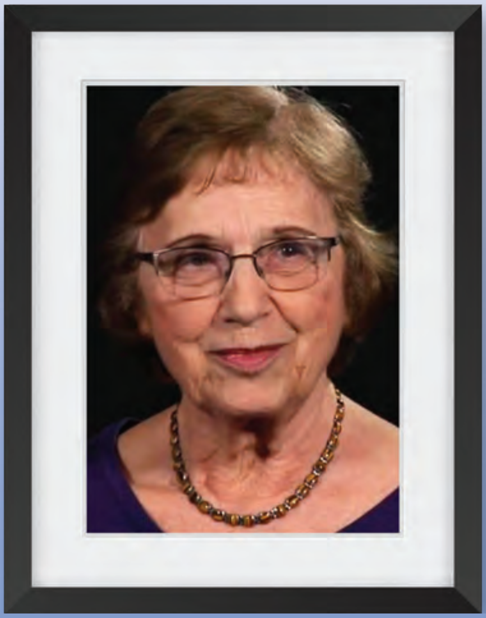
Adobe Stock

Seminal Researchers & Theorists



National Science Foundation, public domain, via Wikimedia

Anne Treisman (1935—2018)

Treisman was educated at Oxford University, where she began her work on selective attention in hearing and vision. She became a leading contributor to attention theory through a career that took her to the Universities of British Columbia and California, Berkeley, and eventually, Princeton. She developed several of the most important theories in the field of attention, particularly Feature Integration Theory (FIT), which generated decades of research attempting to resolve knotty issues such as how basic features are assembled into object representations in vision.

Chapter 5

Attention

> “Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatter-brained state . . .”
>
> William James, *Principles of Psychology, 1890* (1952), p. 261

Chapter Outline

Preview Questions

- *Is there one kind of attention, or does attention do different things in different situations?*
- *What is the purpose of attention? How does it help us to deal with the outside world?*
- *Does attention change over space and time, or does it follow where we are looking and our current level of alertness?*
- *Can we learn something without paying attention to it?*
- *What changes in brain activity can be observed as attention changes from one focus to another?*

5.1 Introduction and Definition of Attention

Everyone does know what attention is. When someone tells you to "Pay attention!" you usually have no doubt about what is meant. James further described attention as a process that determines the contents of consciousness. In other words, the only things that we know, in some sense, are those of which we are consciously aware, and attention directs our awareness. Since attention follows or actually determines much of what we are doing, learning itself relies on attention, and poor attention can interfere with learning. Of course, there are many things that we can learn to do expertly that are not conscious or attended to in any specific way. One could think of how we learn to ride a bicycle or learn to walk, speak, swim, or hit a ball with a bat. Perception and language are only two of the immense human capabilities that are present in some form at birth, but develop prodigiously at the hand of experience, with or without our attention to many details of what is being learned and how. As these important skills become so well-learned that much of their execution becomes automatic (that is without our needing to attend to them), attention can be freed to deal with other tasks, such as understanding and responding to what we see and hear.

Attention

A process that selects information from the environment and from memory for entry into limited-capacity conscious awareness

Attention has been defined in several related ways. Researchers have distinguished between various states of attention, such as vigilance (maintaining focus on a single task or information source, such as a screen for an air traffic controller), arousal or alertness (readiness for any new input), selection and avoidance of distraction, orientation, divided or switching attention, and filtering (Banich & Compton, 2018; Dehaene, 2020). However, the most commonly researched aspects of attention in psychology are in terms of selection and capacity. Selection is necessary because the environment is simply too full of information for us to attend to all of it simultaneously and to process it in any meaningful way. Similarly, we are subject to an enormous number of memories, plans, and worries that can compete for our attending to the environment. For example, as you are reading this sentence, you might be trying to attend to its meaning while trying not to be distracted by background noise from conversations or the TV. It is also possible to be distracted by a sudden itch or by noticing that the room is cold. Some of these interruptions of our attentional focus are important—in special circumstances our survival can depend on being able to avoid an obstacle in the road, or reaching out for a handrail when tripping on the stairs. At other times such interruptions are annoying, as when our train of thought is broken into by some recurring worry about relationships, school, or finances. In general, attention serves to select some important information for us to process deeply, while the vast majority of sensory inputs and recurrent thoughts are not so selected and have only fleeting effects on our conscious awareness, if we are even aware of them at all.

Capacity limitations are related to selection because humans are notoriously incapable of doing two or more demanding things at once. Selection is used to channel the information that we need for accomplishing current tasks into conscious, working memory while preventing information overload from irrelevant information. Kahneman (1973) proposed that such capacity limits place strong constraints on how well we can do two or more things at a time. Wickens (1980) refined this idea by defining capacity limits at different levels in the human attentional system. He argued that attentional capacity could be viewed as having separate pools of processing resources for perceptual processing, cognitive/decision processing, and verbal versus spatial processing. The idea is that resources are consumed more rapidly if two tasks both demand the same type of perceptual or cognitive resources, than if one is primarily perceptual and the other is primarily cognitive. Further, with practice, an attention-demanding process can become automatized such that it requires fewer resources to operate than when it was first being learned (e.g., Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977).

Sometimes we are capable of doing two well-learned activities at once, such as driving an automobile along a familiar route while conversing with a passenger. If, however, a car swerves in front of us or a small child runs into the street ahead, the quality of our conversation is likely to take a sudden downturn. Likewise, when our attention is directed more toward an ongoing conversation than the simultaneous task of driving, then it is the quality of our driving that is likely to suffer. This is one reason that many governments around the world are restricting the use of mobile phones by car drivers. Some research has indicated that talking on a phone while driving increases the risk of accidents to about the same extent as alcoholic intoxication (i.e., a fourfold increase; Redelmeier & Tibshirani, 1997; see also Strayer et al., 2003), and there is no safety advantage in limiting use to hands-free telephones (e.g., Strayer & Johnston, 2001). Anything that requires our full attention literally leaves no capacity for doing much of anything else at the same time.


Adobe Stock

Research has indicated that using a mobile phone while driving increases the risk of accidents to nearly the same extent as alcoholic intoxication.

5.2 Attention and Consciousness

Attention acts as a gatekeeper to determine the contents of consciousness. Our conscious processes are limited (who can profitably listen to a lecture and read an email from a friend at the same time?), so attention works to keep us focused on the task at hand while preventing unwanted thoughts and perceptions from interrupting us. Attention deficit disorders in children and adults as well as the **inhibition deficit hypothesis** of normal aging (e.g., Hasher & Zacks, 1988) indicate that we all have problems to varying degrees in attempting to maintain the focus of attention in the face of numerous distractions. In a class on reading taught by one of the authors (Juola), undergraduates sometimes reported that they had read a page or more of a book before noticing that the mind had wandered onto other topics, yet their eyes had methodically continued to plow through the text. Unwanted thoughts as well as annoying environmental stimuli can be sources of distraction when we are trying to maintain the focus of attention on the task at hand.

Inhibition deficit hypothesis

A theory of cognitive aging proposed by Hasher & Zacks that attributes one of the difficulties that older persons have with the inability to ignore irrelevant information, such as in trying to follow a specific conversation at a noisy party

James's definition of attention was based on an introspective appeal ("Everyone knows . . . ") that rang as true to laypersons of his day as it does to most contemporary researchers in cognition. Attention determines our conscious experiences and recollections. Yet, we do not seem able to agree on a definition of **consciousness**.

Consciousness

The subjective feeling that we have an inner self that experiences the physical world, recalls past experiences, and acts as an agent in making decisions and carrying them out

We can talk freely about consciousness that can be raised, lost, or come to again, but precise definitions escape us, as do related issues of whether animals are conscious or whether computers will ever develop consciousness. We will address such issues later (in Chapter 12) when we will consider how philosophers have viewed consciousness as aided by research in biology, psychology, and artificial intelligence.

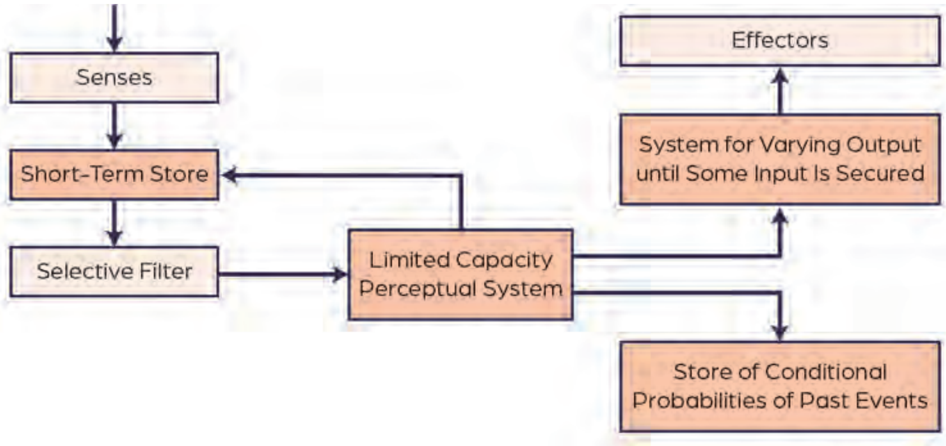
For now, let's equate consciousness with most of what goes on in working, short-term memory (see *For Discussion 1–3: A General Theory of Human Memory* in Chapter 1; see also Chapter 6), as we attempt to remain focused on current, important tasks. As the capacity of working memory is limited, attention serves many roles in maintaining focus and determining what is important enough to override the current contents of memory with new information. Broadbent (1958) described how attention controls working memory contents in his classic **filter theory of attention** (see Figure 5–1). He argued that since the environment is filled with many sources of information that cannot possibly be encoded fully, we need some kind of mechanism to sort things out. For Broadbent, attention presumably acts like a filter or gate that can be set at will to allow some information source to feed directly into working memory, whereas non-selected sources are literally filtered out and the information is lost and gone forever. That is, attention is presumably necessary for learning new information, and inattention in class or while watching a demonstration can produce poor retention, as we all know very well. Much of Broadbent's theory was based on the results of selective listening experiments (see *For Discussion 5–1: The Cocktail-Party Phenomenon*), but the theory can easily be generalized to other sensory modalities. To begin, let's focus on visual attention, and review how it was first studied empirically.

Filter theory of attention

Broadbent's theory assumes that there is an early filter placed between most sensory channels and perceptual awareness, such that typically only one information source at a time is selected for entry into conscious awareness and working memory; contrasted with late filter theories that allow selection after information has been channeled to long-term memory for identification

Figure 5–1 **A Schematization of Broadbent's (1958) Early Filter Theory of Attention**

The sensory systems feed information into a short-term sensory store (e.g., iconic memory for vision; echoic memory for hearing) that briefly holds information while an adjustable filter selects a single message for entry into a limited-capacity perceptual (or working memory) system. The outputs from the perceptual system adjust readouts from the sensory stores (selective attention), consult memory for past conditional probabilities (e.g., environmental correlations), and determine the eventual responses made by the effectors. The filter theory is known as one of the first information-processing theories in psychology.



Source: Adapted from *Vision Science*, by S. E. Palmer, 1999. Cambridge, MA: MIT Press, p. 63.

For Discussion 5–1

The Cocktail-Party Phenomenon

You have surely experienced the problem of trying to carry on a conversation in a noisy environment. The noise can be due to traffic, a rock concert, a sporting event, or even several people who are talking at once. The latter situation was chosen as the paradigm case for a large body of research on a topic known as the cocktail-party phenomenon. The results led to the first theories of how we use attention to filter unwanted information from impinging on conscious awareness.

At a party, there are typically many people speaking simultaneously, and you might attempt to follow or join in on one of these conversations to the attempted exclusion of others. With attention, this can be done, but it is sometimes difficult. Difficulties arise in situations in which you might be trapped with a bore while you notice that a much more attractive and interesting person has entered the room. In other cases, a conversation away from the one to which you are supposed to be attending becomes suddenly salient, for example, if you hear your own name mentioned behind you.

The cocktail party was simulated in simplified laboratory situations in which stereo speakers or headphones were used to present two different spoken messages simultaneously. The participant was instructed to attend to one of the messages and warned that memory for the attended message would be tested later. To ensure that participants were indeed attending to one of the messages throughout, one of the messages was to be "shadowed," that is, reported back aloud as it was heard. Results from experiments by Broadbent (1952, 1954) and Cherry (1953) showed that people were able to shadow one message and repeat it back fairly accurately while ignoring the other message. The more surprising result, however, was the almost complete lack of memory or even awareness of the nonshadowed message. Almost nothing could be recalled from the unattended message, and people did not even notice if the nonshadowed message was changed to a different language or replaced by backward speech!

These results formed the basis of Broadbent's early filter model of attention (see Figures 5–1 and 5–2). The term "early" refers to the idea that unattended messages are filtered out before they can activate memory representations to produce recognized words in conscious awareness. The early filter theory was soon challenged by other theories that acknowledged our abilities to encode many different, familiar items in the external world to the level of recognition. Then, a late filter would select those items relevant for the ongoing task, resulting in decay or active suppression of unwanted, irrelevant codes in memory. Late selection models (e.g., Deutsch & Deutsch, 1963; Norman 1968) are based on the assumption that familiar stimuli, like spoken or written words and common objects, are encoded automatically, so it does not matter if we are attending to them or not. The role of attention in late selection models is to channel the most important and relevant meanings of items into working memory so that we can maintain focus on a single task.

Some ingenious experiments by Anne Treisman (1960, 1964) offered proof that early selection is incomplete as a theory of human attention. Rather, she argued that if an early filter exists, it operates only to attenuate responses to some stimuli while emphasizing the encoding of others. Then, important or relevant information always has an opportunity to get through the filter, even if it comes from an unattended channel. She demonstrated the effects of relevance by switching the two messages in midstream between the attended and unattended channels. Participants in her

(continues)

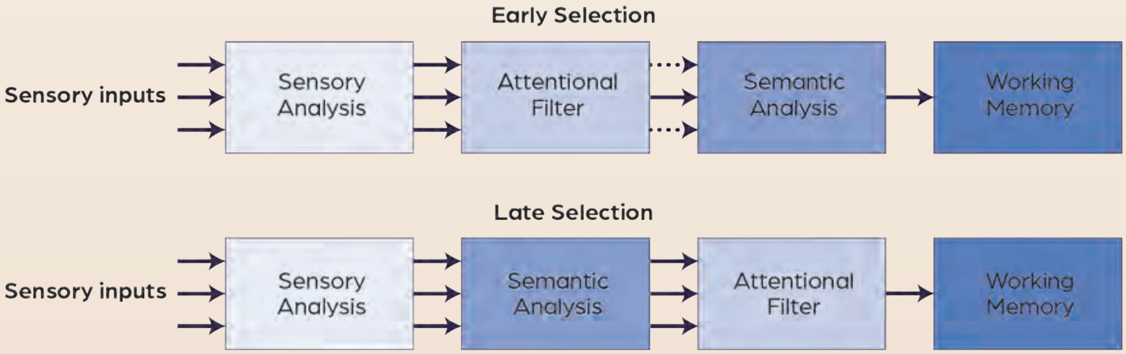
(continued)

research often followed the message, rather than the channel, thus indicating a level of control based on meaning (a late filter effect), rather than on the physical feature of location (early filter).

