Transsaccadic Integration. The fact that people explore a complex visual scene by making a sequence of many fixations on different informative parts amply demonstrates that normal perception is a dynamic, temporally extended process. But the perceptual result of this sequence of saccades does not resemble even remotely the piecemeal series of images it produces on the retina. Rather, the perception seems to be that of a single, unified scene. This perception must therefore be constructed by integrating the information extracted from the sequence of exploratory eye movements into some coherent internal representation of a single scene (Hochberg, 1970). Moreover, this perception must involve some brief form of memory to bridge the gap between fixations.

One possibility for integrating the contents of sequences of saccades is that they are mapped into a larger, spatially organized memory array according to their positions. The result would be an integrated, composite representation of the visual environment, each fixation being superimposed in its proper location of the array, as illustrated in Figure 11.1.11. As plausible as this spatiotopic fusion hypothesis might sound, experimental results show that the visual system does not make use of it (Irwin, 1992), as we will discuss in some detail in Section 12.1.3. Rather, the integration appears to be performed at the level of more abstract representations of objects within the scene.

In addressing this problem of saccadic integration, Hochberg (1970) postulated what he called a schematic map: a representation consisting of possible samplings of a spatially extended scene together with contingent expectancies of what will be seen as a result of those samplings. Hochberg never described the internal structure of his schematic maps in detail, but it seems quite likely that structural descriptions, as described in Chapters 8 and 9, would fit the requirements. They explicitly encode the spatial relations among the various parts of an object, effectively specifying the direction and distance at which various configurations of features would be found. In Palmer's face schema (Figure 8.2.16), for example, vectors specify the information required to get from one facial feature to another. This information could be used to support recognition via overt eye movements or covert shifts of visual attention to appropriate places in the image. This information is also encoded redundantly enough to support a variety of different scan paths. Thus, structural descriptions appear to be well suited to serve as the internal structure that accumulates information gathered from a number of different fixations and that integrates it into a coherent, unified whole. This description could then be used to recognize the same object or scene despite a different sequence of eye fixations, provided the proper components and spatial relationships were verified in the input.

11.2 Visual Attention

Even while our eyes are fixated on a particular location, it does not appear that the visual system passively processes all the information available within the image. Rather, we selectively attend to different aspects of it at different times. Sometimes we attend globally to the whole scene; at other times we attend to a selected object or set of objects; at still other times we attend locally to a specific object part. We may even concentrate on a
particular property of a particular object, such as the color or texture of a shirt we are considering buying. Our ability to engage in these flexible strategies for processing different information within the visual field—generally referred to as attention—is therefore an important component of vision. Indeed, recent experiments suggest that attention may be required for us to consciously perceive anything at all (Mack & Rock, 1998).

Overt eye movements determine what optical information is available to the visual system; covert selective attention determines what subset of this information gets full processing. Attention is such a complex process (or set of processes) that it is difficult to define adequately. For the purposes of this book, however, we will consider visual attention to be those processes that enable an observer to recruit resources for processing selected aspects of the retinal image more fully than nonselected aspects.

Notice that this definition implies two different but related functions of attention: recruiting resources and focusing them on selected aspects of visual information. These correspond to two different properties of attention that theorists often distinguish:

1. Capacity. Capacity is the amount of perceptual resources that is available for a given task or process. Attentional capacity can vary with a number of factors, such as alertness, motivation, and time of day (Kahneman, 1973).

2. Selectivity. Even at a given moment, when the total capacity is fixed, the amount of attention paid to different subsets of visual information can be allocated flexibly to some degree. This ability allows attention to be selective in terms of what gets processed and what does not.

Of these two aspects, selectivity has been more intensively studied by vision scientists. In the remainder of this chapter, we will mainly be discussing selectivity, although capacity issues will sometimes arise as well.

Complex visual scenes like the ones that we normally look at contain a staggering amount of information, far more than we can be aware of at one time. As a result, we have to sample visual information over time in a series of distinct perceptual acts, each of which is inherently selective. As we discussed in the previous section of this chapter, voluntary eye movements are the first line of visual selection. But even when our eyes are stationary and we are processing a single retinal image, we selectively sample the information it contains for further processing.

This second level of selectivity does not consist of overt, physical acts of orienting such as turning our heads or eyes toward objects of interest, but of covert, internal acts of orienting toward different information available within the retinal image. You can demonstrate the selective effect of attention without eye movements simply by focusing on some small object in your field of view and then attending to (or noticing) various nearby objects—without moving your eyes. This is not particularly easy because eye movements and shifts in attention are normally performed together, but they can be separated with effort. The fact that they can is evidence for the existence of selective attention in vision, independent of eye movements.

A more compelling and rigorous demonstration of our ability to attend to different things without fixating them can be achieved with the help of a camera flash. At night or in a windowless room, turn off the lights and adapt to the dark for a few minutes. Face the flash toward the room (away from yourself) and press the button. This brief burst of intense light will “paint” a single unmoving image of the environment on your retina. It cannot be selectively sampled by foveal fixations due to eye movements, because the afterimage necessarily moves with your eyes whenever you make an eye movement. Even so, you will be able to direct your attention to a number of different objects within the scene before it fades from view. This covert sampling is the work of spatial selection, the process of concentrating attentional resources on information from a restricted region of the visual field. At least to a first approximation, such attentional changes seem to be like movements of an internal eye—the so-called “mind’s eye”—that can sample different locations within the stationary afterimage. Given the retinotopic organization of much of visual cortex (see Section 4.1.3), spatial selection can be thought of as internally sampling information from a restricted portion of a cortical map.

Spatial selection is only one aspect of visual attention, however. Attention is also at work when we selectively perceive different properties or features of the same object. Keeping a steady gaze on a complex object, for example, you can focus your attention sequentially on its color, its shape, its texture, its size, and so on. This dem-
onstration illustrates property selection. It is at first difficult to understand property selection in terms of attention being an internal "mind’s eye," for eyes have no physical structure that enables them to select among properties other than space. If different properties have spatially distinct representations within the brain, however, as suggested by the discovery of many different retinotopic maps in visual cortex to encode different visual properties (see Section 1.3.3), then property selection may also be understood in terms of covertly sampling information in different locations of these maps.

11.2.1 Early versus Late Selection

Why do we have the ability to selectively attend to different aspects of visual information? A plausible answer is that it protects the visual system from being overloaded by the massive amount of information available in the visual field. To be effective, however, attention must somehow manage to focus on the most important information given the organism’s current goals, needs, and desires. Otherwise, selective sampling would be essentially random, and random selection is not very useful to an organism. Attention is therefore likely to have some means of selecting the most relevant information to process further so that only irrelevant information is rejected.

But how can the visual system choose the most important information without first processing all the information to determine what is most important? This is the paradox of intelligent selection. If attention operates very early in the visual system, before much processing has been done, it is unclear how the attentional system can determine what is important. If attention operates relatively late, after a good deal of processing has already been done, it is easy to determine what is important, but much of the advantage of selection would have been lost because most of the irrelevant information has already been processed to perform the selection.

Selective attention to important information is possible by using heuristics based on either innate principles or ones learned through individual experience. It is evolutionarily advantageous to attend to some kinds of information before others. Moving objects are generally important for survival, for example, especially objects that are coming toward you. It therefore makes sense for moving objects to attract your attention for further analysis and for objects that are headed in your general direction to have priority. This attentional heuristic might even be hard-wired at birth through evolutionary processes of natural selection, and it seems plausible that it could be performed very early in visual processing.

In other cases, however, the importance of information is highly specific to an individual. Most people have had the experience of seeing their own name "pop out" of a page of text, for example, and grab their attention before any of the other words. This kind of selection clearly cannot be innate; it must be learned through experience by the individual. By the same token, it seems unlikely that it could be selected at a very early stage of processing, for it seems to presuppose the identification of the letters that make up the name. These theoretical considerations are suggestive, but whether (or how much) attentional selection takes place at early versus late stages of processing is an empirical question toward which many experiments have been directed.

Auditory Attention. The first research on whether attention operates early or late in human perception was conducted in the auditory domain. We will describe it briefly because many of the key questions and theoretical issues were originally explored there.

Auditory researchers began studying attention by asking subjects to perform a shadowing task in which they had to repeat aloud the message coming through either the left or right channel of a pair of headphones. The question of interest was what information subjects perceived about the other, unattended channel while performing this shadowing task. Initial results showed that they could perceive gross sensory features without attention, such as whether the unattended channel contained speech sounds or not and whether the voice was male or female. More specific features, such as what was being said or even whether the message was in English or French, were not perceived unless attention was diverted to the unattended channel (Cherry, 1953; Cherry and Taylor, 1954).

On the basis of such findings, British psychologist Donald Broadbent (1958) proposed that auditory attention operated early, analyzing the input to both ears only for gross sensory features and then selecting one ear for further processing of higher-level features to reach the level of meaning. This theory was called filter theory because it assumed that selection was due to an
all-or-none blocking mechanism (or filter) that passed only the selected channel (see Figure 2.2.5).

Subsequent studies showed that auditory attention was not quite this complete or simple, however. Moray (1959) found that subjects were very likely to hear their own name if it was presented in the unattended channel. This phenomenon may be familiar to you from personal experience. If you are at a party talking with one person, for example, and someone nearby says your name, you are very likely to notice it and to shift your attention to find out why your name was mentioned. Note that this fact causes problems for an early selection theory of auditory attention because it suggests that recognition of your name occurs before selection, not after it, as Broadbent’s filter theory would predict.

