happens in airports. **Semantic regularities** are the characteristics associated with the functions carried out in different types of scenes.

One way to demonstrate that people are aware of semantic regularities is simply to ask them to imagine a particular type of scene or object, as in the following demonstration.

DEMONSTRATION

Visualizing Scenes and Objects

Your task in this demonstration is simple—visualize or simply think about the following scenes and objects:

1. An office
2. The clothing section of a department store
3. A microscope
4. A lion ▮

Most people who have grown up in modern society have little trouble visualizing an office or the clothing section of a department store. What is important about this ability, for our purposes, is that part of this visualization involves details within these scenes. Most people see an office as having a desk with a computer on it, bookshelves, and a chair. The department store scene may contain racks of clothes, a changing room, and perhaps a cash register.

What did you see when you visualized the microscope or the lion? Many people report seeing not just a single object, but an object within a setting. Perhaps you perceived the microscope sitting on a lab bench or in a laboratory, and the lion in a forest or on a savannah or in a zoo.

An example of the knowledge we have of things that typically belong in certain scenes is provided by an experiment in which Andrew Hollingworth (2005) had observers study a scene, such as the picture of the gym in Figure 5.39 (but without the circles), that contained a target object, such

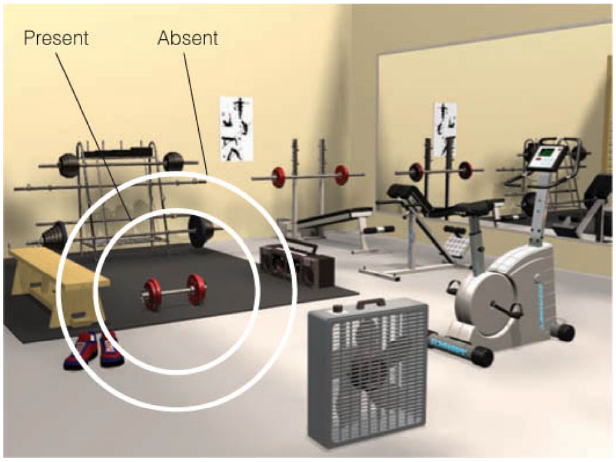


Figure 5.39 ▮ Hollingworth's (2005) observers saw scenes like this one (but without the circles). In this scene the target object is the barbell, although observers do not know this when they are viewing the scene. "Non-target" scenes are the same but do not include the target. The circles indicate the average error of observers' judgments of the position of the target object for trials in which they had seen the object in the scene (small circle) and trials in which the object had not appeared in the scene (larger circle). *(From A. Hollingsworth, 2005, Memory for object position in natural scenes.* Visual Cognition, 12, *1003–1016. Reprinted by permission of the publisher, Taylor & Francis Ltd., http://www.tandf.co.uk/journals.)*

as the barbell on the mat, or the same scene but without the target object, for 20 seconds. Observers then saw a picture of a target object followed by a blank screen, and were asked to indicate where the target object was in the scene (if they had seen the picture containing the target object) or where they would *expect* to see the target object in the scene (if they had seen the same picture but without the target object).

The results are indicated by the circles, which show the averaged error of observers' judgments for many different objects and scenes. The small circle shows that observers who saw the target objects accurately located their positions in the scene. The large circle shows that observers who had not seen the target objects were not quite as accurate but were still able to predict where the target objects would be. What this means for the gym scene is that observers were apparently able to predict where the barbell would appear based on their prior experience in seeing objects in gyms.

This effect of semantic knowledge on our ability to perceive was illustrated in an experiment by Stephen Palmer (1975), using stimuli like the picture in Figure 5.40. Palmer first presented a context scene such as the one on the left and then briefly flashed one of the target pictures on the right. When Palmer asked observers to identify the object in the target picture, they correctly identified an object like the loaf of bread (which is appropriate to the kitchen scene) 80 percent of the time, but correctly identified the mailbox or the drum (two objects that don't fit into the scene) only 40 percent of the time. Apparently Palmer's observers were using their knowledge about kitchens to help them perceive the briefly flashed loaf of bread.

The effect of semantic regularities is also illustrated in Figure 5.41, which is called "the multiple personalities of a blob" (Oliva & Torralba, 2007). The blob is perceived as different objects depending on its orientation and the context within which it is seen. It appears to be an object on a table in (b), a shoe on a person bending down in (c), and a car and a person crossing the street in (d), even though it is the same shape in all of the pictures.

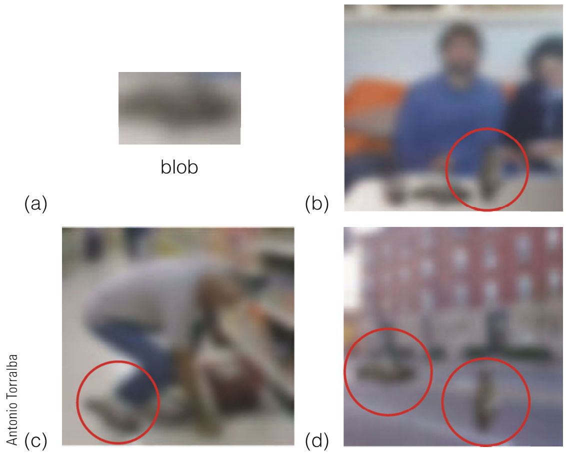


Figure 5.41 ▌ What we expect to see in different contexts influences our interpretation of the identity of the "blob" inside the circles. *(Part (d) adapted from* Trends in Cognitive Sciences, *Vol. 11, 12, Oliva, A., and Torralba, A., The role of context in object recognition. Copyright 2007, with permission from Elsevier.)*

The Role of Inference in Perception

People use their knowledge of physical and semantic regularities such as the ones we have been describing to *infer* what is present in a scene. The idea that perception involves inference is nothing new; it was proposed in the 18th century by Hermann von Helmholtz (1866/1911) who was one of the preeminent physiologists and physicists of his day.

Helmholtz made many discoveries in physiology and physics, developed the ophthalmoscope (the device that an optometrist or ophthalmologist uses to look into your eye), and proposed theories of object perception, color vision, and hearing. One of his proposals about perception is a principle called the **theory of unconscious inference**, which states

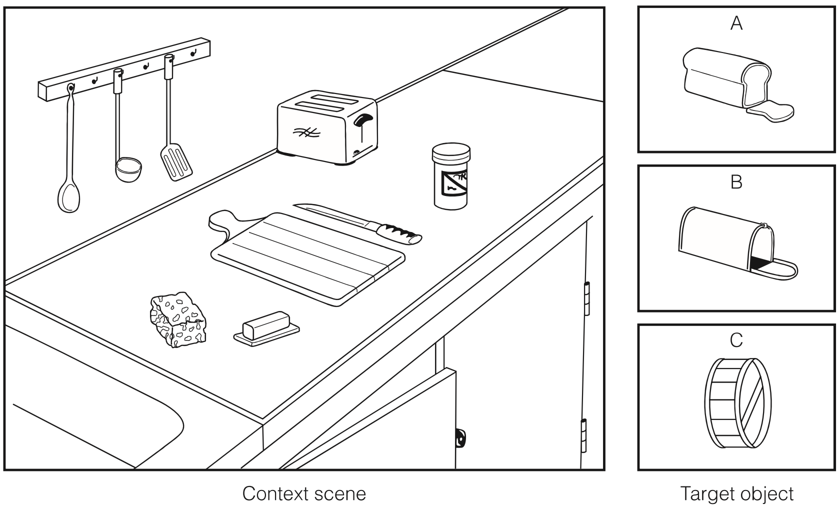


Figure 5.40 ▌ Stimuli used in Palmer's (1975) experiment. The scene at the left is presented first, and the observer is then asked to identify one of the objects on the right.

