CHAPTER 2

# Perception

## Learning Objectives

- Describe the role of the primary visual cortex in the occipital lobe in visual consciousness. Explain how blindsight contributes to understanding this role.
- Define the process of pattern recognition and explain the two kinds of agnosia that represent failures of this process.
- Explain the contributions of conceptually driven and data-driven processes in object recognition.
- Understand why distinctive features are only part of the answer to how we identify objects.

The mind comes to know the world through sensing and perceiving the environment. Sensation refers to the transduction of physical energy, such as sound waves or electromagnetic radiation, into an initial mental representation that can be further processed and transformed over time. Transduction means the conversion of one kind of energy into another kind. For example, in vision, electromagnetic energy is converted into an electrical signal in neurons. As a result of this processing, the objects and events that are present in the environment are perceived, in the sense of being detected. With still more processing, the objects and events are recognized, in the sense of being categorized as meaningful. Even to recognize your own mother involves a sequence of processing stages that is complex and can take as long as a half second.

It is difficult to grasp that a process as rapid and effortless as perception involves multiple stages and transformations of mental representations. Perception is the result of processes that construct mental representations of the information available in the environment. Such representations draw on information stored in memory as well as that present in the environment. As various examples in this chapter will demonstrate, perception is always driven in part by expectations of how the world ought to look or sound based on knowledge stored in long-term memory. In a nightly dream or in the waking hallucinations of a psychotic individual, bizarre perceptions may be fabricated out of whole cloth, secreted from memory alone.

The constructive aspect of mental representation is perhaps easiest to grasp when an illusion is perceived. In illusions, the perceptual processes construct a mental representation that does not accurately mirror the object in the environment. For example, consider the well-known moon illusion. The moon looks much larger on the horizon than it does when overhead. When you see a full moon in its zenith position high in the night sky, in particular, the celestial body looks markedly smaller than it does when the moon is just coming up and is seen near the horizon. Yet why does the moon appear to change size, depending on its position in the sky? The distance of the moon from Earth—more than 240,000 miles—is fixed and invariant, regardless of whether the moon is rising on the horizon or has reached its zenith position.

The knowledge that a familiar object is supposed to appear smaller at a distance affects our perception of the moon. Even though its true size does not change, the mental representation that we construct depends on the location of the moon in the sky. The size illusion stems from our expectation that a familiar object normally looks large up close and small at a distance. Depth information—the apparent distance of objects from the viewer—is computed automatically by a module of perception. The illusion is caused by an error in the output of this depth module. When the moon is on the horizon, there are several cues that the module uses to assign a distance to the object. For example, the small trees or buildings that might be seen far off on the horizon provide information that the horizon moon is at a great distance from the viewer. By contrast, the sky directly overhead where the moon is positioned at its zenith appears to be closer to the viewer. It is as if the heavens overhead are flattened like a bowl, so that observers see the horizon position as further away than the zenith position (Kaufman & Rock, 1962; Rock & Kaufman, 1962).

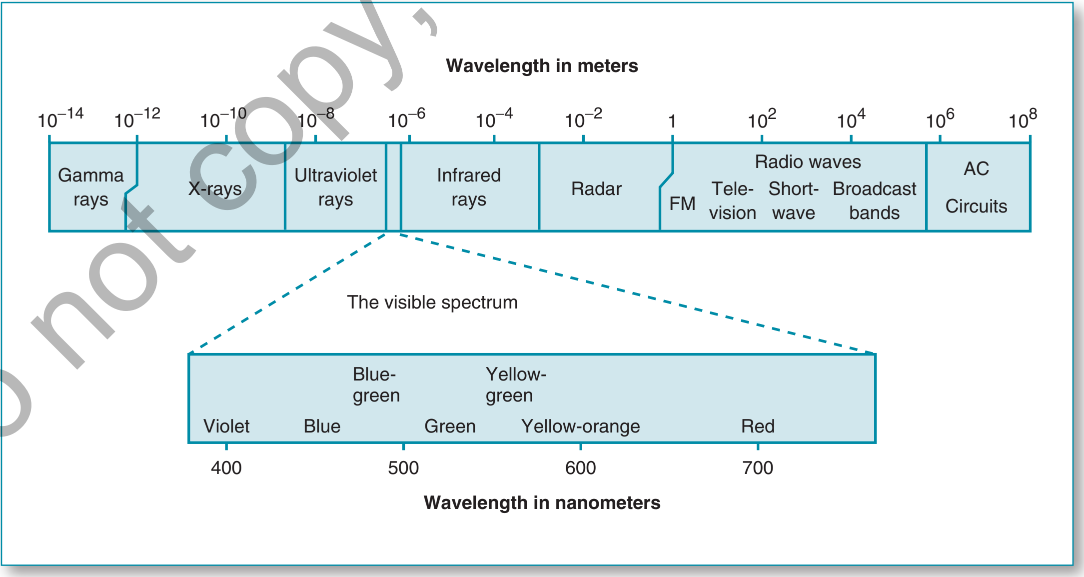
The unconscious inference made by the depth module is that the horizon moon is far off in the distance, whereas the zenith moon is closer by. Normally, a distant object would present a smaller image size on the retina of the eye than it would up close. Of course, the size of the moon as input to the retina in the eye is a constant—the retinal image size never changes, no matter where the moon is positioned in the sky. Consequently, the moon violates what would be expected of an object on Earth that varies in distance from the eye. Consider what would happen to the retinal image of a friend who is standing in front of you and then turns around and walks away a full city block. You are able to accurately judge his distance by the fact that his image gets progressively smaller as he walks away. A moon illusion occurs because the depth-perception module is confounded by contradictory information: The moon looks farther away than on the horizon, but the retinal image size remains unchanged rather than getting smaller. To resolve the contradiction, the depth-perception module infers that the moon at the horizon must in fact be bigger than normal! It constructs a conscious mental representation of a larger-than-normal moon. The module takes the fact that the retinal image size remains unchanged for an object at a distance as evidence that somehow the object got bigger than it normally would be when seen overhead in its zenith position.

Perception is a large subject that lies well beyond the scope of a chapter in a book on cognitive psychology. Yet the basic concepts of detecting and recognizing external stimuli are fundamental to the field. To focus the discussion, four related problems in perception are addressed: Why is it that you can see anything at all, regardless of its identity? Given that you can see something, how do you recognize it as a person instead of, say, a hat rack? Even more specifically, how do you recognize that you are perceiving the face of a person rather than the back or side of the person's head? Finally, how do you recognize what the person is saying to you as the person moves their lips, uttering familiar sounds? Visual sensation, object recognition, face recognition, and speech recognition will illustrate important concepts in perception that lie at the foundation of cognitive psychology as a whole.

## VISUAL CONSCIOUSNESS

Visible light is a narrow band of electromagnetic energy. The wavelengths of light that may be sensed by the human visual system range from 400 to 700 nanometers, where 1 nanometer = $10^{-9}$ meters. As shown in Figure 2.1, the full spectrum of electromagnetic energy dwarfs this tiny band of visible light. Ultraviolet rays, X-rays, and gamma rays are progressively shorter in wavelength and are not sensed by the visual system. The longer wavelengths of infrared rays, radar, radio waves, and AC circuits also go undetected. The visual system cannot construct a mental representation of an object without first transducing electromagnetic energy into a neural signal, and it is sensitive only to wavelengths within the visible spectrum.

**Figure 2.1** The spectrum of electromagnetic radiation includes a narrow band of visible light.



Visible light from the sun or other light sources, such as an indoor table lamp, is reflected off the objects in the environment. This light, in what is called the visual field, is structured in accordance with the structures of the objects themselves. Without structured patterns of light in the visual field, it would be impossible to see in the sense of both detecting a stimulus and recognizing its identity. The detection process begins with the transduction of electromagnetic energy by photoreceptors in the retina of the eye. Photoreceptors are neurons specialized to convert visible light into electrical signals that may be propagated by the neurons of the visual system.

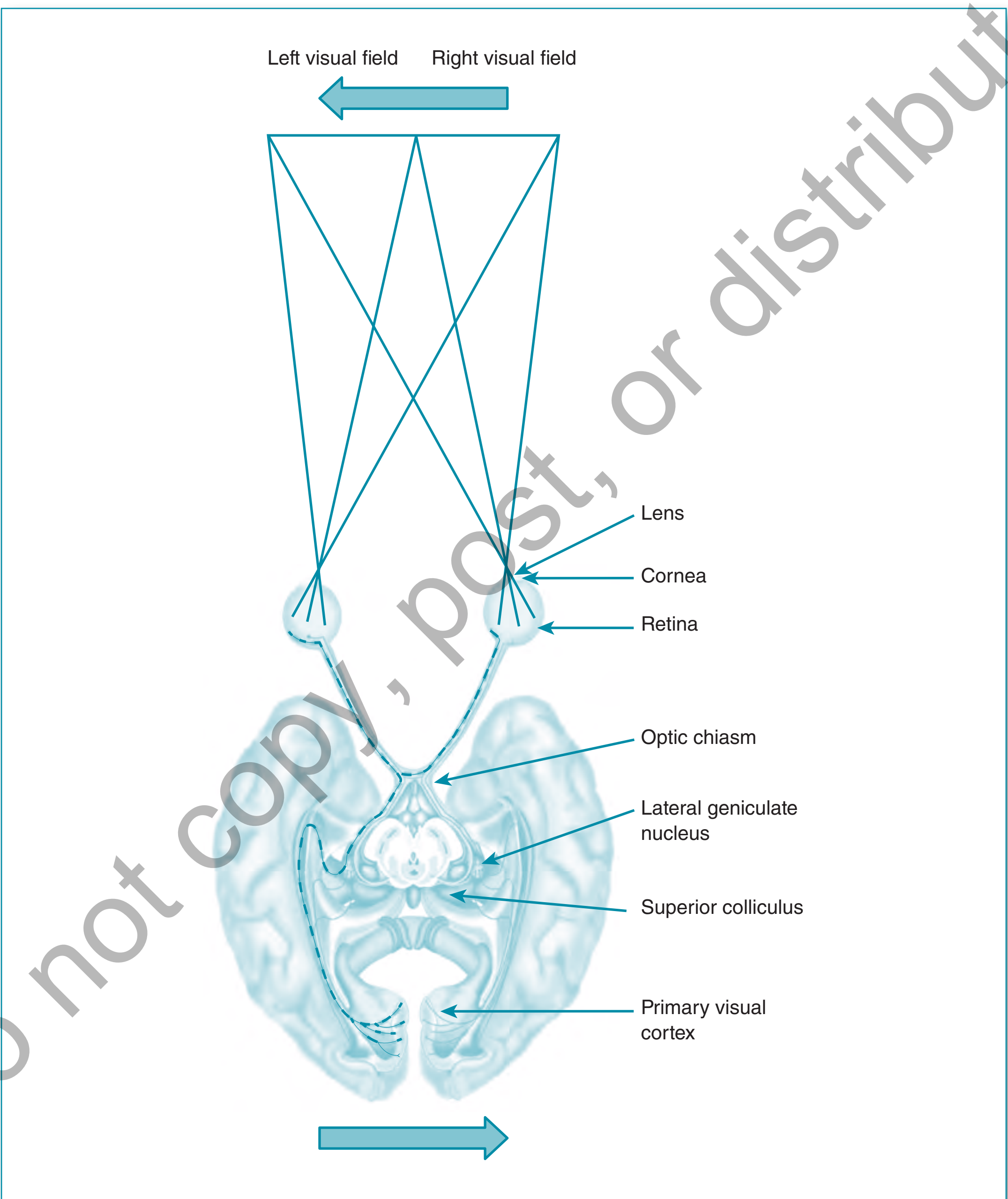
The human visual system has evolved to construct color differences from the variations in wavelength in the visible spectrum. Long wavelengths are red because of the manner in which the brain processes the light and represents it differently from the short wavelengths of blue or violet. In a similar way, the human brain has developed the capacity to perceive objects, faces, and the sounds of speech. All of these are significant perceptual inputs that are necessary for human survival. It is important to keep in mind, however, that the brain developed these capabilities early in human evolution. This means that other kinds of perceptual abilities may be beyond our capacity today, even though they are now important for survival. To return to the driving example, the human visual system is remarkably poor at detecting small changes in the deacceleration of fast-moving objects. Thus, the car ahead on the freeway may be slowing, or even coming to a full stop, but the brain cannot readily pick up this information until a collision is imminent (Evans, 1991). Our poor sensitivity to deacceleration obviously was irrelevant in prehistoric human existence, when everyone was walking. But, on the freeway today, it is potentially a matter of life and death. This weak link in human perception is why center-mounted brake lights are installed on vehicles. The center mounts typically place the light within foveal vision (sharp central vision), and the red light attracts the attention of the driver behind.