This difficulty was overcome by supposing that selection operates both early and late in auditory processing (see Figure 11.2.1). According to the most widely accepted theory, often called attenuator theory (Treisman, 1960), the initial phase of selection based on gross physical properties is only partial. That is, in contrast to Broadbent’s filter theory, early selection merely attenuates (or reduces) the signals in the unattended channels rather than blocking them completely.\(^2\) Attenuator theory can therefore be thought of as a “leaky” version of filter theory.

The second phase of attentional selection in attenuator theory operates during the process of identifying auditory events. Input information first activates dictionary units: internal representations of meaningful words and sounds that enable them to be identified. According to attenuator theory, words from both channels activate their corresponding dictionary units, but to different degrees. The units whose input comes from the attended channel are strongly activated, whereas those from the unattended channel are more weakly activated, owing to prior attenuation in the early selection phase. Many dictionary units can thus be active to different degrees at the same time.

The mechanism of late selection within attenuator theory is that dictionary units have dynamic thresholds that must be exceeded for conscious perception to occur. Highly salient items, such as one’s own name, have dictionary units with permanently lowered thresholds. Even weak activation from the unattended channel will therefore be sufficient to exceed its threshold and attract attention. This enables attenuator theory to account for Moray’s (1959) finding that subjects often hear their name when it occurs in the unattended channel. Less meaningful items will have higher thresholds, so they will tend to be perceived only if they arrive on the attended channel. Treisman also suggested that the thresholds of dictionary units could vary dynamically over time according to context. This would enable contextually expected words to be more easily identified than unexpected words, as reported in other auditory attention experiments (Gray & Wedderburn, 1960).

**The Inattention Paradigm.** Psychologists Arien Mack and Irvin Rock have recently begun investigating

---

\(^2\) Equivalently, the information arriving from the attended channel might be amplified rather than those from the unattended channels being attenuated. It is also possible that both amplification of attended information and attenuation of unattended information might occur. The important proposal is that some attentional mechanisms result in relatively more activation from the attended channel than from the unattended ones, regardless of how this effect is achieved.
similar questions about early versus late selection in the visual domain (Mack & Rock, 1998). They began by asking what visual features could be perceived without attention. To find out, they had to develop a procedure—which they called the **inattention paradigm**—in which attention would not be focused on the object whose properties they wished to study, even though the object would be clearly visible within the visual field if it had been attended. The task they used was a relatively difficult discrimination in which subjects had to determine whether the vertical or horizontal line of a large cross was longer (see Figure 11.2.2A). The cross was presented briefly (200 ms) and followed by a mask. Several trials of this task were presented with nothing except the cross on the computer screen so that subjects would not expect anything else to be presented.

Then, on the third or fourth trial, the experimenters presented an **inattention trial** in which an additional object was displayed near the cross (Figure 11.2.2B). After the subject said which line of the cross was longer, the experimenter asked whether the subject had seen anything in addition to the cross. Perception of various properties of the unattended object was assessed in this inattention trial through a recognition test in which subjects were asked to pick out the alternative that corresponded to the extra object. An example is given in Figure 11.2.2C for an experiment studying shape perception under conditions of inattention.

Figure 11.2.2  The inattention paradigm. Subjects are instructed to determine whether the horizontal or the vertical line of a briefly presented cross is longer (A), but on the inattention trial, an extra unexpected element is presented (B). Subjects are asked whether they saw anything besides the cross and are then given a recognition test (C) to evaluate their perception of the extra element.

Figure 11.2.3  Results from the inattention paradigm. Subjects perform better than chance at recognizing location, color, and number of elements but not shape. (Data replotted from Rock et al., 1992, to equate chance levels.)

After a few more trials of just presenting the cross task, subjects were again shown an extra object and again asked whether they had seen anything. This **divided attention trial** was included because subjects would have been alerted to the possibility of the extra object by the questioning that they received after the inattention trial shortly before. Subjects were given a final trial in which they were told to forget about the cross task entirely and to focus on perceiving anything else that might be present in the display. This **full attention trial** was designed to determine the perceptibility of the extra object under the same presentation conditions as the inattention and divided attention trials but without having to divide attention.

Initial results suggested that simple sensory properties were perceived without attention but that more complex ones were not. Specifically, the results indicated that color, position, and approximate number of objects could be perceived without attention but that shape could not (Rock, Linnett, Grant, & Mack, 1992). This conclusion was based on the fact that subjects performed no better than chance at picking out the correct shape alternative after the critical inattention trial but were almost perfect at doing so after the divided attention trial (see Figure 11.2.3). With properties such as color and position, however, subjects were nearly as good at choosing the correct alternative following the inattention trial as the divided attention trial. These results are reminiscent of early selection phenomena in auditory
attention, much like being able to tell whether the unattended voice was male or female without being able to understand what was being said (Cherry, 1953). Further experiments, to be described shortly, led Mack and Rock (1998) to believe that late selection was also occurring, however.

Other questions that have been examined to clarify the role of attention in perception were whether perceptual grouping or texture segregation occurs under conditions of inattention (Mack, Tang, Tuma, Kahn, & Rock, 1992). In the latter case, for example, the displays contained a texture of randomly placed vertical lines in the background of the cross for all of the initial trials (see Figure 11.2.4A). Then, on the crucial inattention trial, the orientation of all the lines in one quadrant was changed (Figure 11.2.4B), and subjects were asked whether they saw anything different. They did not. On the full attention trial, however, all subjects correctly reported the different quadrant. This indicates that the conscious perception of texture segregation requires attention, contrary to Julesz’s (1984) claim that it is preattentive. Gestalt grouping by proximity and lightness were also found to be absent in the inattention trial but clearly present in the full attention trial.

One surprising result from the early studies with single objects was that about 25% of the subjects reported not perceiving anything at all on the inattention trial (Mack & Rock, 1998; Rock & Mack, 1994). Mack and Rock refer to this phenomenon as inattentional blindness. It cannot be attributed to sensory factors because virtually everyone reported seeing the target on the divided attention and full attention trials. These trials were optically identical to the unattended trial but differed in terms of the subject’s expectation. On the inattention trial, they expected only the cross, but on the divided and full attention trials, they were also monitoring for anything else that might be presented. The much higher incidence of missing the target on the inattention trials therefore strongly suggests that expectation is an important component of inattentional blindness.

Subsequent studies demonstrated a number of even more surprising effects concerning the degree of inattentional blindness. For instance, the amount of attentional blindness was actually greater (typically 50–75%) when the extra object was presented foveally at fixation than when it was presented about 2 degrees off center, as in the usual procedure. Most surprising of all, however, was the finding that the degree of attentional blindness depends greatly on the personal meaningfulness of the extra stimulus. As in the auditory domain, Mack and Rock found that only about 5% of subjects were blind to their own name when it was presented under conditions of inattention. Presenting someone else’s name under the same conditions led to 35% inattentional blindness, and presenting letter strings with only one different letter—e.g., “Koi” instead of “Ken” or “Jeck” instead of “Jack”—led to about 60% inattentional blindness. Similar, but weaker effects of superior perception for meaningful visual stimuli under conditions of inattention were obtained for certain words (such as RAPE and STOP) and for a standard cartoon “happy face” (but not a sad, neutral, or scrambled face).

Clearly, these results from the inattention paradigm suggest that some form of late selection must be at work in visual as well as auditory attention. Unattended objects must be receiving fairly detailed visual processing for inattentional blindness to be so sensitive to the difference between one’s own name and a slight modification of it. Notice that this conclusion actually seems to contradict the earlier empirical finding that shape information is not perceived without attention (Rock et al., 1992). It is not yet clear what the resolution of this conflict will be. Perhaps shape does get processed without attention but does not become consciously perceived unless attention is then drawn to the object because of its high salience, as would be the case with one’s own name. Meaningless shapes would seldom attract attention and therefore fail to become conscious. If their activation dissipates over a matter of seconds, no trace would remain when subjects were asked to report whether they saw anything different. (We will consider some further evidence from the inattention paradigm supporting this
view in Chapter 13 when we tackle the topic of visual awareness.) It is puzzling from this hypothesis that properties such as color and position of meaningless shapes are sometimes perceived even when subjects are not expecting them.

Although these and other results suggest that late selection is possible in vision, the early/late question may not have a single solution. Lavie (1995) has recently proposed that both early and late selection occur, but under different conditions. When the task places a high load on visual processing, she finds evidence that selection operates at an early stage of processing, effectively blocking out stimuli other than those within the current focus of attention. When perceptual load is low, however, selection appears to operate at a later stage, allowing the processing of stimuli outside the focus of attention. The task and stimulus conditions in the inattention paradigm are consistent with a low perceptual load, which may explain why Mack and Rock’s results largely conform to the predictions of late selection.

Mack and Rock’s hypothesis that people are literally blind to unattended visual information, even though such information may be processed extensively at a non-conscious level, is a radical one. Is there any other evidence that bears on it? One compelling source of evidence comes from patients with certain kinds of brain damage who appear not to consciously perceive some of the objects in their visual field because of an inability to attend to them. We will discuss these conditions, known as unilateral neglect and Balint’s syndrome, a bit later in this chapter when we consider neurological mechanisms of attention. But there is also relevant evidence from other studies of normal perceivers who fail to see what would otherwise be clearly perceivable under conditions in which their attention has been captured by some other object or event. We will now examine two of these phenomena, known as the attentional blink and change blindness.