Treisman also showed that if the same message was presented in both channels, but was out of sync by several seconds, listeners did not notice that they were the same. However, if the relative speeds of the two tapes were changed so that the two messages gradually caught up with each other, listeners would eventually notice that the same message was being presented to the two ears. She also found an interesting asymmetry of some theoretical importance. If the unattended message was leading, the two messages had to be played within about 1.5 seconds of each other before the identity was noticed. However, if the attended message was leading, a lag of as much as 4.5 seconds could pass, yet listeners still noticed the identity. This asymmetry points out one fact of early versus late (or deep) processing: Stimulus inputs that are ignored may be processed to the level of meaning, but they remain in memory for only a second or two before they are lost and forgotten. With attention, however, inputs are processed to a deeper level and remain in active, working memory for at least 4.5 seconds, on average, in the shadowing task, and even longer in a more permanent memory system as demonstrated by later tests for recall of the contents of the shadowed message.

Figure 5–2 **Early and Late Selection Models**

The upper figure demonstrates the early selection theories of Broadbent (1958) and Treisman (1960); the dotted lines represent information attenuation suggested by Treisman (1960). The lower figure demonstrates the late selection theories of Deutsch & Deutsch (1963) and Norman (1968).



Sources: *Perception and Communication*, by D. E. Broadbent, London: Pergamon; "Contextual cues in selective listening," by A. M. Treisman, *Quarterly Journal of Experimental Psychology*, 12, pp. 242–264; Attention: Some theoretical considerations, by J. A. Deutsch and D. Deutsch, 1963, *Psychological Review*, 70, pp. 80–90; Toward a theory of memory and attention, by D. A. Norman, 1968, *Psychological Review*, 75, pp. 522–536.

In summary, we have evolved with an attentional system that serves two purposes. First, there is a deliberate attempt to seek out what is important for the task at hand and to maintain focus on this information while attempting to attenuate irrelevant information. At the same time, there is a tendency for information to impinge on conscious awareness if it is related to the current task or if it is potentially important for our survival (e.g., a sudden sound or appearance of a new object).

5.3 Visual Attention

Most often we can freely attend to any part of the visual environment, and this freedom can lead to the illusion that everything is visible and perceptible at once (i.e., that we have a photographic record of the visual scene). This phenomenon has been called "the grand illusion of complete perception" (e.g., Enns, 2004). It is merely an illusion, and it is clearly false. People can and do miss things that are quite obviously present, most commonly because our eyes record more information than we can attend to or report. Change blindness (Rensink, 2001; Simons & Levin, 1997) and inattentional blindness (Mack & Rock, 1998) are two well-known phenomena showing how people can completely miss an otherwise easily detectable target if their attention is focused elsewhere.

The classic change blindness paradigm involves viewing a sequence of pictures in which one scene is alternated with another in rapid succession, and there is one obvious difference between them (e.g., the presence or absence of a tree in the background). Observers often take several seconds or even minutes to detect the difference, even though it is obvious once they notice it. It appears that contrary to the grand illusion most of the visual scene is represented in a sketchy, gist form, and details are available for only a small area on which attention is focused (see Figure 5–3).

Figure 5–3 **An Example of Change Blindness**

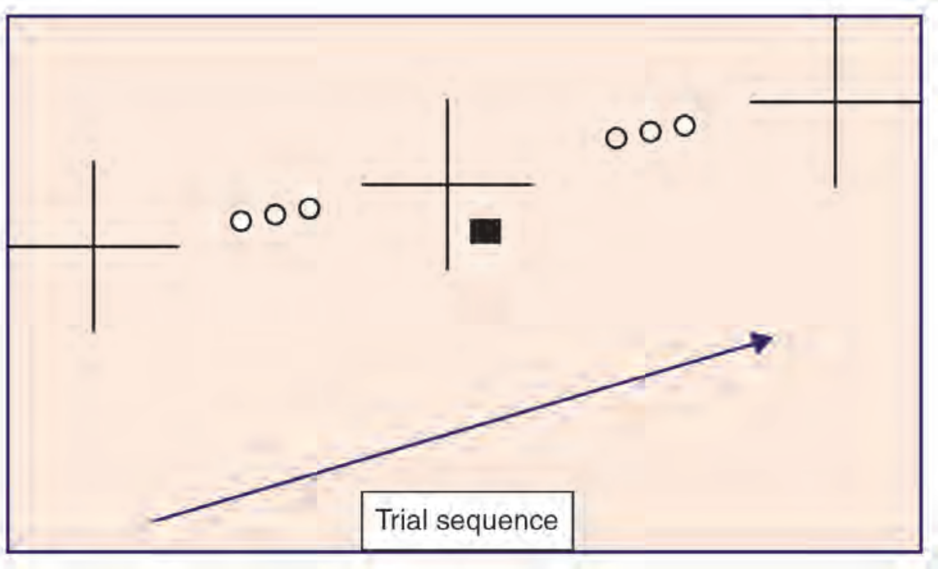
The two photographs represent successive shots in a film used in a change blindness experiment (typically the two scenes are alternated with a short blank or mask between them). Most such changes go unnoticed by observers unless they are specifically attending to them when the change occurs.

Source: Shutterstock

In the inattentional blindness demonstration, subjects are given a difficult perceptual task, such as a brief presentation of two lines that cross in the center of the display, and they have to judge which is longer—the horizontal line or the vertical one. On one of the trials, a geometric form is presented along with the lines, and the subjects are asked afterwards if they could identify it (see Figure 5–4). In fact, most subjects are surprised that there was anything other than the lines presented at all! The conclusion is that we have the illusion of complete perception of a complex stimulus only because we can use our attention to select any component of the visual or auditory scene and experience it at an arbitrary level of detail. In this way, the stimulus functions as a back-up memory from which we can select and direct our conscious awareness to any of a number of specific components (see O'Regan & Noë, 2001).

Figure 5–4 **Example of the Inattentional Blindness Paradigm Used by Mack and Rock (1998)**

Several trials of the main task (judging which of two crossing lines is longer) had a single frame embedded in a series with an unexpected geometric form included in the periphery. Subjects performed poorly on a delayed form memory test, indicating a lack of stimulus recognition without attention.



Another famous example of inattentional blindness was provided by Simons and Chabris (1999). They had subjects watch a video clip of two teams of players passing a basketball back and forth, one ball for a team dressed in black uniforms, and another for a team dressed in white. The task was to count the number of successful passes completed by either the white team or the black team (for different subjects). This task was complicated enough that the observers' attention was completely taken up by counting one set of passes while filtering out the others. The filter worked so well, that most people did not notice that during the video, a person dressed in a gorilla suit walked through the middle of the players involved in the passing game! Most subjects were astonished afterwards to be informed of this fact and could hardly believe they could have missed it when they viewed the same film a second time and were asked to look for the gorilla.

To test the grand illusion of complete perception idea, an important question to ask is: how much information is recognizable in a single eye fixation (e.g., Cattell, 1900)? Eye fixations occur at a more-or-less regular pace of two to five per second, interspersed between eye movements called saccades, but shorter and longer ones can be present in any task. Information presented clearly for about 200 milliseconds (ms) is presumed to represent what is visible in a single eye fixation. Of course, in the real world, our fixations follow a path determined by internal goals or strategies as well as by salient external events. Our perceptions are built up cumulatively over a succession of fixations to generate an internal 3-D model of the environment. Nevertheless, as we look closely at each successive step in an old building while walking up a dimly lit staircase, our eyes can dart to the location of a sudden movement in a corner. We can also trace a regular pattern of fixations separated by rapid saccadic eye movements in certain tasks like scanning an instrument panel or reading a paragraph.

For Discussion 5–2

Measuring the Duration of a Visual Image

Many experiments have been devised to answer the question: How much can we see in a single eye fixation? This question was addressed in Wundt's lab by James McKeen Cattell and also by Raymond Dodge and Benno Erdmann in Germany more than 100 years ago (see Huey, 1908/1968, for a thoughtful review of early research on word perception and reading). They used various types of apparatus to present words or letters very briefly that were to be reported. The general conclusion was that the amount reported depended on the meaning and familiarity of the material. For unrelated consonants, however, only four or five items could reliably be reported even if there were many more letters in the display.

From the behaviorist tradition, the number of letters reported should be taken as the number seen, since a correct report is a reliable behavioral response. Yet almost everyone who has ever participated in these types of experiments has complained that many letters can actually be seen, but they gradually fade away before they can be reported. What can an experimental psychologist do if observers' introspective reports disagree with the data that are collected? What was done for almost a century was to ignore the subjective reports and record the actual data. The number of letters correctly reported after a single brief exposure of a multi-letter display was taken as a measure of what was actually seen.

In 1960, George Sperling published the results from his doctoral dissertation in a paper that became one of the hallmarks of the cognitive revolution. He took his observers' introspections to heart and tried to invent a way to measure how much information was actually available immediately after a brief visual display. He hit on the clever idea that rather than asking his participants to report all the letters in the display (whole report task), he would only ask for part of the display to be reported (partial report task). The trick was that the observers never knew which part was to be reported until after the display was turned off!

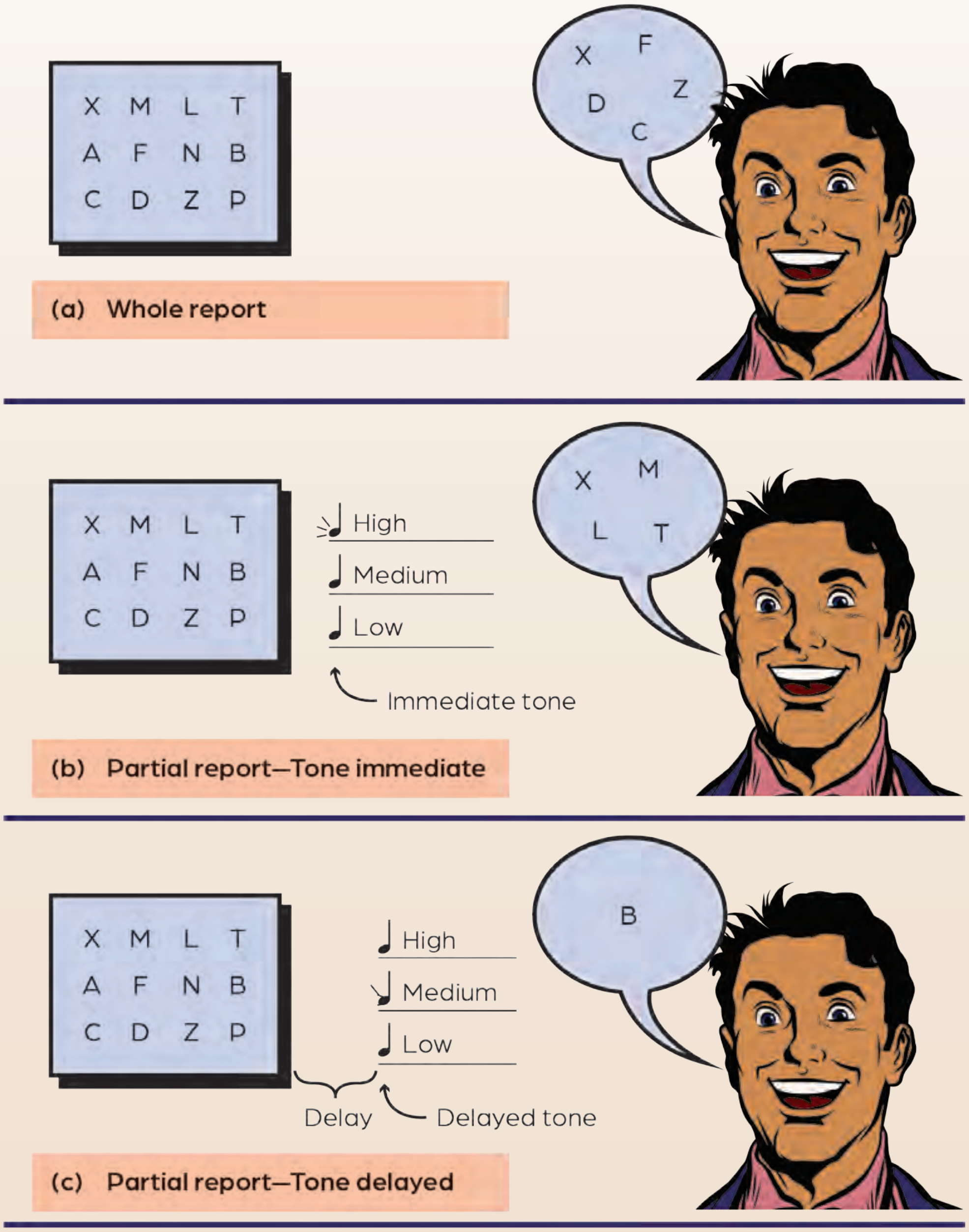
Sperling used displays of unrelated consonants in rows and columns centered on the display. For example, the display could include four consonants in each of three rows. Sperling used a tone, sounded shortly after the display ended, to cue the observers which row to report. A high tone indicated the top row, a medium pitch indicated the middle row, and a low tone indicated the bottom row. He found that people could accurately report letters from the cued row if the tone was presented within a fraction of a second after the display was shut off. However, if the tone was delayed further, observers could only guess from the four or five letters that they could remember in short-term memory (whole report level). That is, the partial report technique showed an advantage over whole report that was statistically significant for cue delays up to almost a half-second after the display was physically gone.

(continues)

(continued)

Figure 5–5 **Sperling's (1960) Experiment**

The procedure used in Sperling's (1960) experiment. (a) In the whole report task, subjects report as many letters as possible from a brief (50 ms) display—about 4.5 is the average number. (b) If a cue is given to report the top row (high tone), middle row (medium tone), or bottom row (low tone), performance is almost perfect (4 out of 4) if the tone occurs as soon as the display is offset. (c) If the tone is delayed for as little as half a second, however, performance drops precipitously.



Source: Adapted from *Cognitive Psychology* (p. 144), by E. B. Goldstein, 2005, Belmont, CA: Thomson/Wadsworth.

The important deduction that Sperling made was that since his subjects could report three to four letters from the single cued row, without knowing which row was to be cued, then they must have had three to four letters available for a brief time from each of the rows.