## Visual Pathways

The cornea and the lens within the eye work together to bring the light reflected from an object into focus on the retina, the structure containing the photoreceptors. This focus occurs in the fovea of the retina, where detailed vision occurs. Failure to achieve a focused image is the cause of vision problems such as an inability to see clearly a close object (farsightedness) or a distant object (nearsightedness). The neural signals generated in the retina are sent via the optic nerve to a portion of the thalamus lying deep in the brain called the lateral geniculate nucleus, as shown in Figure 2.2. The thalamus receives inputs from auditory and other sensory channels, in addition to vision. The pathway continues to the primary visual cortex in the occipital lobe.

The pathway from the optic nerve exiting the left eye projects to both the left and right hemispheres of the occipital lobe. Similarly, inputs to the right eye are sent to and processed by both hemispheres. As shown in Figure 2.2, the axons of the optic nerve cross over to the opposite side of the brain at the optic chiasm. Here, the axons of the optic nerve from the inner, or nasal, half of each retina cross over to the opposite side of the brain; those from the outer, or temporal, half remain on the same side of the brain. This arrangement results in a division of labor in vision so that the objects in the left visual

**Figure 2.2** The visual pathways result in the representations of stimuli from the left visual field projecting to the right visual cortex and vice versa.



field are processed by the right hemisphere and those in the right visual field are processed by the left hemisphere. Representations of stimuli presented to the left visual field project to the right visual cortex, whereas those to the right visual field project to the left visual cortex.

Although most signals follow the pathways just described, about 20% of the signals leaving the retina are projected to another structure lying at the top of the midbrain called the superior colliculus (Schiffman, 2000). This region controls eye movements. Importantly, then, some signals from the retina are processed by regions that do not terminate in the primary visual cortex. The significance of this pathway will soon become clear.

## Visual Cortex

Seeing something rather than nothing depends on the processes that occur in the primary visual cortex (Crick, 1994). Visual consciousness hinges on more than a functional retina, an optic nerve, and a lateral geniculate nucleus. The occipital cortex must also function properly for one to be aware of an object in the visual environment. Two findings strongly support this conclusion.

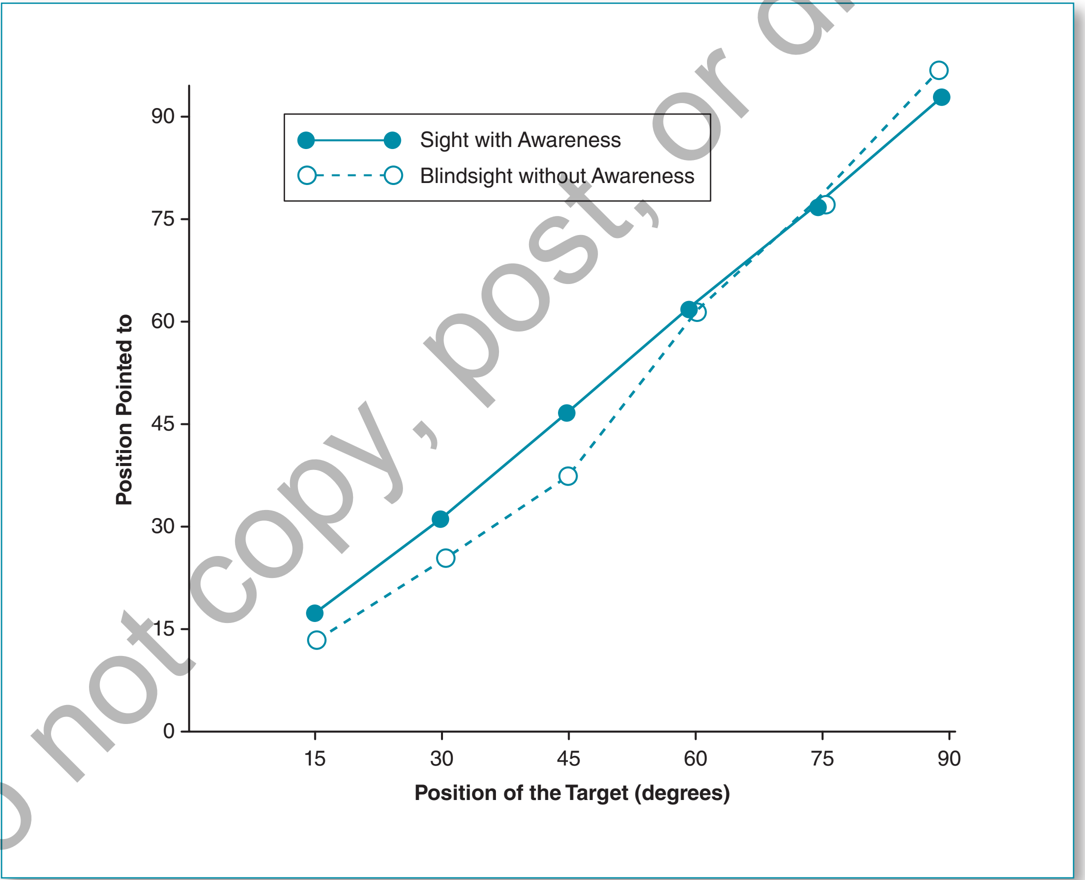
Experiments on the development of the neurons in the visual cortex have shown that there are critical periods during which stimulation must be received for normal development. In cats, the critical period begins during the first few weeks of life and lasts about three or four months. In human beings, the critical period may extend much longer, to four or five years (Schiffman, 2000). To illustrate, Blakemore and Cooper (1970) raised kittens in an environment that restricted the kind of visual stimulation they received. The kittens were kept in darkness for all but about five hours a day. During this time, they lived in an environment consisting only of horizontal lines for one group of kittens and vertical lines for another group. After about five months of this selective exposure, the kittens were tested for their visual awareness of horizontal and vertical lines.

Some tests used single-cell recordings from the primary visual cortex. The cortical cells responded to the orientation of lines received early in life. In the kittens that had been exposed only to horizontal lines, for example, the cells of the primary visual cortex fired at above baseline rates only to horizontal stimuli. Of critical importance, when a black bar was held horizontally, the kittens initially exposed to horizontal lines batted at the bar in play. Their behavior suggested that they could see the bar. By stark contrast, the kittens raised in the vertical line environment ignored the horizontal bar, which implied that the cortical cells were needed for visual awareness.

A second result from a neuropsychological case study confirmed the conclusion that the cortex is necessary for visual consciousness. The patient, known as "D. B.," was a 34-year-old male who suffered from severe migraine headaches. All treatments failed, and the migraines became so severe that surgeons took the extreme step of removing part of his occipital cortex. The surgery was successful in reducing the intensity of the migraines, but it left D. B. blind in about a quarter of his visual field, specifically with respect to objects presented to his left. A test light was presented on a screen situated in front of D. B. The location of the target was varied from trial to trial in a random way, and D. B. was asked to point to its location.

As shown in Figure 2.3, when the target was presented to his normally sighted visual field to the right, D. B.'s pointing responses tracked the actual location of the target, producing a straight line for sight with awareness. Astonishingly, D. B. performed nearly as well when the target was presented to his blind left visual field. Although D. B. reported no visual awareness of anything on these trials, his pointing responses closely (but not perfectly) tracked the target (Weiskrantz, 1986). Despite his lack of conscious perception in these regions, when D. B. was encouraged to guess where the test light had occurred, he was remarkably accurate.

**Figure 2.3** Blindsight enabled a patient to point to the location of a target unaccompanied by any visual awareness of it.



*SOURCE:* From Weiskrantz, L., Warrington, E. K., Sanders, M. D., & Marshall, J., Visual capacity in the hemianopic field following a restricted occipital ablation, in *Brain*: *A Journal of Neurology,* copyright © 1974. Reprinted with permission of Oxford University Press.

Vision without awareness as a result of lesions in the occipital cortex is called **blindsight.** It demonstrates the necessity of intact cortical regions for visual consciousness. Apparently, D. B. succeeded in the location task even in his blind field of vision by using information processed in the superior colliculus. This structure deep in the midbrain controls eye movements and seems to have allowed D. B. to identify the location of an object not consciously seen.