The Attentional Blink. The attentional blink refers to the fact that perception of a second target item is greatly reduced if it is presented within a half second of a first target item (Raymond, Shapiro, & Arnell, 1992; Shapiro, Raymond, & Arnell, 1994). It is typically studied in a rapid serial visual presentation (RSVP) search task, in which subjects are shown a very rapid sequence of visual stimuli, all at fixation where acuity is greatest, and are asked to report targets of a specific type (Forster, 1970). For example, subjects might be shown a series of 15 alphanumeric characters at fixation in a period of only 1.5 s (100 ms per character), 13 of which were digits and 2 of which were letters. Their task would be to report the identity of any letters that they saw in the RSVP stream.

If the rate of presentation is no faster than about 11 items/s, the first target can almost always be correctly identified. If the second target is presented more than about 500 ms after the first target, it too is well perceived and reported. But if the second target is presented within about 200–500 ms of the onset of the first target, subjects are very likely to miss the second one completely. They appear simply not to see it.

This phenomenon has been dubbed the attentional blink because one interpretation is that after the first target captures the subject’s attention, there is a period during which no attention is available for processing the incoming items that immediately follow it, much as blinking keeps visual information from being perceived while the eye is shut. If attention is indeed completely absorbed by processing the first target, then subjects’ failure to identify or even detect a second target can be counted as further evidence for Mack and Rock’s theoretical interpretation of inattentiveness. People do not see the second target because it cannot be attended to at the same time the first target is being processed. Other interpretations are possible, however, including ones based on failure of memory rather than perception (e.g., Wolfe, 1999). We will discuss this proposal shortly.

Subsequent results indicate that during an attentional blink an unperceived target nevertheless receives non-conscious processing to the level of meaning. This fact has been demonstrated both behaviorally and electrophysiologically. Behaviorally, the target that appears during the blink has been shown to prime (facilitate) a semantically related third target item that occurs after the blink is over (Shapiro, Driver, Ward, & Sorenson, 1996). Electrophysiologically, the target during the blink has also been shown to influence a component in the evoked potential (called N400) that is known to be sensitive to semantic factors (see Shapiro & Luck, 1999). It therefore appears that the attentional blink, like inattentional blindness, operates at a fairly late stage of selection.
Change Blindness. Another phenomenon that may support Mack and Rock's hypothesis that lack of attention to an object causes failure to perceive it comes from a series of recent studies on what has come to be called change blindness. The basic finding is that people are surprisingly poor at detecting even gross changes in a visual stimulus if they occur in objects that are not the focus of attention (e.g., Rensink, O'Regan, & Clark, 1995a, 1995b).

A basic change blindness experiment goes like this. Subjects are alternately shown two complex scenes that are identical except for one object or feature that changes. If the two pictures are spatially aligned and presented right after one another with no black interval or distracting event, this task is extremely easy. Attention is immediately called to the stimulus change, which can then be accurately reported. But if a brief blank interval is inserted between the presentations, the task becomes extremely difficult. The same difference that was found effortlessly without a blank interval can now take 20 s or more of repeated alternations while the subject laboriously searches, object by object, for the change.

To experience a version of change blindness for yourself, look back and forth between Figures 11.2.5A and 11.2.5B until you notice the difference between them. If you are like most people, you will find this task surprisingly hard. Once you find the change, you will probably be amazed at how long it took you to spot such a big difference. In this case, there is no blank interval, but the eye movements you have to make between fixations on the target pictures appear to serve the same function and cause blindness to such changes (Blackmore, Breland, Nelson, & Troscianko, 1995; Grimes, 1996). Other abrupt stimulus modifications, such as the sudden appearance of a distracting visual “mudsplash” will produce the same effect (Resnink, O'Regan, & Clark, 1997).

Further experiments show that this insensitivity to change over a short period of time is not a mere laboratory curiosity, but can occur even under normal perceptual conditions. In one study, a subject was approached by a stranger with a map who was asking directions to a location on campus (Simons & Levin, 1997). During their conversation, two workmen, carrying a door lengthwise, walked between the subject and person asking directions. In the few seconds during which the subject could not see the questioner, one of the workmen carrying the door deftly switched places with the person so that after the door passed, the subject was talking to a different person wearing different clothes. Only about 50% of the subjects noticed that any change had occurred when they were asked about it moments later.

One interesting interpretation of this literature on the various conditions under which we fail to see things that are quite clearly visible—including inattentional blindness, attentional blinks, and change blindness of various sorts—is that our impression of normal conscious perceptions of our environment as being rich, complete, and detailed is just a grand illusion (O'Regan, 1992). In
fact, it is claimed, we experience only the things to which we specifically attend for whatever purposes we currently have in mind, and the rest is simply not perceived because of the lack of focused attention. In this view, the unattended portion of the world seems to be there in our perceptions, at least under normal circumstances, because when we examine any given object to see whether it is fully represented in our perception, we necessarily attend to it. Once we have done so, the richly detailed information becomes part of our conscious perception, and seems as though it must have been there all along—even though it hasn’t been. The piecemeal nature of our perceptions can therefore be revealed only under ecologically unusual circumstances when the target object changes quickly.

A different interpretation is possible, however, and in some ways preferable. Wolfe (1999) has argued that these phenomena are evidence of inattentive amnesia rather than inattentional blindness. He claims that all of these supposedly unseen objects and changes are actually experienced perceptually, albeit very briefly. But without the benefit of focussed attention, he suggests, there is absolutely no memory of them, even over very short time intervals. This account has the advantage of not having to explain away the fact we have conscious perceptual experiences everywhere in our visual field rather than just where we are attending: It is simply because we do have conscious perceptions of unattended objects. The problem comes in reporting these fleeting and fragile perceptions when any sort of memory is required. The unattended perceptions are simply gone by the time we can divert attention to them from either the task in the inattention paradigm, the first target in the attentional blink paradigm, or the blank interval, mudsplash, or interrupting door in the change blindness experiments.

The proper interpretation of these intriguing and important findings is not yet clear. What is clear is that attention somehow plays a very important role in our conscious perception of visual events, by enabling nonconscious visual processing to reach consciousness and/or by creating durable representations in memory that can be used to report fleeting conscious perceptions that would otherwise disappear without a trace.

**Intentionally Ignored Information.** Thus far, we have concentrated on what happens to objects that are not attended because of lack of expectation (as in the inattention paradigm), lack of resources (as in the attentional blink) or some sort of distraction (as in change blindness). But what about objects that are **intentionally ignored**? What properties, if any, are perceived under these circumstances, and what is their fate?

The seminal experiment on this topic was performed by Rock and Gutman (1981). They constructed displays containing two novel outline figures that overlapped spatially, one red and the other green, as illustrated in Figure 11.2.6. Half of the subjects were told to attend just to the red figures and to rate them for aesthetic appeal; the other half were told to do the same for the green ones. After seeing a series of such displays, subjects were tested on their memory for the presented shapes using black versions of both sets of figures. Memory for the shape of the attended figures was quite good, but memory for the shape of unattended figures was essentially at chance. Rock and Gutman (1981) concluded that shape was not perceived unless a figure was attended.

There are alternative interpretations of this result, however. Perhaps shape was perceived perfectly well but forgotten so quickly that it was not recognized in the memory test. Or perhaps it was perceived initially but then was suppressed in order to process the attended figure. Although these hypotheses are more complex than assuming that shape was not perceived initially, both are consistent with Rock and Gutman’s results.

Using sophisticated experimental procedures, Tipper and his associates have found evidence that the ignored
shape is perceived at some level but not remembered because of active suppression (Allport, Tipper, & Chmiel, 1985; Tipper, 1985; Tipper & Cranston, 1985). The evidence supporting this interpretation is fascinating. Subjects were presented with two meaningful objects using Rock and Gutman’s overlapping-contours method (see Figure 11.2.7 for examples in this task) and instructed to name just the ones in a given color (e.g., red), ignoring the other. On repetition trials, the target object on a given trial was preceded by a trial in which the same object was presented in the unattended color (see Figure 11.2.7). On unrelated trials, the target object was preceded by a trial in which an unrelated object was presented in the unattended color. The surprising result was that naming the target object took significantly longer in the repetition trials (when the same object had been ignored on the immediately preceding trial) than on the unrelated trials (when it had not). This increase in response time has come to be known as the negative priming effect because the result is the opposite of what is usually found when a to-be-perceived object has been primed by prior exposure.

The existence of negative priming strongly suggests that subjects must have registered the shape of the unattended object, but suppressed its perception in order to correctly name the target object on that trial. The effect of this suppression is then measured on a repetition trial by slowing the process of naming the same object when it is presented in the attended color. Further experiments showed that the unattended figure appears to be processed at least to the level of meaning, because the slowing of responses also occurred for semantic associates of the suppressed object. For example, if a picture of a dog were presented in the unattended color on one trial, the time to name a cat figure presented in the attended color on the next trial would also be measurably slowed (Tipper & Driver, 1988).

These findings are theoretically interesting for at least two reasons. First, they indicate that intentionally ignored objects receive extensive perceptual processing, at least to the semantic level, because for them to slow responses to a related object on the next trial, they must have been identified on the previous trial. Second, they suggest that attention operates not only by facilitating processing of the attended object, but also by inhibiting processing of ignored objects. Although it has always been acknowledged that selection can occur either by facilitating the attended object or by inhibiting the unattended ones, Tipper’s results have provided strong evidence that both mechanisms are at work, at least in this situation.

What is responsible for the difference between the findings of Rock and Gutman versus Tipper and his associates? There are several possibilities:

1. **Long versus short retention interval.** Rock and Gutman typically assessed perceptual effects by measuring memory after a long delay, after several figures had been presented. Tipper’s negative priming paradigm assessed them much more quickly, in the very next trial after only a few seconds had passed.

