Getty Images/JurgaR

2 PATTERN RECOGNITION

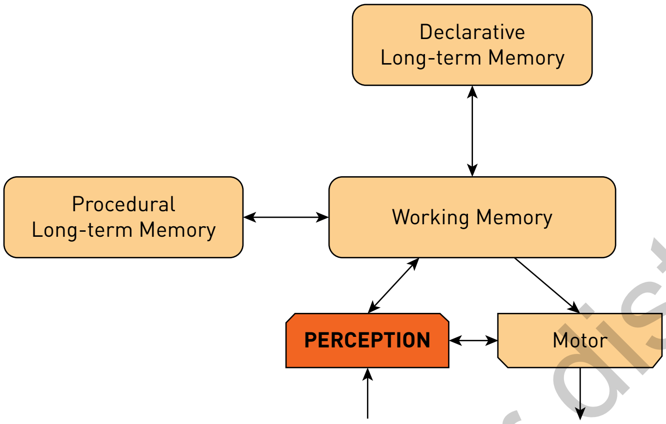
LEARNING OBJECTIVES

1. Contrast feature and structural theories of pattern recognition.
2. Explain how Sperling's partial-report technique contributed to understanding characteristics of the visual sensory store.
3. Explain how the word superiority effect determines why a letter in a word is better recognized than a letter by itself.
4. Discuss the goals of understanding scenes and the applications of deep neural networks.
5. Describe how visual disorders have increased our knowledge of neural pathways.

The study of **pattern recognition** is primarily the study of how people identify the objects in their environment. Pattern recognition, which is discussed in this chapter, and attention, in the next chapter, play lead roles in the perception component of the standard model of cognition (Figure 2.1). We focus on visual pattern recognition in this chapter to provide continuity. Other chapters, such as the next chapter on attention, contain material on speech recognition.

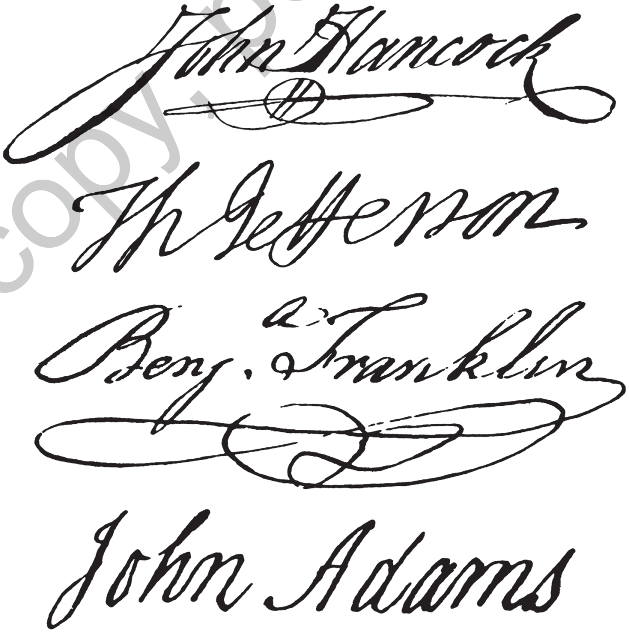
pattern recognition The stage of perception during which a stimulus is identified

FIGURE 2.1 ■ The Perception Component of the Standard Model of the Mind.



Our ability to recognize patterns is impressive if we stop to consider how much variation there is in different examples of the same pattern. Figure 2.2 shows various styles of handwriting. Not all people have the same style of writing, and some handwriting styles are much less legible than others. However, unless it is very illegible, we usually are successful in recognizing the words.

FIGURE 2.2 ■ Variations in Handwriting.



Source: istockphoto.com/DNY59

Our superiority over computers as pattern recognizers has the practical advantage that pattern recognition can serve as a test of whether a person or a computer program is trying to gain access to the Internet. If you have spent much time on the Internet you might have encountered a situation that required you to identify a distorted word before you were allowed to enter a site. The mangled word is easy for people to identify but difficult for computer search programs.

A large part of the literature on pattern recognition is concerned with alternative ways of describing patterns. The first section of this chapter discusses three kinds of descriptions that represent different theories of pattern recognition. The second section is about information-processing models of visual pattern recognition. The next two sections focus on word recognition and scene recognition. The last section on visual agnosia describes how studying brain disorders has contributed to establishing the neural basis of recognizing patterns.

DESCRIBING PATTERNS

Consider the following explanation of how we recognize patterns. Our long-term memory (LTM) contains descriptions of many kinds of patterns. When we see or hear a pattern, we form a description of it and compare the description against the descriptions stored in our LTM. We can recognize the pattern if its description closely matches one of the descriptions stored in LTM. Although this is a plausible explanation, it is rather vague. For example, what form do these descriptions take? Let us consider three explanations that have been suggested: (1) templates, (2) features, and (3) structural descriptions.

Template Theories

Template theories propose that patterns are really not "described" at all. Rather, **templates** are holistic, or unanalyzed, entities that we compare with other patterns by measuring how much two patterns overlap. Imagine that you made a set of letters out of cardboard. If you made a cutout to represent each letter of the alphabet and we gave you a cutout of a letter that we had made, you could measure how our letter overlapped with each of your letters—the templates. The identity of our letter would be determined by which template had the greatest amount of overlap. The same principle would apply if you replaced your cardboard letters with a visual image of each letter and used the images to make mental comparisons.

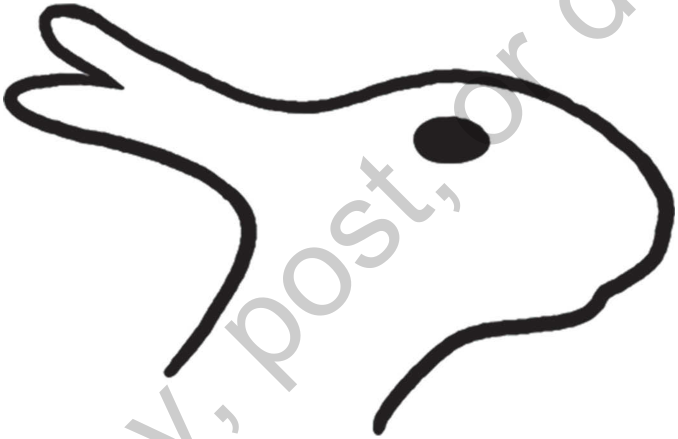
There are a number of problems with using the degree of overlap as a measure of pattern recognition. First, the comparison requires that the template is in the same position and the same orientation, and is the same size as the pattern you are trying to identify. Thus, the position, orientation, and size of the templates would have to be continuously adjusted to correspond to the position, orientation, and size of each pattern you wanted to recognize. A second problem is the great variability of patterns, as was illustrated in Figure 2.2. It would be difficult to construct a template for each letter that would produce a good match with all the different varieties of that letter.

template An unanalyzed pattern that is matched against alternative patterns by using the degrees of overlap as a measure of similarity

Third, a template theory doesn't reveal how two patterns differ. We could know from a template theory that the capital letters *P* and *R* are similar because one overlaps substantially with the other. But to know how the two letters differ, we have to be able to analyze or describe the letters.

A fourth problem is that a template theory does not allow for alternative descriptions of the same pattern. You may have seen ambiguous figures that have more than one interpretation, such as a duck or a rabbit in Figure 2.3. The two interpretations are based on different descriptions; for example, the beak of the duck is the ears of the rabbit. A template is simply an analyzed shape and so is unable to make this distinction. By contrast, a feature theory allows us to analyze patterns into their parts and to use those parts to describe the pattern.

FIGURE 2.3 ■ An Ambiguous Figure that can be Perceived as Either a Duck or a Rabbit.



Source: "What an image depicts depends on what an image means," by D. Chambers & D. Reisberg, 1985, *Cognitive Psychology, 24*, 145–174.

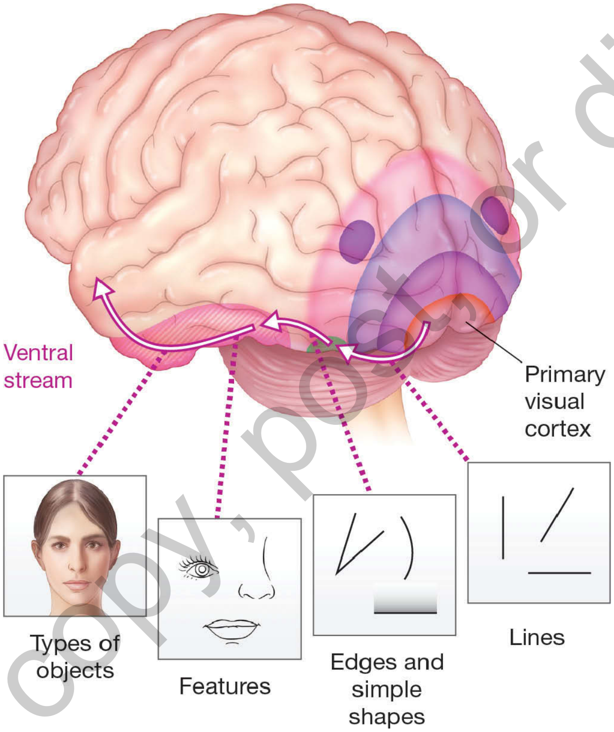
Feature Theories

Feature theories allow us to describe a pattern by listing its parts, such as describing a friend as having long blond hair, a short nose, and bushy eyebrows. Part of the evidence for feature theories comes from recording the action potentials of individual cells in the visual cortex. By placing microelectrodes in the visual cortex of animals, Hubel and Wiesel (1962, 1963) discovered that cells respond to only certain kinds of stimuli. Some cells might respond to a line of a certain width, oriented at a correct angle and located at the correct position in its visual field. Other cells are even concerned about the length of the line. In 1981 Hubel and Wiesel received a Nobel Prize for this work.

feature theory A theory of pattern recognition that describes patterns in terms of their parts, or features

Figure 2.4 shows the neural processing of visual information. Light is initially detected by photoreceptor cells in the retina to extract meaningful information about the visual world. This information is projected to the thalamus and areas of the primary visual cortex where the cells discovered by Hubel and Wiesel respond to features such as lines and simple shapes. These simple shapes are then combined in the ventral stream into more complex features to identify objects. We will learn more about visual features in the next section on perceptual learning and more about neural pathways in the last section on visual disorders.