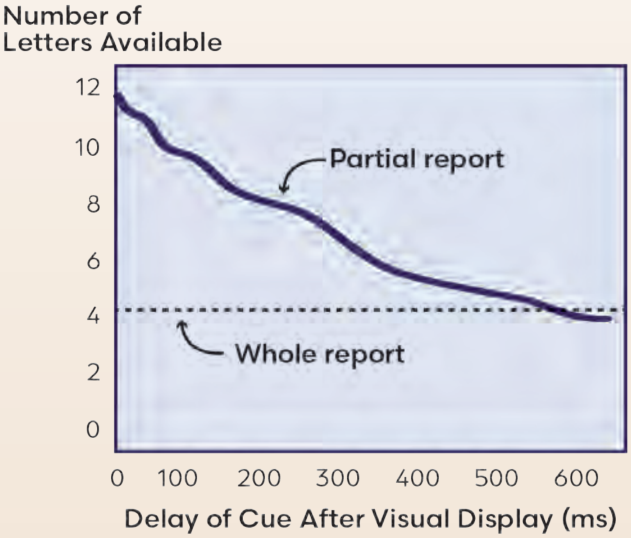
This very-short-term visual memory came to be called **iconic memory** to indicate its sensory nature. The icon is a brief image of a display that persists for some time in sensory memory after the display is gone. The iconic image can be scanned with attention much like a physical display can be scanned with eye movements, but its duration is very brief. Once the icon is gone, and Sperling found that little useful information persisted beyond about 300 ms in most conditions, only those letters that remain have been coded into a more durable form, such as letter names in verbal short-term memory.

Iconic memory

A visual buffer that can hold a large amount of information for a very brief duration (up to 500 ms, if attention is not allocated), involving the primary visual cortex (V1) and early visual areas (V2, V3); its contents are typically considered not consciously rehearseable because they exist in pre-attentive forms that are not maintainable unless augmented by selective attention

Figure 5–6 Partial Report Advantage Over Whole Report

Partial report shows an advantage over whole report while a sensory (iconic) image remains for a fraction of a second after the stimulus is turned off.



Source: Data from "The information available in brief visual presentations," by G. Sperling, 1960, *Psychological Monographs, 74*(11, Whole No. 498), pp. 1–29.

The importance of Sperling's research rests on its demonstration that although introspections might be unreliable, they nonetheless can be sources of theories to be tested in the laboratory. In this case, people's introspections are correct. A brief visual display is stretched out in time by persisting sensory activity, and as long as it lasts, the sensory image can be looked at by the "mind's eye" in much the same way that we examine a concrete image with eye movements. Since the time of Sperling's research, it has been shown that other sensory modalities, such as hearing and touch, also have brief memories. Sensory memories serve the likely function of lengthening the time that sensory information is available so that we can attend to selected parts and code them into a more permanent form for later use. The image of a tiger illuminated by a flash of lightning or the flickering light of a campfire would be useless if it could not be retained long enough for us to figure out what it is. Similarly, speech sounds cannot be recognized as words or sentences unless we have some type of memory available to integrate them over time. The first operation performed on selected sensory information is to try to match or categorize it with respect to known information in memory. We call this categorization process perception, and it is the topic of Chapter 3.

The classic experiment that showed beyond any doubt that people could see more than they can report, even if they cannot see everything clearly, was Sperling's (1960) study of partial report (*For Discussion 5–2: Measuring the Duration of a Visual Image*). In such controlled environments, the effects of attention on what we see and what we are able to report can be cleanly measured. Attention can be focused on some aspects of a visual display, and the region attended will result in better performance for recognizing objects in that region (see *For Discussion 5–3: The Spotlight Model of Attention*). Although we usually fixate (look directly at) the current object of our attention, much research has shown that attention serves a role in deciding where to fixate next. That is, attention operates covertly (without any outward signs) in determining what object or area in the visual field is worthy of more detailed inspection. The selected location then becomes a target for the next overt eye movement. An advantage for information processing in the attended region is often profited at a loss for unattended regions of the field. Thus, in agreement with Broadbent's (1958) theory, Posner (1980) argued that attention's most important role seems to be the selection of what information should be directed to conscious, working memory where it can be encoded more deeply for recognition and response decisions. This selection results in improved detection and recognition even before the eyes start to move, and even if the item of interest disappears before the eyes can fixate it (see *For Discussion 5–3: The Spotlight Model of Attention*).

Besides selection, attention serves as a monitor of conscious processing, always at the alert to detect new or important information. There is nothing so powerful in attracting one's attention as a sudden, unexpected sound or the sudden appearance of a new object in the field of view (Jonides & Yantis, 1988). Other unattended stimuli that have high personal salience, such as one's name, frequently result in a change of attention, even in defiance of one's current focus (e.g., Moray, 1959). Such selection is limited, however, by the ongoing processing capacity of short-term working memory.

For Discussion 5–3

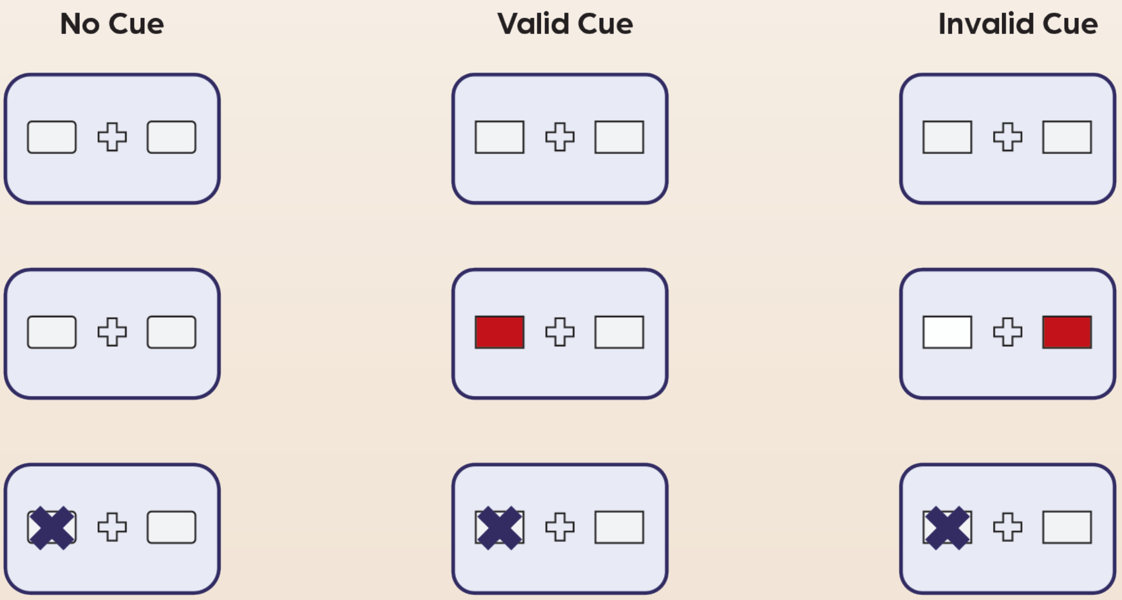
The Spotlight Model of Attention

Helmholtz (1850/1962) was among the first to speculate that attention and the direction of one's gaze are not necessarily correlated. He claimed that if he entered his office in the dark and looked straight ahead, he could prepare himself to attend either to the left or right of his point of fixation. To prove this fact to himself, he then briefly lit his office with an electrical spark and satisfied himself by being able to name more objects visible to the left or right of fixation, in the direction he had been attending. It was as if attention functions like a functional fovea to temporarily emulate the physiological fovea in the center, back part of the eye onto which things that we are looking at directly are imaged. Although the foveated object normally is represented in greater detail in the brain than any peripheral object, attention can function to override this predominance, such that items presented to the fovea are sometimes seen less well than objects presented to attended peripheral locations (e.g., Juola et al., 1991).

The first empirical work to study Helmholtz's claim in controlled situations was designed by Posner and his colleagues (Posner, 1980; Posner et al., 1980). They defined two areas to the left and right of a fixation point where a single target could appear. The defined areas were marked by boxes, and the target was a small X that could appear in one of the boxes. The task was to hit a response button as quickly as possible when the X appeared (a simple RT task). On some trials, one of the boxes brightened briefly before the target was presented (indicated in red in Figure 5–7). This was a cue that the target was likely to occur within that box, although the cue was not functioning to override this predominance, such that items presented to the fovea are sometimes seen less well than objects presented to attended peripheral locations (e.g., Juola et al., 1991).

Figure 5–7 **Three Kinds of Trials (shown in top-to-bottom sequence)**

Task: Hit a left-hand key if the target (X) appears on the left side and the right-hand key if it appears on the right side.



An important point to keep in mind is that the entire time for cue presentation followed by target onset and then target offset was kept below 200 ms, so the observers did not have time to move their eyes before the cue and display were shown and disappeared. Nonetheless, people's behavior indicated that something moved in the direction of the cue.

Posner asserted that this something was attention, and that it moves like a spotlight moves across a stage to illuminate part of the visual field, while leaving other parts in comparative darkness. If a target is presented within the spotlight of attention, it is detected and responded to quickly. If, however, the target is presented elsewhere, the spotlight usually has to be redirected to the target location, which takes time and slows responses. On no-cue trials, there is no spotlight, as observers attempt to spread their attention in a diffuse way over the entire display in preparation for a target that could occur anywhere.

(continues)

(continued)

Later research has shown that the spotlight metaphor, while useful, does not capture the full flexibility of how we distribute attention over the visual field. Depending on display characteristics, and task demands, attention can be distributed more or less evenly over the field, concentrated around central vision (i.e., the fovea), concentrated at some location or over some object in space, or even distributed over rings around the fixation point or in different locations at the same time. Such a high degree of flexibility emphasizes two important characteristics of visual attention. First, attentional systems have evolved to enable us to attend to a variety of different objects and events that are important for our survival. Second, what we see and respond to is determined not only by what is present in the visual world but also by what we are expecting and are determined to find.

5.4 Theories of Visual Attention

5.4a Exogenous versus Endogenous Attention

Exogenous attention

The bottom-up, automatic attraction of attention to a salient aspect of the environment, most certainly to a sudden sound or the sudden appearance of a new object

Endogenous attention

The top-down, voluntary direction of attention toward information that is likely to be related to one's goals or needs

Posner (1980) differentiated between what he called **exogenous** (environmentally generated, or bottom-up) and **endogenous** (self-generated, or top-down) controls of attention. Exogenous controls are mediated by lower brain systems, such as the colliculi in the midbrain, and they can result in rapid and automatic orienting of attention to an exogenous cue. Stimuli that most clearly attract attention automatically are sudden sounds or the appearances of new objects in the visual field. These primitive orienting responses have evolved early in animals, since being aware of new objects is an important determinant of survival. In fact, the superior colliculi of animals, such as fish and frogs, are the main processors of visual stimuli in their central nervous systems. It has been said that a frog will starve to death in a cage filled with dead flies, because they literally are not visible as food. Similarly, characters in the movie *Jurassic Park* were warned to remain still so that they would not be spotted by the *tyrannosaurus rex*. In mammals, and especially primates, these primitive brain-stem systems remain functional, but they are largely augmented, especially in conscious awareness, by visual centers in the cerebral cortex.

Endogenous controls of attention include deliberate searching for specific items, such as looking for a pair of red socks in a drawer full of multicolored stockings and other articles. Stimuli with features that match what we are looking for attract attention, but in a slower and less certain way than do exogenous cues (e.g., Jonides, 1980). This difference is primarily due to the fact that exogenous cues trigger responses in a reflexive, bottom-up way under direction from automatic, hardwired systems in the brain stem. Endogenous responses, on the other hand, are determined by search strategies under top-down control of executive systems and the frontal eye fields in the frontal lobes of the brain that direct the actions of lower visual systems.

5.4b Treisman's Feature Integration Theory (FIT)

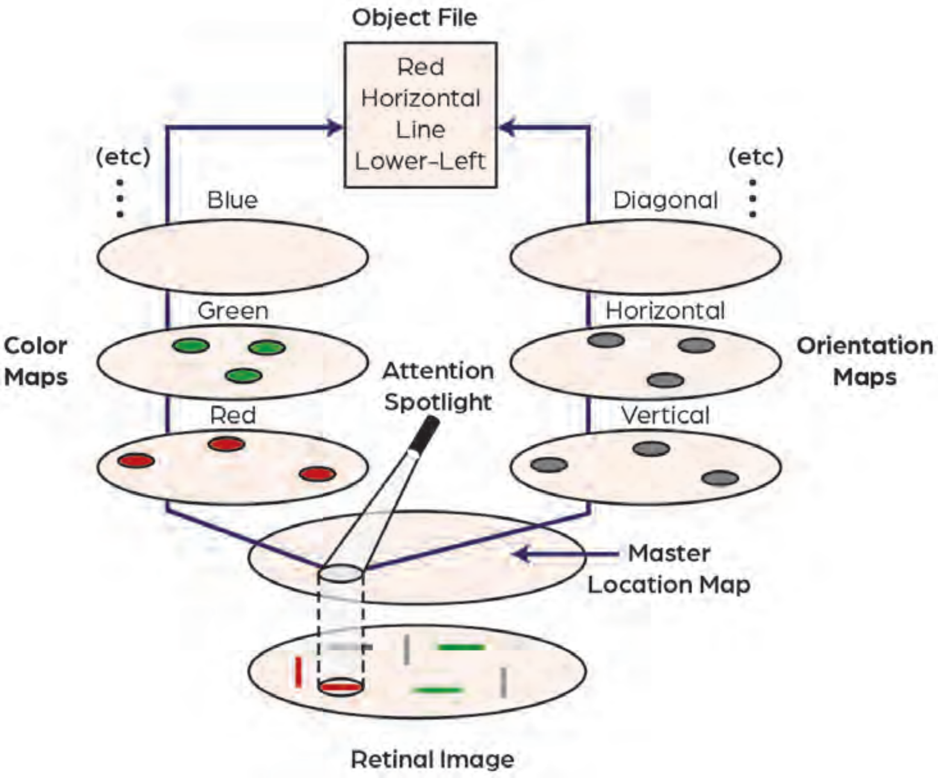
Many theorists have expanded on the argument that attention operates at different levels in the cognitive system. For example, Treisman and Gelade (1980) developed Neisser's (1967) argument that attention follows two successive stages in selecting items to be examined closely. The first stage is a **preattentive stage** in which a broad swath of the environment is processed in parallel. Most features in the field of view, including such things as color, orientation, size, depth, and motion, are coded by separate modules into feature maps that represent those specific aspects of the visible stimuli. In addition, there is a master map of locations to which all of the individual feature maps can be related. This relation, however, is achieved only in a second stage, called focal attention (see Figure 5–8).

Preattentive stage

In FIT, essentially the information available in sensory (e.g., iconic) memory; features are briefly available in different modules or "feature maps" of the environment that are registered and maintained in parallel

Figure 5–8 **Treisman's (e.g., Treisman & Gormican, 1988) Feature Integration Theory (FIT) of Attention**

All features in the visual field are registered in parallel in separate "maps" for color, orientation, motion, size, and other fundamental features. A master map of locations can be scanned with an attentional spotlight. Wherever attention is focused on the location map, features in the other modules at that location are bound together into an object file. In other words, attention acts like a glue that binds features into a recognizable object.