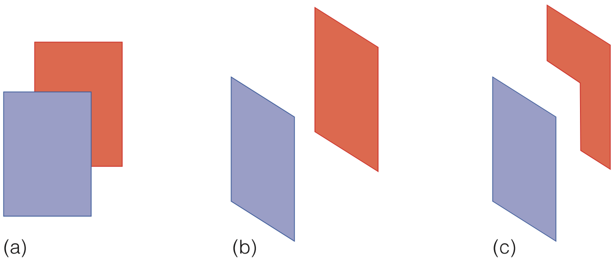


Figure 5.42 ▌ The display in (a) is usually interpreted as being (b) a blue rectangle in front of a red rectangle. It could, however, be (c) a blue rectangle and an appropriately positioned six-sided red figure.

that some of our perceptions are the result of unconscious assumptions we make about the environment.

The theory of unconscious inference was proposed to account for our ability to create perceptions from stimulus information that can be seen in more than one way. For example, what do you see in the display in Figure 5.42a? Most people perceive a blue rectangle in front of a red rectangle, as shown in Figure 5.42b. But as Figure 5.42c indicates, this display could have been caused by a six-sided red shape positioned either in front of or behind the blue rectangle. According to the theory of unconscious inference, we infer that Figure 5.42a is a rectangle covering another rectangle because of experiences we have had with similar situations in the past. A corollary of the theory of unconscious inference is the **likelihood principle**, which states that we perceive the object that is *most likely* to have caused the pattern of stimuli we have received.

One reason that Helmholtz proposed the likelihood principle is to deal with the ambiguity of the perceptual stimulus that we described at the beginning of the chapter. Helmholtz viewed the process of perception as being similar to the process involved in solving a problem. For perception, the task is to determine which object caused a particular pattern of stimulation, and this problem is solved by a process in which the observer brings his or her knowledge of the environment to bear in order to infer what the object might be. This process is unconscious, hence the term *unconscious inference.* (See Rock, 1983, for a modern version of this idea.)

Modern psychologists have quantified Helmholtz's idea of perception as inference by using a statistical technique called **Bayesian inference** that takes probabilities into account (Kersten et al., 2004; Yuille & Kersten, 2006). For example, let's say we want to determine how likely it is that it will rain tomorrow. If we know it rained today, then this increases the chances that it will rain tomorrow, because if it rains one day it is more likely to rain the next day. Applying reasoning like this to perception, we can ask, for example, whether a given object in a kitchen is a loaf of bread or a mailbox. Since it is more likely that a loaf of bread will be in a kitchen, the perceptual system concludes that bread is present. Bayesian statistics involves this type of reasoning, expressed in mathematical formulas that we won't describe here.

Revisiting the Science Project: Designing a Perceiving Machine

We are now ready to return to the science project (see pages 4 and 100) and to apply what we know about perception to the problem of designing a device that can identify objects in the environment. We can now see that one way to make our device more effective would be to program in knowledge about regularities in the environment. In other words, an effective "object perceiving machine" would be able to go beyond processing information about light, dark, shape, and colors that it might pick up with its sensors. It would also be "tuned" to respond best to regularities of the environment that are most likely to occur, and would be programmed to use this information to make inferences about what is out there.

Will robotic vision devices ever equal the human ability to perceive? Based on our knowledge of the complexities of perception, it is easy to say "no," but given the rapid advances that are occurring in the field of computer vision, it is not unreasonable to predict that machines will eventually be developed that approach human perceptual abilities. One reason to think that machines are gaining on humans is that present-day computers have begun incorporating humanlike inference processes into their programs. For example, consider CMU's vehicle "Boss," the winner of the "Urban Challenge" race (see page 101). One reason for Boss's success was that it was programmed to take into account common events that occur when driving on city streets.

Consider, for example, what happens when a human driver (like you) approaches an intersection. You probably check to see if you have a stop sign, then determine if other cars are approaching from the left or right. If they are approaching, you notice whether they have a stop sign. If they do, you might check to be sure they are slowing down in preparation for stopping. If you decide they might ignore their stop sign, you might slow down and prepare to take appropriate action. If you see that there are no cars coming, you proceed across the intersection. In other words, as you drive, you are constantly noticing what is happening and are taking into account your knowledge of traffic regulations and situations you have experienced in the past to make decisions about what to do.

The Boss vehicle is programmed to carry out a similar type of decision-making process to determine what to do when it reaches an intersection. It determines if another car is approaching by using its sensors to detect objects off to the side. It then decides whether an object is a car by taking its size into account and by using the rule "If it is moving, it is likely to be a car." Boss is also programmed to know that other cars should stop if they have a stop sign. Thus, the computer was designed both to sense what was out there and to go beyond simply sensing by taking knowledge into account to decide what to do at the intersection.

The problem for computer vision systems is that before they can compete with humans they have to acquire a great deal more knowledge. Present systems are programmed with just enough knowledge to accomplish specialized tasks like driving the course in the Urban Challenge. While Boss is programmed to determine where the street is and to always stay on the street, Boss can't always make good decisions about when it is safe to drive off-road. For example, Boss can't tell the difference between tall grass (which wouldn't pose much of a threat for off-road driving) and a field full of vertical spikes (which would be very unfriendly to Boss's tires) (C. Urmson, personal communication, 2007).[1]

To program the computer to recognize grass, it would be necessary to provide it with knowledge about grass such as "Grass is green," "Grass moves if it is windy," "Grass is flat and comes to a point." Once Boss has enough knowledge about grass to accurately identify it, then it can be programmed not to avoid it, and to drive off-road onto it if necessary. What all of this means is that while it is helpful to have lots of computing power, it is also nice to have knowledge about the environment. The human model of the perceiving machine has this knowledge, and uses it to perceive with impressive accuracy.

The Physiology of Object and Scene Perception

Thousands of experiments have been done to answer the question "What is the neural basis of object perception?" We have seen that object perception has many aspects, including perceptual organization, grouping, recognizing objects, and perceiving scenes and details within scenes. We first consider neurons that respond to perceptual grouping and figure–ground.