FIGURE 2.4 ■ Neural Processing of Visual Features.



Source: Adapted from *NNATE: How the Wiring of Our Brains Shapes Who We Are*, by K. J. Mitchell, 2018, Princeton, NJ: Princeton University Press.

Perceptual Learning

Feature theories are convenient for explaining perceptual development, and one of the best discussions of feature theories is contained in Eleanor Gibson's (1969) *Principles of Perceptual Learning and Development.* Gibson's theory is that perceptual learning occurs through the discovery of features that *distinguish* one pattern from another.

Although most pattern recognition theorists make use of the feature concept, it is often a challenging task to find a good set of features. Gibson (1969) proposed the following criteria as a basis for selecting a set of features for uppercase letters:

1. The features should be critical ones and present in some members of the set but not in others to provide a contrast.
2. The identity of the features should remain unchanged under changes in brightness, size, and perspective.
3. The features should yield a unique pattern for each letter.
4. The number of proposed features should be reasonably small.

Gibson used these criteria, empirical data, and intuition to derive a set of features for uppercase letters. The features consist primarily of different lines and curves that are the components of letters. Examples of lines include a horizontal line, a vertical line, and diagonal lines that slant either to the right or to the left as occur in the capital letter *A*. Examples of curves include a closed circle (the letter *O*), a circle broken at the top (the letter *U*), or a circle broken at the side (the letter *C*). Most letters consist of more than one feature, such as a closed circle and a diagonal line in the letter *Q*.

A set of features is usually evaluated by determining how well it can predict **perceptual confusions,** as confusable items should have many features in common. For example, the only difference in features for the letters *P* and *R* is the presence of a diagonal line for the letter *R;* therefore, the two should be highly confusable. The letters *R* and *O* differ in many features, and so they should seldom be confused.

One method for generating perceptual confusions is to ask an observer to identify letters that are presented very rapidly (Townsend, 1971). It is often difficult to discriminate physically similar letters under these conditions, and the errors provide a measure of perceived similarity. Holbrook (1975) compared two feature models to determine how successfully each could predict the pattern of errors found by Townsend. One was the model proposed by Gibson and the other was a modification of the Gibson model proposed by Geyer and De Wald (1973). The major change in the modification was the specification of the number of features in a letter (such as two vertical lines for the letter *H*) rather than simply listing whether that feature was present.

A comparison of the two models revealed that the feature set proposed by Geyer and De Wald was superior in predicting the confusion errors made both by adults (Townsend, 1971) and by four-year-old children (Gibson et al., 1963). The prediction of both models improved when the features were optimally weighted to allow for the fact that some features are more important than others in accounting for confusion errors. Because the straight/curved distinction is particularly important, it should be emphasized more than the others.

perceptual confusion A measure of the frequency with which two patterns are mistakenly identified as each other

Distinctive Features

Children learn to identify an object by being able to identify differences between it and other objects. For example, when first confronted with the letters *E* and *F*, the child might not be aware of how the two differ. Learning to make this discrimination depends on discovering that a low horizontal line is present in the letter *E* but not in the letter *F*. The low horizontal line is a **distinctive feature** for distinguishing between an *E* and an *F;* that is, it enables us to distinguish one pattern from the other.

Perceptual learning can be facilitated by a learning procedure that highlights distinctive features. An effective method for emphasizing a distinctive feature is to initially make it a different color from the rest of the pattern and then gradually change it back to the original color. Egeland (1975) used this procedure to teach prekindergarten children how to distinguish between the confusable letter pairs *R-P, Y-V, G-C, Q-O, M-N*, and *K-X.* One letter of each pair was presented at the top of a card with six letters below it, three of which matched the sample letter and three of which were the comparison letter. The children were asked to select those letters that exactly matched the sample letter.

One group of children received a training procedure in which the distinctive feature of the letter was initially highlighted in red—for example, the diagonal line of the *R* in the *R-P* discrimination. During the training session, the distinctive feature was gradually changed to black to match the rest of the letter. Another group of children viewed only black letters. They received feedback about which of their choices were correct, but they were not told about the distinctive features of the letters. Both groups were given two tests—one immediately after the training session and one a week later. The "distinctive features" group made significantly fewer errors on both tests, even though the features were not highlighted during the tests. They also made fewer errors during the training sessions.

Emphasizing the distinctive features produced two benefits. First, it enabled the children to learn the distinctive features so that they could continue to differentiate letters after the distinctive features were no longer highlighted. Second, it enabled them to learn the features without making many errors during the training session. The failure and frustration that many children experience in the early stages of reading (letter discrimination) can impair their interest in later classroom learning.

Focusing on distinctive features might aid in distinguishing among faces, as it does in distinguishing among letters. To test this, Brennan (1985) used computer-generated **caricatures** that make distinctive features even more distinctive. For instance, if a person had large ears and a small nose, the caricature would have even larger ears and an even smaller nose than the accurate drawing. When students were shown line drawings of acquaintances, they identified people faster when shown caricatures than when shown accurate line drawings (Rhodes

distinctive feature A feature present in one pattern but absent in another, aiding one's discrimination of the two patterns

caricature An exaggeration of distinctive features to make a pattern more distinctive et al., 1987). Making distinctive features more distinctive through exaggeration facilitated recognition.

Combining Features

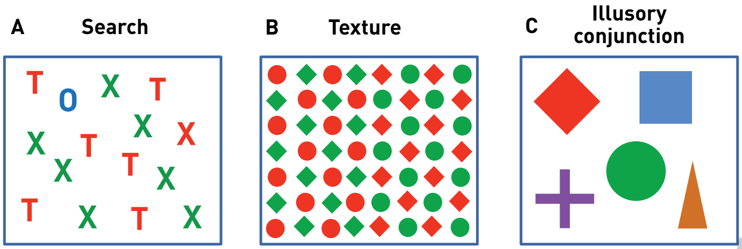
Distinctive features are a key component of our ability to locate an object in our environment. If you ever waited for your luggage at an airport, you may have noticed many people tie colorful ribbons to their luggage to help find their bags more easily because they "pop out" from the crowd. This phenomenon, as illustrated by the red flower at the beginning of the chapter, is a major prediction of feature integration theory (Treisman & Gelade, 1980).

According to this theory, all features across the entire visual landscape are represented simultaneously and pre-attentively. Thus one need only monitor the relevant feature to locate a distinctive item. Treisman and Gelade (1980) found that reaction times to find an object in a *single feature search* were independent of the size of the display, indicating that searching for a single feature is accomplished all at once. However, when two or more features must be combined in a *conjunction search*, each object in a visual scene must be examined for the combined features, which requires using attention. Returning to the airport example—if you have a standard black bag, you will now have to examine each black bag for size, shape, and so forth.

Many of the Treisman's experiments on feature integration theory explored the problem of how a perceiver combines color and shape, as these two features are analyzed by separate parts of the visual system. Figure 2.5 shows several demonstrations of how color and shape interact (Wolfe, 2018). In Panel A, it is easier to find the blue *O*, defined by the unique feature blue, than to find the red *X*. Finding the red *X* requires attending to the conjunction of red and *X* because there are also red *T*s and green *X*s in the display. Treisman found that it did not matter how many other letters were in the display, if people searched for a letter defined by a unique color or shape. The uniqueness made the letter pop out from the rest of the display, as occurs for the blue *O*. However, adding more red *T*s and green *X*s to the display would increase the time to find the red *X* because it requires attending to a conjunction of features.

Panel B illustrates another finding that is predicted by the attention requirements of feature integration theory. It is not immediately obvious that the left half of the display differs from the right half because attention is necessary for perceiving conjunctions of color and shape. The circles and diamonds switch colors, which you can observe by closely attending to the shape and color combinations. Another important implication of Treisman's theory is referred to as the "illusory conjunctions." Following a brief glimpse of the display in Panel C, observers may report seeing an incorrect combination of color and shape, such as a green square. Feature integration theory states that it requires attention to combine features such as color and shape. Insufficient attention, therefore, causes incorrect combinations of features.

FIGURE 2.5 ■ Visual Displays used to Evaluate Feature Integration Theory.



Structural Theories

A limitation of feature theories is that descriptions of patterns often require that we specify how the features are joined together. Describing how features join together to create a structure is a guiding principle of Gestalt psychology. To Gestalt psychologists, a pattern is more than the sum of its parts. Providing precise descriptions of the relations among pattern features was initially formalized by people working in the field of artificial intelligence who discovered that the interpretation of patterns usually depends on making explicit how the lines of a pattern are joined to other lines (Clowes, 1969).