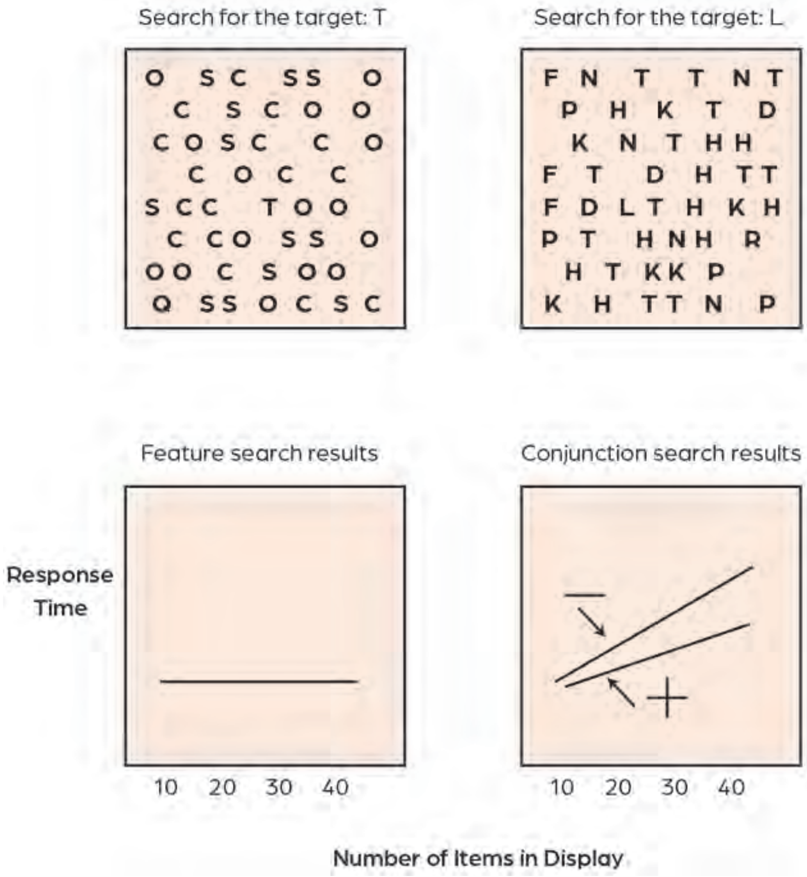


Source: Adapted from *Vision Science* (figure, p. 557), by S. E. Palmer. © 1999 Massachusetts Institute of Technology, by permission of The MIT Press.

In Treisman's theory, focal attention is limited to a succession of small areas in the visual field, much as a spotlight is moved around a large stage to illuminate specific settings. When the spotlight of attention falls on a particular location, all features represented in the preattentive stage are integrated into a single, object-level representation. For this reason, Treisman's theory is called the Feature Integration Theory (FIT) of attention. It is as though specific features are free-floating in a preattentive field, and attention acts as the glue that binds them into recognizable objects. Objects usually can be recognized only one at a time, however, making the second, focal stage of attention a serial process unlike the parallel feature processing that occurs in the earlier, preattentive stage (see Figure 5–9).

Figure 5–9 **A Comparison of Fast, Feature-Based ("Pop Out") Search and Slower Conjunction Search**

There is a difference between a search for a target that differs from the distractors by a single feature (orientation, left panel) versus a search for a target that differs from the distractors by a conjunction of features (i.e., how vertical and horizontal components are combined). Idealized search time data are shown in the lower panels for each type of search. In feature search, neither positive (target present) nor negative (target absent) search times depend on the number of items. The target appears to "pop out" from the distractors. In conjunction search, however, the search appears to be serial, rather than parallel, and it terminates with finding a target if it is present. Thus, on average, the positive search times (+ trials) are about half as long as the negative ones (– trials), since all items must be searched on target-absent trials.



Treisman and Gelade (1980) based their theory on two related findings from studies of visual search. First, it is well known that search is fast and easy for a target that differs from all other items in the visual field by a single feature. Try the tasks shown in Figure 5–9 to see that finding a target T among letters with curved features and no horizontal elements is quite easy. However, finding an L among letters with horizontal and vertical features is not so easy. Treisman and her colleagues argued that if the target differs from background distractor items by a single salient feature, then search should be parallel (all items examined at once), based on information obtainable preattentively. That is, a single red X will pop out from a field of blue Xs when one is looking for a red item. And it will be just as easy to decide that there is no target when there are only blue items in the display. However, if the target is defined by a unique combination of features, such as a red X among blue Xs and red Ns, the search is more difficult. Since the target features exist by themselves in other display items, the target can be found only in the second, focal stage of attention. In other words, if the target is defined by features that have to be conjoined by focal attention, in order to integrate all its components, then search proceeds serially (one at a time) until the target is found or until all items have been examined without finding a target.

Treisman and Gelade proposed a simple test to determine whether a target can be found at the preattentive level or only at the later, focal attentive stages of processing. The test is based on an examination of a plot of response times (RTs) in search tasks against the number of items in the display. In these search tasks, the target is present either once or not at all, so the person's task is to respond positively as soon as the target is found and to respond negatively when the search is completed without finding the target. The results are generally straightforward: If the target is defined by a single, unique feature, such as color, motion, or orientation, that differs from all the other items in a scene (**feature search**), search times are independent of the number of displayed items. (i.e., the response time by display size [RT by set size—see Figure 5–9] functions are flat). If, however, finding the target is possible only after focal attention has integrated the display items' features, such as finding the one red horizontal line among blue horizontals and red verticals (**conjunction search**), the slope of the RT by set size function is greater than zero. In some search tasks, especially those with more than about five items in the display, it seems as if we quit as soon as the target item is found, and the slope is about twice as large on negative (no target) trials as on trials containing the target. This result is expected, since if the display items are searched serially, the search should terminate as soon as the target is found, which would be, on average, about halfway through the search. On no-target trials, however, the search would have to cover all display items until a negative response could be made.

Feature search

The search for a target object that differs from the background distractors by a single, salient feature; the target appears to pop out, and search time is independent of the number of distractors in the display

Conjunction search

The search for a target object that differs from the background distractors by a unique combination of features that are individually present among the distractors; search time increases linearly with the number of distractors in the display

A second result Treisman used to support FIT was the occasional report of **illusory conjunctions** in a search display. That is, if a display is presented too briefly, or if an observer's attention is focused elsewhere, some display items will be processed only preattentively. If they do not undergo focal attentive processing, their features should be free-floating, and they should be almost as likely to combine with features from neighboring items as with features from their own item. If a person is asked to name an object from an unattended region containing red Xs and blue Os, for example, the response "Blue X" is likely to be made, even though no such stimulus was shown.

Illusory conjunctions

According to FIT, features that are not bound together by focal attention can be combined incorrectly, and subjects might report seeing objects that were not actually present

5.4c Modifications of FIT

Like any good theory, FIT generated much research that eventually led to revisions of the theory. (See the special issue of Attention, Perception, & Psychophysics, 2020, devoted to honoring Treisman's theory and its recent updates and extensions published 2 years after her death.) The main reasons for the revisions were the unreliability of some findings that were its main supports. That is, parallel search, as defined by a flat RT by set size function, is not limited to feature search. Some conjunctions apparently pop out as well (e.g., Nakayama & Silverman, 1986; Wolfe et al., 1989). Nakayama and Silverman showed that target items differing from distractors in depth or motion could be found quickly, without having to search through irrelevant items at a different depth or relative motion from the target. The crucial aspect in visual search appears to be whether or not the display can be segmented into potential targets and items that can be easily rejected. If some nontargets can be rejected as a set by having a salient feature or two that eliminates them from being possible targets, then the remaining items can be searched more efficiently. Searching through the remaining items can now be parallel if the target is defined by a single feature among these items (e.g., Egeth et al., 1984).

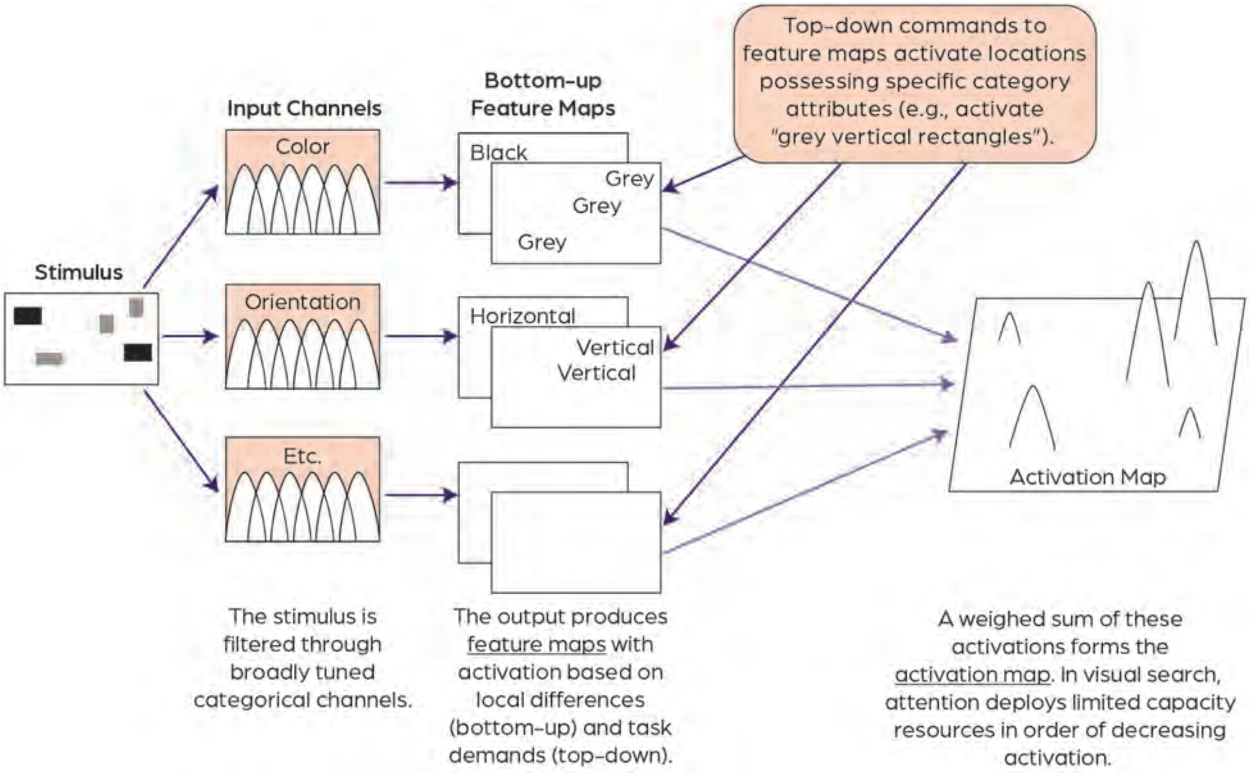
In a different rebuttal to FIT, Cohen and Ivry (1989) challenged the notion that illusory conjunctions occur frequently. They found that subjects reported such conjunctions only rarely, and when they occurred, they usually were miscombined features from neighboring items. They claimed that the effects of attention are small, at best, and the results are inconsistent with the idea that features are bound into an object only with focal attention and freely dissociated without it. Donk (1999) has also claimed that when features are processed, they include coarse, absolute location information as well as good relative location information. When feature conjunction errors are reported, they are most likely to be confusions between targets and nearby distractors in the display, rather than random miscombinations of free-floating features.

5.4d Wolfe's Guided Search Theory

Wolfe (1994, Wolfe et al., 1989) has proposed a revision of FIT that he calls the guided search model. In guided search, there are two sources of information that aid selection of potential targets. One source is bottom-up information from the stimulus. This information is represented as a two-dimensional salience map of the visual environment. Salience is determined primarily by contrast; features or objects are highly salient if they contrast strongly from their neighbors in brightness, color, orientation, motion, or other primitive features. The second source of information is the top-down activation of objects or features that closely match a target template. Top-down activation is a deliberate, strategic process that can call attention to objects that belong to the target category. Bottom-up and top-down information are combined to produce an overall activation map (see Figure 5–10). The peaks on the map are objects that are most salient and match most closely with what we are looking for. Attention then selects items one at a time from those most active in the overall map in a serial search for the target. If salience and top-down information activate the target most highly, it pops out, and the search is fast and accurate. If, however, many distractors are also highly salient and they appear to be similar to the target, then the search can be a slow, tedious examination of many items before the target is found. As an example, searching for the letter T among a background of Ls does not allow for guidance, since all items contain the same features, and only their configuration differs. Search then is serial through all items until the target is found or the observer gives up (Wolfe, 2020). Thus, different types of search results (including serial or parallel at the limits; see Figure 5–9) can be produced by special cases of a single model in which search difficulty is determined by item salience and item similarity to a desired target (see Duncan & Humphreys, 1989, for a similar model).

Figure 5–10 **Wolfe's (1994) Guided Search Model of Visual Attention**

Similar to Treisman's FIT, the visual stimulus is assumed to be analyzed by separate modules into feature maps for color, orientation, etc. Bottom-up activation of item locations is determined by feature salience (i.e., the degree to which features contrast with their immediate neighbors). In addition, top-down activation is added to the feature maps in proportion to their similarities to the target of the search. Bottom-up salience and top-down target-similarity activations combine to produce an overall activation map. Attention then scans this map by beginning at the highest peak, and continuing to lower activated regions until a desired target is found or the search is abandoned.



Source: Adapted from "Guided search 2.0: A revised model of visual search," by J. Wolfe, 1994, *Psychonomic Bulletin and Review*, 1, p. 205.