Neurons That Respond to Perceptual Grouping and Figure–Ground

Many years after the Gestalt psychologists proposed the laws of good continuation and similarity, researchers discovered neurons in the visual cortex that respond best to displays that reflect these principles of grouping. For example, Figure 5.43a shows a vertical line in the receptive field (indicated by the square) of a neuron in a monkey's striate cortex. The neuron's response to this single line is indicated by the left bar in Figure 5.43d. No firing occurs when lines are presented outside the square (Zapadia et al., 1995).

But something interesting happens when we add a field of randomly oriented lines, as in Figure 5.43b. These lines, which fall outside the neuron's receptive field, cause a decrease in how rapidly the neuron fires to the single vertical line. This effect of the stimuli that fall outside of the neuron's receptive field (which normally would not affect the neuron's firing rate), is called **contextual modulation**, because the context within which the bar appears affects the neuron's response to the bar.

[1]Chris Urmson is Director of Technology, Tartan Racing Team, Carnegie Mellon University.

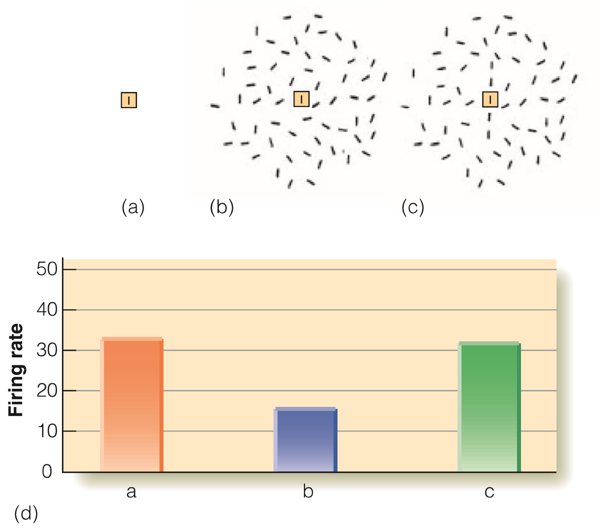


Figure 5.43 How a neuron in the striate cortex (V1) responds to (a) an oriented bar inside the neuron's receptive field (the small square); (b) the same bar surrounded by randomly oriented bars; (c) the bar when it becomes part of a group of vertical bars, due to the principles of similarity and good continuation. *(Adapted from Zapadia, M. K., Ito, M., Gilbert, C. G., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys.* Neuron, 15, *843–856. Copyright © 1995, with permission from Elsevier.)*

Figure 5.43c shows that we can increase the neuron's response to the bar by arranging a few of the lines that are outside the receptive field so that they are lined up with the line that is in the receptive field. When good continuation and similarity cause our receptive-field line to become perceptually grouped with these other lines, the neuron's response increases. This neuron is therefore affected by Gestalt organization even though this organization involves areas outside its receptive field.

Another example of how an area outside the receptive field can affect responding is shown in Figure 5.44. This neuron, in the visual cortex, responds well when leftward-slanted lines are positioned over the neuron's receptive field (indicated by the green bar in Figure 5.44a; Lamme, 1995). Notice that in this case we perceive the leftward slanting bars as a square on a background of right-slanted lines. However, when we replace the right-slanted "background" lines with left-slanted lines, as in Figure 5.44b, the neuron no longer fires.

Notice that when we replaced the right-slanted background lines with left-slanted lines the *stimulus* on the receptive field (left-slanted lines) did not change, but our *perception* of these lines changed from being part of a *figure* (in Figure 5.44a) to being part of the background (Figure 5.44b). This neuron therefore responds to right-slanted lines only when they are seen as being part of the figure. (Also see Qui & von der Heydt, 2005).

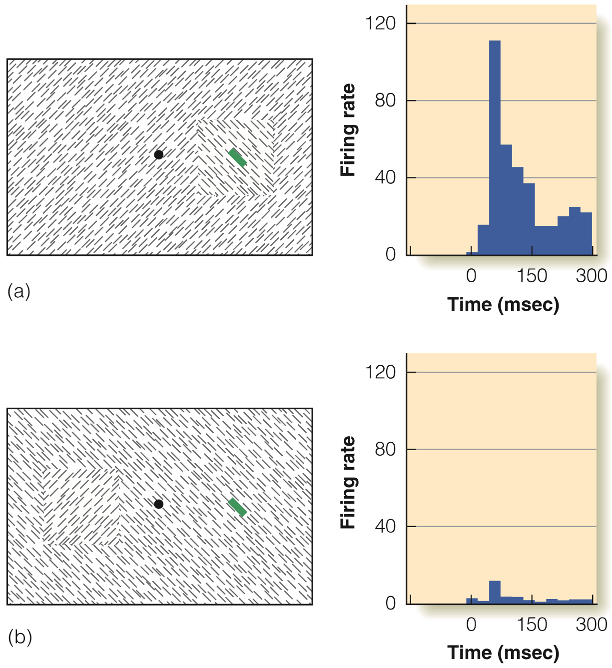


Figure 5.44 ▮ How a neuron in V1 responds to oriented lines presented to the neuron's receptive field (green rectangle). (a) The neuron responded when the bars on the receptive field are part of a figure, but there is no response when (b) the same pattern is not part of a figure. *Adapted from Lamme, V. A. F. (1995). The neurophysiology of figure–ground segregation in primary visual cortex.* Journal of Neuroscience, 15, *1605–1615.*

How Does the Brain Respond to Objects?

How are objects represented by the firing of neurons in the brain? To begin answering this question, let's review the basic principles of sensory coding we introduced in Chapters 2 and 4.

Review of Sensory Coding In Chapter 2 we described *specificity coding,* which occurs if an object is represented by the firing of a neuron that fires *only* to that object, and *distributed coding,* which occurs if an object is represented by the *pattern* of firing of a number of neurons. In Chapter 4 we introduced the idea that certain areas are specialized to process information about specific types of objects. We called these specialized areas *modules.* Three of these areas are the fusiform face area (FFA), for faces; the extrastriate body area (EBA), for bodies; and the parahippocampal place area (PPA), for buildings and places. Although neurons in these areas respond to specific types of stimuli, they aren't totally specialized, so a particular neuron that responds only to faces responds to a number of different faces (Tsao et al., 2006). Objects, according to this idea, are represented by distributed coding, so a specific face would be represented by the pattern of firing of a number of neurons that respond to faces.

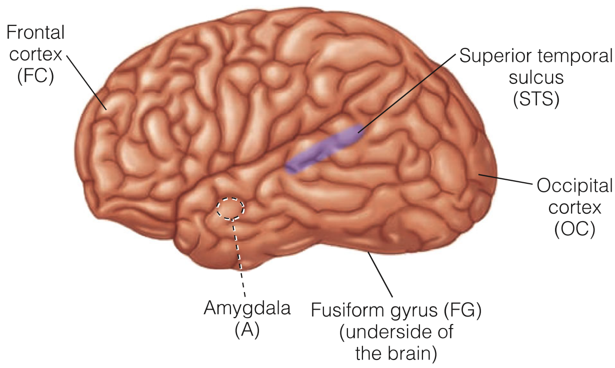


Figure 5.45 ▮ The human brain, showing some of the areas involved in perceiving faces. Some of the perceptual functions of these areas are: OC = initial processing; FG = identification; A = emotional reaction; STS = gaze direction; FC = atttractiveness. The amygdala is located deep inside the cortex, approximately under the ellipse.