Structural Descriptions

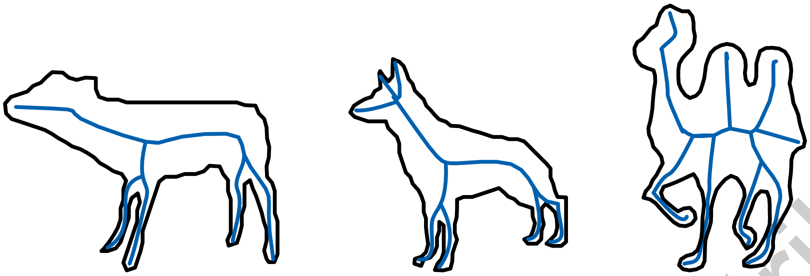
Structural theories describe the relations among the features by building on feature theories. Before we can specify the relation among features, we have to specify the features. A structural theory allows specification of how the features fit together. For example, the letter **H** consists of two vertical lines and a horizontal line. But we could make many different patterns from two vertical lines and a horizontal line. What is required is a precise specification of how the lines should be joined together—the letter **H** consists of two vertical lines connected at their midpoints by a horizontal line.

Figure 2.6 illustrates shape skeletons for different animals that are based on structural descriptions originally proposed by Blum (1973) as a method for distinguishing among biological forms. Wilder et al. (2011) adapted Blum's methods to make predictions about how people would classify novel shapes into categories, such as *animal* and *leaf.* Their successful predictions support the argument that people use these kinds of descriptions to make classifications. The skeleton shapes of animals have relatively curvy limbs compared to the fewer, straighter limbs of leaves.

Moving from a two-dimensional world to a three-dimensional world creates additional challenges for identifying and describing the relations among features. Figure 2.7 illustrates the problem of identifying features by the relative difficulty of perceiving the three patterns as cubes (Kopfermann, 1930). The left pattern is the most difficult to perceive as a cube, and the pattern in the middle is the easiest. Try to guess why before reading further. (Hint: Think about the challenge of identifying features for each of the three examples.)

structural theory A theory that specifies how the features of a pattern are joined to other features of the pattern

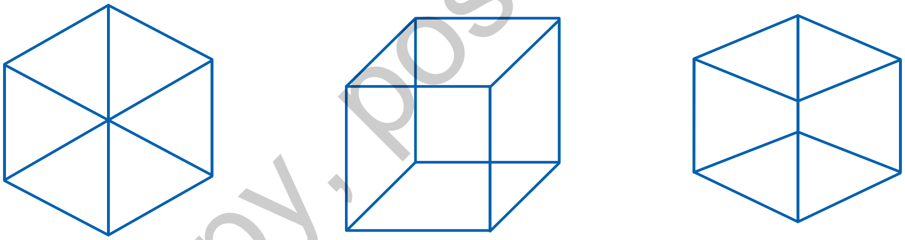
FIGURE 2.6 ■ Examples of Skeleton Structures of Animals.



The theme of Hoffman's (1998) book on visual intelligence is that people follow rules in producing descriptions of patterns. The first of the many rules described in his book is to always interpret a straight line in an image as a straight line in three dimensions. Therefore, we perceive the long vertical line in the center of the right pattern in Figure 2.7 as a single line. However, it is necessary to split this line into two separate lines to form a cube because the lines belong to different surfaces. It is particularly difficult to see the figure on the left as a cube because you also need to split the two long diagonal lines into two shorter lines to avoid seeing the object as a flat pattern.

The pattern in the middle is easy to perceive as a cube, which you may have recognized as the famous Necker cube. The Necker cube is well known because your perception of the front and back surfaces of the cube changes as you view it (Long & Toppino, 2004). It is yet another example that a structural description can change when the features do not change!

FIGURE 2.7 ■ Perceiving Cubes.



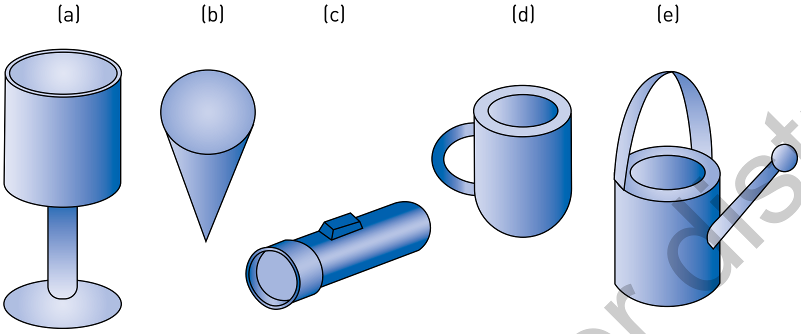
Source: Visual Intelligence, by D. D. Hoffman, 1998, New York: Norton.

Biederman's Component Model

Descriptions of three-dimensional objects would be fairly complicated if we had to describe each of the lines and curves in the object. For example, the cubes in Figure 2.7 each consist of 12 lines (which you may find easier to count in the left and right cubes after splitting the lines than in the reversing Necker cube). It would be easier to describe three-dimensional objects through simple volumes such as cubes, cylinders, edges, and cones than to describe all the features in these volumes.

The advantage of being able to form many different arrangements from a few components is that we may need relatively few components to describe objects. Biederman (1985) has proposed that we need only approximately 35 simple volumes (which he called **geons**) to describe the objects in the world. Some objects contain the same geons, but the geons are arranged differently. The mug (d) in Figure 2.8 would become a pail, if the handle were placed at the top rather than at the side of the container. Add two additional geons, and the pail becomes a watering can (e).

FIGURE 2.8 ■ Different Arrangements of Geons Produce Different Objects.



Research by Biederman et al. (2009) established that it is easier to discriminate one geon from a different geon than to discriminate two variations of the same geon. For example, U.S. college students can more easily discriminate the middle object in Figure 2.9 from the left object (a different geon with straight sides) than from the right object (a variation of the same geon with greater curvature).

A question raised by these findings is whether there are cultural differences in people's ability to discriminate among geons. The distinction between straight lines and curves is fundamental in western culture, as we have already discovered, for discriminating among letters of the alphabet. In contrast, there is less of the need to discriminate between lines and curves by the Himba, a seminomadic people living in a remote region of Namibia. Nonetheless, the Himba also are more able to distinguish different geons from each other (the left two objects) than variations of the same geon (the right two objects).

If pattern recognition consists mainly in describing the relations among a limited set of components, then deleting information about the relations among those components should reduce people's ability to recognize patterns. To test this hypothesis, Biederman removed 65% of the contour from drawings of objects, such as the two cups shown in Figure 2.10. In the cup on the left, the contour was removed from the middles of the segments, allowing observers to see how the segments were related. In the cup on the right, the contour was removed from the vertices so observers would have more difficulty recognizing how the segments were related. When drawings of different objects were presented for 100 msec, subjects correctly named 70% of the objects if the contours were deleted at midsegments. But if the contours were deleted at the vertices, subjects correctly named fewer than 50% of the objects (Biederman, 1985). As predicted, destroying relational information was particularly detrimental for object recognition.

geons Different three-dimensional shapes that combine to form three-dimensional objects

FIGURE 2.9 ■ Discriminating between Different Geons (Middle and Left) is Easier than Discriminating between Different Variations of the Same Geon (Middle and Right).

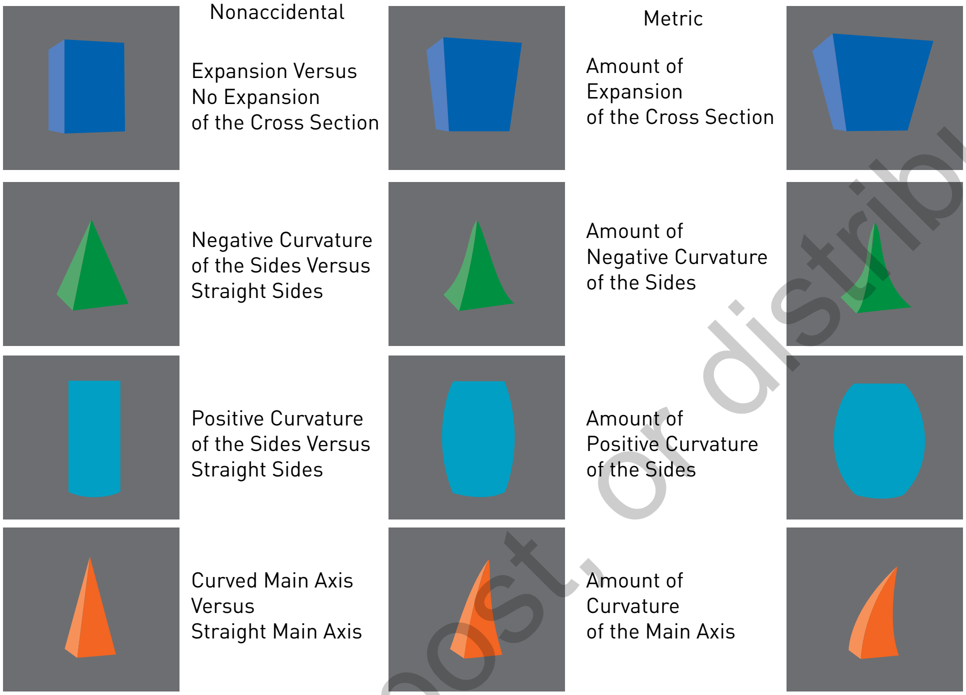
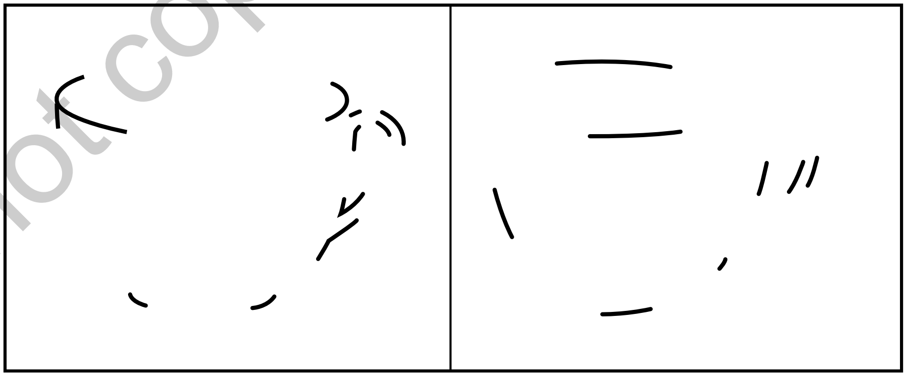

Nonaccidental
Metric
Expansion Versus No Expansion of the Cross Section
Amount of Expansion of the Cross Section
Negative Curvature of the Sides Versus Straight Sides
Amount of Negative Curvature of the Sides
Positive Curvature of the Sides Versus Straight Sides
Amount of Positive Curvature of the Sides
Curved Main Axis Versus Straight Main Axis
Amount of Curvature of the Main Axis

FIGURE 2.10 ■ Illustration of 65% Contour Removal Centered at Either Midsegments (Left Object) or Vertices (Right Object).