5.5 The Effects of Perceptual Load on Attention

5.5a Lavie's Perceptual Load Theory

Lavie's perceptual load theory

The idea that easy perceptual tasks result in irrelevant information being processed deeply, requiring late selection; whereas difficult perceptual tasks result in early selection and filtering of irrelevant information

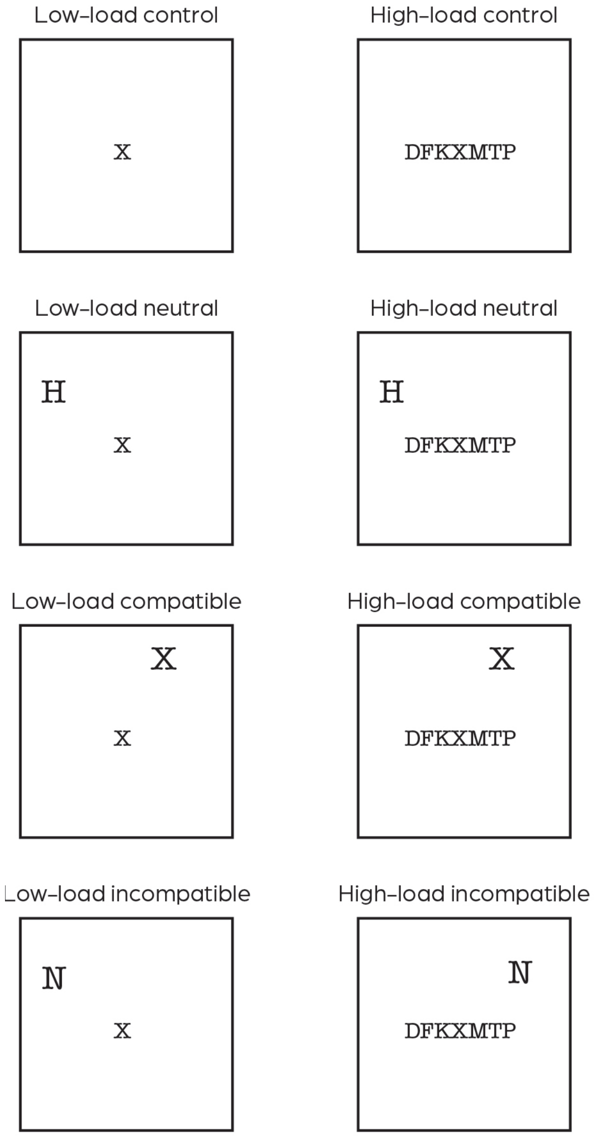
Lavie (1995, 2001) has suggested that the scope of attention depends largely on the perceptual load of any given task. She was interested in addressing the question of whether Broadbent's (1952) early filter or Norman's (1968) late filter were more general descriptors of how selective attention works (see *For Discussion 5–1: The Cocktail-Party Phenomenon*).

Lavie (1995) proposed a hybrid model of attention that combines early and late selection models into a flexible framework. The model acknowledges that human perception has some capacity limits (early selection), but within these limits, perception proceeds automatically to a deep level, at which time late selection mechanisms sort things out (Lavie, 2001). Lavie made use of the Eriksen and Eriksen (1974) flanker paradigm in demonstrating how her theory works. In the flanker paradigm, several letters are presented at once in a visual display with the central one being the target letter. The task is to discriminate between two possibilities for the target, say, whether it is an N or an X, and to hit either of two response buttons as rapidly as possible depending on which target is present. These letters can be surrounded (flanked) by distracting letters that can make the task easy or difficult. If the flankers are neutral letters, like H or Y, the task is more difficult than if the target letter is presented alone. If the flankers are incompatible with the target, that is, the flankers map onto a different response than the central target letter (XNX or NXN), response times and errors increase even more than when the flankers are neutral letters. Finally, if the flanking letters are compatible with the target, that is, they map onto the same response as the central target (XXX or NNN), response times and error rates decrease back to, or even below, the single-letter case.

Lavie (1995) varied the perceptual load of target detection either by presenting the target alone or by including it in a horizontal string with six other letters. It is well known that increasing the number of letters in a search display of similar letters increases the time to find a target (e.g., Atkinson et al., 1969). In either case, one or the other target (N or X) was present in each display. Also included in the display was a much larger letter positioned above or below the horizontal row that contained the target. Participants were told to ignore the large letter, and it was either neutral (H), or it was an N or an X. On compatible trials the large letter matched the target (central letter) shown, and on incompatible trials it did not match the target. Lavie expected that when perceptual load was low (single-target condition), attentional capacity would not be limited, and there would be a strong chance that the large, peripheral letter would be recognized and show compatibility effects on response times. When the perceptual load was high (target embedded among six distractor letters), the search for the target would take more capacity, and less would be left over for processing the peripheral letter (see Figure 5–11).

Figure 5–11 **Manipulation of Perceptual Load in the Flanker Task (Based on Lavie, 1995)**

The task is to hit a left-hand button for one letter (e.g., "N") and to hit a right-hand button for the other letter ("X"). These target letters are always presented in the center of the display. In the high-load condition, six flanker letters are added to the central target. To test the distribution of attention, large peripheral letters (which should be ignored) are shown in some trials that are either neutral, compatible, or incompatible with the central target. Compatibility of the peripheral letters affected response times in the low-load conditions only. Examples of the conditions are shown in Figure 5–11.



The results were exactly as Lavie (1995) predicted. When the peripheral letter was incompatible with the target presented on any trial, the incompatibility effect showed up only on low-load trials. When the perceptual load was high, no compatibility effect was found. Lavie argued that if any task does not use all of the attentional capacity available, other information in the environment might be encoded automatically and unavoidably such that it will compete for later selection. However, if any task is demanding enough to command all of one's attentional resources, irrelevant or extraneous stimuli will not be processed to a deep level—they will be weeded out by early selection. Thus, Lavie holds that both early and late selection theories of attention are true, but a better theory is a hybrid that predicts under what conditions selection will be early or late.

Lavie's theory is relevant for understanding human performance in dual-task situations (see also Johnston & Heinz, 1978). If two tasks are only minimally demanding of attentional capacity, then it is possible to do both at the same time without much loss of performance on either. One might be tempted to say that driving at a moderate speed on a good highway with light traffic would be manageable enough for a skilled driver to engage in a conversation with a passenger (or even with someone via a mobile phone) without much damage to either driving or conversational skills. However, if the conversation turns to a stressful topic with high emotional content, driving performance is likely to suffer, as is conversational quality if the traffic suddenly becomes congested or road conditions become dangerous. In these cases, high loads in one task can reduce performance in the other. Similarly, if you are trying to read in a room with a TV on, you might find that you can share your attention between the two if the reading and TV programs are both light in content. But if the text is highly relevant to an important midterm examination the next day or if the TV suddenly breaks to an important international development, attention is likely to be more clearly focused on one medium or the other, with the unattended information blocked out by an early filter.

5.5b Neurophysiological Support for the Perceptual Load Theory

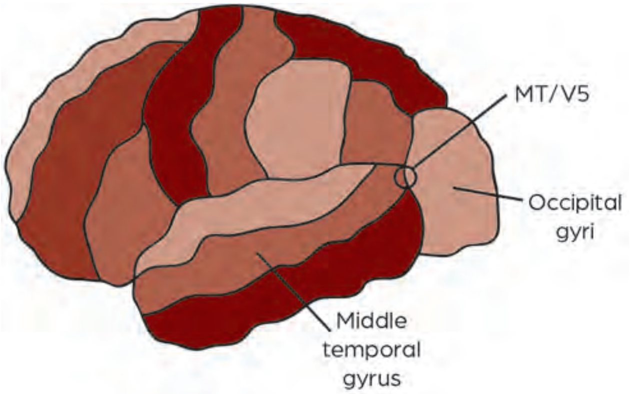
An interesting demonstration supporting Lavie's theory was reported by Rees, Frith, and Lavie (1997). They used familiar words displayed on a background of small dots. On some trials, the dots appeared to move outward from the center of the screen, and on other trials the dots were stationary. While the participants viewed these displays, their brain activity in different regions was measured by fMRI techniques. As expected, different levels of brain activation for stationary versus moving displays were found in known motion processing centers, such as the middle temporal gyrus, **area MT/V5** in the cerebral cortex.

Area MT/V5

The middle temporal gyrus (MT/V5) of the cortex; an area that has many cells that are sensitive to the speed and direction of motion

The main manipulation in Rees et al.'s experiment was the amount of perceptual load demanded by the words presented in the middle of the screen. On some trials, the load was low, as participants were asked to determine simply if the words were printed in upper- or lowercase letters. On other trials, the load was high, as the participants had to hit one response button if the word contained two syllables and another button if it contained either one or three syllables. An important control was that the words shown were exactly the same in the two conditions; the only difference was the task that the subjects had to perform. Therefore, any difference observed between tasks must be due to the different cognitive processes involved and not to the stimulus or response aspects of the task. The results were clear: Motion-sensitive areas of the brain (i.e., area MT) were more responsive in the low-load conditions than in the high-load conditions. It is apparent that the low-load task freed attention to process other areas of the display, and motion detection centers of the brain were highly activated. When the task demanded a high level of attention, however, less capacity was available for processing other information in the display, and motion-sensitive brain areas were less activated. It seems as though attention-demanding tasks can prevent us from noticing irrelevant information through active suppression of normal sensory coding. Although such a mechanism should generally be useful in focusing attention on the task at hand, it can sometimes lead to problems, with the common example being what can happen when one is engrossed in an important phone conversation while driving!

Figure 5–12 **Location of the Area MT/V5**



Source: Adapted from *The Student's Guide to Cognitive Neuroscience*, 3rd ed. (p. 24), by J. Ward, 2015. Psychology Press.

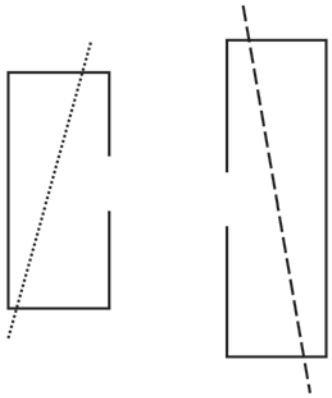
5.6 Location versus Object Theories of Attention

Another theoretical issue of some interest to those who wonder how visual attention works is whether attention is directed to locations in space or to objects that occupy various locations. In other words, does object segregation determine where attention is allocated, or is attention needed to find out where objects are? In Treisman's FIT, attention is allocated to a master map of locations, and object identity is revealed only after attention binds the features at that location into a recognizable whole. In other theories, objects are segregated, and even recognized, preattentively, and attention is directed to objects, not locations.

The debate can obviously become quite complicated because, of course, objects occupy specific locations, and attention to an object necessitates attention to its location. One approach to simplify things has been to ask participants to identify two target features that are equally close to each other, but sometimes fall on the same object and sometimes fall on different objects. For example, Duncan (1984) asked observers to report two features about a superimposed line and rectangle (see Figure 5–13). Two of these features belonged to the line (it was a dotted or a dashed line that was tilted clockwise or counterclockwise from vertical) and two features belonged to the rectangle (it was either tall or short and had a gap on the left or right side). After a brief presentation of a stimulus pair, decisions were consistently more accurate if the two features to be reported were either both on the line or both on the rectangle. Discriminating between one feature on the line and one on the rectangle presumably required attending to two different objects and required more attentional capacity. Even though the lines and rectangles were overlapping such that all relevant features were equidistant, attention seemed to move along one object more easily than between two objects. This result supports the object-based theory of attentional allocation over a space-based theory, since attention directed solely on the basis of location should be indifferent to object relations within its focus.

Figure 5–13 **Stimuli Like Those Used in Duncan's (1984) Experiment**

Subjects had to judge whether the rectangle was tall or short or had a gap on the left or right, or whether the line was dotted or dashed, or oblique to the left or right of vertical. Judgments were more accurate for two features if they were reported from the same figure (rectangle or line) than if one feature was to be reported for each figure.



Duncan's (1984) conclusion was qualified by Lavie and Driver (1996) who showed that spatial cues presented before the onset of a pair of objects could eliminate the object advantage. They used a pair of intersecting lines as objects, and the task was to determine whether two breaks in the lines (which could be either an empty gap or a gap with a dot in the middle) were the same or different. The breaks could occur either both in the same line or one in each line. The same-different judgment was easier when both gaps were in the same line, a result supporting Duncan's findings. However, when spatial cues were given in advance to direct participants' attention to the locations of the gaps, the same-object advantage disappeared. It is clear that attention can be directed to locations in space, and if enough time is given to segregate objects from the background, attention will spread from its initial location to encompass any object near its focus. Thus, space-based descriptions of attention are likely to be true of early stages of attentional allocation, before objects are segregated from their background. Object-based descriptions of attention are more descriptive of processes occurring later in vision, after objects have been segregated and integrated from their component features.

Braun et al. (2001) reached a similar conclusion. They used a task in which a central cluster of letters at various rotations had to be classified as being either all the same letter or a mixture of two different letters. A secondary task was to identify a single peripheral target item presented briefly at the same time as the central letters.

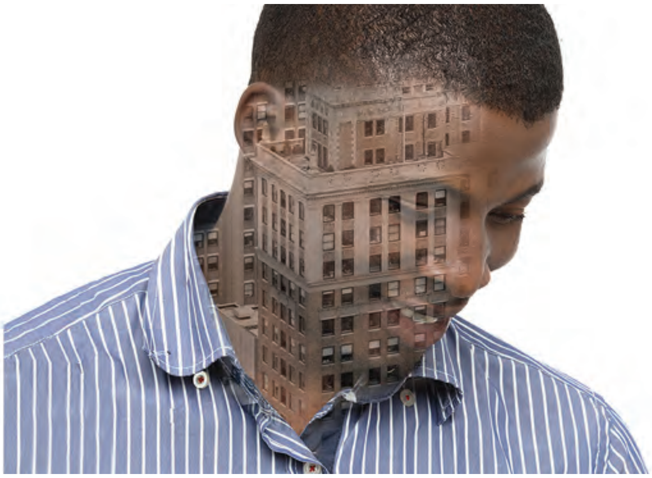
They found that the central task was very attention demanding, such that the peripheral target was essentially processed without attention. In this case, peripheral targets that were defined by single salient features (such as shape or color only) were easily identified, but peripheral targets that were defined by spatial combinations of features (like a red square but not a red circle or a blue square) were identified at chance levels. Braun et al. (2001) supported Treisman's Feature Integration Theory, in the sense that features cannot be integrated at some point in space if attention is directed elsewhere. In this case, object perception depends on attention, and attention is space-based until it can be focused later on objects within this space.

In an fMRI study by O'Craven, Downing, & Kanwisher (1999), participants were shown pictures in which a face and building were superimposed (like that shown in Figure 5–14). They were asked to pay attention to either the face or the building. Face recognition is associated with the Fusiform Face Area (FFA), whereas perception of a building is associated with the Parahippocampal Place Area (PPA). The FFA showed activation when participants paid attention to the face, whereas the PPA showed activation when they paid attention to the building. So, their results showed the neurological foundations of different types of object attention.

ERP studies have also supported the conclusion that attention to specific features, such as color or motion, as well as to specific locations is allocated earlier than attention to specific types of objects. Task goals seem to be mediated by executive processes initiated in the lateral prefrontal cortex, and evoked potentials for brain areas responsible for analyzing location (extrastriate areas of occipital cortex), color (area V4 in the occipital cortex) and motion (MT, or middle temporal lobe) are initiated sooner for these types of featural cues than are object-based cues, such as faces vs. buildings (Banich & Compton, 2018).

Figure 5–14 **Example of an Overlapped Image of a Face and Buildings**

In an fMRI study by O'Craven, Downing, & Kanwisher (1999), participants were shown images similar to Figure 5–14 and asked to pay attention to either the face or the building. The Fusiform Face Area (FFA) was activated when participants paid attention to the face; the Parahippocampal Place Area (PPA) was activated when they paid attention to the building.