Blindsight is vision without awareness that can be observed in patients with lesions in the occipital cortex.

## PATTERN RECOGNITION

The term **pattern recognition** refers to the step between the transduction and perception of a stimulus in the environment and its categorization as a meaningful object. There is more to seeing or hearing than simply perceiving the patterns of light or sound available in the environment. It is necessary to categorize the object on the basis of its perceived features. Look at the drawings in Figure 2.4. Each drawing shows the same object from a different point of view. Although the visual information received by the retina in each case is quite different, the same object is easily recognized. The visual features are perceived and then used to categorize the object as a dog. For recognition to succeed, the mental representation of dogs, and Shelties in particular, must be retrieved and matched against the visual features that are perceived in the drawing.

The ability to perceive depends on pattern recognition—categorizing objects and events detected in the environment by matching their preliminary representations with patterns stored in long-term memory.

### Agnosia

As a result of a neuropsychological condition called agnosia, a stimulus can be perceived and understood in terms of its properties but not recognized as a meaningful object. Patients suffering from lesions in certain regions of the brain can see objects but not recognize them at all. Such individuals are not blind; nonetheless, they fail to "see" in the fullest sense because for them, pattern recognition has failed.

For example, Sacks (1970) described a man identified as "Dr. P." who suffered from a massive brain tumor or degenerative disease that destroyed portions of his occipital cortex. Dr. P. taught music at a local school and appeared to Sacks as a cultivated man with great charm, humor, and imagination—certainly not someone suffering terribly from a serious brain disorder. However, on closer examination, it became clear that Dr. P. suffered from a form of visual agnosia, specifically an inability to recognize objects clearly from their shapes. For example, during a neurological examination, Dr. P. had removed his shoe as part of a reflex test. When asked to put his shoe back on, Dr. P. seemed baffled as he stared

**Figure 2.4** An example of pattern recognition in which different features in each drawing are categorized as a Sheltie dog.

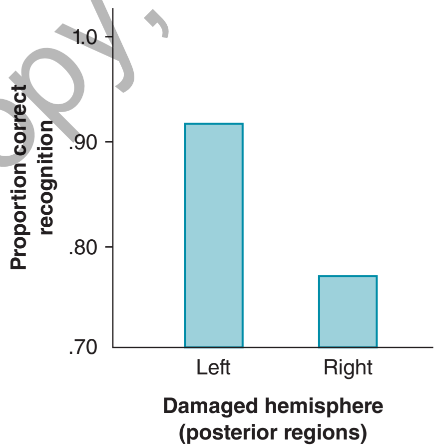
intently at his foot, put his hand to it, and said, “This is my shoe, no?” Stunned, Sacks replied, “No, it is not. That is your foot. There is your shoe.” “Ah!” exclaimed Dr. P., “I thought that was my foot” (p. 9). The damage to Dr. P.’s brain had impaired his ability to pick up the concrete textures and other details of visual experience. Because the outline of his foot matched the outline of his shoe, he could not distinguish between the two. As Dr. P. prepared to leave the examining room, he “reached out his hand and took hold of his wife’s head, tried to lift it off, to put it on. . . . He had apparently mistaken his wife for a hat! His wife looked as if she was used to such things” (p. 10).

Two kinds of visual agnosia have been documented, one resulting from damage to the right hemisphere and the other resulting from damage to the left hemisphere (Gazzaniga et al., 1998). In both cases, the primary visual cortex is intact and supports the ability to see objects in the visual field, but the objects cannot be recognized. Normally, human beings can recognize an object despite wide variations in the details of how the object looks. A dog is a dog, no matter its distance, its orientation, or the angle of viewing. In the case of **apperceptive agnosia,** such ready object recognition fails as a result of difficulties in identifying the visual features that define a perceptual category.

E. K. Warrington (1982) discovered that patients with damage to the rear or posterior region of their right hemispheres made frequent errors in recognizing objects presented at unusual angles. In the example given earlier, a picture of a dog from the front, showing its head in relation to its body, was easily recognizable by all patients in the study. Yet when the picture was taken from behind, without the dog's face or feet in the picture, patients with posterior right hemisphere damage often made mistakes, as shown in Figure 2.5. By contrast, other patients with posterior damage in the left hemisphere were able to succeed on this test with a high level of accuracy. Other data showed that damage to the anterior regions of either hemisphere did not cause a problem on the unusual views test, leading to the conclusion that the right posterior hemisphere is critical for successful perceptual categorization.

In the case of **associative agnosia,** object recognition fails because of difficulties in identifying the functional features that define a semantic category. The problem is not at all perceptual in nature. Instead, the sufferer of associative agnosia cannot categorize objects successfully at an abstract level of meaning. The unusual views test that trips up

**Figure 2.5** Perceptual categorization fails in apperceptive agnosia because of posterior right hemisphere damage.



*SOURCE:* From Warrington, E. K., Neuropsychological studies of object recognition, in *Philosophical Transactions of the Royal Society of London,* 298B (1982). Reprinted with permission.

patients with apperceptive agnosia fails to bother those with associative agnosia. By contrast, a test that requires matching objects in terms of semantic categories while ignoring their appearance causes problems for individuals with associative agnosia. For example, suppose that an individual is shown a cane, a closed umbrella, and an open umbrella, and is asked to identify which two objects have the same function. An individual with perceptual agnosia has no difficulty in seeing the open and closed umbrella as representing the same semantic category. However, individuals with associative agnosia often fail to do so; they cannot see beyond the perceptual similarities of the cane and the closed umbrella.

These two kinds of agnosia demonstrate that the pattern-recognition process involves two separate levels of categorization. The visual features of an object must first be matched against representations in long-term memory that identify perceptual categories. Variations in how an object looks (e.g., its orientation, the angle of viewing) must be ignored, whereas features that do matter (e.g., eyes, ears, fur, tail) are heeded. This perceptual level of categorization appears to be mediated by posterior regions in the right hemisphere and occurs prior to semantic categorization (E. K. Warrington, 1985). As can be seen in patients with associative agnosia, however, it is possible to see two objects as alike perceptually (e.g., a cane and a closed umbrella) and to fail to see that they belong to different semantic categories and have different names. The functional features of an object must also be matched against representations stored in long-term memory to identify semantic categories and names. E. K. Warrington (1985) contended that this second stage is dependent on processes supported by the left hemisphere.

*Apperceptive agnosia* refers to a failure of pattern recognition caused by an inability to categorize objects at a perceptual level of analysis. *Associative agnosia*, by contrast, is caused by an inability to categorize objects at a functional semantic level of analysis.

## Top-Down Versus Bottom-Up Processes

A **schema** is a mental representation that organizes knowledge about related concepts. Imagine, for a moment, the classroom that you attend for cognitive psychology. In forming a mental picture of this particular environment, you activate a schema that represents what you know about classrooms in general and their relations to other types of rooms. The schema involves many concepts, such as those of a room, a desk, a table, a computer, an overhead projector, a projection screen, and a video recorder. In imagining each of these objects, you activate its conceptual representation, which represents what you know about the general characteristics of a category of objects—for example, tables. Organized knowledge representations or schemas direct exploration of the environment to sample features of the objects and events to be perceived.

As you walk into the building on campus containing your classroom, your mind unconsciously begins to anticipate the objects and events that you will soon see and hear. These anticipations play a vital role in directing exploration of the environment (Neisser, 1976). The steps you take, the way you turn your head, the objects you reach for and grasp, and the eye movements you make are directed by your expectations. For example, the eye movements made to explore the environment are guided by your immediate goals (Yarbus, 1967). If you anticipate seeing a particular friend in the classroom, for example, then your eyes will quickly scan the faces of people to confirm your expectation. If expectations are violated by novel, surprising events, then these are explored extensively. For example, suppose that a student brings his pet boa constrictor to class one day. People, desks, and books are expected in a classroom—but not snakes. The surprising object would be scrutinized immediately.

Top-down or **conceptually driven processes** reduce the need to sample all of the information available in the environment by providing the perceiver with expectations. Simultaneously, bottom-up or **data-driven processes** analyze the edges, lines, areas of light and dark, colors, sounds, and other physical features available briefly in sensory memory. These processes pick up the features needed to confirm or refute expectations. Through such simultaneous processing from both the bottom up and the top down, people can perceive the features of the environment with remarkable quickness and accuracy.

> Paradoxically, a single letter is identified faster when in the context of an entire word than when isolated. This word superiority effect is caused by top-down or conceptually driven expectations activated by the word.

The contribution of each type of process depends on the perceptual circumstances (Shepard, 1984). Strong bottom-up activation occurs when perceiving under good viewing conditions. In poor, ambiguous viewing conditions, accurate perception depends more strongly on top-down than on bottom-up activation. Very strong top-down activation is responsible for the hallucinations experienced nightly in dreams. The lack of any significant external input during dreaming might be why it is experienced as real (Antrobus, 1991). Daydreaming, or imagining an event while awake and concurrently processing some external events, also depends on top-down activation, but it is less intense and is not experienced as real.

In the laboratory, several experiments have shown that the speed as well as the accuracy with which a person can identify an object depends on the context in which the process occurs (Biederman, Glass, & Stacy, 1973; Friedman, 1979; Palmer, 1975). One expects to see a cow in a farm scene or a fireplug in a city street scene. Putting the cow and the fireplug in the wrong scene measurably slows one's ability to recognize them by pitting top-down processes against bottom-up processes. Preventing the activation of a schema or frame—by removing or scrambling the context so that it looks incoherent—also hinders pattern recognition by requiring that all of the work be done from the bottom up.

Perceiving each word on this page as you read is conceptually driven, in part. Reicher (1969) presented a word (WORK), a nonword (ORWK), or a single letter (K) as a stimulus to participants. A mask (####) then appeared that stopped the processing of the original stimulus by filling the contents of iconic memory with irrelevant visual elements. Probe letters also appeared above (D) and below (K), the fourth element of the mask. The observers then guessed which of these had occurred earlier. Surprisingly, the letter *K* was correctly identified more often when it appeared in the word than when it appeared in isolation. The **word superiority effect** refers to a single letter being recognized faster in the context of a whole word than when presented as an isolated letter. The word activates conceptually driven processes that ease the recognition of each individual letter. The nonword stimulus fails to activate these processes and so supports the same level of identification accuracy as does the single letter.

Conceptually driven processes operate from the top down—from long-term memory to sensory memory—to identify the stimulus. Data-driven processes operate from the bottom up—from sensory memory to long-term memory—to achieve the same goal.