Source: "Human image understanding: Recent research and a theory," by I. Biederman, 1985, *Computer Vision, Graphics, and Image Processing*, *32*, 29–73.

In conclusion, structural theories extend feature theories by specifying how the features are related. Sutherland (1968) was one of the first to argue that if we want to account for our very impressive pattern recognition capabilities, we will need the more powerful kind of descriptive language contained in a structural theory. The experiments in this section show that Sutherland was correct. We now look at how pattern recognition occurs over time.

INFORMATION-PROCESSING STAGES

The Partial-Report Technique

To completely understand how people perform a pattern recognition task, we have to identify what occurs during each of the information-processing stages (pattern recognition, attention, working memory) discussed in Chapter 1. George Sperling (1960) is responsible for the initial construction of an information-processing model of performance on a visual recognition task. We discuss his experiment and theory in detail because it provides an excellent example of how the information-processing perspective has contributed to our knowledge of cognitive psychology.

Subjects in Sperling's task saw an array of letters presented for a brief period (usually 50 msec) and were asked to report all the letters they could remember from the display. Responses were highly accurate if the display contained fewer than five letters. But when the number of letters was increased, subjects never reported more than an average of 4.5 letters correctly, regardless of how many letters were in the display.

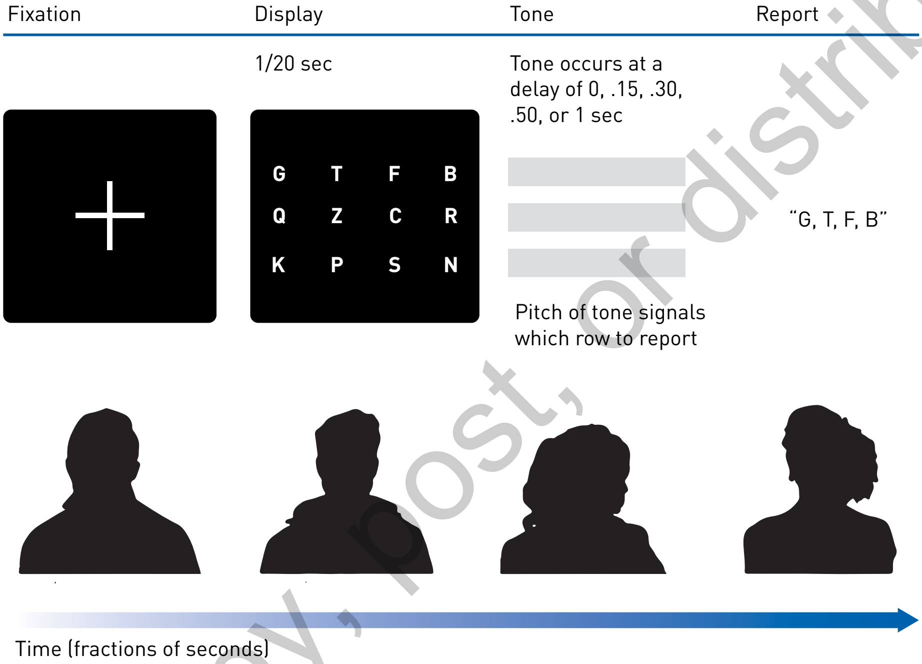
A general problem in constructing an information-processing model is to identify the cause of a performance limitation. Sperling was interested in measuring the number of letters that could be recognized during a brief exposure, but he was aware that the upper limit of 4.5 might be caused by an inability to remember more than that. In other words, subjects might have recognized most of the letters in the display but then forgot some before they could report what they had seen. Sperling, therefore, changed his procedure from a **whole-report procedure** (report all the letters) to a **partial-report procedure** (report only some of the letters).

In the most typical case, the display consisted of three rows, each containing four letters. Subjects would be unable to remember all 12 letters in a display, but they should be able to remember four letters. The partial-report procedure required that subjects report only one row. The pitch of a tone signaled which of the three rows to report: the top row for a high pitch, the middle row for a medium pitch, and the bottom row for a low pitch. The tone sounded just after the display disappeared, so that subjects would have to view the entire display and could not simply look at a single row (Figure 2.11). Use of the partial-report technique is based on the assumption that the number of letters reported from the cued row equals the average number of letters perceived in each of the rows because the subjects did not know in advance which row to look at. The results of this procedure showed that subjects

whole-report procedure A task that requires observers to report everything they see in a display of items

partial-report procedure A task in which observers are cued to report only certain items in a display of items could correctly report three of the four letters in a row, implying that they had recognized nine letters in the entire display.

FIGURE 2.11 ■ Sperling's (1960) Study of Sensory Memory. After the Subjects had Fixated on the Cross, the Letters were Flashed on the Screen Just Long Enough to Create a Visual Afterimage. High, Medium, and Low Tones Signaled which Row of Letters to Report.

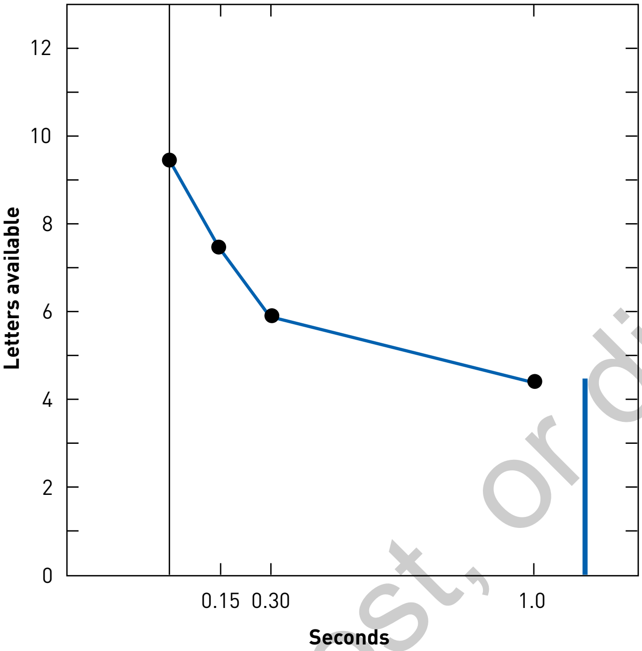

Fixation
Display
Tone
Report
1/20 sec
Tone occurs at a delay of 0, .15, .30, .50, or 1 sec
G T F B
Q Z C R
K P S N
"G, T, F, B"
Pitch of tone signals which row to report
Time (fractions of seconds)

Source: "The information available in brief visual presentations," by G. Sperling, 1960, *Psychological Monographs, 74* (11, Whole No. 498).

Photo credit: iStock/Vectorig

It often happens that what is best remembered about a scientist's work is not what that person originally set out to investigate. Although Sperling designed the partial-report technique to reduce the memory requirements of his task and to obtain a "pure" measure of perception, his work is best remembered for the discovery of the importance of a visual sensory store. How did this come about? The estimate that subjects had perceived nine letters was obtained when the tone occurred immediately after the termination of the 50-ms exposure. In this case, subjects could correctly report approximately three-quarters of the letters, and three-quarters of 12 is 9. But when the tone was delayed until one second after the display, performance declined to only 4.5 letters. That is, there was a gradual decline from nine letters to 4.5 as the delay of the tone was increased from 0 to one second (Figure 2.12).

FIGURE 2.12 ■ Recall as a Function of Delay of a Signaling Tone.



Source: "The information available in brief visual presentations," by G. Sperling, 1960, *Psychological Monographs*, *74* (11, Whole No. 498).

The most interesting aspect of the number 4.5 is that it is exactly equal to the upper limit of performance on the whole-report task, as represented by the blue bar in Figure 2.12. The partial-report procedure has no advantage over the whole-report procedure, if the tone is delayed by one second or more. To explain this gradual decline in performance, Sperling proposed that the subjects were using a **visual sensory store** to recognize letters in the cued row. When they heard the tone, they selectively attended to the cued row in the store and tried to identify the letters in that row. Their success in making use of the tone depended on the clarity of information in their sensory store. When the tone occurred immediately after termination of the stimulus, the clarity was sufficient for recognizing additional letters in the cued row. But as the clarity of the sensory image faded, it became increasingly difficult to recognize additional letters. When the tone was delayed by one second, the subjects could not use the sensory store at all to focus on the cued row, so their performance was determined by the number of letters they had recognized from the entire display that happened to be in that row. Their performance was therefore equivalent to the whole-report procedure, in which they attended to the entire display.