Source: Adapted Shutterstock

5.7 Distribution of Attention Over Time

5.7a Attending to Important Moments

Initial studies of attention were directed towards discovering how attention is distributed over space or directed to certain input channels, features, or objects. It is clear, however, that our attention is distributed over time as well as space. Most people have experienced a frightening event, such as a near or actual auto accident, where there might have been a point in which time appeared to stand still or at least slow down as one tried to take in the full importance of what is happening at a single moment (e.g., Tse et al., 2004). Similarly, in certain sporting events, many minutes can pass with little action, but suddenly a crucial play occurs that can determine the outcome of the entire game (e.g., a penalty shot in ice hockey or soccer, or a close play at home plate in baseball). Most people share an introspective awareness that such an experience can create a vivid image of the scene, with events preceding and following the special event literally being crowded out of memory.

The introspective feeling that some particular experiences jump out, seem to occur more slowly, and are retained better than other events is supported by laboratory studies of how attention changes over time. It is well known from studies of worker and student performance that attention cannot remain focused at a high level on any task indefinitely. Lecturers are encouraged by their students to offer breaks in classes that run for an hour or more. Similarly, people have difficulty maintaining attention during demanding tasks, like managing an air-traffic control station or engaging in simultaneous translations of one language into another as is done at large international meetings and at the United Nations. Periodic swings in attention over time have been studied in the experimental literature on human **vigilance**. In detecting the presence of rare events, such as radar or sonar signals against a background of noise, performance often shows a significant drop in half an hour or less after beginning the task (Broadbent & Gregory, 1963; Mackworth, 1948).

Vigilance

The attempt to maintain attention at a high level of focus on a single task for extended periods of time

5.7b Attending to a Rapid Sequence of Events

Perhaps of even more interest than long periods of attentional variation are the moment-to-moment changes in memory for events that are attended to and remembered, whereas others apparently escape attention and are lost from memory. Lawrence (1971) studied how well people could identify words when they were presented one at a time at a high rate to a single place in the middle of a visual display. This research procedure has come to be called **Rapid Serial Visual Presentation (RSVP)**. In RSVP presentations, the time between successive items is measured in milliseconds (ms). The stimulus onset asynchrony (SOA) is the time between the onsets of successive items, and the interstimulus interval (ISI) is the blank time (if any) between two successive items. Thus, SOA = (any item's exposure duration + ISI).

RSVP (rapid serial visual presentation)

The presentation of a succession of different visual stimuli to a single, central point in the field of view. Rates of presentation can vary, but are usually set at about 100 ms per item (10 items/sec) to prevent masking found at faster rates and naming and rehearsal likely at higher rates. Items typically are seen clearly enough to be identified, but are gone too quickly to be consolidated into memory.

The SOA is the crucial variable in much RSVP research. At very short SOAs, on the order of 50 ms or less, successive items tend to run together and become difficult to differentiate, much less to identify and remember. This phenomenon is known as integration masking, and it is due to the fact that the visual system tends to integrate information over time in intervals approaching 100 ms. If the SOA is set to about 100 ms or so, each item appears distinctly visible and recognizable, but each of them occurs too rapidly to be remembered for very long. Still, Lawrence (1971), Potter (1976; Potter & Levy, 1969), and others have shown that even though very few of the items in an RSVP sequence could be reported from memory after the sequence was completed, perception and recognition of each of the items was fairly complete. This conclusion was reached from experiments in which a specific target (e.g., a picture of a boat) or a target category (e.g., an animal name) was given before an RSVP sequence of pictures or words. Even at rates of 10 items/sec (100 ms/item), target detection is often very good, with report accuracy of 90% or more for young adult participants. Lawrence even noted that search for targets in an RSVP stream could be 50% more efficient than a similar search through a list of words printed down the center of a piece of paper. This result prompted some people to wonder if tasks such as reading could also be made more efficient if the text were presented in the RSVP mode rather than being read with normal eye movements (see *For Discussion 5–4: Reading Text Presented in the RSVP Format*).

For Discussion 5–4

Reading Text Presented in the RSVP Format

Text is normally read with a sequence of fixations, lasting about 150 to 350 ms, interspersed with saccadic eye movements, which are considerably shorter, about 20 to 50 ms. Most of these saccades are progressive, leading to fixations that land in the next word or two after the one previously fixated, but some are regressive (to earlier parts of the text), some are to other parts of the word currently fixated, and others are return sweeps to the next line of text. Thus, the visual input to a reader is a sequence of snapshots of text segments occurring at a rate of about 4 to 5 per second. Research by Rayner and others (e.g., McConkie & Rayner, 1975; Rayner, 1998) has shown that we are sensitive to only a small part of the text at any time. McConkie and Rayner used a computer display that changed with each eye fixation to reveal text in only a narrow window around the point of fixation. The rest of the text was represented by rows of Xs. This window can be arbitrarily small, with reading speeds and comprehension rates falling precipitously for very small windows of only a few characters. They found that reading is fairly normal in speed and comprehension if the window is expanded to about 15 to 20 character spaces around the fixation point, with most of the window extending to the right. Further, it appears that only the leftmost word or two is processed to the level of meaning, whereas the rightmost part of the window is primarily used for determining the landing point of the next eye movement.

These facts about reading, as well as the high levels of performance that can be achieved in searching through RSVP lists of words, suggest that reading sentences and paragraphs in the RSVP mode might promote efficient reading for comprehension. That is, entire passages of text could be presented one or a very few words at a time to a central location on

(continues)

(continued)

a computer screen, and the text could be read without the need for eye movements. Indeed, work by Forster (1970) and Gilbert (1959) showed that text could be read when presented this way by adults and children, respectively.

Most people, when first hearing about reading text presented in the RSVP mode, object that it would be both unnatural and difficult. On the other hand, it could be argued that reading with eye movements is hardly a natural act, since the ability to read has been limited to only relatively few members of our species in all our history. Besides, eye movements could be inefficient due to the lack of perfect oculomotor control, the loss of information during saccades (e.g., saccadic suppression, see Volkmann, 1976), and the fact that some attention must be diverted from the main process of text comprehension in order to control eye movements themselves.

In the lab at the University of Kansas, a number of students and one of the authors (Juola) spent about 10 years trying to discover the optimal means of enhancing readability of text presented in the RSVP mode. The group varied the presentation rate and window size of text selected from standardized tests of reading comprehension and found the optimal presentation format. The optimum method is to present one long word or two or more short words for about 250 ms each in single windows left-justified, so that people tend to fixate a point a few character positions from the left-hand border. Further, the text should be parsed as much as possible so that short syntactic phrases are preserved rather than being split between two successive windows (i.e., "The girl threw the red ball into the street" would be parsed: The girl / threw / the red ball / into the street). Any phrases more than 15 characters in length were arbitrarily split near the middle in order to prevent showing windows too long to be read in a single eye fixation.

The combination of an average of two words every 250 ms equals eight words per second, or 480 words per minute (WPM). This is almost twice as fast as a normal reader usually processes text, so the presentation rate can be slowed as needed to match readers' abilities. Several studies have shown that text presented in this optimal RSVP format can be read just as easily and with equivalent comprehension as text presented for the same total time in a normal page format to be read with typical eye movements (Juola et al., 1982). Moreover, we found the RSVP method to result in superior comprehension performance for readers of lower than average verbal ability among college students (Chen, 1986), and for middle-school students who were classified as disabled readers (Juola, 1988). The method has also been shown to be an optimal way of presenting text when the viewing area is extremely limited, as in some small electronic devices (Juola et al., 1995). Reading RSVP text should remain a viable option for electronic displays in various contexts, for instructional purposes for beginning and remedial readers, and for the use as a tool in the study of cognitive processes involved in reading.

Potter later argued that successive items presented at a rate of 10 per second result in complete perception and recognition of each of them, as long as they are familiar items such as letters, digits, words, or pictures of common objects. However, they are quickly replaced in memory by successive items before they can be consolidated into a retrievable form. In other words, RSVP lists of items can be processed well enough to pick out one or two salient target items, but only because these targets receive additional processing do they survive the destructive interference from succeeding items in the stream.

5.7c Chun and Potter's Two-Stage Theory

Chun and Potter (1995, 2001) expanded on these ideas to propose a two-stage model for how items are processed in RSVP streams. In a first stage of processing, each item is processed automatically, as deeply as possible, until it is replaced by the succeeding item. For familiar items, this first stage includes perception and recognition of the item so that its meaning in long-term memory is accessed and understood. However, before any more processing beyond simple recognition can be accomplished, the next item is presented and recognized, and so on to the end of the stream. If a certain item is salient in some way—such as by being presented in a different location, size, or color from the other items—it can be entered into a second stage of memory that is protected from being overwritten by new inputs. In this second stage, more attention is directed to elaborate coding of the item, such that it is maintained in memory longer and is more likely to be recalled at the end of the RSVP sequence than items that had received only stage 1 processing.

Chun and Potter's (1995) model was initially developed to explain a finding that demonstrates our ability to select items from a temporal sequence of events. The selected items are processed more deeply, but other items must necessarily receive less processing, resulting in their loss from memory. This loss due to the attentional selectivity of events distributed over time has come to be called the **attentional blink (AB)**, and it is the topic of the next section.

Attentional blink (AB)

The finding that if two targets are present in an RSVP stream, the second target is frequently missed if it follows too closely on the heels of the first target

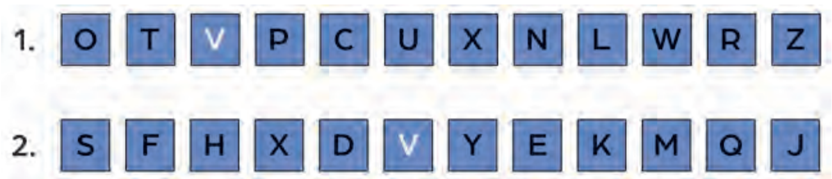
5.8 The Attentional Blink

It is known from the work of Posner (1980; *For Discussion 5–1: The Cocktail-Party Phenomenon*) and others how attention can be distributed over space. Recently, some of the principles that guide our distribution of attention over time have been determined as well. Just as spatial attention selects some region or items in the visual scene for further processing to the detriment of others, certain events occurring in time can be selected with the consequence that other events might be lost from memory. Broadbent and Broadbent (1987) and Raymond et al. (1992) had observers monitor an RSVP stream of words or letters, looking for predefined targets. When two successive targets were separated in the stream by about 180–540 ms (i.e., at a lag of about 2–6 items at 90 ms per item), the second one was often missed, although the first was identified with more than 90% accuracy. Curiously, if the two targets occurred in immediate succession, the two were usually both reported, a phenomenon known as Lag 1 sparing (see Figures 5–15 and 5–16).

Visser et al. (1999) proposed a model for the attentional distribution over time in which specific episodes are defined by critical events that match what we are looking for or are otherwise important to us. Objects or events included in these episodes are selected for further processing, and while they are being handled, new objects or events can be missed, or only processed to a relatively shallow level. However, the selection mechanism can be relatively sluggish, so that if two targets occur in close succession, they are likely both to be included in the same attentional episode, and thus report probability is high for both targets. If there are one or more intervening items between the two targets, however, the episode is unlikely to include both targets, and the second will not be selected for deeper processing. It will then be less likely to be available for later report. After the first target has been recognized and consolidated into working memory, attention is freed to examine the incoming stream for new targets. Theories such as these (Chun and Potter, 1995; Visser et al., 1999) are consistent with the U-shaped decrement in reporting the second target after the first is detected (Figures 5–15 and 5–16).

Figure 5–15 **The Attentional Blink Task**

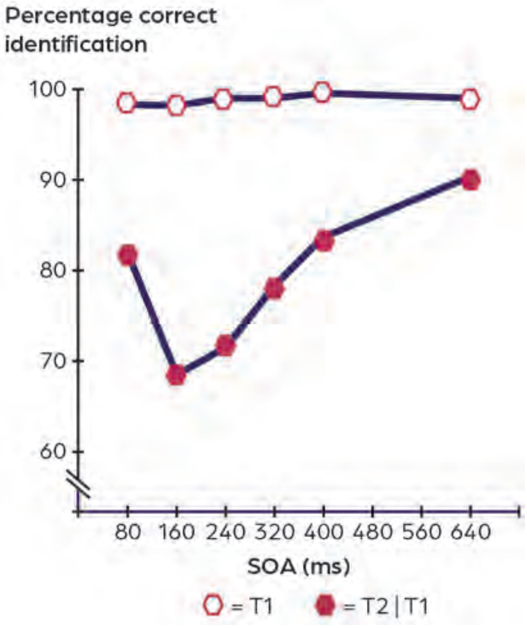
The task originated by Raymond et al. (1992) was detection of one or two targets in a rapid, serial visual presentation (RSVP) stream of characters. The first target (T1) was always present, and it was the only white letter among a stream of black letters. The second target (T2) was the black letter X, and it occurred on only 50% of the trials. When it was present it appeared 1–8 positions after the first target.



Source: Adapted from "Priming the semantic neighbourhood during the attentional blink," by I. M. Harris, and M. J. J. Little, September 14, 2010, *PLOS ONE*. http://dx.doi.org/10.1371/journal.pone.0012645

Figure 5–16 **The Attentional Blink Graph**

The graph shows the percentage of trials on which targets were detected when present. The percentages for T1 (always present) and T2 (present on half the trials) are plotted as a function of the time between the onsets of T1 and T2 (SOA). T1 is almost always detected, but T2|T1 (the percentage of times that T2 was correctly detected given that T1 was correctly identified on the same trial) is much lower. The dip in T2 detection when it follows T1 by about 160 to 320 ms is called the attentional blink.



Source: Data from "Temporal constraints on conscious vision: On the ubiquitous nature of the attentional blink," by M. Nieuwenstein, J. Theeuwes, B. Wyble, and M. Potter, 2009 *Journal of Vision*, 18, pp. 1–14.

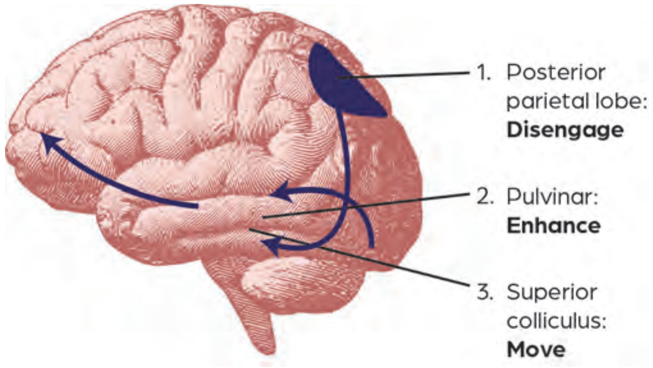
5.9 Brain Mechanisms of Visual Attention

5.9a The Visual Pathway Revisited

Over 50% of the primate brain is involved in visual sensation, perception, recognition, and eye movement control (see Figure 5–17). Projections from the retinas of both eyes make up the optic nerve that travels mainly to the thalamus except for a small proportion of the fibers that break off and head to the superior colliculus in the brain stem. From the lateral geniculate nuclei of the thalamus, there are projections along the optic tract to the far, back regions of the occipital lobes; an area called V1, the primary projection area for vision. The thalamus also receives input fibers from many parts of the cortex as well as from the cerebellum. The superior colliculus connects directly with the eye movement control centers in the brain stem and seems to be involved with rapid and automatic movements of attention and the eyes. All these visual areas, and many others, show retinotopic mapping (i.e., there is a one-to-one correspondence between areas on the retina and areas on the thalamus, superior colliculus, and area V1 in the cerebral cortex). Although these areas show distortions, mainly to reflect greater analysis of central vision near the foveal areas of the retinas, they all have a more-or-less veridical map of the visual world laid over them.