We also noted that even though modules are specialized to process information about specific types of stimuli such as faces, places, and bodies, objects typically cause activity not only in a number of neurons within a module, but also in a number of different areas of the brain. Thus, a face might cause a large amount of activity in the FFA, but also cause activity in other areas as well. Firing is, therefore, distributed in two ways: (1) across groups of neurons within a specific area, and (2) across different areas in the brain.

More Evidence for Distributed Activity Across the Brain We begin our discussion where we left off in Chapter 4—with the idea that objects cause activity in a number of different brain areas. Faces provide one of the best examples of distributed representation across the brain. We know that the fusiform face area (FFA) is specialized to process information about faces, because the FFA responds to pictures of faces but not to pictures of other types of stimuli.

But perceiving a face involves much more than just looking at a face and identifying it as "a face," or even as "Bill's face." After you have identified a face as, say, your friend Bill, you may have an emotional reaction to Bill based on the expression on his face or on your past experience with him. You may notice whether he is looking straight at you or off to the side. You may even be thinking about how attractive (or unattractive) he is. Each of these reactions to faces has been linked to activity in different areas of the brain.

Figure 5.45 shows some of the areas involved in face perception. Initial processing of the face occurs in the occipital cortex, which sends signals to the fusiform gyrus, where visual information concerned with identification of the face is processed (Grill-Spector et al., 2004). Emotional aspects of the face, including facial expression and the observer's emotional reaction to the face, are reflected in activation of the amygdala, which is located within the brain (Gobbini & Haxby, 2007; Ishai et al., 2004).

Evaluation of where a person is looking is linked to activity in the superior temporal sulcus; this area is also involved in perceiving movements of a person's mouth as the person speaks (Calder et al., 2007; Puce et al., 1998). Evaluation of a face's attractiveness is linked to activity in the frontal area of the brain.

The fact that all, or most, of these factors come into play when we perceive a face has led to the conclusion that there is a distributed system in the cortex for perceiving faces (Haxby et al., 2000; Ishai, 2008). The activation caused by other objects is also distributed, with most objects activating a number of different areas in the brain (Shinkareva et al., 2008).

Connecting Neural Activity and Perception

The results we have been describing involved experiments in which a stimulus was presented and brain activity was measured. Other experiments have gone beyond simply observing which stimulus causes firing in specific areas to studying connections between brain activity and what a person or animal *perceives.*

One of these experiments, by Kalanit Grill-Spector and coworkers (2004), studied the question of how activation of the brain is related to whether a person recognizes an object by measuring brain activation as human observers identified pictures of the face of a well-known person—Harrison Ford. They focused on the fusiform face area (FFA). To locate the FFA in each person, they used a method called the region-of-interest (ROI) approach.

METHOD | Region-of-Interest Approach

One of the challenges of brain imaging research is that although maps have been published indicating the location of different areas of the brain, there is a great deal of variation from person to person in the exact location of a particular area. The **region-of-interest (ROI) approach** deals with this problem by pretesting people on the stimuli to be studied before running an experiment. For example, in the study we are going to describe, Grill-Spector located the FFA in each observer by presenting pictures of faces and nonfaces and noting the area that was preferentially activated by faces. Locating this ROI before doing the experiment enabled researchers to focus on the exact area of the brain that, *for each individual person,* was specialized to process information about faces.

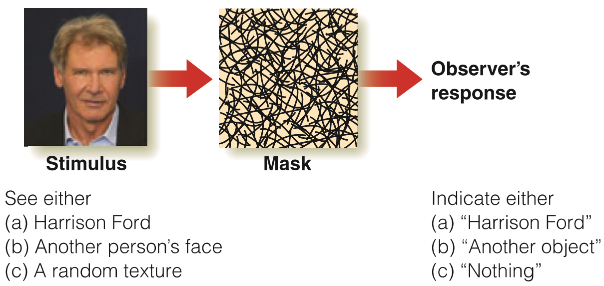
Once Grill-Spector determined the location of the FFA for each observer, she presented stimuli as shown in Figure 5.46. On each trial, observers saw either (a) a picture of Harrison Ford, (b) a picture of another person's face, or (c) a random texture. Each of these stimuli was presented briefly (about 50 ms) followed immediately by a random-pattern mask, which limited the visibility of each stimulus to just 50 ms (see Method: Using a Mask to Achieve Brief Stimulus Presentations, page 114).

The observer's task in this experiment was to indicate, after presentation of the mask, whether the picture was "Harrison Ford," "another object," or "nothing." This is the "observer's response" in Figure 5.46. The results, based on presentation of 60 different pictures of Harrison Ford, 60 pictures of other faces, and 60 random textures, are shown in Figure 5.47. This figure shows the course of brain activation for the trials in which Harrison Ford's face was presented. The top curve (red) shows that activation was greatest when observers correctly identified the stimulus as Harrison Ford's face. The next curve shows that activation was less when they responded "other object" to Harrison Ford's face. In this case they detected the stimulus as a face but were not able to identify it as Harrison Ford's face. The lowest curve indicates that there was little activation when observers could not even tell that a face was presented.

Remember that all of the curves in Figure 5.47 represent the brain activity that occurred not when observers were responding verbally, but *during presentation* of Harrison Ford's face. These results therefore show that neural activity that occurs *as a person is looking at a stimulus* is determined not only by the stimulus that is presented, but also by how a person is processing the stimulus. A large neural response is associated with processing that results in the ability to *identify* the stimulus; a smaller response, with *detecting* the stimulus; and the absence of a response with missing the stimulus altogether.

Connections between neural responses and perception have also been determined by using a perceptual phenomenon called **binocular rivalry**: If one image is presented to the left eye and a different image is presented to the right eye, perception alternates back and forth between the two eyes. For example, if the sunburst pattern in Figure 5.48 is presented only to the left eye, and the butterfly is presented only to the right eye, a person would see the sunburst part of the time and the butterfly part of the time, but never both together.

D. L. Sheinberg and Nikos Logothetis (1997) presented a sunburst pattern to a monkey's left eye and a picture such as the butterfly or another animal or object to the monkey's right eye. To determine what the monkey was perceiving, they trained the monkey to pull one lever when it perceived the sunburst pattern and another lever when it perceived the butterfly. As the monkey was reporting what it was perceiving,



© Stephane Cardinale/People Avenue/Corbis

Figure 5.46 | Procedure for the Grill-Spector et al. (2004) experiment. See text for details.