In 1963, Sperling proposed an information-processing model of performance on his visual report task. The model consisted of a visual information store, scanning, rehearsal, and an auditory information store. The **visual information store (VIS)** is a sensory store that preserves information for a brief period lasting from a fraction of a second to a second. The decay rate depends on such factors as the intensity, contrast, and duration of the stimulus and also on whether exposure to the stimulus is followed by a second exposure. Visual masking occurs when a second exposure, consisting of a brightly lighted field or a different set of patterns, reduces the effectiveness of the VIS.

For pattern recognition to occur, the information in the sensory store must be scanned. Sperling initially considered scanning to occur for one item at a time, as if each person had a sheet of cardboard with a hole in it just large enough for a single letter to appear.

The next two components of the model were **rehearsal** (saying the letters to oneself) and an **auditory information store** (remembering the names of the letters). To remember the items until recall, subjects usually reported rehearsing the items. Additional evidence for verbal rehearsal was found when recall errors often appeared in the form of auditory confusions—in other words, producing a letter that sounded like the correct letter. The advantage of the auditory store is that subvocalizing the names of the letters keeps them active in memory. Sperling's auditory store is part of short-term memory (STM), a topic we will consider later in the book.

Sperling revised his initial model in 1967. By this time, evidence had begun to accumulate suggesting that patterns were not scanned one at a time but were analyzed simultaneously. This distinction between performing one cognitive operation at a time (**serial processing**) and performing more than one cognitive operation at a time (**parallel processing**) is fundamental in cognitive psychology. Sperling, therefore, modified his idea of the **scan component** to allow for pattern recognition to occur simultaneously over the entire display, although the rate of recognition in a given location depended on where the subject was focusing attention.

Sperling's model was the first that attempted to indicate how various stages (sensory store, pattern recognition, and STM) combined to influence performance on a visual processing task. It contributed to the construction of information-processing models and led to the development of more detailed models of how people recognize letters in visual displays.

visual information store (VIS) A sensory store that maintains visual information for approximately one-quarter of a second

rehearsal Repeating verbal information to keep it active in short-term memory (STM) or to transfer it into long-term memory (LTM)

auditory information store In Sperling's model, this store maintains verbal information in short-term memory (STM) through rehearsal

serial processing Carrying out one operation at a time, such as pronouncing one word at a time

parallel processing Carrying out more than one operation at a time, such as looking at an art exhibit and making conversation

scan component The attention component of Sperling's model that determines what is recognized in the visual information store (VIS)

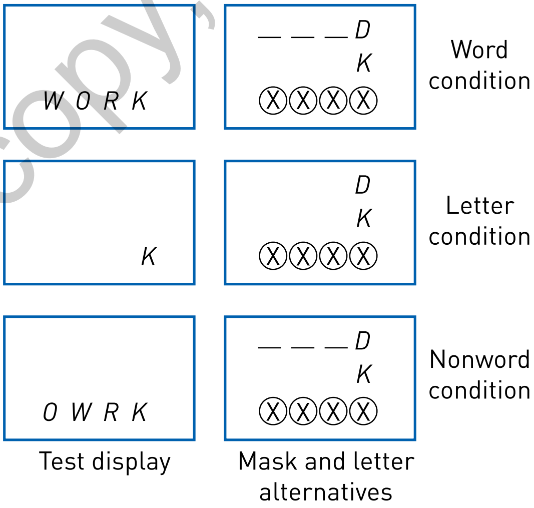
WORD RECOGNITION

The Word Superiority Effect

Much of the research on pattern recognition during the 1970s shifted away from how people recognize isolated letters to how people recognize letters in words. This research was stimulated by a finding that was labeled the *word superiority effect.* Reicher (1969), in his dissertation at the University of Michigan, investigated a possible implication of the scan component in Sperling's 1967 model. If the observer tries to recognize all the letters in a word simultaneously (Alderman et al., 2010), is it possible to recognize a four-letter unit in the same amount of time as it takes to recognize a single letter?

To answer this question, Reicher designed an experiment in which observers were shown a single letter, a four-letter word, or a four-letter nonword. The task was always to identify a single letter by selecting one of two alternatives. The exposure of the stimulus was immediately followed by a visual masking field with the two response alternatives directly above the critical letter. For example, one set of stimuli consisted of the word WORK, the letter *K*, and the nonword OWRK. The two alternatives, in this case, were the letters *D* and *K*, which were displayed above the critical *K* (Figure 2.13). Observers indicated whether they thought the letter in that position had been a *D* or a *K*.

FIGURE 2.13 ■ Example of the Three Experimental Conditions in Reicher's (1969) Experiment. The Mask and Response Alternatives Followed the test Display. The Task was to Decide which of the Two Alternatives had Appeared in the Test Position.



Source: "Perceptual recognition as a function of meaningfulness of stimulus material," by G. M. Reicher, 1969, *Journal of Experimental Psychology, 81*, 275–280.

This example illustrates several characteristics of Reicher's design. First, the four-letter word has the same letters as the four-letter nonword. Second, the position of the critical letter is the same for the word and the nonword. Third, both of the response alternatives make a word (WORD or WORK) for the word condition and a nonword for the nonword condition. Fourth, the memory requirements are minimized by requiring that subjects identify only a single letter, even when four letters are presented.

The results showed that subjects were significantly more accurate in identifying the critical letter when it was part of a word than when it was part of a nonword or when it was presented alone (the **word superiority effect**). Eight of the nine subjects did better on single words than on single letters. The one subject who reversed this trend was the only subject who said that she saw the words as four separate letters, which she made into words; the other subjects said that they experienced a word as a single word, not as four letters making up a word.

The word superiority effect is an example of top-down processing. It demonstrates how our knowledge of words helps us to more rapidly recognize the letters within a word. Top-down processing, based on knowledge stored in LTM, can aid pattern recognition in different ways. Top-down processing also helps us recognize words in sentences because the sentence constrains which words can meaningfully fit into the sentence.

A Model of the Word Superiority Effect

One of the great challenges for psychologists interested in word recognition has been to explain the reasons for the word superiority effect (Pollatsek & Rayner, 1989). A particularly influential model, the **interactive activation model** proposed by McClelland and Rumelhart (1981), contains several basic assumptions that build on the assumptions of Rumelhart's earlier model of letter recognition. The first assumption is that visual perception involves parallel processing. There are two different senses in which processing occurs in parallel. Visual processing is spatially parallel, resulting in the simultaneous processing of all four letters in a four-letter word. This assumption is consistent with Sperling's parallel scan and with Rumelhart's model of how people attempt to recognize an array of letters.

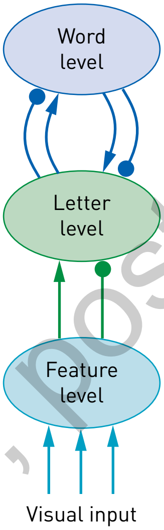
Visual processing is also parallel in the sense that recognition occurs simultaneously at three different levels of abstraction. The three levels—the feature level, the letter level, and the word level—are shown in Figure 2.14. A key assumption of the interactive activation model is that the three levels interact to determine what we perceive. Knowledge about the words of a language interacts with incoming feature information to provide evidence about which letters are in the word. This is illustrated by the arrows in Figure 2.14, which show that the letter level receives information from both the feature level and the word level.

word superiority effect The finding that accuracy in recognizing a letter is higher when the letter is in a word than when it appears alone or is in a nonword

interactive activation model A theory proposing that both feature knowledge and word knowledge combine to provide information about the identity of letters in a word

There are two kinds of connections between levels: excitatory connections and inhibitory connections. **Excitatory connections** provide positive evidence, and **inhibitory connections** provide negative evidence about the identity of a letter or word. For example, a diagonal line provides positive evidence for the letter *K* (and all other letters that contain a diagonal line) and negative evidence for the letter *D* (and all other letters that do not contain a diagonal line). Excitatory and inhibitory connections also occur between the letter level and word level, depending on whether the letter is part of the word in the appropriate position. Recognizing that the first letter of a word is a *W* increases the activation level of all words that begin with a *W* and decreases the activation level of all other words.

FIGURE 2.14 ■ The Three Levels of the Interactive Activation Model, with Arrows Indicating the Excitatory Connections and Circles Indicating Inhibitory Connections.



The interactive activation model was the first step for McClelland and Rumelhart in their development of neural network models of cognition. They referred to such models as **parallel distributed processing (PDP)** models because information is evaluated in parallel and is distributed throughout the network. A **neural network model** consists of several components

excitatory connection A positive association between concepts that belong together, as when a diagonal line provides support for the possibility that a letter is a *K*

inhibitory connection A negative association between concepts that do not belong together, as when the presence of a diagonal line provides negative evidence that a letter is a D

parallel distributed processing (PDP) When information is simultaneously collected from different sources and combined to reach a decision

neural network model A theory in which concepts (nodes) are linked to other concepts through excitatory and inhibitory connections to approximate the behavior of neural networks in the brain

(Rumelhart et al., 1986), some of which we have already considered in the interactive activation model. These include the following:

1. A set of processing units called **nodes**. Nodes are represented by features, letters, and words in the interactive activation model. They can acquire different levels of activation.
2. A pattern of connections among nodes. Nodes are connected to one another by excitatory and inhibitory connections that differ in strength.
3. Activation rules for the nodes. **Activation rules** specify how a node combines its excitatory and inhibitory inputs with its current state of activation.
4. A state of activation. Nodes can be activated to various degrees. We become conscious of nodes that are activated above a threshold level of conscious awareness. For instance, we become consciously aware of the letter *K* in the word *WORK* when it receives enough excitatory influences from the feature and word levels.
5. Output functions of the nodes. The output functions relate activation levels to outputs—for example, what threshold has to be exceeded for conscious awareness.
6. A learning rule. Learning generally occurs by changing the weights of the excitatory and inhibitory connections between the nodes, and the learning rule specifies how to make these changes.