Figure 5–17 **Brain Areas Associated with Visual Processing and Visual Attention (Based on Posner & Raichle, 1994)**

The three main centers of visual attention include the posterior parietal cortex, the pulvinar nucleus of the thalamus, and the superior colliculus in the brain stem. These have been respectively identified with the disengagement of attention from the current focus of attention, the selective enhancement of attended information, and the movement of attention to a new focus of attention.



Source: Adapted Shutterstock

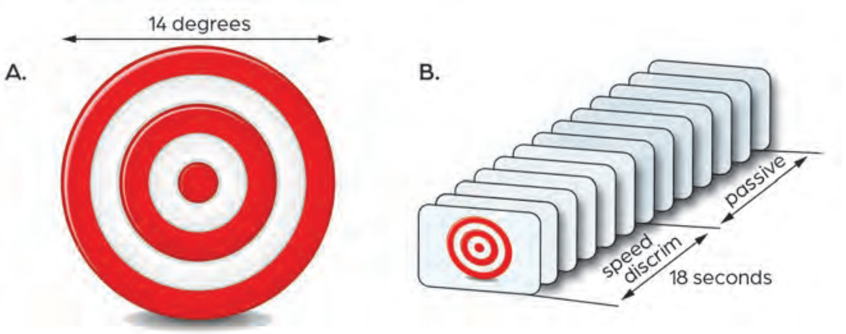
From V1, visual pathways lead to association areas in the occipital lobe, including areas labeled V2, V3, and V4. From there, visual information follows the dorsal pathway to the parietal lobe of the brain for functions mainly related to spatial analysis, spatial orienting, and motor programming. Other pathways follow a ventral pathway to the temporal lobe for processes related to object recognition and naming (see Chapter 2; Mishkin et al., 1983). Further connections travel in both directions to and from the frontal lobes, including areas involved in executive control of visual search and the frontal eye fields that are involved in voluntary attention and direction of eye movements to specific objects and locations in space. Research has shown that activities in all of these brain areas can be modulated by attention (Braun et al., 2001). Further, different areas of the brain that are involved in attending to, and abstracting additional information from, specific visual objects and events apparently enhance their cooperation by synchronizing rhythmic patterns of oscillations in their neural firing (e.g., Gregoriou et al., 2009).

5.9b Neurophysiological Evidence for Attentional Effects in Vision

Although there is some controversy about whether early visual processing centers in the brain, like the thalamus and area V1 in the occipital lobe, can be influenced by attention, recent research has supported the view that attentional influences exist. For example, changes in blood flow measured using the fMRI technique can be observed when subjects alternatively view a display passively or view it in order to make some visual discrimination. Heeger et al. (2001) used this technique while presenting a series of ring-like stripes that cycled between inward and outward movement, continuously, around a fixation point. In the passive condition, participants simply observed the display, and in an experimental condition, observers had to discriminate between subtle changes in the rings' speed of motion. Measurements of blood flow using fMRI showed that during the discrimination phase, neural activity was higher in both areas V1 and MT (the middle temporal gyrus that has many cells responsive to motion) than when the same displays were viewed passively. However, it is likely that, rather than being modulated directly by attention, activity in the primary visual cortex (V1) is prepotentiated by feedback from higher-level, extrastriate areas (Hopf et al., 2009).

Figure 5–18 **Experiment Using Concentric Rings**

Heeger et al.'s (2001) experiment used concentric rings that moved either inward or outward from a central fixation point. When subjects had to look for changes in the speed of the rings' motion, brain areas involved in motion perception showed higher levels of activity than when the same displays were to be viewed passively.



Source: Adapted from "*Neuronal correlates of attention in human visual cortex*," by D. J. Heeger, S. P. Gandhi, A. C. Huk, and G. M. Boynton, 2001, In J. Braun, C. Koch, & J. Davis (Eds.). *Visual Attention and Cortical Circuits*. Cambridge, MA: MIT Press. Art from Shutterstock.

Studies using single-cell response measurements in monkeys have also indicated that individual neurons respond differently to the same stimulus when it is attended and when it is ignored. McAdams and Maunsell (1999) recorded the activities of neurons in area V4 of the occipital lobe in macaque monkeys when they viewed a display containing color patches on one side and striped patterns on the other. The monkeys were trained to look straight ahead but attend to either the left side or the right side of the display on individual trials. They then alternated between performing either a color discrimination task or an orientation discrimination task. In the color task, the monkeys were to push a lever in one direction or the other to indicate whether or not two successive color patches were the same hue. In the orientation task, they indicated whether or not two successive bands of stripes were oriented in the same direction. The important point is that on all trials, color patches and stripes were shown in the two successive displays, so that comparisons could be made of responses to the same stimuli when the monkeys were attending to the color patches or to the stripes.

McAdams and Maunsell (1999) recorded from individual cells in area V4 that were tuned to various orientations. That is, if a cell that shows a maximum response to a stripe aligned 70 degrees clockwise from vertical, it will generate diminished response to stripes aligned at 50 or 90 degrees. The important result is that all such cells showed an average increase in firing rate of 22% when the stripe discrimination task was performed as when the same stripes were presented to the same cells, but the monkeys were attending to the simultaneous color discrimination task (see also Maunsell & McAdams, 2001).

Similar studies of individual neurons' response rates in the frontal eye fields (FEFs) of monkeys have shown that these cells are involved in attentional selection and motor control. Thompson et al. (2001) trained monkeys to discriminate between a colored target and an array of distractors in another color. The task for the monkeys was to fixate the uniquely colored targets on some trials (go trials) and to withhold the fixations on other trials (no-go trials). When the target item appeared in a FEF cell's receptive field, it showed a more vigorous response than when a distractor was present. The level of responding was less vigorous for both kinds of items on no-go trials than on go trials. Thompson concluded that cells in the frontal eye fields show greater responses for items that are to be the targets of subsequent fixations than for similar items for which a target response has not been trained. Further, an additional increase in responding was recorded when this selection was followed by an actual eye movement than when the movement was suppressed. These results support the notion that cells in the frontal eye fields of primates, including humans, are involved both in covert orienting (selecting possible targets for subsequent inspection) and overt orienting (directing a fixation toward the selected target).

In humans, brain-imaging studies have revealed that many parts of the brain, including subcortical structures, are involved in visual attention. Endogenous, or top-down attention, seems to be controlled by executive functions of the left frontal lobe, whereas exogenous, or bottom-up attentional capture, is a function of the superior colliculus in the midbrain and projections to the parietal lobe. These roles are reflected in the hypotheses that the prefrontal cortex executes control functions for focusing on task-relevant information, whereas the posterior parietal cortex has a visual salience map; a distinction reminiscent of the top-down and bottom-up components of Wolfe's (1994) guided search model. Both frontal and parietal lobe structures seem to be involved in alerting and orienting to new stimuli, in response to inputs from the superior colliculus and the pulvinar nucleus of the thalamus. These latter structures are also involved in automatic control of eye movements to novel stimuli, thus suggesting that human attentional systems might have evolved from more primitive mechanisms designed for eye movement control (Baars, 2007).

Braun et al. (2001) and Banich and Compton (2018) provide cogent summaries of many studies supporting the notion that attention can modulate the activity of individual neurons and concentrated populations of neurons in all known visual centers of the brain. Traditionally, early brain-imaging studies identified different aspects of attention with widely divergent cortical processes. That is, alerting is presumably signaled by activity in the frontal and parietal lobes, primarily in the right hemisphere; orienting is associated with activity in the superior parietal lobes, the temporal-parietal junction, and the frontal eye fields; and executive control is maintained by the anterior cingulate and lateral prefrontal cortex. Certain automatic responses that influence attention, like orienting to a highly salient stimulus, are mediated subcortically (e.g., by the superior colliculi) (Carrasco, 2011; Knudsen, 2007; Posner & Petersen, 1990). Some of these changes are due to hardwired systems that have evolved to alert us to important changes in the environment, others are due to learned associations between simple stimulus patterns and arbitrary responses. It is clear that attention is a highly evolved mechanism for enhancing neural responses to important stimuli while inhibiting responses to stimuli that are irrelevant in a given context (Desimone & Duncan, 1995; Tsotsos et al., 2001).

Attention also seems to affect different cortical areas when it is directed to spatial locations and when it is directed to specific features or objects in the field of view, regardless of their location. Spatial attention has its effects earlier both in time and in levels of the visual system than feature-based attention, particularly when signaled by an external, peripheral cue. Exogenous cues can result in enhanced signal processing in the cued area, and suppressed responses to noise in the areas surrounding the target that peak at about 100 to 120 ms after the cue. However, these responses are transient, and the attentional response can dissipate and even be suppressed in the cued area, a phenomenon known as inhibition of return (Posner & Cohen, 1984). Desimone and Duncan (1995) formulated the biased-competition hypothesis to describe attentional effects at many levels of the visual system. Within any level, neurons associated with the cued location are activated, whereas neurons with receptive fields in adjacent areas are inhibited. Therefore, one of the primary roles of spatial attention is to increase the signal-to-noise ratio by enhancing signal processing while also excluding external noise in the target region—in effect, setting up spatial filters. Attention has also been shown to increase the amplitude of neural impulses in attended regions, as well as to increase the synchronization of neural responses in successive layers of the visual system tuned to the target location (Fries et al., 2001). The result is that attended regions appear as if they are brighter and more salient than unattended areas.

Feature-based attention, on the other hand, is more likely to be based on endogenous cues related to basic visual primitives, such as color, orientation, and direction of motion. Attending to different features can modulate activity in cortical areas specialized for processing those dimensions (Carrasco, 2011).

That is, attending to motion results in enhanced activity in the medial-temporal region (MT), color results in modulation in extrastriate area V4, and orientation affects responses in V1 and V2. Bichot et al. (2005) showed that individual neurons in the monkey cortex responded more strongly to an attended feature presented to their receptive fields than when the same feature was shown but not attended. Similarly, aftereffects resulting from prolonged exposure to specific colors, orientations, and motion directions are stronger when the relevant features are attended than when they are viewed but not attended. Feature-based attention ". . . has the remarkable property that its effects are not constrained to the locations of the stimuli that are voluntarily attended; they spread across space" (Carrasco, 2011, p. 1510). Neuronal modulations have even been observed in cells responding to spatial locations where no stimuli are present (McManis et al., 2007; Serences & Yantis, 2007). Even when observers are directed to attend to one or the other visual hemifield, brain activity to items presented to the unattended field show larger responses if they match the feature attended on the other side. Further, the motion aftereffect (seeing illusory motion in a direction opposite to that actually present in an adapting field) can be found in nonadapted areas of the visual field (Arman et al., 2006), indicating widespread selection for relevant visual features.

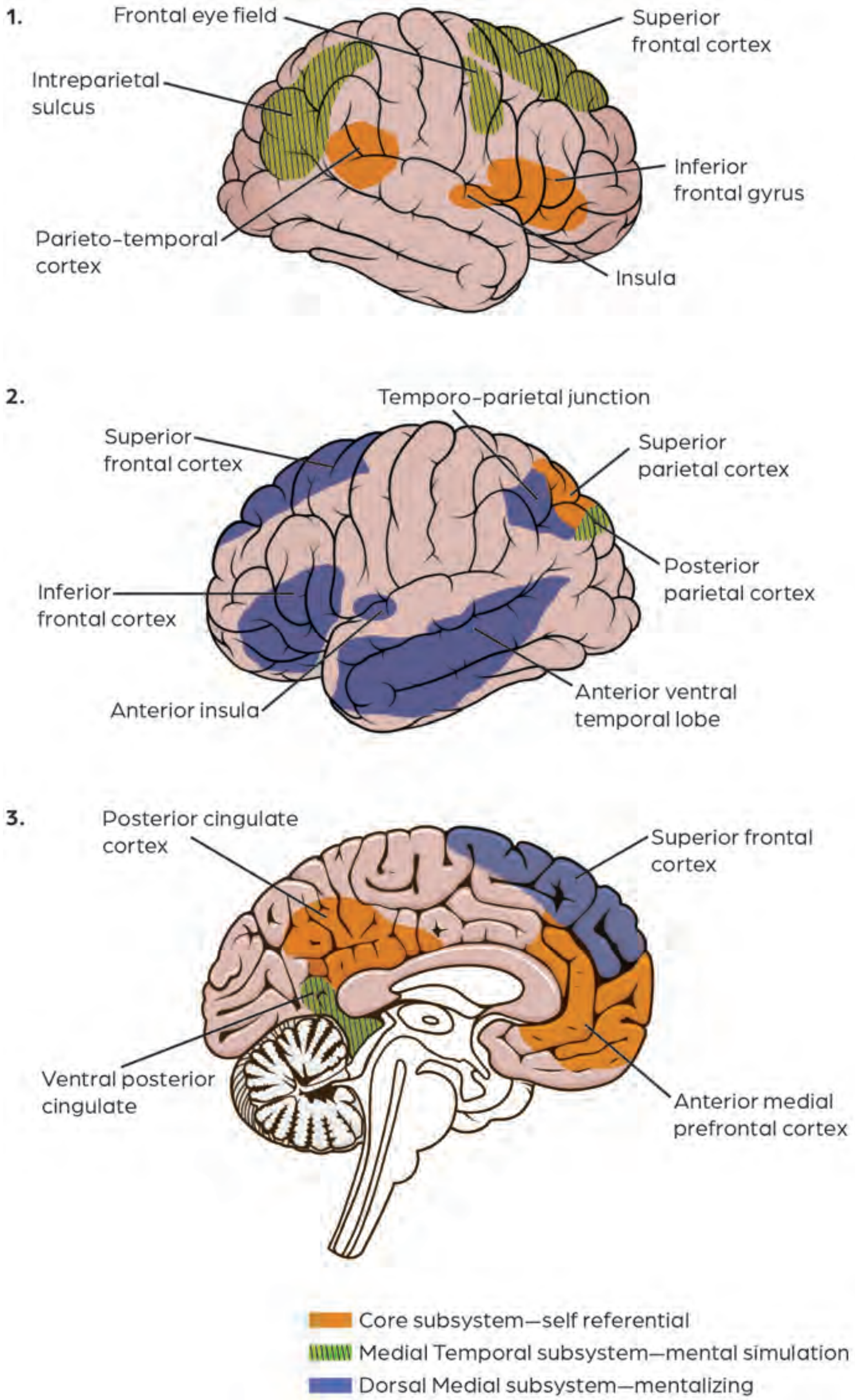
These results affirm the main assertion held by modern theories of attention—it is not a unitary construct that can be identified with a simple underlying brain mechanism. Further, attentional effects can be found both far and wide in the visual field as well as in its various cortical representations.