Tulving, Mandler, and Baumal (1964) showed how varying amounts of context provided in reading speeds word recognition. They presented a target word either with no context (0 words), as the last word of a phrase (4 words), or as the last word of a sentence (8 words). The more context provided, the more conceptually driven processes should aid recognition of the target. As shown in what follows, the participant first read the context, if given, and then briefly viewed a target word such as *opponent*:

opponent

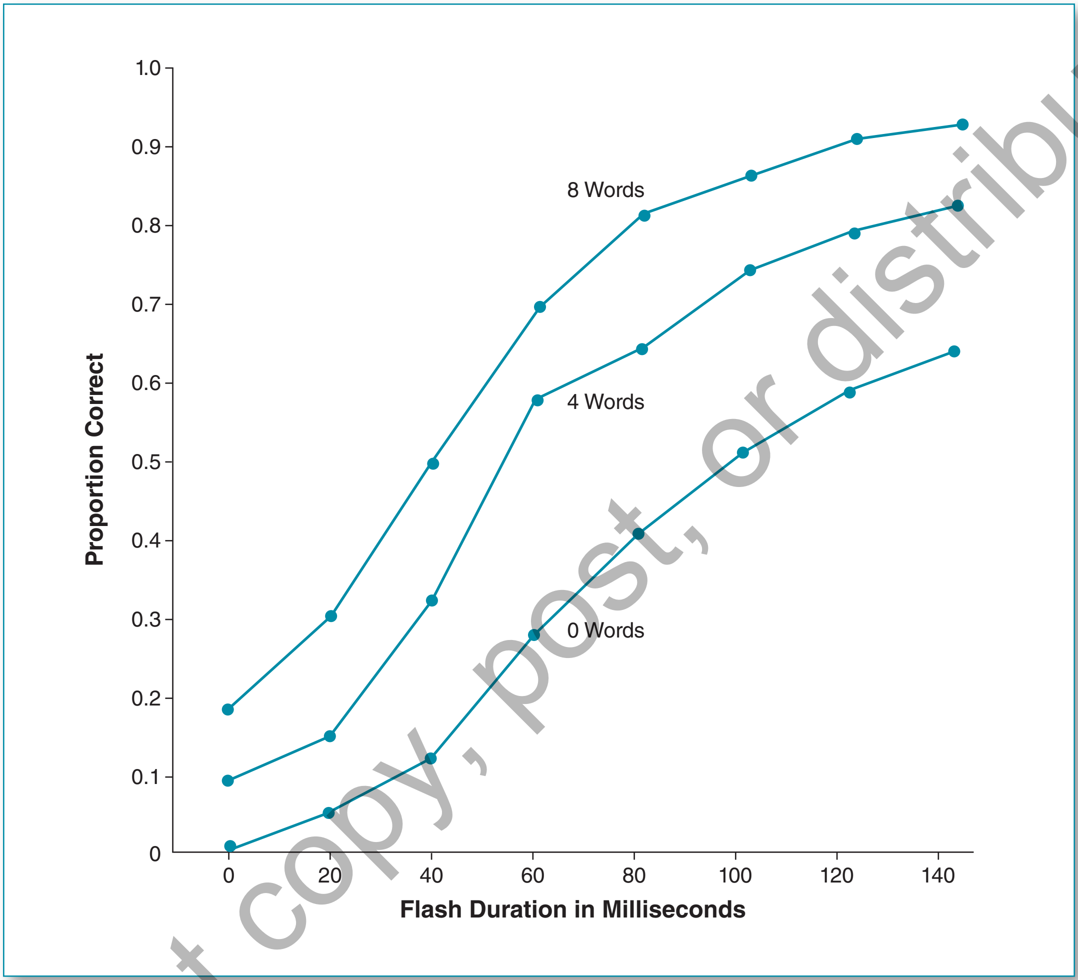
challenged by a dangerous opponent

The political leader was challenged by a dangerous opponent

Tulving et al. (1964) also varied the exposure duration of the final word, *opponent*, from 0 to 140 milliseconds. The longer the exposure, the more data-driven processes should aid recognition. Note that in the zero condition, only conceptually driven processes are at work, allowing perhaps a correct guess about the target word. As can be seen in Figure 2.6, with eight words of context, the proportion of correct recognition averaged nearly .20. The systematic increases with longer exposure durations show the role of making more data available from the bottom up. The differences among the eight-, four-, and zero-word curves show the role of more precise expectations working from the top down.

Another laboratory phenomenon that is at least partly explained by conceptually driven processes is **change blindness.** Suppose that as you spoke with a person, a different person were surreptitiously substituted. Would you notice? Suppose that in viewing a photograph of two people, the heads were surreptitiously exchanged as your eyes sampled features from one part of the picture to another. Would you notice that? People assume they would, but the results from experimentation show otherwise. Nearly 50% of observers missed these kinds of changes in the visual environment (Simons & Ambinder, 2005). Change blindness refers to the phenomenon that people fail to notice large changes in visual scenes. Limited attention and other factors probably are also involved, but it is clear that expectations that observers have about their visual environment play a major role in the features that they sample from the environment. Data-driven processes do not notice these changes because they are not expected.

**Figure 2.6** Word recognition varies with the amount of context provided.



*SOURCE:* From Tulving, E., Mandler, G., & Baumal, R., Interaction of two sources of information in tachistoscopic word recognition, in *Canadian Journal of Psychology, 18,* © 1964. Reprinted with permission.

## Object Representations

If pattern recognition requires matching perceived information against perceptual representations stored in long-term memory, then what is the nature of these representations? Research into this question has examined several possibilities, but a firm answer to the question remains elusive. One possible solution is that perceptual concepts are stored as lists of **distinctive features.** For example, the distinctive features of block letters can be specified readily. Some features are straight lines at particular angles (e.g., *E, M*) and others are curves (e.g., *O, C*). A list of a relatively small number of distinctive features allows a complete specification of the printed alphabet (Gibson, 1969).

## Learning Activity 2.1

As fast as you can, count the number of times that the letter *X* occurs in the first block of letters shown below. Next, again count the number of times *X* occurs in the second block.

BCGQOXPSXQPBGUPXQRBCQPRGBUXPQPSRUCBPUPXQURSQPUCBCGXUQPUCBXUS

TYAZTXIKFWFTMNXLIAZLXVFTELNAWLXWYTAZXMKLNHWHVZLATXMZVFITAXZY

Which block of letters required more time to count the number of *X*s? Why was it more difficult?

**Feature Detectors.** There is powerful evidence that the visual cortex of mammals is organized to detect the presence or absence of simple features. Hubel and Wiesel (1959, 1963) presented an edge, a slit of light, or a darkened bar at different orientations to the eyes of a cat or a monkey. At the same time, they recorded the neural activity in single nerve cells in the occipital lobe of the lightly anesthetized animal. Hubel and Wiesel discovered that the cells were tuned to respond maximally to bars of a particular orientation. For instance, some cells fired rapidly in response to a vertical bar, whereas others preferred a horizontal bar.

Distinctive features differentiate objects during pattern recognition. Neural cells in the occipital cortex are tuned to fire when stimulated by simple lines presented at a particular orientation.

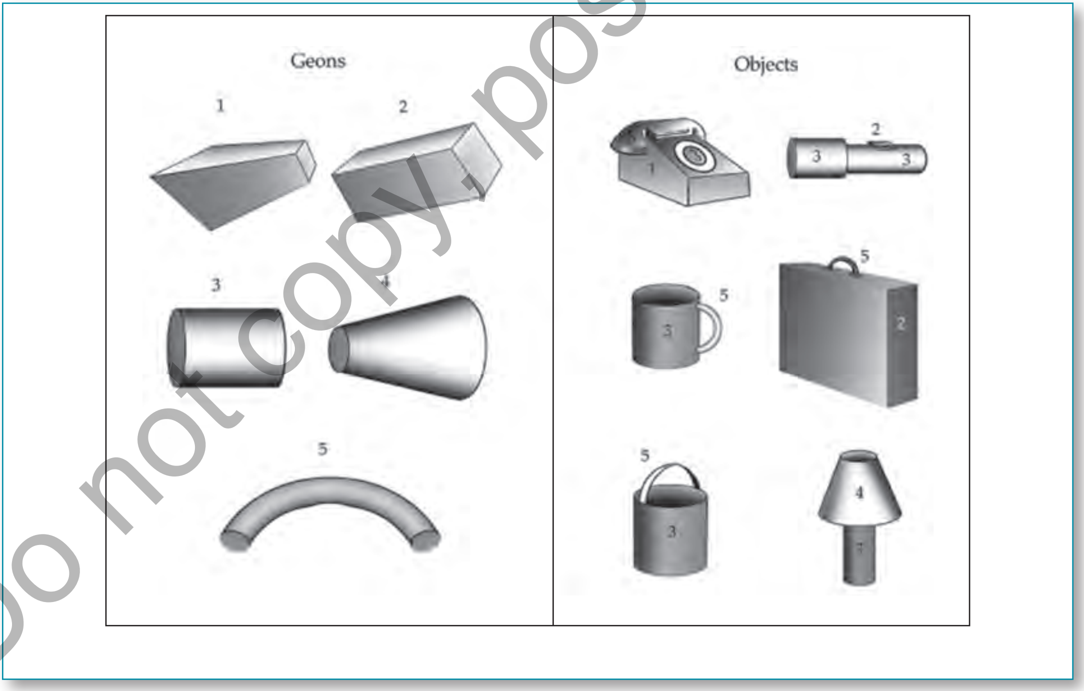
In human vision, evidence for feature detection can be seen in visual search tasks. Neisser (1963) asked for the participant to search for a particular letter among a long list of lines of printed letters, similar to the task illustrated in Learning Activity 2.1. In one condition, the letter shared many features with the distractors; for example, the subject was asked to search for *Z* among *T, L, K, M, V,* and other letters with straight lines. In another condition, the target letter, say *Z*, stood out clearly from the distracters such as *O, Q, P, B, D,* and other rounded letters. The more rapid search times obtained by Neisser in the second condition, in which the target stood out, suggest that the human visual cortex analyzes stimuli in terms of simple component features. Note that if people compared each letter to a unique, unanalyzed template in memory, then their search time ought to be the same whether the distracters were straight or rounded. This is one of several experimental results at odds with a theory holding that mental representations of objects are simply unanalyzed templates (Hummel & Biederman, 1992). Instead, the distinctive features that make up a given object are an important part of its mental representation.