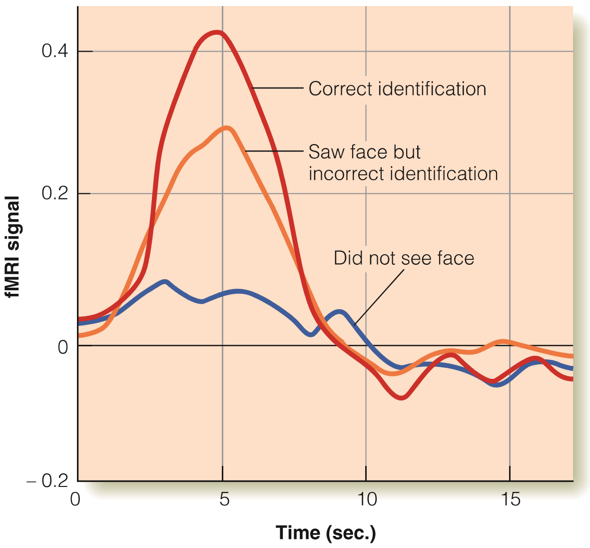


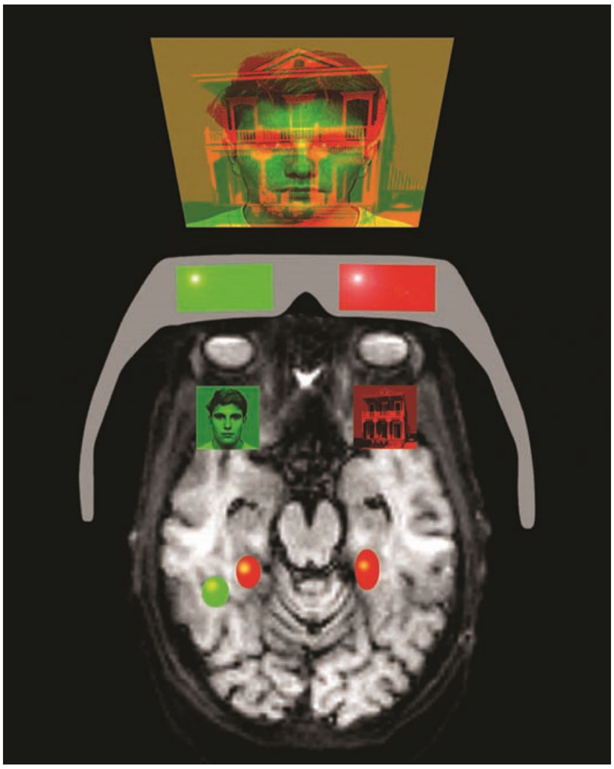
Figure 5.47 ▮ Results of Grill-Spector et al. (2004) experiment for trials in which Harrison Ford's face was presented. Activity was measured in the initial part of the experiment, when Harrison Ford's face was presented. *(From Grill-Spector, K., Knouf, N., & Kanwisher, N., The fusiform face area subserves face perception, not generic within-category identification,* Nature Neuroscience, *7, 555–562. Reprinted by permission from Macmillan Publisher Ltd. Copyright 2004.)*

they simultaneously recorded the activity of a neuron in the inferotemporal (IT) cortex that had previously been shown to respond to the butterfly but not to the sunburst. The result of this experiment was straightforward: The cell fired vigorously when the monkey was perceiving the butterfly and ceased firing when the monkey was perceiving the sunburst.

Image not available due to copyright restrictions

Consider what happened in this experiment. The images on the monkey's retinas remained the same throughout the experiment—the sunburst was always positioned on the left retina, and the butterfly was always positioned on the right retina. The change in perception from "sunburst" to "butterfly" must therefore have been happening in the monkey's brain, and this experiment showed that these changes in perception were linked to changes in the firing of a neuron in the brain.

This binocular rivalry procedure has also been used to connect perception and neural responding in humans by using fMRI. Frank Tong and coworkers (1998) presented a picture of a person's face to one eye and a picture of a house to the other eye, by having observers view the pictures through colored glasses, as shown in Figure 5.49. The images are shown as overlapping in this figure, but because each eye



Courtesy of Frank Tong

Figure 5.49 ▮ Observers in the Tong et al. (1998) experiment viewed the overlapping red house and green face through red-green glasses, so the house image was presented to the right eye and the face image to the left eye. Because of binocular rivalry, the observers' perception alternated back and forth between the face and the house. When the observers perceived the house, activity occurred in the parahippocampal place area (PPA), in the left and right hemispheres (red ellipses). When observers perceived the face, activity occurred in the fusiform face area (FFA) in the left hemisphere (green ellipse). *(From Tong, F., Nakayama, K., Vaughn, J. T., & Kanwisher, N., 1998, Binocular rivalry and visual awareness in human extrastriate cortex.* Neuron, *21, 753–759.)*

received only one of the images, binocular rivalry occurred. Observers perceived either the face alone or the house alone, and these perceptions alternated back and forth every few seconds.

Tong determined what the observers were perceiving by having them push a button when perceiving the house and another button when perceiving the face. As the observer's perception was flipping back and forth between the house and the face, Tong measured the fMRI response in the parahippocampal place area (PPA) and the fusiform face area (FFA). When observers were perceiving the house, activity increased in the PPA (and decreased in the FFA); when they were perceiving the face, activity increased in the FFA (and decreased in the PPA). This result is therefore similar to what Sheinberg and Logothetis found in single neurons in the monkey. Even though the image on the retina remained the same throughout the experiment, activity in the brain changed, depending on what the person was experiencing.

Something to Consider: Models of Brain Activity That Can Predict What a Person Is Looking At

When you look at a scene, a pattern of activity occurs in your brain that represents the scene. When you look somewhere else, a new pattern occurs that represents the new scene. Is it possible to tell what scene a person is looking at by monitoring his or her brain activity? Some recent research has brought us closer to achieving this feat and has furthered our understanding of the connection between brain activity and perception.

Yakiyasui Kamitani and Frank Tong (2005) took a step toward being able to "decode" brain activity by measuring observers' fMRI response to grating stimuli—alternating black and white bars like the one in Figure 5.50a. They presented gratings with a number of different orientations (the one in Figure 5.50a slants 45 degrees to the right, for example) and determined the response to these gratings in a number of fMRI voxels. A *voxel* is a small cube-shaped area of the brain about 2 or 3 mm on each side. (The size of a voxel depends on the resolution of the fMRI scanner. Scanners are being developed that will be able to resolve areas smaller than 2 or 3 mm on a side.)