2. **Indirect versus direct measures.** Rock and Gutman asked their subjects to make direct assessments of perceptual memory, whereas Tipper assessed perceptual memory indirectly by measuring performance on a task in which the effects of a previously seen stimulus could be measured as an increase in response time. The latter may be a much more sensitive index of visual processing than simply asking what subjects remember.

3. **Novel versus familiar stimuli.** To study negative priming effects on naming latencies, Tipper used figures that could be named, and this required familiar, meaningful
stimuli rather than the novel, meaningless ones that Rock and Gutman typically studied.

Surprising answers to these and other questions have come from a series of experiments by Treisman and DeSchepper (1996; DeSchepper & Treisman, 1996). To address the third possibility, they looked for negative priming effects using novel figures like Rock and Gutman’s. They did so by changing the task from naming individual figures to deciding whether pairs of figures were the same or different. Each trial’s display contained three figures: an attended figure in the target color (green in Figure 11.2.8) overlapping with an unattended figure in the unattended color (red in Figure 11.2.8), and a comparison figure (always white). Subjects had to indicate whether the green and white figures had the same shape or not, and reaction time was measured. The important question was whether the unattended red figure in the “prime” trial would affect performance on a later “probe” trial in which it reappeared as the attended (green) figure.

The critical experimental conditions were defined by the relation between the attended (green) figure on the probe trial and the unattended (red) figure on a previous prime trial after some number of intervening trials. When they were on consecutive trials (lag = 1), a negative priming effect of 55 ms was obtained for trials in which the same shape was repeated (versus control trials in which a different shape preceded it). This result shows that the difference between Rock and Gutman’s findings and Tipper’s was not due to the novelty/familiarity of the stimulus materials because negative priming was obtained with novel meaningless figures very much like Rock and Gutman’s.

Next, DeSchepper and Treisman investigated the effects of delay. Would the negative priming effect last for only a few trials, as Tipper (1985) had found with his familiar shapes, or would it last as long as the memory delays in Rock and Gutman’s experiments? Being careful to show each critical figure in only two trials, they found the same amount of negative priming at lags of 1, 100, and 200 trials. Additional experiments showed that measurable effects of a previous exposure could be obtained up to one month later!

These results indicate that the processes underlying negative priming can be very long-lasting indeed if the figures in question are novel. Moreover, they demonstrate how sensitive indirect measures of memory can be. When explicit memory was tested at comparable delays of 72–104 trials using four-alternative forced-choice recognition procedures—that is, picking the one previously shown figure from among four alternatives—memory for unattended shapes was 26%, no higher than chance (25%). Even attended novel figures were recognized only a bit better than chance (34%) at these delays. The primary reason for the difference between the results of Rock and Gutman and those of Tipper and his associates therefore appears to be the use of direct versus indirect measures of visual memory.

### 11.2.2 Costs and Benefits of Attention

We now turn to the nature of spatial selection under conditions of explicit attention when the observer is expecting the possibility of some event that contains needed information. We have been presuming that if such an event occurs in a location that is attended, it is processed in ways that are somehow different from how it would be if it were not attended. But precisely what are the consequences of explicitly attending to one object or place rather than another?
Selective attention certainly sounds like a good thing if it enables an organism to focus the bulk of its visual processing capacity on objects, locations, and properties of interest. But this concentration of visual resources presumably comes at a price: Unattended objects and/or properties receive correspondingly less processing. This is the "double-edged sword" of selective attention: It may have significant costs as well as significant benefits. For it to be evolutionarily useful, the benefits should outweigh the costs.

The Attentional Cuing Paradigm. The question of how to measure the costs and benefits of selective attention has been studied most extensively by psychologist Michael Posner and his colleagues at the University of Oregon. Posner, Nissen, and Ogden (1978) developed an attentional cuing paradigm that has proven to be particularly well suited to examining costs and benefits. The task is simplicity itself: Subjects must press a button as soon as they detect a brief flash of light. The light is presented either to the left or to the right of a central fixation point, as shown in Figure 11.2.9.

The crucial manipulation that allowed the costs and benefits of selective attention to be studied was a cue presented in the center of the visual field before the test flash. This cue gave subjects information about where the test flash was likely to appear. A left-pointing arrow (←) indicated that the flash would occur to the left of fixation on 80% of the left-arrow trials. A right-pointing arrow (→) indicated that the flash would occur to the right of fixation on 80% of the right-arrow trials. A plus (+) indicated that the flash was equally likely to occur on either side of the fixation point. The cue was shown 1 second before the test flash appeared, and the subject's reaction time (RT) to respond to the flash was measured. Subjects were instructed to keep their eyes fixated on the center of the screen. To be sure that eye movements did not contaminate the results, however, all trials on which subjects moved their eyes were discarded.

The relation between the central cues (←, →, or +) and the position at which the test flash appeared (on the left or right side of the screen) defines three attentional conditions of interest:

1. **Neutral trials.** When the + cue was presented, subjects got no prior information about the position of the test flash, so they presumably attended equally to both locations. RTs in this divided attention condition constitute a baseline against which RTs in the other two conditions can be compared to evaluate costs and benefits of focused attention.

2. **Valid trials.** On 80% of the arrow-cued trials, the flash appeared on the side to which the arrow pointed. On these "valid" trials, subjects are presumed to move their attention to the location cued by the arrow. If there are measurable benefits of selectively attending to the cued location, detection of the test flash should be faster on these valid trials than on the neutral trials.

3. **Invalid trials.** On 20% of the arrow-cued trials, the flash appeared on the side opposite the arrow, where subjects were not expecting it. If there are measurable costs of selectively attending to the cued location, detection of the test flash should be slower on these invalid trials than on both the neutral and the valid trials.

The results of this study are shown in Figure 11.2.10. Posner, Nissen, and Ogden (1978) found that RTs to the valid cues were about 30 ms faster than RTs to the neutral cues and that RTs to the invalid cues were about 30
**Figure 11.2.10** Costs and benefits of attention. The results in the cuing paradigm show that valid trials are faster than neutral trials (the benefit of correctly directing attention), whereas invalid trials are slower than neutral trials (the cost of misdirecting attention). (Data from Posner, Nissen, & Ogden, 1978.)

ms slower. This indicates that both costs and benefits are present in this particular task and that they are about equal. Given that the 30-ms benefit was obtained on 80% of the cued trials (the valid ones) and the 30-ms cost was obtained on only 20% (the invalid ones), the net benefits outweighed the net costs, at least in this objective sense.

Beyond measuring the basic costs and benefits due to attention, Posner and his associates also wanted to measure how long it takes subjects to shift attention to the cued location. They did this by performing a second experiment in which they varied the time interval between the presentation of the arrow cues and the test flashes (Posner, Nissen, & Ogden, 1978). The shortest interval was 50 ms, and the longest was 1000 ms. The experimenter reasoned that if the test flash were presented too soon after the cue, subjects would not have enough time to shift their attention to the cued location, and neither costs nor benefits would result. As the interval between cue and test increases, however, subjects would be increasingly likely to have completed the shift of attention by the time the test flashed. Thus, Posner and his colleagues predicted that both costs and benefits would increase as a function of the cue-to-test interval until some maximum level was reached, indicating the completion of the attentional shift.

The results of this experiment are shown in Figure 11.2.11A. Look first at the neutral trials in the middle. RTs to test flashes after the + cues were not much affected by the cue-to-test interval, presumably because they provided no information about the location of the test flash. This is the baseline to which performance in the other two conditions should be compared. As in the first experiment, performance on valid trials was faster than that on neutral trials. The difference between these two curves therefore measures the benefit of selective attention (Figure 11.2.11B). Notice that the magnitude of this benefit increases steadily as the cue-to-test interval increases, reaching its highest level at about 400 ms. Performance on invalid trials is again slower than that on neutral trials. The difference between these two curves therefore measures the cost of selective attention (Figure 11.2.11B). Notice that the magnitude of this cost increases as the cue-to-test interval increases, reaching its highest level by 200 ms. Thus, we conclude that attentional shifts from one location to another accrue benefits to the attended location and costs to the unattended location. Moreover, we infer that it takes about 400 ms for people to complete such an attentional shift and that the costs seem to accrue slightly before the benefits.

**Voluntary versus Involuntary Shifts of Attention.** Attention researchers have extended this experi-
mental paradigm to study the effects of different kinds of attentional cues. You have probably noticed that when there is a sudden change in your field of view, such as the appearance of a new object, it seems to draw your attention automatically. This involuntary summoning of attention appears to be quite different from the voluntary, effortful process of directing attention according to the arrow cues in the experiments just described.

Jonides (1981) extended the cost/benefit paradigm to find out what kind of differences there might be between voluntary and involuntary shifts of attention. He examined voluntary shifts of attention using centrally presented symbolic cues such as the arrows at fixation as described above. These are sometimes called push cues because attention must be “pushed” from the symbolic cue to the cued location. He also examined involuntary shifts of attention by presenting peripheral arrows right next to the cued location. Because it was expected that these peripheral cues could effectively summon attention directly to that location, they are sometimes called pull cues. Valid and invalid trials for each cue type were constructed by presenting the target object in the cued location or some other location, respectively.

Several differences have been reported between voluntary and involuntary shifts of attention using push and pull cues:

1. **Pull cues produce benefits without costs.** Push cues produced both benefits and costs relative to the neutral condition. In contrast, pull cues produced benefits without corresponding costs.

2. **Pull cues work faster.** When the cue-to-test interval was varied, the results indicated that an equivalent shift of attention took only about 100 ms instead of 200–400 ms.