The last component—the learning component—is one of the most important features of a neural network model because it enables the network to improve its performance. An example would be a network model that learns to make better discriminations among letters by increasing the weights of the distinctive features—those features that are most helpful for discriminating.

By 1992, the neural network approach had resulted in thousands of research efforts and an industry that spends several hundred million dollars annually (Schneider & Graham, 1992). The excitement of this approach can be attributed to several reasons. First, many psychologists believe that neural network models more accurately portray how the brain works than other, more serial models of behavior. Second, adjusting the excitatory and inhibitory weights that link nodes allows a network to learn, and this may capture how people learn. Third, the models allow for a different kind of computing in which many weak constraints (such as evidence from both the feature and word levels) can be simultaneously considered. Neural network models have continued to be developed into one of the most powerful learning methods in AI, as indicated by their application to recognizing scenes.

nodes The format for representing concepts in a semantic network

activation rule A rule that determines how inhibitory and excitatory connections combine to determine the total activation of a concept

SCENE RECOGNITION

Word recognition differs from letter recognition because words are composed of interacting letters. Similarly, scene recognition differs from object recognition because scenes are composed of interacting objects that are typically arranged in a meaningful spatial layout. Recognizing objects in scenes is often driven by accomplishing goals, as explained in the next section.

Goal-Driven Scene Understanding

Although our physical environment is usually stable, our goals can change and determine how we interact with the environment. Figure 2.15(A) illustrates four goals of scene understanding based on recognition, visual search, navigation, and action. Recognition determines whether a scene belongs to a certain category (a beach) or depicts a particular place (my living room). Visual search involves locating specific objects within the scene, such as sand, a bridge, or a lamp. Navigation determines whether it is possible to reach a particular location, such as crossing a stream. Action encompasses a broad set of activities, such as swimming, hiking, and watching television.

The four questions at the top of the figure are examples of questions we might ask for each of the different scenes (Malcolm et al., 2016). The first question "What is the scene?" requires scene recognition. It begins with gist—the perceptual and semantic information acquired from a single glance. Gist can include a conceptual understanding (a party), the spatial layout of the environment, and a few objects. It depends on the familiarity of stored representations, such as furniture is found in a living room. Unfamiliar scenes require more processing time than a brief glance to achieve scene understanding.

The second question "Where is X?" requires visual search using eye movements rather than a quick glance. Eye fixations focus on particular objects rather than the overall environment. They are required to answer the third question "How do I get from A to B?" Answering this question requires finding paths and potential obstacles that could block navigation, such as approaching objects. The last question "What can I do here?" determines actions, the topic of Chapter 6. Figure 2.15(B) shows scene properties that are needed to fulfill these goals. Low-level features, such as edges, establish the identity of objects. Object identities determine semantic categories and the actions that can be performed in those environments.

Deep Neural Networks

Computer scientists continued to develop the neural network models of the 1980s and connectionist models based on deep neural networks, which are some of the great success stories of AI (Sejnowski, 2018). **Deep neural networks** utilize the same principles as simpler networks but have added multiple layers of connections to fine-tune the weights of thousands of connections.

Figure 2.16 illustrates the application of deep neural networks to image recognition. The input begins with pixels from the image, and the output classifies the image as one of 1000

deep neural networks Networks that learn by adjusting thousands of connections in multiple layers

FIGURE 2.15 ■ Goal-driven Scene Recognition.

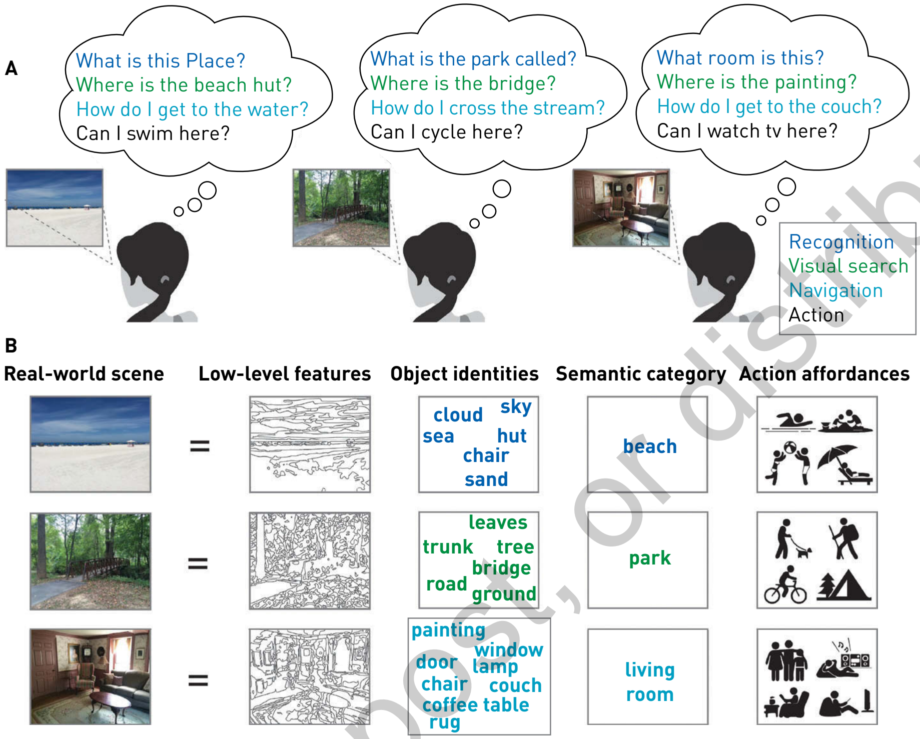
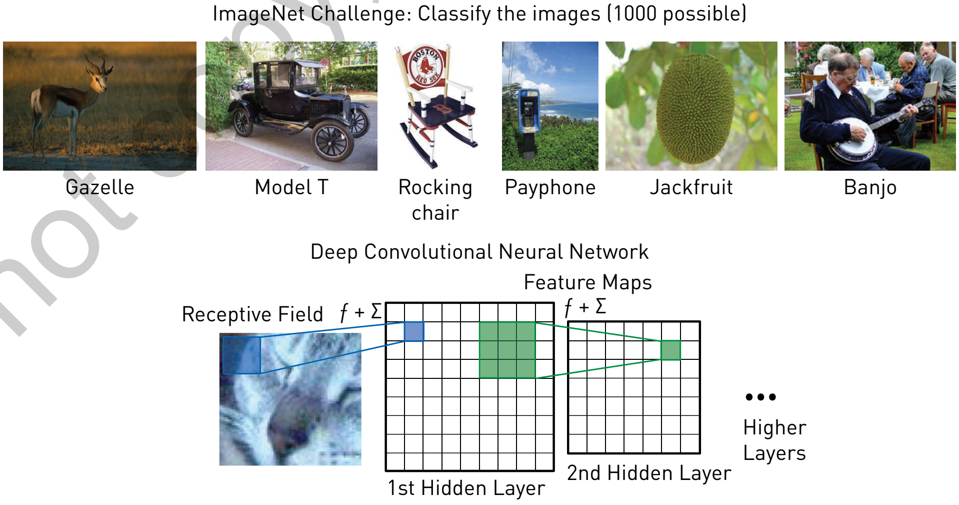


FIGURE 2.16 ■ Application of a Deep Neural Network to Classify Images.



Source: "Comparing the visual representations and performance of humans and deep neural networks," by R. A. Jacobs & C. J. Bates, 2019, *Current Directions in Psychological Science, 28*, 34–39.

possible pictures. In between are many hidden layers in which each layer receives input from a small number of units in the previous layer to establish more global connectivity. The layers are hidden because, in contrast to the three layers in Figure 2.14, their function can be difficult to interpret.

The authors of this article, Robert Jacobs and Christopher Bates at the University of Rochester's Department of Brain and Cognitive Sciences, review evidence that people are still superior at recognizing images under adverse conditions. The authors list several reasons for our perceptual advantage over machines. We learn to recognize objects in perceptually rich, dynamic, interactive environments whereas networks are trained on static images. We can take advantage of three-dimensional features whereas networks are more limited to two dimensions. A disadvantage for people, however, is our capacity limits because we cannot visually perceive and represent all aspects of a scene. These limits can nonetheless occasionally be an asset when we learn to focus on the more discriminative features.

Artificial intelligence continues to improve and the use of drones illustrates how pattern recognition can be shared by people and machines (Morris & Chakrabarty, 2019). Drones are limited by small payload capabilities and onboard processing power so some of the computational demands are offloaded to a ground computer (Figure 2.17). Joint activity between the equipment and the operator requires sensing, planning, and communication based on coordination between people and machines.

FIGURE 2.17 ■ Commanding a Drone.