One of the properties of fMRI voxels is that there is some variability in how different voxels respond. For example, the small cubes representing voxels in Figure 5.50a show that the 45-degree grating causes slight differences in the responses of different voxels. A grating with a different orientation would cause a different pattern of activity in these voxels. By using the information provided by the responses of many voxels, Kamitani and Tong were able to create an "orientation decoder," which was able to determine what orientation a person was looking at based on the person's

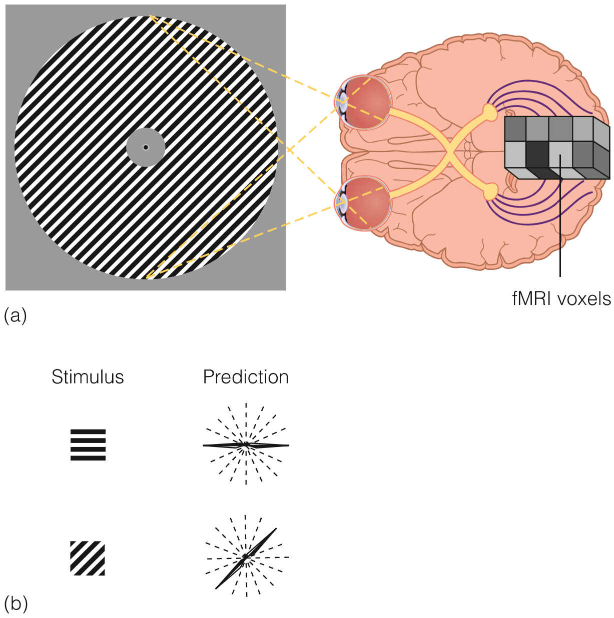


Figure 5.50 (a) Observers in Kamitani and Tong's (2005) experiment viewed oriented gratings like the one on the left. The cubes in the brain represent the response of 8 voxels. The activity of 400 voxels was monitored in the experiment. (b) Results for two orientations. The gratings are the stimuli presented to the observer. The line on the right is the orientation predicted by the orientation decoder. The decoder was able to accurately predict when each of the 8 orientations was presented. *(From Kamitani, Y., & Tong, F., Decoding the visual and subjective contents of the human brain,* Nature Neuroscience, 8*, 679–685. Reprinted by permission of Macmillan Publishers Ltd. Copyright 2005.)*

brain activity. They created this decoder by measuring the response of 400 voxels in the primary visual cortex (V1) and a neighboring area called V2 to gratings with eight different orientations. They then carried out a statistical analysis on the patterns of voxel activity for each orientation to create an orientation decoder designed to analyze the pattern of activity recorded from a person's brain and predict which orientation the person was looking at.

Kaminiti and Tong demonstrated the predictive power of their orientation decoder by presenting oriented gratings to an observer and feeding the resulting fMRI response into the decoder, which predicted which orientation had been presented. The results, shown in Figure 5.50b, show that the decoder accurately predicted the orientations that were presented.

In another test of the decoder, Kaminiti and Tong presented two overlapping gratings, creating a lattice like the one in Figure 5.51, and asked their observers to pay attention to one of the orientations. Because attending to each orientation resulted in different patterns of brain activity, the decoder was able to predict which of the orientations the person was paying attention to. Think about what this means.

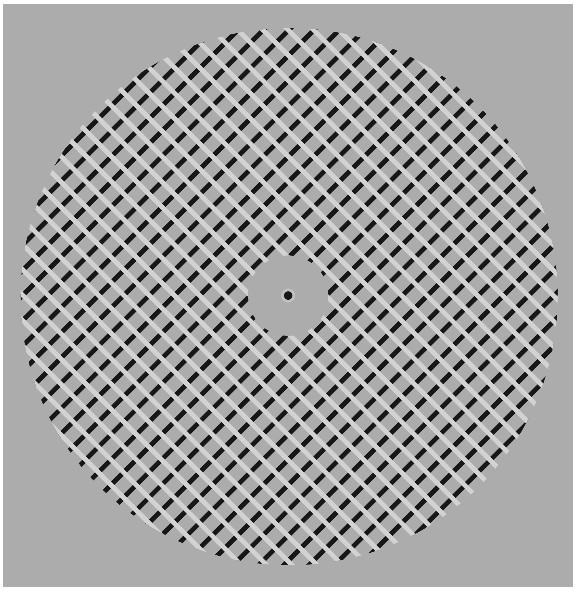


Figure 5.51 ▌ The overlapping grating stimulus used for Kaminiti and Tong's (2005) experiment, in which observers were told to pay attention to one of the orientations at a time. *(From Kamitani, Y., & Tong, F., Decoding the visual and subjective contents of the human brain,* Nature Neuroscience, *8, 679–685. Reprinted by permission of Macmillan Publishers Ltd. Copyright 2005.)*

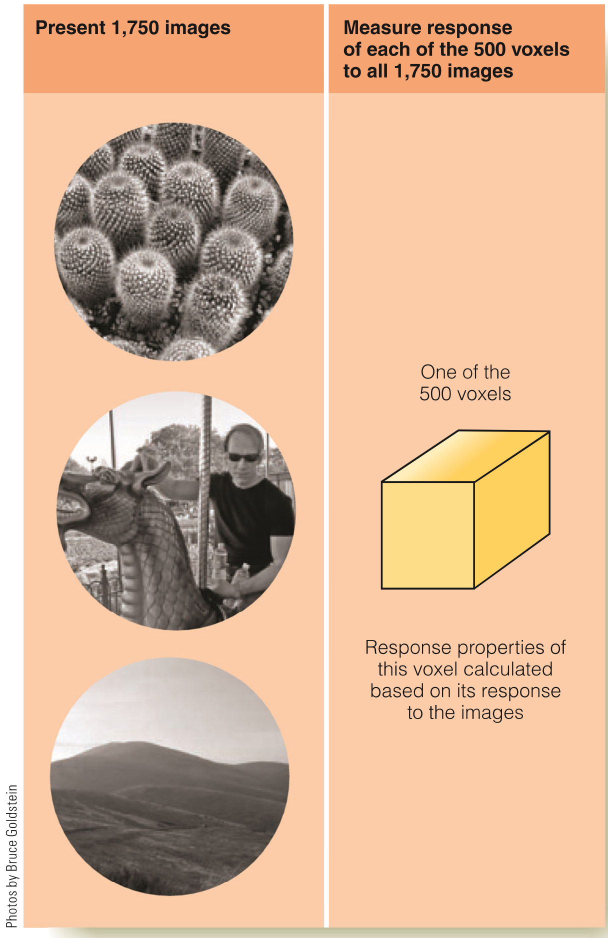


Figure 5.52 ▌ The first part of the Kay et al. (2008) experiment, in which the scene decoder was created. They determined the response properties of 500 voxels in the striate cortex by measuring the response of each voxel as they presented 1,750 images to an observer. Three images like the ones Kay used are shown here. The cube represents one of the 500 voxels. The scene decoder was created by determining how each of the 500 voxels responded to an image's position in space, orientation, and level of detail.

If you were in Kaminiti and Tong's laboratory looking over their observer's shoulder as he or she was observing the overlapping gratings, you would have no way of knowing exactly what the person was perceiving. But by consulting the orientation decoder, you could find out which orientation the observer was focusing on. The orientation decoder essentially provides a window into the person's mind.