3. **Pull cues cannot be ignored.** When the validity of push cues was lowered to chance level (50%), subjects were able to ignore them. When the validity of pull cues was comparably reduced, they still produced significant benefits. Indeed, they did so even when subjects were instructed to actively ignore them.

**Three Components of Shifting Attention.** The results of these experiments on attentional cuing clearly demonstrate that attention has measurable effects on a task as simple as detecting the onset of a visual signal. They also show that it can be moved under either voluntary or involuntary control. Moving attention from one object to another seems intuitively simple enough, but how exactly does it happen?

Posner has suggested that a sequence of three component operations is required to shift attention from one object to another (e.g., Posner & Petersen, 1990; Posner, Walker, Friedrich, & Rafal, 1984):

1. **Disengagement.** Since attention is normally focused on some object, the first thing that must happen is to disengage it from that object.

2. **Movement.** Once it is disengaged, attention is free to move and must be directed toward the new object.

3. **Engagement.** After reaching the target, attention must be reengaged on the new object.

Moving attention from one object to another seems so simple that it is hard to believe that it is composed of these separate processes. However, evidence from neuropsychology suggests not only that they are distinct operations, but that they are controlled by three widely separated brain centers.

Patients with damage to parietal cortex (see Figure 11.2.12) show a pattern of costs and benefits on the cuing task, indicating that they have difficulty disengaging their attention from objects. Patients with damage to the superior colliculus in the midbrain show a different pattern of results, suggesting that they have difficulty moving their attention. (These patients are also severely impaired in making voluntary eye movements, a fact that suggests an important connection between attention and eye movements to which we will return at the end of this chapter.) Finally, patients with damage to certain centers in the thalamus, including the lateral pulvinar nucleus, appear to have difficulty engaging their attention on a new object. Thus, the seemingly simple and unitary operation of shifting attention from one object to another actually requires a coordinated effort among three widely separated regions of the brain. When neural functioning in these areas is impaired, attentional movements fail in predictable ways (see Posner & Raichle, 1994, for a review).

**11.2.3 Theories of Spatial Attention**

How can we understand visual attention theoretically? As is often the case in cognitive science, the first step in theorizing about a mental process is to find an appropri-
Figure 11.2.12 Three brain centers that are involved in orienting attention. Areas of parietal cortex control the disengagement of attention from objects; circuitry in the superior colliculus in the midbrain controls the movement of attention from one location to another; and certain centers in the thalamus, including the lateral pulvinar nucleus, control the engagement of attention on a new object. (From Posner & Raichle, 1994.)

The internal eye metaphor is of limited theoretical utility, however, because there is a sense in which it is the operation of the eyes (i.e., vision) that we are trying to explain in the first place. The eye metaphor therefore has the potential problem of infinite regress: If attention is like an internal eye and if the real eye has an internal eye of attention, then does the internal attentional eye also have its own internal eye? And what about the internal eye of that internal eye? Even if the answers to these questions are negative—there may be just one internal eye of attention—many theorists find it preferable to liken attention to something simpler and better understood than an eye.

The Spotlight Metaphor. Among the most crucial aspects of the internal eye metaphor are the selection of one region on which to concentrate processing and the ability to move from one region to another over time. A simpler metaphor that captures both of these characteristics is that attention is like a spotlight (e.g., Posner, 1978). According to this spotlight theory, the object at the location where the spotlight of attention is focused is “illuminated” so that it stands out and can be processed more effectively than the less illuminated objects in other regions. Once that object has been processed, the attentional spotlight can be shifted to a different location by moving along a path from the present object to the next one, presumably through the disengage/move/engage sequence of elementary attentional operations mentioned earlier.

The spotlight metaphor of attention has proved both popular and productive. Many experimental results can
be understood in terms of it, and new experiments have been devised to test some of its predictions. Consider how Posner's cuing results might be explained, for example. On a cued trial, subjects use the cue to move the attentional spotlight from the central cue to the appropriate location. If the test flash occurs there (a valid trial), it is already in the spotlight of attention, so it can be processed quickly without requiring any subsequent attentional shift. However, if the test flash occurs on the unexpected side (an invalid trial), the spotlight is in the wrong location and must be moved to the correct one before the response can be made. If there is no directional cue (a neutral trial), the spotlight stays in the center and would be moved only half as far as it would on an invalid trial. The spotlight metaphor can thus account for the basic pattern of results shown in Figure 11.2.10. It can also account for the results obtained when the cue-to-test interval is varied, because it will take some amount of time for the attentional spotlight to reach the cued side from the center.

Further experiments have tested a number of predictions derived from the spotlight metaphor. Some have received strong support, but others are controversial. Among the most interesting predictions are the following:

1. **Rate of motion.** The amount of time it takes to shift attention to a target object should increase systematically with the distance over which it must be moved, as though a spotlight were scanning from one place to another. Tsal (1983) has obtained evidence supporting this prediction and has estimated the rate of motion experimentally at about 8 ms per degree of visual angle.

2. **Trajectory.** When a spotlight is moved from one object to another, it illuminates the objects along the path between them. Some evidence suggests that the same is true when attention is moved (e.g., Shulman, Remington, & McLean, 1979).

3. **Size.** Spotlights are generally fixed in size. Eriksen and Eriksen (1974) reported evidence suggesting that the attentional spotlight is about 1 degree of visual angle in size. (As we will see, however, there is also evidence that it can vary in size.)

4. **Unitariness.** A spotlight can be moved from place to place, but it cannot be divided into two or more separate regions. Eriksen and Yeh (1985) reported evidence suggesting that the same is true of attention. (But again there is also evidence for the opposite conclusion.)

Despite its successes, there are a number of problems with the simple spotlight metaphor that have led theorists to consider alternatives. One difficulty is that, despite Eriksen and Eriksen's (1974) conclusion that attention covers only about 1 degree of visual angle in their particular experiment, it seems that under normal viewing conditions attention can cover a much wider area of the visual field, such as when you look globally at a large object or even a whole scene. It also seems that attention can be narrowed to a tiny region of the visual field, as when you scrutinize a small detail. These considerations have led to an alternative metaphor.

**The Zoom Lens Metaphor.** The zoom lens theory likens attention to the operation of a zoom lens on a camera that has variable spatial scope (Eriksen & St. James, 1986). The analogy is not exact, however, for the idea is that attention can cover a variable area of the visual field is usually coupled with the further assumption that varying the size of the attended region changes the amount of visual detail available within it. With a relatively wide attentional scope, only coarse spatial resolution is thought to be possible, whereas with relatively narrow scope, fine resolution is possible.

Shulman and Wilson (1987) tested this idea experimentally. They showed subjects large letters made up of small letters, like the stimuli Navon (1977) used to study global and local processing (see Section 7.6.3), as illustrated in Figure 7.6.9. On some trials, subjects had to identify the large letters, and on other trials the small ones. Shortly after each such trial, they had to respond to a sinusoidal grating that was either low in spatial frequency (wide fuzzy stripes) or high in spatial frequency (thin fuzzy stripes) (see Section 4.2.1). Shulman and Wilson found that responses to low-spatial-frequency gratings were enhanced after subjects had attended to the large global letter and that responses to high-spatial-frequency gratings were enhanced after subjects had attended to the small local letters. This is precisely what would be expected if attention worked like a zoom lens that took time to be adjusted to different sizes and spatial resolutions, large sizes being associated with coarse resolution (low spatial frequencies) and small sizes with fine resolution (high spatial frequencies). These findings
are therefore widely cited as supporting the zoom lens metaphor.

Notice that the spotlight metaphor is actually compatible with the zoom lens metaphor in the sense that they can be usefully combined. One can easily conceive of a spotlight that is variable in size as well as position. If the total power of the spotlight is fixed, then a wide beam will illuminate a large region dimly, and a narrow beam will illuminate a small region intensely. This connection between beam width and brightness is not exactly the same as the presumed relation between attentional scope and resolution, but it provides a relatively simple metaphor for thinking about how attention might be distributed over space in a way that includes position, scope, and effectiveness.

**Space-Based versus Object-Based Approaches.**
The metaphors for attention that we have considered thus far—an internal eye, a spotlight, and a zoom lens—all have one important thing in common: They assume that attention selects a region of space. A spotlight, for example, illuminates whatever lies within its beam, whether it is an object, part of an object, parts of two or more nearby objects, or nothing at all. An important alternative to these space-based theories is the possibility that attention actually selects a perceptual object (or group of objects) rather than a region of space (e.g., Duncan, 1984). Notice that these object-based theories of attention allow a good deal of leeway in how attention might be deployed, because of differences in what constitutes a perceptual object. It could be directed at a single complete object, part of an object, or even an aggregation of objects, as discussed in Chapter 6 when we considered the hierarchical structure of perceptual organization.

Identifying perceptual objects as the domain of selective attention might make object-based accounts seem too ill-defined to be useful, but it does impose significant constraints on the distribution of attention. Unlike space-based theories, for example, object-based theories of attention cannot account for selection of arbitrary portions of two or more different objects, even if they are located within a spatially circumscribed region such as might be illuminated by a spotlight. Also unlike space-based theories, object-based theories can, under certain conditions, account for attentional selection of several discontinuous regions of space. These conditions require that the “object” of attention be a perceptual grouping of several objects whose members are typically defined by some common property (such as color or motion) with other objects interspersed between them. If attention can be allocated just to the set of objects within such a group, it need not occupy a connected region of space. The spotlight and zoom lens metaphors require that a unified region of space be selected.