5.9c The Default Network: An Attentional System Not Based on Sensory Inputs

Most of the human brain is devoted to analyzing sensory inputs, particularly vision, and determining responses to them. Almost all of these areas have also been shown to be influenced by attention. However, there are large areas in the frontal, temporal, and parietal lobes that are relatively inactive in responding to sensory inputs, but rather seem to be most involved in dealing with inward thoughts and memories. These areas have been called the "default network" because they are more active when we are at rest or not actively engaged in perceiving the outside world (Raichle, 2015). Researchers have identified three main components of the default network. These include a self-referential network, a network for memory and imagination, and a third network supposedly involved in reflection on one's own mental states and those of others (Banich & Compton, 2018). The self-referential network, or core subsystem, is composed of the anterior medial (front and center) prefrontal cortex and the posterior cingulate cortex, which are tucked in below the parietal cortex and between the occipital cortex and the midbrain. This network presumably is involved in giving personal meaning and value to present and prior events. The memory and imagination network combines functions of the ventral posterior cingulate, just below the posterior cingulate, and the posterior parietal cortex. The third network is widely distributed over the temporal, frontal, and parietal lobes and is presumably metacognitive in nature. That is, it is supposed to be involved in inference-making and self-reflection as well as developing a "theory of mind" about relations between our own cognitions and those of others (see Andrews-Hanna et al., 2014).

Figure 5–19 **The Default Network**

Three main components of the default network include: 1) the self-referential network (in orange), which is involved in giving personal meaning and value to present and prior events; 2) the memory and imagination network (in green), which combines functions of the ventral posterior cingulate and the posterior parietal cortex; and 3) a network (in purple) that involves inference-making, self-reflection, and developing "theory of mind" about relations between our own cognitions and those of others.



Source: Adapted from *Cognitive Neuroscience*, by M. T. Banich and R. J. Compton, 2018, Cambridge, UK: Cambridge University Press.

5.10 Neurological Disorders of Attention

5.10a Visual Neglect in Patients with Parietal Damage

Since the dorsal pathway from the occipital to the parietal lobes of the brain seems to be specialized for handling spatial information, it should not be surprising that parietal lobe damage results in severe spatial processing deficits. Among the most severe is a syndrome called **visual neglect**, in which the patient seems to be totally unaware of any stimuli presented to the field of view opposite the lesion (i.e., the contralesional field). That is, if the right parietal lobe is damaged, a patient with neglect will appear to be blind to any visual stimulus presented to the left of the centerline. There appears to be an important hemispheric asymmetry to the expression of neglect. The deficit is more common and usually more severe following right parietal damage than after damage to the left parietal lobe (De Renzi, 1982).

Visual neglect

A result most typically of damage to the right parietal lobe that causes an inability to attend to the contralateral (left) visual field as well as to the left sides of imagined objects or scenes

A similar deficit can be produced if there is unilateral damage to the visual sensory pathway. For example, if a lesion exists in the left or right pathway from the lateral geniculate nucleus to the primary visual cortex, or if one half of the occipital cortex is damaged, the visual field on the side opposite the lesion (the contralateral side) will not be seen—a condition known as **hemianopia**. However, there are several critical distinctions between neglect and hemianopic patients. First, neglect is seldom complete or permanent, and the area of neglect usually does not perfectly follow the midline. Second, neglect patients are seldom aware of their deficit, whereas hemianopic patients complain bitterly that they cannot see well on one side of space (Husain, 2001).

Hemianopia

The loss of vision on one side of the visual field due to damage in the visual centers of the brain in one hemisphere

Since sensory processes up to and including primary visual cortex are intact in most neglect patients, the question has been raised as to whether the deficit is attentional rather than perceptual. Support for the attentional deficit notion has come from several sources. Driver (1998), in a review of data from neglect patients, noted that some patients show deficits in reporting the left halves of objects even when the entire object is presented in the nonneglected field of view. Such results clearly show an attentional bias to the right rather than a localized spatial deficit.

Similar conclusions were drawn from a classic study by Bisiach and Luzatti (1978). They demonstrated neglect in patients who were asked to remember a stored image rather than an actual visual scene. They asked two patients with right-hemisphere damage to describe the buildings in the main square in Milan, Italy, a city very familiar to both of them. First, they were asked to imagine themselves standing in front of the cathedral. Consistent with their apparent attentional deficits, they described buildings that would have been visible on the right side of the square and omitted ones on the left. When they were asked to cross over mentally to the other side of the square, turn around, and imagine what they would see, they now described the buildings they had previously omitted, by again concentrating on ones that would have appeared on the right side only. The similar results obtained in neglect patients for real and imagined scenes eliminate a perceptual explanation for the phenomenon. Instead, parietal damage in the right hemisphere clearly reduces attention to the left side of real or imagined images.

5.10b Visual Extinction

Further evidence for an attentional deficit in right-parietal patients comes from individuals showing a similar but less severe deficit called extinction. Indeed, some patients who initially show symptoms of neglect soon after a brain injury can recover to a condition in which only extinction is shown. Like neglect patients, people with visual extinction primarily have damage to the right parietal lobe of the brain, but unlike such patients, they can and do report seeing objects presented to the contralesional (left) visual field. The deficit shows up, however, when two different objects are presented at the same time, one to the left side and one to the right. The classic finding in the extinction literature is that the left-side object now becomes invisible, or at least unreported, whereas the ipsilesional (right) side object is reported (Posner et al., 1984; 1987). It is as though attention is drawn to the right side to such an extent that when there is a competing stimulus, only objects on the ipsilesional side are noticed and reported. When the right-side field is empty, however, attention is free to engage contralesional, left-side objects, and they are then reported accurately.

5.10c Bálint's Syndrome

A related attentional deficit resulting from brain damage also shows a tendency for patients to ignore parts of a display, but in this case, the neglected part or parts cannot be predicted in advance. Rather, all parts of a complex display might be ignored but one, yet this part is reported in great detail. The disorder has come to be called simultagnosia, which is part of **Bálint's syndrome**, after the initial report of symptoms of individuals who had suffered bilateral parietal damage (Bálint, 1909). Patients with simultagnosia usually report seeing only one object at a time, even when many objects are presented simultaneously. Farah (1990) later identified two types of simultagnosia associated with bilateral damage either to the dorsal (parietal) or ventral (temporal-occipital) visual pathways. Bilateral parietal damage can produce Bálint's syndrome, which, rather than being linked to a particular spatial region, as is neglect and extinction, is characterized by the inability to perceive more than one object at a time (Husain, 2001). Damage to the ventral pathway can result in agnosias that are often quite specific, such as prosopagnosia, which is the inability to recognize faces of familiar people (see Chapter 3, Section 3.6c).

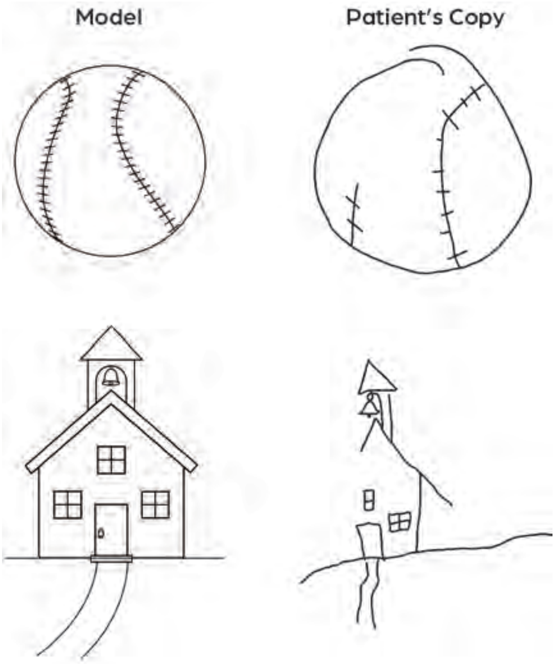
Bálint's syndrome

Bálint's syndrome consists of three major symptoms: (1) simultagnosia; (2) oculomotor apraxia, which is a defect of controlled, voluntary, and purposeful eye movement; and (3) optic ataxia, which is an impairment in visually guided hand movement

Further support for the nonspatial nature of simultagnosia was reported by Humphreys et al. (1994). They studied two patients with Bálint's syndrome, one having suffered from bilateral strokes in the parietal region and the other having suffered from anoxia. When single words or pictures of objects were presented to the central field of view, they could name them easily. However, when the words and pictures were superimposed, the patients either reported the picture alone or both the word and the picture. Words alone were almost never reported, indicating that the word information was sometimes extinguished by a picture, even though they occupied the same spatial location. Husain (2001) used these results to make an important generalization about neglect, extinction, and simultagnosia. He argued that the spatial bias in patients showing neglect or extinction is due to the asymmetric nature of their disability. The field of view contralateral to the lesion suffers from delayed processing, and thus, its representation is inhibited by attention and perceptual processes that reach completion for the ipsilateral side first. In contrast, with bilateral damage, there is no spatial bias, yet the first item processed again tends to inhibit perception and recognition of other objects in the field, regardless of their relative spatial positions. In an intact dorsal brain system, there is less likelihood of an overpowering winner-take-all perceptual process, and we are able to attend to several things at once, or at least in a rapid sequence. Normal perceptual processes are freed from the trap of focusing on one position or one object to the exclusion of information readily available elsewhere in the field of view. Normal perception of a world populated by objects and events centered on the point of fixation depends on our ability to distribute attention over the visible scene. Another related theory explains simultagnosia as an impairment of attentional disengagement (e.g., Farah, 1990). Using Posner's framework (Figure 5–17), simultagnosia can be explained by assuming that patients can engage attention on an object; however, because of their parietal damage, they cannot disengage their attention from the object. Therefore, their attention remains fixated on the object that was engaged first.

Figure 5–20 **Drawings by a Patient with Right Parietal Damage When Asked to Copy the Figures Shown on the Left**

Representations of a copying task by a patient with visual neglect. In this illustration by a patient with right parietal damage, information from the left side is ignored.



Chapter Summary

Attention apparently serves a dual purpose in enabling us to perform many complex tasks in a noisy environment. First, it selects items of importance, either for the task at hand or because they are of potential significance for survival (e.g., sudden onsets of new objects or events). Second, it monitors the capacity and focus of working memory to prevent us from being overloaded with information and incapable of responding appropriately. Attention is guided by endogenous strategies and goals as well as exogenous alerts from the environment. Therefore, attention is bidirectional, in the sense that it determines to a large extent what memories are formed as well as which memories are retrieved to guide subsequent attentional selection (Sherman & Turk-Browne, 2024). Attention also seems to serve as a binding process that unites the various features of an object into a holistic representation when attention is focused on its location. Neurophysiological studies of attention have shown that it can affect most cerebral processes by priming relevant channels and inhibiting irrelevant ones for the task at hand. Further, parietal lobe damage can seriously affect patients' abilities to attend to regions of space or multiple objects in space. Attention has truly evolved as the gateway to consciousness, as envisioned by James.

Review Questions

1. Is there one kind of attention, or does attention do different things in different situations?

Although attention has been defined as the gateway to consciousness, it does not appear to be a unitary concept. Attention can select things from the environment based on different information, such as an item's location, size, color, loudness, or pitch. We can look for red items in a drawer or listen for a flute in a symphony. Further, attention can be distributed over space and time, as certain objects or locations are processed to the exclusion of others, and certain events are selected and processed more deeply than preceding or following ones. Finally, attention can be goal-directed and under conscious control, or it can be captured automatically by salient objects or events in the environment, especially abrupt onsets of new stimuli.

2. What is the purpose of attention? How does it help us to deal with the outside world?

The main purpose of attention is to maintain conscious awareness of objects and events of current interest or importance, including those important for survival. It is an obvious fact that both the physical environment and the contents of memory contain far more information than we can process at any one time. Further, information in both of these sources varies greatly in their current importance values. Attention serves to select and maintain a small set of important information in conscious, working memory so that we can direct our limited central processing capacity to the task at hand. Still, attention can and must be diverted to important changes in the environment so tasks can be interrupted when necessary. It does us no good to carry on a clever conversation while driving if we do not notice the truck crossing into our lane.

3. Does attention change over space and time, or does it follow where we are looking and our current level of alertness?

Since attention controls the contents of working memory, it is likely to be the driving force behind selection of where to look next or when to listen for a critical sound. It is also more likely that attention directs eye movements as attention can shift to new objects or locations in the visual field within 50–150 ms, whereas the response latency of an eye movement to a visual target is around 250 ms. Attention presumably examines the periphery for items of interest and then guides eye movements to those locations one at a time.

4. Can we learn something without paying attention to it?

Although information can be selected by attention or even bypass attentional processing and automatically activate representations in long-term memory, it is believed that learning is something different. The learning of new information involves a structural change or addition to information in long-term memory, not merely the activation of existing memories. Of course there are different kinds of memories, and some of these can be acquired without attention, such as the implicit learning of certain motor skills, but the storage and retention of real-life events and encyclopedic knowledge of the world seems to be laid down through conscious processing in working memory, and its contents are determined by attention.

5. What changes in brain activity can be observed as attention changes from one focus to another?

Just about all areas of the brain involved with perception can have their levels of activity modulated by attention. Attention to a stimulus can enhance the neural activity measured by ERPs or fMRI at the earliest stages of perceptual analysis (i.e., at the primary sensory projection areas of the cortex), and it can result in inhibition of the neural activity associated with brain areas responding to unattended information. It is likely that excitation and inhibition of sensory channels reflects the selective aspect of attention to direct important information to deeper processes of analysis while attempting to filter out unwanted or irrelevant information. Attention has also been shown to modulate brain activity in the dorsal pathway, leading to object selection by location, and in the ventral pathway, leading to object selection by features, such as color or orientation. Finally, selection of objects for inspection is indicated by activity in the frontal lobes of the brain known to control voluntary eye movements; thus, all major lobes of the cerebral cortex show modifications of their activities by changes in attention.

Key Terms

Area MT/V5	**148**
Attention	**128**
Attentional blink (AB)	**155**
Bálint’s syndrome	**164**
Conjunction search	**143**
Consciousness	**129**
Endogenous attention	**140**
Exogenous attention	**140**
Feature search	**143**
Filter theory of attention	**130**
Hemianopia	**163**
Iconic memory	**137**
Illusory conjunctions	**143**
Inhibition deficit hypothesis	**129**
Lavie’s perceptual load theory	**146**
Preattentive stage	**141**
RSVP (rapid serial visual presentation)	**152**
Vigilance	**152**
Visual neglect	**163**