But what about stimuli that are more complex than oriented gratings? Kendrick Kay and coworkers (2008) have created a new decoder that can determine which photograph of a natural scene has been presented to an observer. In the first part of their experiment, they presented 1,750 black and white photographs of a variety of natural scenes to an observer and measured the activity in 500 voxels in the primary visual cortex (Figure 5.52). The goal of this part of the experiment was to determine how each voxel responds to specific features of the scene, such as the position of the image, the image's orientation, and the degree of detail in the image, ranging from fine details (like the two top images in Figure 5.52) to images with little detail (like the bottom image). Based on an analysis of the responses of the 500 voxels to the 1,750 images, Kay and coworkers created a scene decoder that was able to predict the voxel activity patterns that would occur in the brain in response to images of scenes.

To test the decoder, Kay and coworkers did the following (Figure 5.53): (1) They measured the brain activity pattern to a test image that had never been presented before (the lion in this example). (2) They presented this test image and 119 other new images to the decoder, which calculated the predicted voxel activity patterns (shown on the right) for each image. (3) They selected the pattern that most closely matched the actual brain activity elicited by the test image. When they checked to see if the image that went with this pattern was the same as the test image, they found that the decoder identified 92 percent of the images correctly for one observer, and 72 percent correctly for another observer. This is impressive because chance performance for 120 images is less than 1 percent. It is also impressive because the images

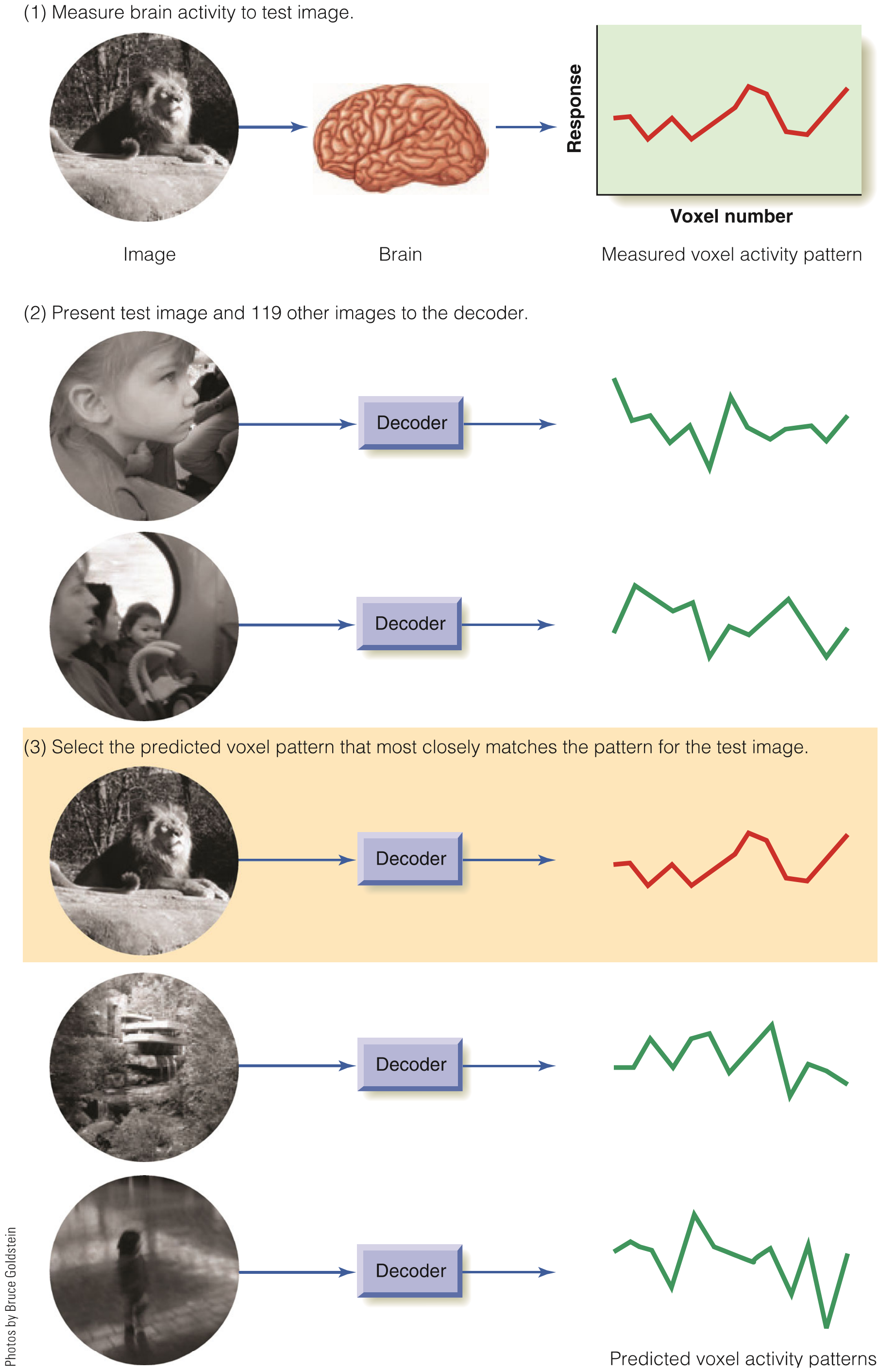


Figure 5.53 To test their scene decoder, Kay and coworkers (2008) first (a) measured an observer's brain activity caused by the presentation of a test image that the observer had never seen, and then (b) used the decoder to predict the pattern of voxel activity for this test image and 119 other images. The highlighted pattern of voxel activity indicates that the decoder has correctly matched the predicted response to the test image with the actual brain activity generated by the test image that was measured in (a). In other words, the decoder was able to pick the correct image out of a group of 120 images as being the one that had been presented to the observer. *(Based on Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L., Identifying natural images from human brain activity,* Nature, 7185, *352–355, Fig 1, top. Reprinted by permission from Macmillan Publisher Ltd. Copyright 2008.)*

presented were new ones, which the decoder had never been exposed to before.

Do these results mean that we can now use brain activity to "read minds," as suggested by some reports of this research that appeared in the popular press? These experiments do show that it is possible to identify information in the activity of the primary visual cortex that can predict which image *out of a group of images* a person is looking at. However, we are still not able to create, from a person's brain activity, a picture that corresponds to what the person is seeing. Nonetheless, this research represents an impressive step toward understanding how neural activity represents objects and scenes.

TEST YOURSELF 5.2

1. What is a "scene," and how is it different from an "object"?
2. What is the evidence that we can perceive the gist of a scene very rapidly? What information helps us identify the gist?
3. What are regularities in the environment? Give examples of physical regularities, and discuss how these regularities are related to the Gestalt laws of organization.
4. What are semantic regularities? How do semantic regularities affect our perception of objects within scenes? What is the relation between semantic regularities and the idea that perception involves inference? What did Helmholtz have to say about inference and perception? What is Bayesian inference, and how is it related to Helmholtz's ideas about inference?
5. What is a way to make a robotic vision device more effective? Why is there reason to think that machines are gaining on humans? What do computer vision systems have to do before they can compete with humans?
6. Describe research on (a) neurons that respond to perceptual grouping and to figure–ground; (b) the distributed nature of the representation of faces in the brain; and (c) connections between brain activity and perception (be sure you understand the "Harrison Ford" experiment and the two binocular rivalry experiments).
7. Describe how fMRI has been used to create "orientation decoders" and "scene decoders" that can predict how the brain will respond to (a) oriented gratings and (b) complex scenes.