Some of the strongest evidence for an object-based view of attention comes from a neurological condition known as Balint’s syndrome, in which patients are unable to perceive more than one object at any time. We will discuss this syndrome later (in Section 11.2.7) when we consider the physiology of attention, but there is also good experimental evidence for object-based attention with normal perceivers. One of the most widely cited studies is an experiment by Duncan (1984). He reasoned that if attention is allocated to objects rather than to regions of space, it should be easier for subjects to detect two different properties of the same object than two properties of different objects that lie within the same region of space. He showed subjects displays like the one illustrated in Figure 11.2.13. Each stimulus consisted of two objects: a box with a gap in one side and a line running through the box. Each object had two relevant attributes. The box was either short or long and had the gap slightly to the left or right of center. The line was either dotted or dashed and tilted clockwise or counterclockwise from horizontal. After a brief presentation, subjects had to report either one or two attributes. When two attributes were tested, they could belong to the same object or different objects. Duncan found that if the two attributes belonged to different objects, subjects were worse at detecting the second property than

![Figure 11.2.13](image)

*Figure 11.2.13* Stimuli for Duncan’s experiment on object-based attention. Subjects had to report two features of a stimulus display that varied on four dimensions: line slant (left or right), line type (dashed or dotted), box length (long or short), and gap placement (left or right). Performance was better when the two features belonged to the same object than to different objects.
the first. But if they belonged to the same object, no such difference was obtained.

These results can easily be explained within an object-based view of attention. When the two properties come from the same object, no shift of attention is required because attention is defined by the single object. When they come from different objects, an attentional shift is required to detect the second property, taking additional time and therefore reducing accuracy. This pattern of results is more difficult to square with space-based theories, however, because the two objects occupy essentially the same region of space. A roughly circular spotlight that illuminates either the box or the line, for instance, will necessarily illuminate the other object. It is therefore not clear how such a difference in detecting properties would arise unless objects were somehow implicated in the allocation of attention. Notice that if a space-based theory allows the attentional spotlight to be shaped tightly around specific objects (e.g., LaBerge & Brown, 1989), they take on a significantly object-based flavor.

There is also recent evidence from a Posner-type cuing experiment suggesting that attention operates at an object-based level. Egly, Driver, and Rafal (1994) showed subjects displays containing two rectangles oriented either horizontally or vertically, as shown in Figure 11.2.14. After the initial presentation, the edges of one end of one rectangle brightened briefly, providing a pull cue to attend to that location. Subjects were then to make a response as quickly as possible when a dark square appeared anywhere in the display. When the cue was valid, the target square appeared briefly in the cued end of the cued object. There were two different types of invalid cues trials, however. In the same object condition, the target square appeared in the opposite end of the cued rectangle. In the different object condition, it appeared in the uncued object but at the same end as the cue.

As usual in the cuing paradigm, responses were faster when the target appeared at the cued location than when it appeared at either of the uncued locations. The results for the uncued locations showed object-based attentional effects, however, in that the same object condition was faster than the different object condition, even though their distances from the cued location were equal. This finding suggests that switching from one object to another incurs an additional cost that cannot be attributed to distance.³

A third finding that lends credence to the object-based view concerns an observer’s ability to keep track of several moving objects at once. Pylyshyn and Storm (1988) showed subjects a random array of static dots and designated several of them as target objects to be attended by flashing them on and off. All the dots then began to move in quasi-random (but continuous) trajectories, and subjects were instructed to try to keep track of the ones that were initially designated as targets. After several seconds of such motion, one of the elements flashed, and subjects were asked whether or not that particular dot was one of the initially designated targets. Pylyshyn and Storm found that subjects could track as many as five dots at once.

If one believes that this tracking task requires attention (and Pylyshyn and Storm do not), these results pose significant problems for space-based theories of attention. First, most space-based theories assume that the region to which attention can be allocated is a unitary,

---
³More recent experiments have shown that the “objects” in question are perceptually completed objects rather than retinally defined objects. Moore, Yantis, and Vaughan (1998) concluded this after finding results similar to those of Egly, Driver, and Rafal when the different ends of the same-object stimuli had been retinally separated by an occluding object. They also found this pattern of results when the two ends of the same object were defined only by illusory contours.
convex, connected area. However, the tracking task seems to require observers to attend to a number of disconnected spatial regions at the same time. Equally important, the trajectories of regions through space are defined only by virtue of the objects that traverse them. How else could attention be allocated to the proper regions of space at the proper times?

These results have another feature that is at least somewhat troubling from the object-based perspective, however: They seem to indicate that attention can be split among multiple objects. One could, of course, extend the object-based view specifically to allow for attention to be divided among some small number of objects. Indeed, this is how Pylyshyn and Storm (1988) interpreted their results. But there is another possibility that does not require giving up the unitary nature of attention. Perhaps observers keep track of the multiple dots by grouping them initially into a single superordinate object and attending to that group as a unitary entity. The designated dots could then be perceived, for example, as the corners of a virtual polygon whose shape changes over time as the dots move. Yantis (1992) has tested predictions from this hypothesis and found support for them in several experiments.

The current debate between object-based and space-based theories of attention often implies that they are mutually exclusive—that one or the other is correct but not both. This would presumably be the case if attention operates at just one level in the visual system. But what if attention operates at multiple levels? At an early image-processing level, such as Marr's primal sketch, a space-based definition of attention is the only thing that makes much sense, because in this low-level representation, coherent "perceptual objects" have not yet been designated. But at a higher level, after organizational processes have identified figures against grounds, objects could certainly be the basis for allocating attention. Both hypotheses may therefore be correct, just at different levels of the visual system.

11.2.4 Selective Attention to Properties

The theories of selective attention we have just discussed—including spotlights, zoom lenses, and even object-based theories—are designed to account for spatial selection. They cannot be the whole story of visual attention if its capabilities extend to selection of different properties, however. When you inspect a prospective purchase at a clothing store, for example, you seem to be able to focus selectively on its color, style (shape), texture, and size as you consider the garment. This ability appears to imply that attention must have important nonspatial components that select for other sorts of properties. Such evidence is anecdotal at best, however. Can people really attend to different properties of the same object independently or does attending to one necessarily result in perceiving them all? In this section we will consider experimental evidence that bears on this question.

The Stroop Effect. Early experiments seemed to indicate that if an object is attended, certain properties are processed automatically, even if the observer is trying to ignore them. This implies that selection by properties is either nonexistent or incomplete.

The best-known evidence for this conclusion comes from the Stroop effect, named for the psychologist, J. Ridley Stroop, who discovered it in 1935. The Stroop effect refers to the fact that when subjects are required to name the color of ink in which color words are printed, they show massive interference when the color word itself conflicts with the ink color to be named (Stroop, 1935). Examples of stimuli that produce this effect are shown in Color Plate 11.1.

You can demonstrate the Stroop effect for yourself by timing how long it takes you to name the ink colors in the column of X's on the left (the control condition) versus the conflicting color-word condition in the center versus the compatible color-word condition on the right. Even without timing yourself with a stopwatch, you will find it much more difficult to get through the middle column than the other two. This fact indicates that shape information is being processed automatically whenever the color of the word is attended and that the response to the identity of the word interferes with naming the ink color. This finding seems to imply that, unlike our intuitions based on everyday experience, selective attention to color, independent of shape, may not be possible after all.

One might wonder whether this interference is specific to color. Further research has shown that it is not. Stroop interference occurs, for instance, if the subject's task is to name an object that is drawn in outline around the name of a different object, as illustrated in Figure 11.2.15A. It also occurs when subjects have to name the
Figure 11.2.15 Additional Stroop effects. Stroop compatibility effects are also found in naming drawings of objects when object names are presented inside them (A) and in naming spatial locations when location names are the objects to be located (B).

location of letter strings (LEFT, RIGHT, or XXXX) when they are presented on either the left or right side of the display, as shown in Figure 11.2.15B (Clark & Brownell, 1975, 1976).

Interestingly, interference in the reverse direction does not occur in the standard Stroop task: Color words can be read just as quickly when they are printed in a conflicting color of ink as when they are printed in the compatible color. This shows that some selective attention to properties is indeed possible. It also suggests that there may be something special about reading printed words. One important consideration is that reading is a salient and highly practiced perceptual task. Educated people spend a significant portion of their waking hours reading printed words, and this degree of practice may result in highly automatic processing of letters and words. Another relevant factor is that there is a much more direct association between the printed color word and the sound of the color name (which is what the subject must produce) than there is between the color itself and the sound of the color name. That is, the letter-to-sound correspondences of English may provide more direct access to the name than the color does. The importance of this fact is supported by the finding that Stroop interference is reduced significantly when subjects must press different buttons to indicate ink color rather than when they must say color names. It also explains why noncolor words that begin with the same letters as color words also produce Stroop interference, such as ROD instead of RED or BLOB instead to BLUE. (See MacLeod, 1991, for a review of Stroop effects.) These factors suggest that we should consider evidence from other kinds of tasks that do not depend so heavily on the special properties of reading printed words before we draw conclusions from what may be a very special case.

Integral versus Separable Dimensions. Psychologist Wendell Garner of Yale University has also attempted to answer questions about selective attention to different properties, including both discrete features and continuous dimensions. From an extensive series of experiments, he concluded that there is no single, simple answer. Different patterns of results arise, depending on the particular pair of dimensions being studied. His findings led him to distinguish between two different relations that can hold between pairs of properties or dimensions: separability and integrality (Garner, 1974).