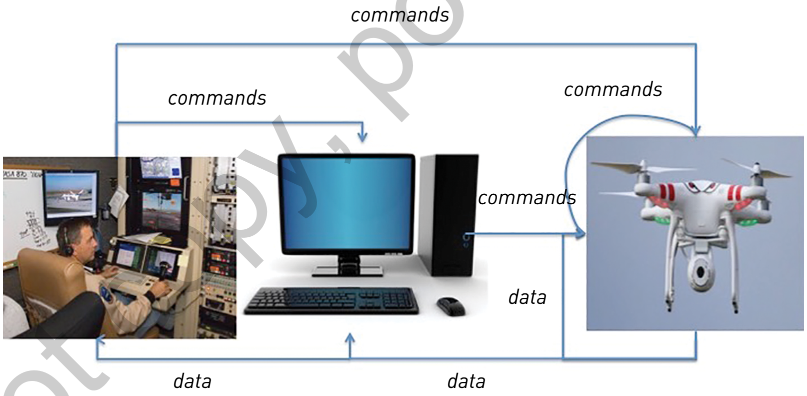
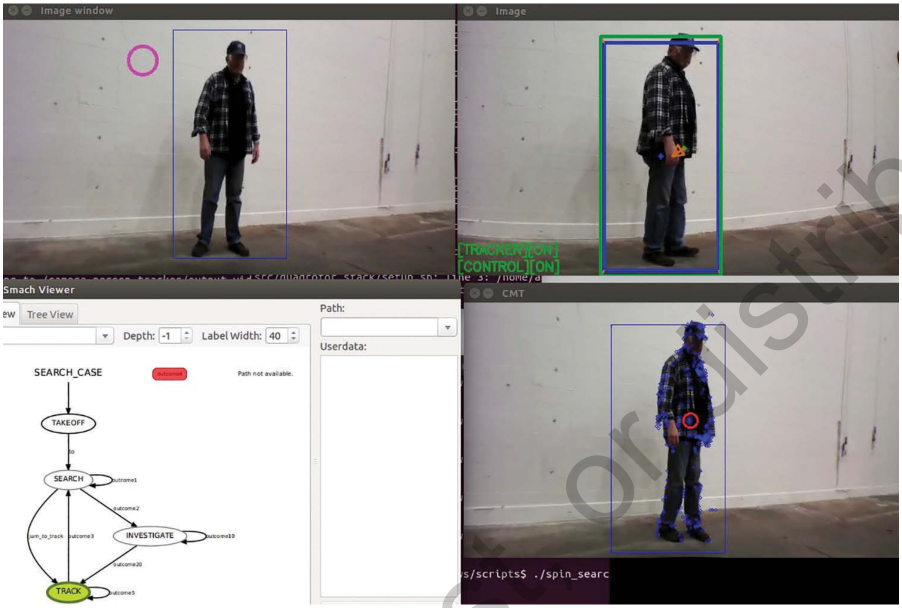


Figure 2.18 shows an application to a search and track task at the Ames Research Center in California. A controller monitors the search for a target and switches to a track mode if the target is found. At some point, the target may take evasive maneuvers that require the controller to switch back to the search mode. A key component of this interaction is the boundary between human and machine decisions. For instance, humans at the console may control the search phase and allow the drone to conduct the tracking phase. Although Morris and Chakrabarty (2019) focus on searching and tracking, they argue that many of the design principles apply to integrating human decision-making with various types of devices. An important decision related to the ethical use of AI discussed in Chapter 1 requires determining who should be tracked.

FIGURE 2.18 ■ Testing the Search and Track Mode at the Ames Research Center Indoor Facility.



```
src/quadrotor_stack/setup.sh: line 3: /home/a
s/scripts$ ./spin_searc
```

APPLICATIONS

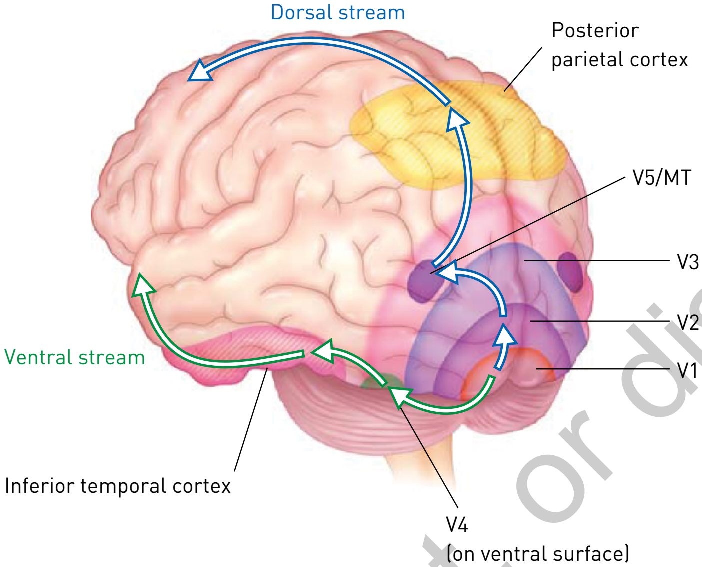
Brain Pathways

People's remarkable ability to recognize patterns occasionally falls victim to various types of visual disorders, and studying these disorders has contributed to our understanding of visual perception (Haque et al., 2018). For example, patients with brain damage have revealed a dissociation between knowing what an object is and knowing where the object is located. Damage to one part of the brain results in an impairment of the ability to recognize visual stimuli, whereas damage to another part of the brain results in an impairment of the ability to indicate their spatial location.

These impaired aspects of vision are similarly impaired in visual imagery (Levine et al., 1985). A patient with object identification difficulties was unable to draw or describe the appearance of familiar objects from memory, despite being able to draw and describe in great detail the relative locations of landmarks in his neighborhood, cities in the United States, and furniture in his hospital room. A patient with object localization difficulties could not use his memory to perform well on the spatial localization tasks but could provide detailed descriptions of the appearance of a variety of objects.

Figure 2.19 illustrates the two pathways that support the localization and the identification of objects. The *where* pathway is primarily associated with object location and spatial attention. It is often referred to as the dorsal pathway because it is located in the dorsal (or upper) part of the brain. The dorsal pathway runs upward to the parietal lobes and has strong connections with the frontal lobe that coordinates limb and eye movements.

FIGURE 2.19 ■ Brain Pathways for Spatial (Dorsal) and Object (Ventral) Identification.



The other pathway, which results in object recognition, is known as the *what* pathway. It travels from the primary visual cortex in the occipital lobe and processes information such as shape, size, and color, as previously illustrated in Figure 2.4. It is primarily located in the temporal lobes and is often referred to as the ventral pathway because it is located in the ventral (or lower) part of the brain.

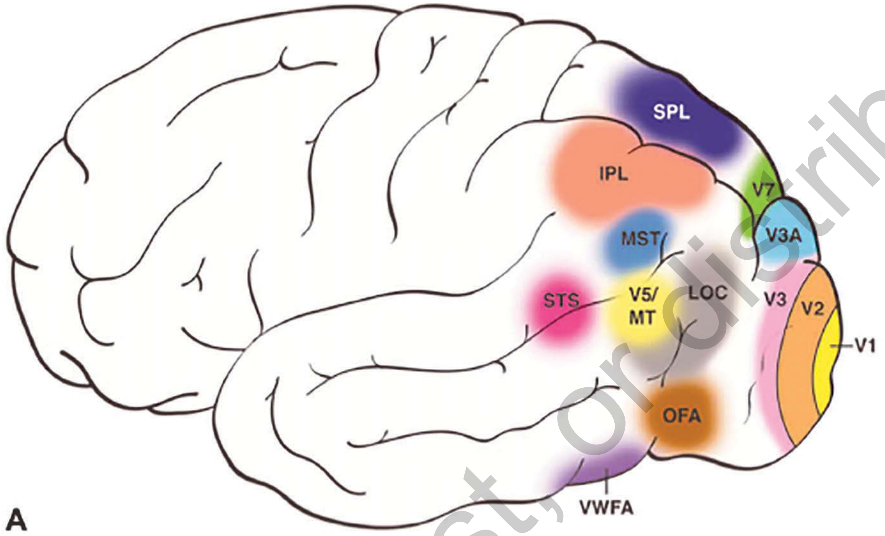
Figure 2.20 shows the approximate locations of specialized areas for object recognition. Some parts of the brain—the occipital face area (OFA)—respond more to faces than to other types of objects. Although this area is best activated by faces, it can also be activated by other objects, particularly if the person has acquired previous expert knowledge about those objects (Haque et al., 2018). The visual word form area (VWFA) is activated during reading.

Visual Disorders

Much of our knowledge of how the brain recognizes patterns comes from studies of patients with visual agnosia. Visual agnosia is a general disruption in the ability to recognize objects. Agnosia patients have normal visual acuity and generally show no memory deficits. The disability is also limited to a single sensory modality—for example, if you show a patient a set of keys, he will not be able to recognize them; however, if you hand him the keys to feel, he will easily identify them as keys. There are specialized forms of this disorder, such as an inability to recognize faces or familiar places.

visual agnosia An impairment in the recognition of visual objects

FIGURE 2.20 ■ Specialized Areas of the Brain. Areas Discussed in the Text Include the Parietal Lobe (SPL and IPL), the Temporal Lobe (MST and MT), the Visual Cortex (V1-V7), the Occipital Face Area (OFA), and the Visual Word Form Area (VWFA).