THINK ABOUT IT

1. This chapter describes a number of perceptual heuristics, including the Gestalt "laws" and the light-from-above heuristic. Think of some other heuristics—either perceptual or from some other area—that help you solve problems quickly using "best guess" rules. (p. 109)
2. Consider this situation: We saw in Chapter 1 that top-down processing occurs when perception is affected by the observer's knowledge and expectations. Of course, this knowledge is stored in neurons and groups of neurons in the brain. In this chapter, we saw that there are neurons that have become tuned to respond to specific characteristics of the environment. We could therefore say that some knowledge of the environment is built into these neurons. Thus, if a particular perception occurs because of the firing of these tuned neurons, does this qualify as top-down processing? (p. 116)
3. Reacting to the results of the recent DARPA race, Harry says, "Well, we've finally shown that computers can perceive as well as people." How would you respond to this statement? (p. 119)
4. Biological evolution caused our perceptual system to be tuned to the Stone Age world in which we evolved. Given this fact, how well do we handle activities like downhill skiing or driving, which are very recent additions to our behavioral repertoire? (p. 115)
5. Vecera showed that regions in the lower part of a stimulus are more likely to be perceived as figure. How does this result relate to the idea that our visual system is tuned to regularities in the environment? (p. 108)
6. We are able to perceptually organize objects in the environment even when objects are similar, as in Figure 5.54. What perceptual principles are involved in perceiving two separate zebras? Consider both the Gestalt laws of organization and the geons of RBC theory. What happens when you cover the zebras' heads, so you see just the bodies? Do these priciples still work? Is there information in addition to what is proposed by the Gestalt laws and RBC theory that helps you perceptually organize the two zebras? (p. 105)
7. How did you perceive the picture in Figure 5.55 when you first looked at it? What perceptual assumptions in-

Barbara Goldstein

Figure 5.54 Which principles of organization enable us to tell the two zebras apart?

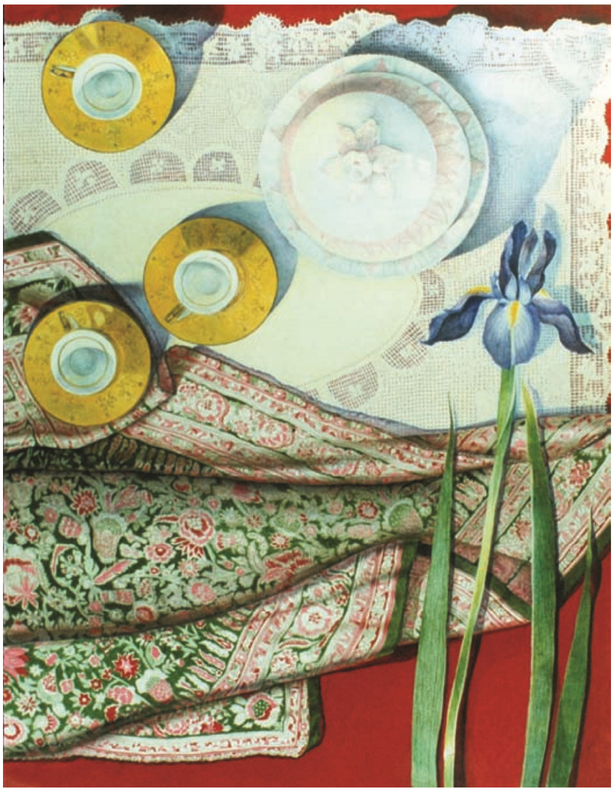


Figure 5.55 ▌ *The Scarf,* a drawing by Rita Ludden.

fluenced your response to this picture? (For example, did you make an assumption about how flowers are usually oriented in the environment?) (p. 118)

IF YOU WANT TO KNOW MORE

1. *Robotic vehicles.* To find out more about the DARPA race, go to www.grandchallenge.org or search for DARPA on the Internet. (p. 101)

2. *Perceiving figure and ground.* When you look at the vase-face pattern in Figure 5.21, you can perceive two blue faces on a white background or a white vase on a blue background, but it is difficult to see the faces and the vase simultaneously. It has been suggested that this occurs because of a heuristic built into the visual system that takes into account the unlikelihood that two adjacent objects would have the same contours and would line up perfectly. (p. 108)

   Baylis, G. C., & Driver, J. (1995). One-sided edge assignment in vision: I. Figure-ground segmentation and attention to objects. *Current Directions in Psychological Science 4,* 140–146.

3. *When does figure separate from ground?* The Gestalt psychologists proposed that figure must be separated from ground before it can be recognized. There is evidence, however, the meaning of an area can be recognized before it has become separated from the ground. This means that recognition must be occurring either before or at the same time as the figure is being separated from ground. (p. 108)

   Peterson, M. A. (1994). Object recognition processes can and do operate before figure-ground organization. *Current Directions in Psychological Science, 3,* 105–111.

4. *Global precedence.* When a display consists of a large object that is made up of smaller elements, what does the nervous system process first, the large object or the smaller elements? An effect called the *global precedence effect* suggests that the larger object is processed first. VL 21

   Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9,* 353–383.

5. *Experience-dependent plasticity and object recognition.* A person's experience can shape both neural responding and behavioral performance related to the recognition of objects. (p. 116)

   Kourtzi, Z., & DiCarlo, J. J. (2006). Learning and neural plasticity in visual object recognition. *Current Opinion in Neurobiology, 16,* 152–158.

6. *Boundary extension effect.* When people are asked to remember a photograph of a scene, they tend to remember a wider-angle view than was shown in the original photograph. This suggests that visual mechanisms infer the existence of visual layout that occurs beyond the boundaries of a given view. There is evidence the parahippocampal place area may be involved in boundary extension. (p. 118)

   Intraub, H. (1997). The representation of visual scenes. *Trends in Cognitive Sciences, 1,* 217–222.

   Park, S., Intraub, H., Yi, D.-J., Widders, D., & Chun, M. M. (2007). Beyond the edges of view: Boundary extension in human scene-selective cortex. *Neuron 54,* 335–342.

7. *Identifying cognitive states associated with perceptions.* Research similar to that described in the Something to Consider section has used fMRI to identify different patterns of brain activation for tools and dwellings. (p. 124)

   Shinkareva, S. V., Mason, R. A., Malave, V. L., Wang, W., Mitchell, T. M., & Just, M. (2008). Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PLoS ONE, 3*(1), e1394.