1. Separable dimensions. Pairs of dimensions are separable if people can selectively attend to one or the other at will, without interference from the unattended property. The internal representations of separable dimensions therefore appear to be completely independent. Classic examples of separable dimensions are the color and shape of an object, both of which Garner found could be perceived selectively.

2. Integral dimensions. Pairs of dimensions are integral if people cannot selectively attend to one without also perceiving the other. Classic examples of integral dimensions are the saturation and lightness of a color. These two dimensions seem to be processed together whenever one attends to the color of an object.

The integral/separrable dichotomy arose in a number of different experimental paradigms that Garner developed. The most powerful and widely studied method concerned a set of closely related tasks requiring speeded classification. Subjects were presented with various subsets of four stimuli defined by two different dimensions and were asked to classify them in different ways on different tasks. The speed of the perceptual discrimination was measured by reaction time. We will illustrate these conditions using the example of figures
The unattended property varies with respect to the attended one within different blocks of trials, as illustrated in Figure 11.2.16.

1. **Unidimensional variation condition.** Subjects are told to classify the stimuli according to their value on one of the two dimensions while the other dimension is held constant.

On a unidimensional block of trials subjects would classify white squares and black squares according to lightness; on another block of trials they would classify white circles and black circles according to lightness. Thus, the shape of the figures is constant within each block of trials in which lightness is the dimension of classification. The same procedure would be followed for the other dimension on further blocks of trials—for example, white squares and white circles would be classified by shape in one block, and black squares and black circles by shape in another block. Note that in these unidimensional variation conditions, only two of the four stimuli are presented in any given block.

2. **Correlated variation condition.** Subjects again are told to classify the stimuli according to the value on just one dimension, but the other dimension varies in a correlated fashion.

For instance, white squares would be discriminated against black circles in one block of trials. In another block, black squares would be discriminated against white circles. In the correlated condition, therefore, subjects could use either of the two properties—or both simultaneously—to perform the classification task. If performance in these correlated conditions is faster than in either of the two unidimensional conditions, the difference in reaction time is called a **redundancy gain** because the two properties in the correlated condition are redundant with each other. As in the unidimensional case, only two of the four stimuli are presented in any one block.

3. **Orthogonal variation condition.** Subjects again have to classify according to a single specified dimension, but this time the other dimension varies independently (orthogonally) so that all four stimuli are presented within each block of trials.

On one block of trials subjects would have to classify all four stimuli according to shape alone (square versus

---

*Figure 11.2.16 Three speeded classification tasks studied by Garner. Subjects were required to discriminate between two values of one dimension when the other dimension was held constant (unidimensional variation), when the other dimension co-varied (correlated variation), or when the other dimension varied independently (orthogonal variation).*
Figure 11.2.17 Characteristic patterns of results for separable versus integral dimensions. (C = correlated, U = unidimensional, O = orthogonal, IL = interference loss, RG = redundancy gain; see text for details.)

circle), and in another block according to lightness alone (black versus white). In these orthogonal conditions, subjects have to actively ignore the second dimension to perform accurately. It therefore might take additional time to effectively separate the relevant dimension from the irrelevant one. If performance in these orthogonal conditions is slower than for the corresponding unidimensional case, the difference in reaction time is called an interference loss.

Figure 11.2.17A shows idealized results that might be found for these three conditions. Reaction times in the unidimensional conditions (U) give the baseline data for each dimension and show that discriminating shape takes longer than discriminating lightness. Comparing these times to those of the correlated conditions (C) shows no significant differences from the faster unidimensional case (in this case, lightness), indicating that there is no redundancy gain when either or both properties could be used to make the discrimination. The results of the orthogonal condition also show no significant difference from the corresponding unidimensional case, indicating no interference loss. This is the pattern of results expected for separable dimensions, assuming they can be selectively attended at will: neither redundancy gain nor interference loss. Such results therefore support the view that selective attention to properties is possible.

When the stimuli are single color chips that vary in saturation and lightness, however, the pattern of results changes dramatically to that characteristic of integral dimensions. As shown in Figure 11.2.17B, the correlated (C) condition produces a significant redundancy gain (RG) compared to the unidimensional (U) condition in both cases. The orthogonal (O) condition also produces a significant interference loss (IL) relative to the corresponding unidimensional (U) condition in both cases. It is as though subjects cannot pay attention to either one of the properties without automatically perceiving the other. If the two properties vary together, they help performance; if they vary independently, they hurt performance. Quite a different pattern of results emerges, however, if the same two dimensions of color are used in two spatially separated color chips. There is now no redundancy gain and no interference loss (that is, the same pattern as in Figure 11.2.17A). Thus, when lightness and saturation are spatially separated, they are attentionally separable. This fact reflects the efficiency of spatial selection.

Garner developed several other tasks using pairs of properties whose results he expected to support the distinction between separable and integral dimensions. One such task was to have subjects make a particular kind of similarity judgment. On each trial they were presented with three stimuli and asked to indicate which two seemed most similar. The three stimuli differed on two dimensions, as shown in Figure 11.2.18. The question was which pair subjects would see as most similar: Y and Z, which were closer together in the two-dimensional space of stimulus attributes but differed on both dimensions, or X and Y, which had exactly the same value on one dimension but were farther apart in the stimulus space. The results of many experiments showed that subjects tend to choose the “close” pair (Y and Z) as being most similar when the two dimensions were found to be integral in the speeded classification tasks described above. This situation is illustrated in Figure 11.2.18B by the dimensions of a rectangle’s width and height. Most people see Y as being more similar to Z than to X in this example. In contrast, people tend to
Garner’s similarity classification task. Three two-featured stimuli are presented such that one pair (X and Y) share the same value on one dimension, whereas another pair (Y and Z) are closer together in the dimensional space. Integral dimensions tend to produce classification based on overall similarity (Y and Z are judged most similar), whereas separable dimensions produce classification based on identity on one dimension (X and Y are judged most similar).

Garner’s findings as well as Stroop interference effects show that selective attention to different properties is possible for some pairs of dimensions but not for others. They therefore indicate that metaphors based solely on the spatial distribution of attention are inadequate to account for selective phenomena of perception. A broader theoretical framework is needed to integrate spatial and property aspects of attention. Such an approach has been developed and tested by psychologist 1969). This idea has been tested most rigorously within Garner's framework of integral versus separable dimensions and has received considerable support. Children younger than about 6 years of age process almost all dimensions integrally, whereas older children process some integrally and others separably. The usual task for assessing this developmental trend is similarity judgment with triads of stimuli, as shown in Figure 11.2.18. The critical result is that for the same separable pair of dimensions, adults and older children choose the dimensionally identical pair (X and Y) as being more similar, whereas young children choose the closest pair (Y and Z) as being more similar (Shepp & Swartz, 1976; Smith & Kemler, 1977; Smith & Evans, 1989).

Figure 11.2.18 Garner's similarity classification task. Three two-featured stimuli are presented such that one pair (X and Y) share the same value on one dimension, whereas another pair (Y and Z) are closer together in the dimensional space. Integral dimensions tend to produce classification based on overall similarity (Y and Z are judged most similar), whereas separable dimensions produce classification based on identity on one dimension (X and Y are judged most similar).
Anne Treisman in her feature integration theory. To understand how she arrived at and tested this theory, we must first consider how visual processing might differ when attention is spatially distributed versus when it is focused.

11.2.5 Distributed versus Focused Attention

We claimed at the outset of this section that attention can function as a mechanism of perceptual selection. We reasoned that because there is far more information in a single visual scene than any observer can perceive at once, only a portion of it is selected for further processing to avoid sensory overload. This view of the function of attention implies that visual processing before attentional selection is somehow fundamentally different from the processing that occurs after it. This distinction is classically referred to as preattentive versus attentive (or sometimes postattentive) processing (e.g., Treisman, 1985). We will avoid this terminology because Mack and Rock (1998) have made a convincing case that all experimental tasks normally used to define so-called “preattentive” vision constitute conditions in which attention is explicitly deployed to perceive an expected target at an unknown location. Instead, we will call the mode of processing that occurs when subjects are prepared for the target to appear in any location distributed attention and we will call that which occurs when they have selected a single perceptual object focused attention. We thus take the view that both distributed attention and focused attention constitute examples of attentive processing, but ones that nevertheless differ in important respects.

The key difference between distributed attention and focused attention is parallel processing versus serial processing. The distributed processing that occurs before focused attention occurs in parallel, that is, simultaneously over the whole visual field. Processing that occurs after focused attention is serial, that is, a sequence of attentional fixations, each of which covers a limited region of space. Image-processing operations—such as finding edges and regions defined by luminance, color, texture, motion, and so forth—are generally thought to be performed automatically and in parallel. They would be sufficient to construct all the structures in Marr’s (1982) primal sketch, for example. It is less clear whether information about surfaces in depth is also processed in this way. But certainly by the time discrete perceptual objects have been created, attention can be selectively focused on one object (or a group of them) to the exclusion of other objects in the visual field. The closer scrutiny afforded by focused attention is thought to allow more detailed information about the chosen object to be processed.

Visual Pop-Out. A question of considerable theoretical interest is: What kinds of information can be processed in parallel without focused attention? In a series of influential papers, Treisman and her colleagues proposed that the phenomenon of visual pop-out indicates distributed attentional processing (e.g., Treisman, 1985; Treisman & Gelade, 1980; Treisman & Gormican, 1988). Pop-out occurs when an observer is looking at a field of similar objects, and one of them appears simply to “pop out” of the rest phenomenologically because it is very different from the others. In effect, the different item calls attention to itself, much as a moving object does in an otherwise stationary visual environment or a single red element does among a field of blue ones. Consider Figure 11.2.19, for example. All except one of the objects are circles, but the one cross effortlessly pops out from everything else. No laborious scrutiny of individual elements is required to detect its presence.