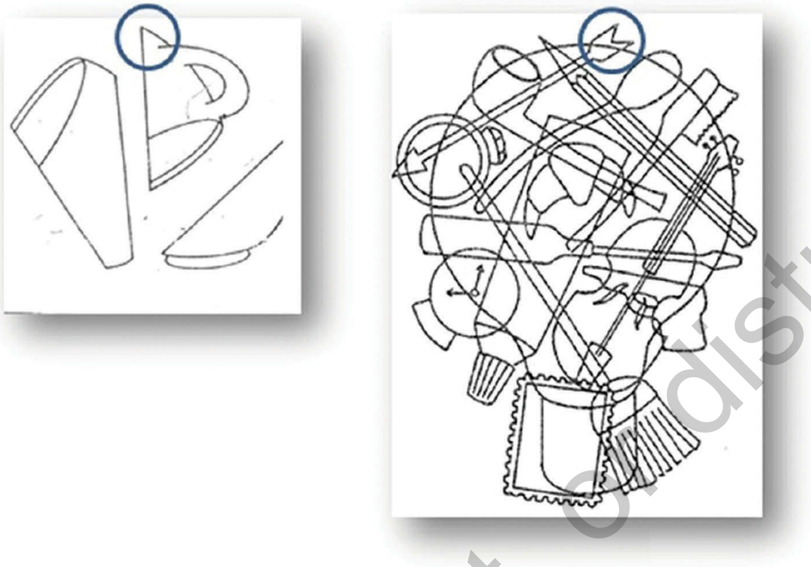


Two general categories of agnosia disorders are **apperceptive agnosia** and associative agnosia (Farah, 2004). Apperceptive agnosia disrupts the ability of patients to group visual elements into contours, surfaces, and objects (Farah, 2004). Evidence from these patients demonstrates the pattern recognition is normally hierarchical—starting with simple cells in the primary visual cortex and then combining these features to form a perception of the whole object, such as the face in Figure 2.4. The fusiform gyrus in the temporal lobe is of critical importance to this process (Konen et al., 2011), as is the lateral occipital cortex (Ptak et al., 2014). It is the last stage of combining features that impairs people with visual agnosia.

Inadequate eye movements contribute to the failure to combine visual features (Raz & Levin, 2017). A patient with apperceptive agnosia identified an object as a bird from a visual organization test, shown in the left panel of Figure 2.21. He identified the circled fragment as a beak but ignored the rest of the picture. In the right panel from an overlapping-figures test, a patient hesitated in deciding whether the circled fragment was an arrow or the ear of a cat. His eye movements did not track the length of the object to determine its identity.

apperceptive agnosia An inability to combine visual features into contours, surfaces, and objects

FIGURE 2.21 ■ Object Identification Tests.



Taken from the *Hooper visual organization*

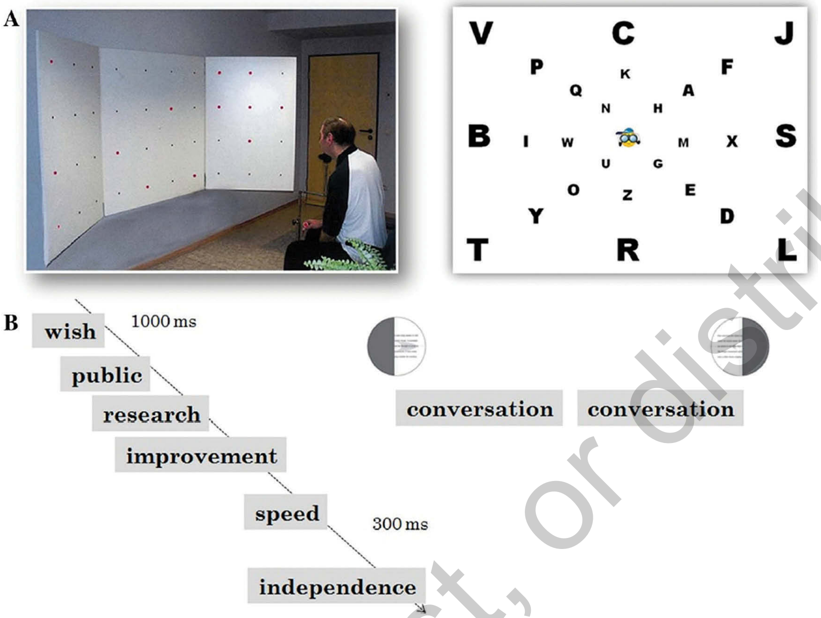
Taken from the *15-object test*

Inadequate eye movements also occur in reading when patients perform shorter and delayed eye movements that limit their ability to integrate letters (Raz & Levin, 2017). A training task, shown in the right panel of Figure 2.22A, requires them to track letters in an alphabetical sequence. Another training task, shown in Figure 2.22B, provides practice in reading words. The number of letters in the words increases while the presentation time decreases as training progresses. To perceive the entire word, patients are trained to fixate on either the beginning or the end of the word depending on their particular deficit. Training tasks also exist for large visual fields. The person in the left panel of Figure 2.22A is searching for a square composed of four red dots.

In contrast to apperceptive agnosia, **associative agnosia** patients can combine visual elements into a whole perception but are unable to identify that perception. The most curious fact about these patients is they can accurately copy a line drawing but are unable to recognize what they have drawn! Essentially these patients can perceive the object but can no longer associate their perception with its meaning.

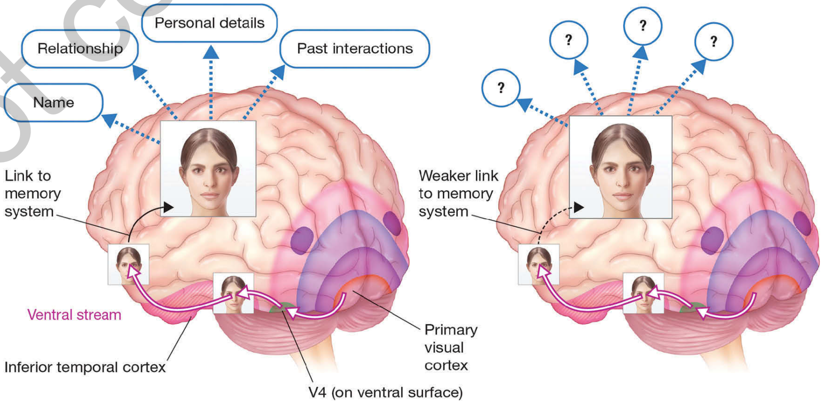
associative agnosia An ability to combine visual features into a whole object but an inability to recognize that object

FIGURE 2.22 ■ Training Eye Movements (A) and Reading Words (B).



Face blindness provides an informative case study of how the "what" stream fails to connect to other parts of the brain. An area of the cortex known as the *fusiform face area* is responsive to recognizing that an object is a face, even for people with face blindness (Mitchell, 2018). Although the brain performs the initial stage of face processing perfectly well, it fails to communicate that information with the frontal cortex for people with face blindness (Figure 2.23). The link to the frontal cortex is necessary to recall information such as the person's name, personal details, relationship, and past interactions.

FIGURE 2.23 ■ Disruptive Pathways Causing Face Blindness.



Source: Adapted from *INNATE: How the Wring of Our Brains Shapes Who We Are*, by K. J. Mitchell, 2018, Princeton, NJ: Princeton.

The condition can be so debilitating that patients may not recognize close family members or even their own face. The famous neuropsychologist Oliver Sacks suffered from face blindness prior to his death. His book *The Man Who Mistook His Wife for a Hat* is based on a clinical case study of such a patient (Sacks, 1985). Interestingly, many patients suffering from this disorder can recognize the faces of loved ones after they hear them speak.

SUMMARY

Pattern recognition is a skill that people perform very well. Three explanations of pattern recognition are template, feature, and structural theories. A template theory proposes that people compare two patterns by measuring their degree of overlap. A template theory has difficulty accounting for many aspects of pattern recognition. The most common theories of pattern recognition, therefore, assume that patterns are analyzed into features. Perceptual discrimination requires discovering distinctive features that distinguish between patterns. Treisman's experiments on feature integration theory explored how a perceiver combines two features that are analyzed by separate parts of the visual system. Structural theories state explicitly how the features of a pattern are joined together. They provide a more complete description of a pattern and are particularly useful for describing patterns consisting of intersecting lines.

Sperling's interest in the question of how many letters can be perceived during a brief exposure resulted in the construction of information-processing models for visual tasks. Sperling proposed that information is preserved very briefly in a visual information store, where all the letters can be simultaneously analyzed. When a letter is recognized, its name is verbally rehearsed and preserved in an auditory store that is a part of short-term memory.

Recognition of letters in a word is influenced by perceptual information and the letter context. The finding that a letter can be recognized more easily when it is part of a word than when it is part of a nonword or is presented by itself has been called the *word superiority effect.* An influential model of this effect is the interactive activation model proposed by McClelland and Rumelhart. Its major assumption is that knowledge about the words of a language interacts with incoming feature information to provide evidence regarding which letters are in the word. Scenes are composed of interacting objects that are typically arranged in a meaningful spatial layout. Recognizing objects in scenes is often driven by accomplishing goals. Deep neural networks, used in scene recognition and many other complex AI tasks, utilize the same principles of simpler networks but have added multiple layers of connections to fine-tune the weights of thousands of connections.

Visual agnosia is a disruption in the ability to recognize objects. There are specialized forms of recognition disorders, such as an inability to recognize objects or familiar places. The "where" pathway is located in the upper parietal area and is primarily associated with object location and spatial attention. The "what" pathway supports object recognition and is primarily located in the lower temporal lobes. Patients with apperceptive agnosia are unable to combine visual features into a complete pattern whereas associative agnosia patients can, but these patients can not identify the pattern.

RECOMMENDED READING

Hoffman's (1998) book, *Visual Intelligence*, provides both a readable and scholarly analysis of how we construct descriptions of objects. Fallshore and Schooler (1995) argue that verbally describing faces can lower later recognition because verbal descriptions ignore configural information. Kristjansson and Egeth (2019) provide an extensive history of how feature integration theory integrated relevant research in cognition, perception, and neuropsychology. Experts provide an overview of the theoretical contributions of neural networks (McClelland et al., 2010). For a history of neural networks that has resulted in the exciting accomplishments of deep networks read *The Deep Learning Revolution* (Sejnowski, 2018). A very readable introduction to genetics and brain circuits is Kevin Mitchell's (2018) book *INNATE: How the Wiring of Our Brains Shapes Who We Are.*