**Structural Descriptions.** Other researchers have explored a problem with the feature detection theory of pattern recognition. Specifically, they have shown that the relations among features are as important to recognition as the features themselves. A letter *Z* is not simply three independent lines at certain angles. The lines must be structured in accordance with the rules for constructing the letter *Z*. A face, for instance, is not simply a collection of features positioned haphazardly—an eye here, a nose there, a mouth over there. The relations among features must conform to the rules that define the structure of the face. In other words, the whole object is not simply a list of independent features. The relations among features are equally important. One therefore needs a grammar or set of rules for how to put the features together properly (Reed, 1973; Sutherland, 1968). Several other studies have shown that people process the relations among features in perception (Hummel & Biederman, 1992; Reed, 1974; Reed & Johnson, 1975).

Structural descriptions consider not just features, but also the relations among features, to facilitate pattern recognition.

Biederman (1987) proposed a set of 26 basic geometric features (geons) that, when put into

**Figure 2.7** Examples of geons or basic sub-objects involved in scene perception.



*SOURCE:* From Biederman, I., Human image understanding: Recent research and a theory. *Computer Vision, Graphics, and Image Processing*, *32*, 29–73. Copyright © 1985. Reprinted with permission from Elsevier.

proper structural relationships, constitute all visual objects. A sample of these is presented in Figure 2.7. A simple object, such as a cup, comprises the geons numbered 3 and 5. The telephone involves these geons and more that are related very differently than in the cup.

**Figure 2.8** Perception of the object depends on the availability of structural relations at the vertices.

*SOURCE:* From Biederman, I., Human image understanding: Recent research and a theory. *Computer Vision, Graphics, and Image Processing, 32,* 29–73. Copyright © 1985. Reprinted with permission from Elsevier.

An interesting prediction of Biederman's theory is that not only is relational information needed, but it may be more critical to perception than the features themselves. Biederman (1985) deleted 65% of the contours (features) from drawings of common objects, as illustrated in Figure 2.8. For example, the cup on the left retains the vertices so that an observer can pick up the relations among the remaining contours. The cup on the right destroys this relational information by removing contours from the vertices. Biederman found that observers, after a 100-millisecond exposure, could accurately identify the left-hand cup 70% of the time as compared with only 50% for the right-hand cup. The color, texture, and other details that add such richness to our perceptual experience are less relevant to recognition than are the vertices. In support of this, Biederman and Ju (1988) found that schematic line drawings are indeed recognized as quickly as color photographs of objects.

## MODULARITY

Thus far, it is clear that explaining how one is able to see and recognize a familiar object is a nontrivial problem. The visual system constructs a mental representation that allows the object in the environment to be both seen and understood as meaningful. The final two sections of this chapter show that particular kinds of pattern recognition invoke specialized processes. The evidence to date suggests that the perception of faces and that of speech each draws on processes that have evolved to cope with the particular demands of the task. A module refers to a set of processes that are automatic, fast, encapsulated apart from other cognitive systems, and instantiated in a localized area of the brain (Fodor, 1983). There may be several modules, each dedicated to the perception of an important class of stimuli such as faces or speech.

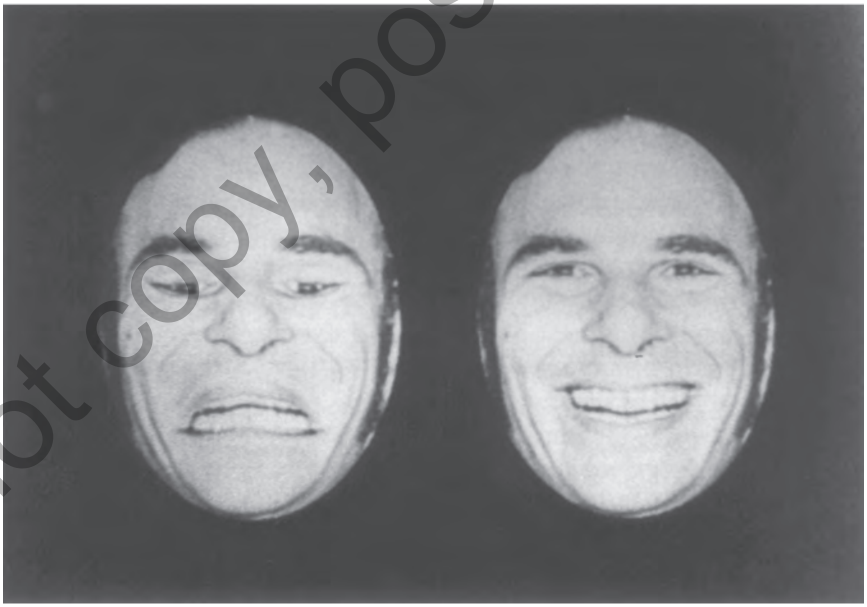
Social interactions are crucial to our survival and reproduction, and these depend on the ability to recognize faces and speech. Therefore, perhaps it is not surprising that the cognitive system includes specialized modules for processing these categories of stimuli. Obviously, speech is central to communication between two or more human beings. Less obviously, facial expressions provide a key means for communicating emotional states. Through body language and particularly facial expressions, human beings communicate whether they feel happiness, sadness, or anger, for example. The role of holistic processing and modularity in face processing is presented next, and then the chapter concludes with a consideration of speech perception.

### Holistic Versus Analytic Processing

Although the theorist can identify the features or parts that make up a whole object, an observer may perceive only the meaningful object. Several factors control the extent to which perception is dominated by the whole versus the parts, including the type of stimuli presented to the observer and the task required of the observer (Treisman, 1987). **Holistic processing** refers to perceiving the whole object; **analytic processing** refers to perceiving the features that compose the whole. Holistic processing, then, involves the spatial-relational aspects of the features of the whole face. Analytic processing targets the nose, the eyes, the lips, and other specific features instead of their relations. Faces, perhaps more than any other object, are perceived holistically rather than analytically. An intriguing demonstration of this fact comes from an illusion that occurs when the normal orientation of facial features is inverted (P. G. Thompson, 1980).

First, study the pair of faces in Figure 2.9 in the normal orientation. In the face on the left, the eyes and mouth have been turned upside down. As you can see, the face takes on a grotesque appearance as a consequence. Now, turn the book upside down and study the two faces again. Notice that when viewing the faces in an unusual orientation, the grotesqueness disappears. Both faces take on a normal appearance. This demonstrates that in the normal orientation, holistic processing heavily influences face perception. The individual features are encoded, but so, too, are their spatial relations that together compose the whole face. When the normal relations among the eyes, nose, mouth, and eyebrows are rearranged, the face looks grotesque. But the holistic processing of the face can be disrupted by inverting the face 180 degrees, a position that we rarely encounter in everyday perception. The face as a whole no longer dominates perception; the individual parts of the face are taken on their own terms and appear perfectly normal to the eye.

**Figure 2.9** A demonstration of holistic processing of faces.



*SOURCE:* From Bartlett, J. C., & Searcy, J., Inversion and configuration of faces. *Cognitive Psychology, 25,* 281–316. Copyright © 1993. Reprinted with permission from Elsevier.

Several studies have shown that face perception is more vulnerable to inversion than is perception of other kinds of objects (Searcy & Bartlett, 1996; Valentine, 1988). Murray, Yong, and Rhodes (2000) found that as a face is rotated from 0 to 180 degrees, there is a discontinuity in its appearance. Up to rotations of 90 degrees, a normal face looks increasingly bizarre, whereas a distorted face looks less and less bizarre. Between 90 and 120 degrees of rotation, the distorted face begins to look fine and continues to do so until it is inverted a complete 180 degrees. This is not the case with the normal face, which continues to look more bizarre as it is rotated up to 180 degrees. Try rotating the page of this book and notice what happens to the distorted face somewhere between 90 and 120 degrees. Inversion disrupts the holistic processing of spatial-relational information more than it disrupts the analytic processing of features.

## Face Perception

Perception of faces is unique in that it is more strongly influenced by holistic processing than by analytic processing.

Why is it that upright faces are perceived more holistically than analytically? Farah (1990, 1998) presented several lines of converging evidence pointing to the existence of a specialized module for face recognition. **Prosopagnosia** is a selective inability to recognize faces that does not involve other kinds of vision difficulties. A prosopagnosic patient cannot recognize the photographs of famous individuals, but when the patient is tested on other kinds of complex visual discrimination tasks, no deficit is found (MacNeil & Warrington, 1993). For example, a sheep farmer had no problem in distinguishing photographs of his own sheep from pictures of other sheep despite their close similarity in appearance. Yet his recognition of individual human faces was profoundly impaired.

If the holistic processing of spatial relations in faces is driven by a specialized module, then what would happen if this module were damaged, as in prosopagnosia? Normal controls have more difficulty recognizing inverted faces as compared with upright faces because the module constructs an accurate representation of the test face. If damaged, the module would provide inaccurate information and disrupt performance. Farah (1990) discovered that a prosopagnosic patient actually correctly identified more faces when they were inverted (72%) than when they were upright (58%). Normal controls showed the expected pattern of more correct identifications with upright faces (94%) than with inverted faces (82%). By inverting the face, the damaged module in the prosopagnosic patient was removed from play, thus improving performance.

Farah (1990) also discovered that damage to the occipital and temporal cortices, usually bilateral damage in both hemispheres, was correlated with prosopagnosia. A localized region of the temporal lobe seems to be crucial for face recognition. One way to demonstrate this localization is to examine other neurological disorders and compare their effects on performance on different kinds of tests. Reading and face recognition tests both deal with complex visual tasks that reveal a double dissociation. There are a variety of kinds of lesions in the brain caused by head injuries or strokes that are collectively called acquired dyslexia. Relative to normal controls, dyslexic patients perform poorly on a reading test but show no impairment on a test of face recognition. By contrast, prosopagnosic patients show deficits on the face recognition test but not on the reading test, relative to normal controls. Because brain damage can be extensive, some of these patients had trouble recognizing objects of any kind—that is, they suffered agnosia in addition to prosopagnosia or dyslexia. However, patients rarely suffered from a combination of dyslexia and prosopagnosia. In short, prosopagnosia was uncorrelated with acquired dyslexia, suggesting that reading and face recognition are handled by different structures in the brain and can be selectively damaged.

Another source of evidence in favor of a face-recognition module comes from normal college students in an object versus part recognition task. Participants learned the names of normal upright faces and objects during the first phase of the experiment. Next, they were asked to recognize the faces or objects in the whole condition. For example, they either were shown a face and asked, "Is this Jim's face?" or were shown a house in the object condition and asked, "Is this Jim's house?" In the part condition, they were tested on particular features of the studied faces or objects. For example, they might be shown a nose in isolation and asked, "Is this Jim's nose?" or they might be shown a door and asked, "Is this Jim's door?" Recognition was just as good for the parts of houses as for whole houses. However, recognition was substantially less accurate in the part condition than in the whole condition for faces. The participants had difficulty processing the faces in an analytic manner—zeroing in on, say, the nose by itself—during study or test. This outcome is understandable if faces are processed holistically by a specialized module.

Individual differences in holistic processing ability are a reliable predictor of face recognition ability, too (Richler, Cheung, & Gauthier, 2011). People who are especially sensitive to the holistic configuration of a face are most bothered by a task where the top half and the bottom half are misaligned. These individuals turn out to be the best at recognizing faces in a standardized assessment called the Cambridge Faces Memory test.

Studies using fMRI have revealed a network of brain regions in the occipital and temporal lobes that together comprise the face-processing module. These include the fusiform face area in the temporal cortex and the occipital face area in the inferior occipital cortex. Although these regions are bilateral, they show a strong preference for facial stimuli in the right hemisphere (Steeves et al., 2009). The precise contribution of these regions to face perception is still uncertain, but an important clue came from a comparison of two patients with prosopagnosia. Both patients had a lesion in the occipital face area of the right hemisphere, but the right fusiform face area showed a normal range of fMRI activation in response to the general category of faces. Steeves et al. suggested that the occipital face area provides the detailed information needed to perceive an individual face, once it has been categorized by the fusiform face area.

A specialized module is responsible for face perception. Other modules may exist for specific kinds of perception such as speech recognition.

A module is adaptive because it quickly extracts perceptual information from the environment. For example, the rapid speed of the face-processing module enables one to extract enough information from even a very brief exposure to a face to form a first impression of the individual's personality. Willis and Todorov (2006) proposed that making inferences about personal traits—such as attractiveness, likeability, trustworthiness, and competence—is a fast, automatic, intuitive process. They found that with only 100 milliseconds' exposure to a face, participants' judgments about a face were highly correlated with the judgments made when participants were given an unlimited time to examine the face. Increasing the time up to 500 milliseconds did not change these correlations at all, although participants became more confident in their judgments, and their judgments of all the test faces became somewhat more negative with additional exposure time. Finally, doubling the exposure time from 500 milliseconds to a full second did little but increase confidence still further. This is not to say that slow, reflective, and conscious deliberation does not have a role in helping us to discriminate between individuals on the basis of, say, attractiveness versus competence. Also, conscious processes help build confidence, but the first impressions are mediated by unconscious processes.

The amygdala's role in the rapid, unconscious, and automatic sensing of fearful stimuli may be at work in the detection of threatening faces. Winston, Strange, O'Doherty, and Dolan (2002) discovered a higher degree of amygdala activation in viewers looking at faces rated as untrustworthy compared with the activation obtained for trustworthy faces. The mechanisms involved in face perception, then, play an important social function. Human beings rapidly and intuitively make judgments about whether an individual is friendly and approachable versus threatening.

Face perception, then, offers a good example of a module and illustrates how its output can combine with processing by an emotional network involving the amygdala. The face module rapidly and automatically processes faces, but its localized brain regions may not be totally dedicated to faces. It has been discovered that after 10 hours of learning to discriminate completely novel complex visual objects—called "Greebles" by the researchers—participants began to show evidence of processing them holistically, just as they would faces. Moreover, the fusiform face area and the occipital face area became selectively more active, as seen in fMRI measurements, to the Greebles (Bukach, Gauthier, & Tarr, 2006). Thus, it may be most appropriate to think of these brain regions as being specialized to process complex, holistic visual objects that are repetitively perceived because of their importance for survival. Because we are a highly social species, human faces plainly fit this description.

## Speech Perception

Another example in perceptual modularity may well be the most remarkable feat of pattern recognition performed by human beings. Having recognized the face of a person standing before you, you then must decipher the sounds being emitted by the person's vocal tract. Spoken language and its comprehension is a vast subject that will be addressed in Chapter 8. The issue at hand here is how the complex, information-packed auditory signals of spoken language are perceptually recognized at an extraordinarily rapid rate, enabling you to understand the words spoken to you.

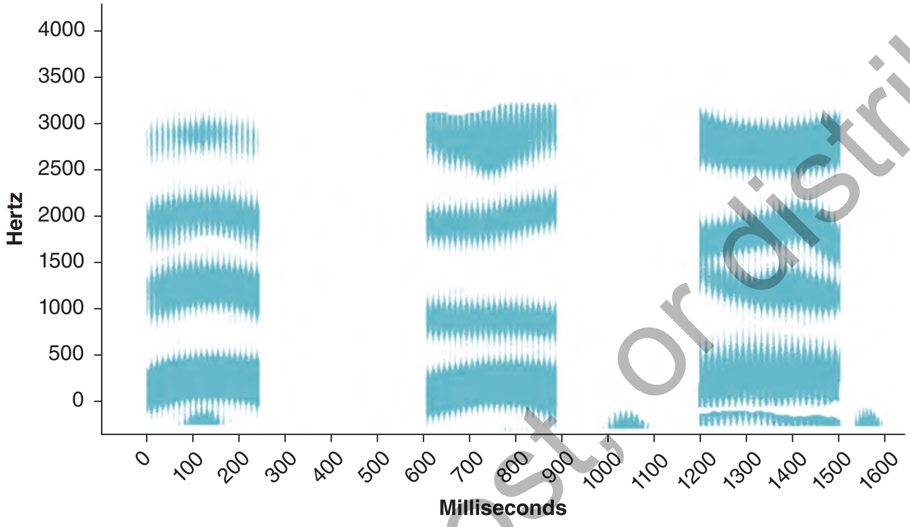
The term *speech perception* does not refer to the comprehension of language, a topic that will be addressed in Chapter 8. Comprehension processes are the same whether the modality of language input is speech that is heard, written language that is read, or sign language that is seen. Rather, the perception of speech refers specifically to the pattern recognition computations that make contact with a word stored in long-term memory from the auditory signal of spoken language. In other words, a phonological representation of the word must be activated by the incoming auditory signal. Just as the brain has dedicated a module to face perception, a module for speech perception allows words to be identified in a fast, automatic fashion without any effortful intention on the part of the listener. The speech-perception module is intricately designed to cope with an astonishingly complex auditory signal that conveys the sounds of language.

Consider that oral language uses basic speech sounds to distinguish words with different meanings. A speech sound or phonological segment that makes a difference in meaning is called a **phoneme.** Each phoneme is pronounced in a distinctly different manner from all others, and this difference in pronunciation signals a difference in the meaning. For example, *pill* and *kill* differ with respect to initial phoneme, and this signals a difference in the meaning of the two words. Now, consider that normal speech unfolds at a rate of about 12 phonological segments per second. The speech perception system handles this rate with ease and can, in fact, cope very well with speech artificially accelerated to 50 phonological segments per second (Foulke & Sticht, 1969). A listener can even understand a speaker who whispers a sentence, despite the fact that whispering alters not only the intensity of the acoustic signal but also its frequency. It has been estimated that the brain must process 40,000 bits of information per second to recognize the phonemes that are the building blocks of spoken language (Fodor, 1983). The fast, automatic extraction of speech signals suggests that it is the work of a module dedicated to the task.

The fact that everyday speech is riddled with noise and indeterminacy makes the task of speech perception all the more daunting (McClelland & Elman, 1986). Unless the speaker formulates complete sentences and articulates them clearly and slowly in a quiet setting, the speech signal is fragmentary. Yet listeners somehow manage to understand speakers who rapidly utter incomplete sentences and even distorted words in noisy environments. A speaker might not articulate clearly, but the listener uses top-down recognition processes to fill in the gaps. Warren (1970) presented listeners with tape recordings of a sentence with a single phoneme deleted, such as "The state governors met with respective legi*latures convening in the capital city." The asterisk marked the spot where the /s/ was removed and replaced with a cough lasting 0.12 seconds. Warren presented the recording to 20 listeners and asked them whether any sounds were missing. Only one individual heard a missing sound, and that person identified the wrong sound as missing. Clearly, the listeners had restored the missing phoneme. Even when the missing phoneme comes at the beginning of the word (e.g., *eel) and is disambiguated by a later word in the sentence (e.g., shoe), listeners rarely report any perception of a gap.

The module dedicated to speech perception is adapted to process a very complex series of acoustic signals arriving at the ears that do not correspond in a one-to-one fashion to the critical speech sounds. To see this, consider a **speech spectrogram,** which represents the physical acoustic energy of an utterance by plotting frequency in hertz or cycles per second on the $y$-axis and time in milliseconds on the $x$-axis. Examples are shown in Figure 2.10 for "bab," "dad," and "gag" spoken with a British accent (Ladefoged, 1982). The darker the

**Figure 2.10** Speech spectrograms for (left to right) "bab," "dad," and "gag" spoken with a British accent.

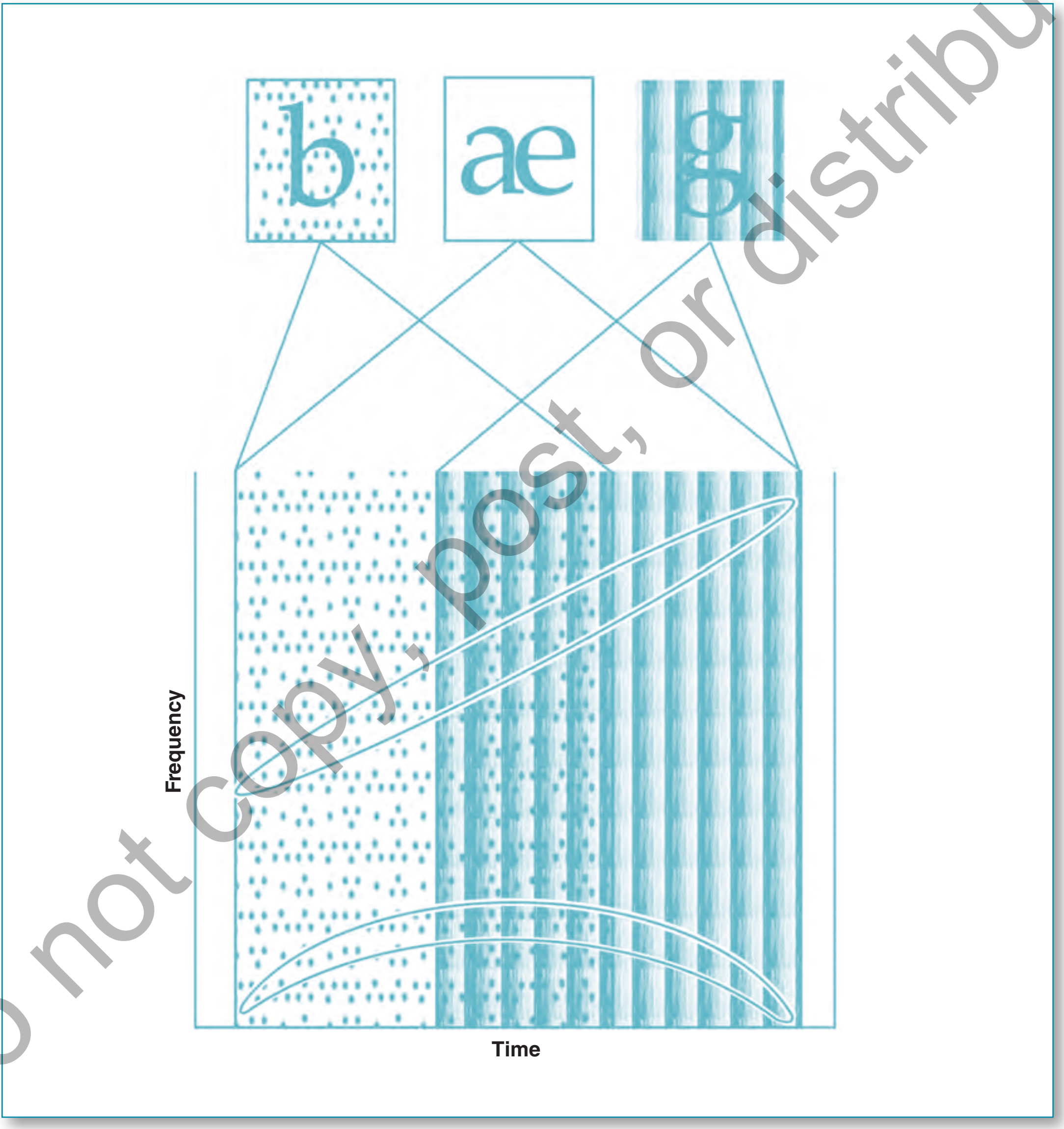


*SOURCE:* From *Course in Phonetics, 2nd edition,* by Ladefoged, P., 1982. Reprinted with permission of Heinle, a division of Thomson Learning: http:// www.thomsonrights.com, fax 800-730-2215.

band of energy at a particular frequency, the greater its amplitude. Notice that the energy clusters at low-, medium-, and high-level frequencies. These bands are called **formants.** The first formant is the lowest frequency band, the second formant is at the next higher frequency band, and so on. One might expect that the spectrum for, say, "gag" could be neatly divided into three time segments, with the early segment providing an invariant feature for the phoneme /g/ followed by one for /a/ and then returning to the one for /g/. It turns out, however, that the three time segments of the speech spectrogram do not match up with the three phonemes /g/, /a/, and /g/.

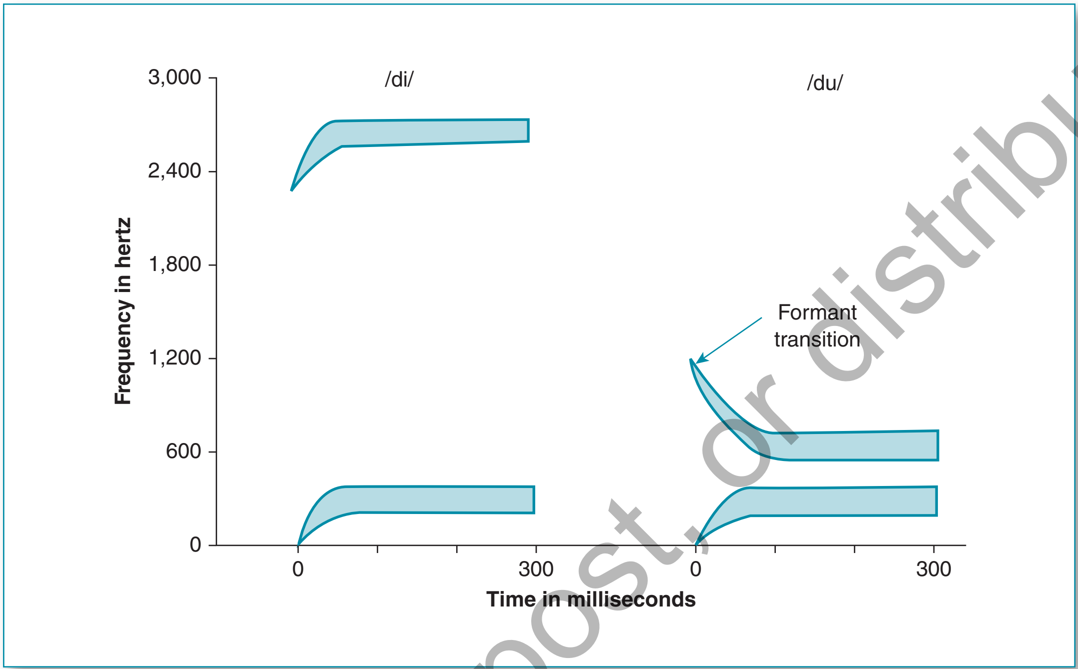
Instead, each segment of the acoustic signal provides clues about the identity of more than one phoneme (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). This is called **coarticulation.** As shown in Figure 2.11, each of the three phonemes of "beg" are being transmitted simultaneously. They are not separated in time, with /b/ followed by /ae/ and then /g/. Instead, the acoustic energy corresponding to the phonetic segment of /b/ overlaps that of the other phonemes. Phrased differently, before you have articulated /b/, the vocal track is already taking shape to articulate /ae/. Notice, too, from Figure 2.11 that you begin to articulate /g/ even before finishing the articulation of /b/. The key point about coarticulation is that multiple phonetic segments are being articulated in parallel at each point in time.

**Figure 2.11** Coarticulation as parallel transmission of phonemes.



*SOURCE:* From Liberman, A. M., Cooper, F., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431–459. Reprinted with permission.

**Figure 2.12** Spectrograms for /di/ and /du/.



*SOURCE:* From Liberman, A. M., Cooper, F., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431–459. Reprinted with permission.

Moreover, the acoustic spectrum fails to reveal a distinctive invariant feature for a particular phoneme that stays the same in all contexts (Liberman et al., 1967). Phonemes lack invariant distinctive features. As illustrated in Figure 2.12, the spectrogram for /di/ versus /du/ reveals different formants for the phoneme /d/, depending on whether it is followed by either the phoneme /i/ or /u/. The first formant is the same in each case. But look at the second formant containing the higher frequencies. As the speaker enunciates the /d/ phoneme, a remarkable change occurs at about 200 milliseconds: The formant turns to higher frequencies when followed by /i/ and to lower frequencies when followed by /u/. Consequently, a listener could not zero in on the acoustic spectrum and identify the phonetic segment of /d/ by matching it with a distinctive feature that remains the same in all contexts.

Both coarticulation and the lack of invariance imply that listeners must process the context in which a given acoustic signal occurs. The relations among features are just as critical as the features themselves. Recall that the same is true in understanding the recognition of visual objects; only a structural theory that specifies both features and their relations is adequate. In speech recognition, a remarkably large number of features and relations must be processed in a fraction of a second simply to identify a single phoneme. Furthermore, unlike the recognition of static visual objects, speech must be recognized over time. Both the sounds that precede a given phonetic segment and those that follow it influence perception (Salasoo & Pisoni, 1985). The contextual nature of the acoustic speech signal enormously complicates the job of the listener. To illustrate, a speaker can produce the phonemes /b/ and /p/ by changing only one feature during articulation. For the /p/ phoneme, the vocal cords do not vibrate, whereas for the /b/ phoneme, they do. The listener must detect this difference in discriminating words such as "pad" versus "bad." To do so, the listener must process 16 acoustic features that bear on the correct identification of /p/ versus /b/ (Lisker, 1986).

Third, the acoustic signals composing the speech stream are virtually continuous throughout a sentence (Foss & Hakes, 1978). Few pauses occur, and astonishingly, the pauses that do occur generally fall in the middle of words, not between words. Pauses mark boundaries between words less than 40% of the time (R. A. Cole & Jakimik, 1980). This phenomenon is illustrated with a portion of the speech spectrogram for the sentence "John said that the dog snapped at him," shown in Figure 2.13.

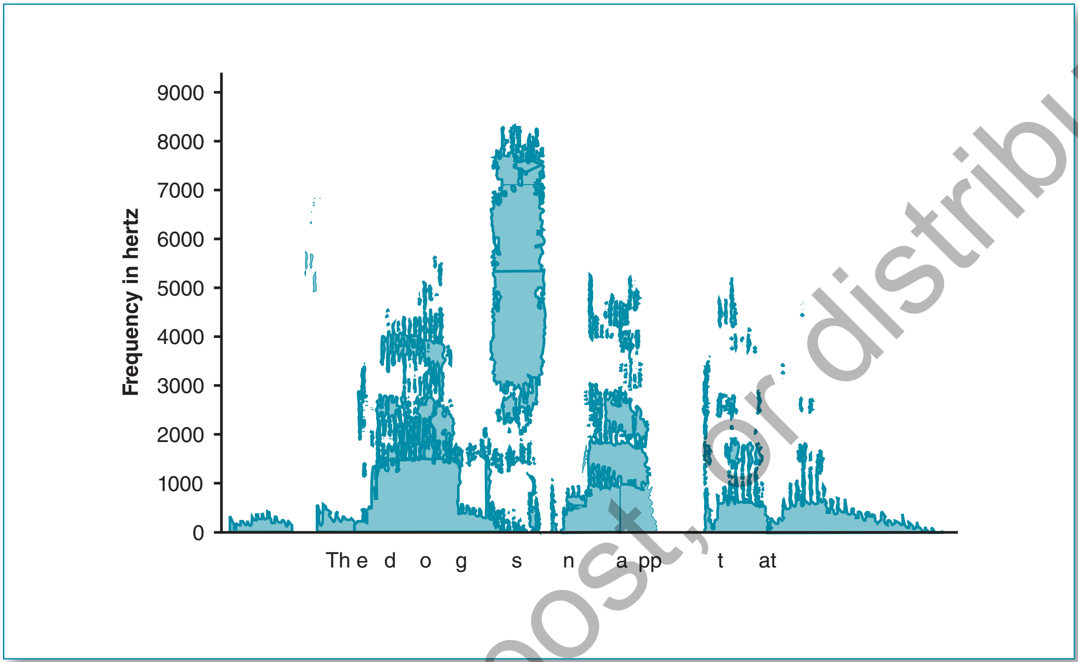
Phonemes are coarticulated, meaning that each segment of the acoustic signal provides clues about the identity of more than one phoneme. As a consequence, the signal lacks an invariant feature for a particular phoneme that stays the same in all contexts.

To conclude, the module for speech perception thus operates on multiple fronts to extract the phonemes accurately with astonishing rapidity. The phonological processing capabilities of the temporal superior gyrus is an important contributor to the neural computations involved, but the module must be thought of as distributed across a network of multiple brain regions (Poeppel & Monahan, 2008). A ventral pathway maps the phonological representation of how a word sounds to the concepts that underlie the meaning of the word. As will be seen in Chapter 7, these conceptual representations are stored throughout the temporal lobe. At the same time, a dorsal pathway links the phonological representation to the articulatory/motor programs that can produce the sounds in speech.

Notice the pauses in acoustic energy between the /s/ and the /n/ and between the /p/ and the /t/. The listener hears pauses between the words and phrases of the sentence, but the acoustic energy fails to provide them. Instead, they are inserted by the speech-recognition processes that categorize the acoustic input; these divide the sounds into the words, phrases, clauses, and sentences through top-down or conceptually driven recognition processes. Conversely, the pauses in acoustic energy that do not signal an important linguistic unit, such as those in the word *snapped,* are not perceived by the listener.

It is easier to appreciate the role of conceptually driven processes in speech perception when listening to a foreign language. The continuous nature of the acoustic speech stream is perceived as it really is. The pauses that occur in the middle of a word are heard correctly, while one word streams into another. The coarticulation effect discussed earlier applies across word boundaries as well as within word boundaries. Thus, the speaker is sending acoustic clues at any given moment about the identity of phonemes that belong to adjacent words. The true complexity of the stream, if analyzed solely from bottom-up or data-driven processes, can be readily heard when listening to a native speaker of a language that is

**Figure 2.13** Portion of the speech spectrogram for “John said that the dog snapped at him.”



*SOURCE:* From Foss, D. J., & Hakes, D. T., *Psycholinguistics: An Introduction to the Psychology of Language, 1st edition,* copyright © 1978. Reprinted with permission of Pearson Education, Inc., Upper Saddle River, NJ.

foreign to us. But when listening to our own first language, we hear the speech stream as a sequence of neat and tidy packages of sound that specify meaning.

A critical function of the speech-processing module is the categorization of speech input at the phonemic level, a phenomenon called **categorical perception.** Subtle variations in the acoustic signal are ignored unless they mark a boundary between one phoneme and another. For example, /b/ and /p/ differ in terms of the amount of time that elapses between the release of the lips and the onset of voicing. The voice onset time for /b/ is immediate (0 seconds). For /p/, the voice onset time is 0.06 seconds. Within this narrow window of time lies the boundary between hearing one phoneme versus another.

Lisker and Abramson (1970) demonstrated the phenomenon of categorical perception by continuously varying voice onset time from –0.15 to +0.15 seconds using computer-synthesized speech. For the 31 stimuli, the acoustic signal differed by 0.01 seconds in voice onset time, yet only two phonemes were heard. Listeners identified all sounds as /b/ over a large range of variation in the acoustic signal, from –0.15 seconds up to just over 0 seconds. As soon as the voice onset time slightly exceeded 0 seconds, the listeners began to hear /p/ instead of /b/ and continued to do so for all remaining stimuli. What matters, then, is not the degree of change in voice onset time; all variations of 0.15 seconds are heard as the same phoneme. Instead, what matters is whether the change in acoustic signal crosses a sharply defined boundary. There is, therefore, a sharp decision boundary that distinguishes the perception of one phoneme from the perception of another related one.

Although speech phonemes exhibit well-defined categorical boundaries, it is mistaken to conclude that the auditory system cannot sense gradual transitions in voice onset time. Data-driven sensory processes plainly pick up these differences (Massaro, 1994). The sensory system detects continuous changes in the speech signal, but a decision process assigns the signal to one phonemic category or another. Repp and Liberman (1987) found that the boundary between two phonetic categories is flexible, to a degree. Precisely where a given listener locates the boundary depends on the context provided by other stimuli.

Speech signals are assigned to phonemes on the basis of well-defined categorical boundaries. The continuous speech stream is heard as separate words and phrases as a result of conceptually driven recognition processes.

Despite the complexity involved in extracting phonetic segments from the speech stream, infants between the ages of 1 and 4 months can detect the acoustic features that distinguish one phoneme from another. Indeed, it appears that at this age, infants are prepared to identify not only the phonemes of their native language but virtually all possible phonetic segments used in human languages (Eimas, Miller, & Jusczyk, 1987). Such evidence is consistent with the idea that speech is perceived by a special processing module (Eimas & Miller, 1992).

Infants cannot, of course, report what they hear. Yet by ingeniously monitoring the rate infants suck on a pacifier, developmental psychologists can infer changes in attention to a stimulus. The sucking schema is well established in a 1-month-old infant. In fact, sucking is one of a small number of reflexes present at birth. This basic sensorimotor schema develops with experience in nursing and displaces the reflex. It turns out that infants suck faster when attending to a novel stimulus. With repeated presentations of the stimulus, the sucking rate slows down as the infant habituates to the stimulus. If the stimulus is abruptly changed in a way that is noticed by the infant, then dishabituation occurs (i.e., the sucking rate suddenly increases). The difference between preshift stimuli and postshift stimuli can be measured by the difference in rates of sucking.

Using this method, Eimas (1974) found categorical perception of speech by infants. The infants dishabituated when a change in the acoustic signal crossed a phonetic boundary. Subsequent research has shown that infants can, in fact, discriminate among the stimuli that fall within a phonemic boundary (Miller & Eimas, 1983). Like adults, however, infants appear tuned to pick up the critical differences that separate one phoneme from another and to process the context in which the acoustic signals occur.

Furthermore, infants are not born with full capabilities in categorizing speech. Newborns can detect differences among syllables that contain different phonemes, but their representations at this early stage of development might not be full-fledged phonetic segments (cf. Eimas & Miller, 1992). Instead, over the first 1 or 2 months of life, the infant may progress from a global representation of the syllable to the specific phonemic-level representations (Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, & Mehler, 1988).

## SUMMARY

1. Perception begins with the transduction of the physical energy of a stimulus into an initial neural representation of the stimulus. As a result, the objects and events that are present in the environment are perceived, in the sense of being detected. With still more processing, the objects and events are recognized, in the sense of being categorized as meaningful. Visual consciousness depends on representations being processed in the visual cortex. Patients with blindsight lack any visual awareness but are able to guess accurately about the actual locations of objects in space.
2. The ability to perceive depends on pattern recognition—that is, categorizing objects and events detected in the environment by matching their preliminary representations with patterns stored in long-term memory. A stimulus can be perceived and understood in terms of its properties but not recognized as a meaningful object—a neuropsychological condition called agnosia. Patients suffering from lesions in certain regions of the brain can see objects but not recognize them at all. Apperceptive agnosia refers to a failure of pattern recognition caused by an inability to categorize objects at a perceptual level of analysis. Associative agnosia, by contrast, is caused by an inability to categorize objects at a functional semantic level of analysis.
3. Schemas generate expectations about the objects and events that will be encountered. These expectations direct exploration of the environment in the form of eye movements and other bodily movements that pick up the information available. The sampled information either confirms or modifies the original expectations, in turn leading to renewed exploration. Top-down or conceptually driven pattern recognition refers to the use of expectations to ease the process of finding a match between incoming stimuli and schemas that store our knowledge about the world in long-term memory. Bottom-up or data-driven pattern recognition refers to the use of the features picked up from the environment. Both data and expectations play a critical role in rapid, accurate, and adaptive perception.
4. The representation of objects in long-term memory has been viewed theoretically as feature lists and as structural descriptions. An object can be represented in terms of a list of distinctive features that discriminate it from other objects. The problem with this view is that two objects might include the same features but differ in terms of the relationships of their features. A structural description takes into account both the distinctive features and their relations.
5. Holistic processing refers to perceiving the whole object; analytic processing refers to perceiving the features that compose the whole. Perception of faces is unique in that it is more strongly influenced by holistic processing than by analytic processing. A specialized module is responsible for face perception. Face perception is automatic, fast, encapsulated from other cognitive systems, and instantiated in a localized area of the brain. Prosopagnosic patients suffering from damage to the module are unable to recognize faces despite intact object recognition in general.

6. Speech perception is challenging because the acoustic signal for the basic sounds of speech that communicate meaning—phonemes—is highly complex. Phonemes are coarticulated, meaning that each segment of the acoustic signal provides clues about the identity of more than one phoneme. As a consequence, the signal lacks an invariant feature for a particular phoneme that stays the same in all contexts. Speech signals are assigned to phonemes on the basis of well-defined categorical boundaries. Gradual variations in the acoustic signal are perceived categorically. Finally, the acoustic energy in speech is often continuous across word boundaries. The continuous speech stream is heard as separate words and phrases as a result of conceptually driven recognition processes.

## KEY TERMS

blindsight
pattern recognition
apperceptive agnosia
associative agnosia
schema
conceptually driven processes
data-driven processes
word superiority effect
change blindness
distinctive features
holistic processing
analytic processing
prosopagnosia
phoneme
speech spectrogram
formants
coarticulation
categorical perception

## QUESTIONS FOR THOUGHT

- In driving a car, it is necessary to identify numerous objects and traffic signs. Explain how conceptually driven pattern recognition helps to achieve this readily. In what circumstances are data-driven processes most important?
- How does the categorical perception of speech contribute to the fast processing of phonemes by the speech-recognition module?
- In what ways are apperceptive agonosia, associative agnosia, and prosopagnosia similar? Specifically, how do these three forms of agnosia differ?