The phenomenon of visual pop-out for a given pair of mutually exclusive properties (such as a diagonal line within a field of many vertical lines or a red square in a field of many green squares) implies that the visual system can detect these properties in parallel over the whole visual field. If it did not, the different element would not pop out immediately but would require a sequential search through individual elements in the
A. PARALLEL SEARCH

<table>
<thead>
<tr>
<th>Number of Distractors</th>
<th>Response Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Absent</td>
</tr>
<tr>
<td>32</td>
<td>Present</td>
</tr>
</tbody>
</table>

B. SERIAL SEARCH

<table>
<thead>
<tr>
<th>Number of Distractors</th>
<th>Response Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Absent</td>
</tr>
<tr>
<td>32</td>
<td>Present</td>
</tr>
</tbody>
</table>

Figure 11.2.20 Patterns of response times for parallel and serial search. In parallel search, response times do not vary as the number of distractors increases. In serial search, response time increases linearly, the slope of the target-absent displays being twice that of target-present displays.

display. Effortful serial search is a process that is believed to characterize focused attentional processing.

This line of thinking led Treisman and Gelade (1980) to develop a specific experimental method for determining whether properties are processed serially or in parallel. Its logic is that if a property can be analyzed simultaneously over the whole visual field (that is, with distributed attention), a single target object with that property should be detected in the same amount of time, no matter how many other objects are in the display. However, if the display must be searched sequentially for a target object having that property (that is, with focused attention), the time required to find it should increase linearly with the number of items in the display.

These two patterns of predicted results are illustrated in Figure 11.2.20. Example stimuli for such a task are shown in Figure 11.2.21, in which line orientation is the relevant search dimension within displays containing 2, 8, and 20 distractor elements. Detecting the presence of the diagonal target line within these stimuli produces a flat response time function like that shown in Figure 11.2.21 An example of parallel search. The tilted line segments can be detected as quickly among 20 distractors as among 2, indicating a parallel search process.

11.2.20A. Such results indicate that the diagonal line can be detected among vertical lines simultaneously over all spatial locations. Making such a discrimination thus appears to be within the capabilities of distributed attentional vision. Similar results indicate that pop-out occurs when subjects search for targets that differ from distractors in color, shape, size, and a host of other simple visual properties.

Treisman and her colleagues interpret visual pop-out of a given feature as evidence that it is elementary in visual processing, that is, that it is one of the basic properties used by the visual system in constructing its initial representation of the image. They further argue that the findings from visual pop-out experiments are broadly consistent with what might be expected from the known physiology of the visual system. Explorations of visual cortex using single-cell recording and autoradiographic techniques (see Section 2.2.3) have identified many different regions that code distinct visual properties (Zeki, 1978). We have already discussed the fact that cells that are sensitive to different orientations and spatial frequencies are present in V1. Other cortical regions appear to code properties such as color, binocular disparity, and motion (see Section 4.4). One of the striking aspects of the architecture of these cortical regions is that their cells are arranged in clear retinotopic maps: sheets of cells whose spatial arrangement corresponds topographically to the 2-D positions of their receptive fields on the retina. Thus, the visual system appears to analyze the visual image into separate representations for different properties, each of which is organized by position. Activity within such maps could account for the spatially parallel detection of the odd element in a pop-out experiment if unique activity in a retinotopic map calls focused attention to that location.
Search Asymmetry. Treisman and Souther (1985) extended the investigation of visual pop-out by asking whether this phenomenon was symmetrical. For example, would a single circle with a line through its contour pop-out of a display containing a background of many circles without such lines (Figure 11.2.22A) as easily as a single circle does against a background of many circles with intersecting lines (Figure 11.2.22B)?

Treisman and Souther measured the time required to detect targets among distractors for several pairs of features, systematically varying which feature was the target and which was the distractor. In many cases, they found a marked asymmetry in the detection time. For example, when subjects had to detect a circle with a line through it against a background of simple circles, the target popped out, as evidenced by flat search functions (Figure 11.2.22C). When they had to detect a simple circle against a background of circles with lines through them, however, the search function increased steeply as a function of the number of distractors (Figure 11.2.22D), the pattern indicative of serial search. Similar asymmetries have been reported for other dimensions: A tilted line can be detected more quickly against a background of vertical lines than the reverse; a circle with a gap in it can be detected more quickly against complete circles than the reverse; an ellipse can be detected more quickly against circles than the reverse; and a curved line can be detected more quickly against a background of straight lines than the reverse (Treisman & Gormican, 1988; Treisman & Souther, 1985).

Treisman has interpreted such results as indicating that a target defined by the presence of a feature can be detected against a background of distractors defined by its absence more quickly than the reverse. To understand why this might be true, consider the case of the circle with a line through it (Figure 11.2.22A). Here the discriminative feature is the straight line segment. When a circle-with-line is the target, the retinotopic map for straight vertical lines contains neural activity at only one location, regardless of the number of distractors. This activity calls focused attention to that location, and the target can thus be detected as quickly among many distractors as among few. But when a circle-without-line is the target (Figure 11.2.22B), the feature map for straight vertical lines contains activity at all locations in which a display element is present. The wide distribution of activity in this feature map cannot lead to visual pop-out of the location in which there is no such line, and indeed it does not.

One important implication of this account is that there is no feature map defined by the absence of a line. If there were, then no search asymmetry would be found; the circle-without-line would pop out due to unique activity at the target’s location in the hypothetical no-straight-line map. Assuming that something like this explanation in terms of presence versus absence of features is correct, it can then be used to infer how elements are coded in terms of visual features.

11.2.6 Feature Integration Theory

Physiological evidence for the presence of retinotopic feature maps contains substantial support for the proposal that the visual system divides input information into many distinct subsystems that analyze different properties. But if this is true, how do all these properties get put back together into unified perceptual objects? This is a problem because we do not perceive unconnected attributes at different positions in the visual field (e.g., disconnected redness and verticalness); we perceive
their conjunction in unified perceptual objects (a red horizontal line). The process of conjoining different properties into visual objects is called binding. Binding is an important theoretical problem because without some mechanism to bind properties into objects properly, an observer would not be able to tell the difference between a display containing a red circle and a blue triangle and a display containing a blue circle and a red triangle. Because people seldom bind features in the wrong way—at least under normal conditions—there must be some mechanism that does the job of producing conscious perceptions of unified objects.

Anne Treisman recognized the importance of the binding problem and proposed a theory of focused visual attention to solve it. The theory is called feature integration theory because it concerns how unitary perceptual objects are constructed by attentional acts that conjoin features into objects (Treisman, 1988, 1993; Treisman & Gelade, 1980; Treisman & Gormican, 1988). Consistent with the physiological evidence that different retinotopic maps code different visual properties, Treisman began with the assumption that before focused attention, the visual system contains separate representations of stimulus features such as redness, greenness, verticality, horizontality, and so on, as illustrated in Figure 11.2.23. Each of these feature maps is retinotopically organized according to locations in space and is constructed independently of all others. Thus, when there are red horizontal lines in the visual field, activity occurs at the appropriate corresponding locations in both the “red” feature map and the “horizontal” feature map. But simultaneous activity in both maps is not sufficient, according to feature integration theory, for an observer to perceive them as red horizontal lines. These two separate features must be conjoined by some additional process to bind them together.

Treisman theorized that feature binding is accomplished by selectively attending to specific locations in the visual field, as illustrated in Figure 11.2.23. This is done by focusing an attentional “spotlight” on a region of a master location map, which results in the conjunction of all features that are registered at the corresponding location in all the various feature maps. This conjunction then constitutes the unified perceptual object we consciously experience.

To bind features correctly, the spotlight of attention must be narrowly focused on a limited region of the location map so that just the corresponding features are selected and bound together. Perceiving multiple objects in the field of view therefore requires that focused attention be moved sequentially from one location to another to construct the complex, multifaceted objects of conscious perceptual experience. This is the point at which visual processing changes from being parallel and nonspecific (distributed attention) to being serial and selective (focused attention).

Feature integration theory has made a number of interesting experimental predictions that have been confirmed in extensive empirical testing. Here we list three of the most important findings: conjunction search, texture segregation, and illusory conjunctions.

**Conjunction Search.** The first prediction of feature integration theory is that pop-out should occur in visual search for target objects that can be discriminated from distractor objects by an elementary feature but not for objects that require conjunctions of features (Treisman & Gelade, 1980). For example, a black vertical line
Figure 11.2.24 Examples of elementary and conjunction search displays. In part A, the target differs from the distractors only in orientation. In part B, it differs only in color. In part C, it differs by a conjunction of orientation and color: a black vertical target among white vertical and black horizontal distractors.

should pop out against a set of distractors consisting of just black horizontal lines because of its orientation (see Figure 11.2.24A) or among just white vertical ones because of its color (see Figure 11.2.24B), but it should not pop out against distractors consisting of both black horizontal lines and white vertical ones (see Figure 11.2.24C).

Treisman and Gelade (1980) tested this prediction about conjunction search using the visual search paradigm described in the previous section. Targets defined by the conjunction of two features (e.g., black and vertical in Figure 11.2.24C) could be distinguished from the distractors (white vertical and black horizontal lines) only by attending to both features at the same time. These conjunction trials produced the increasing response time pattern predicted by serial search using focal attention (Figure 11.2.20B). Targets defined by the disjunction of the same two features (vertical or black in Figures 11.2.24A and 11.2.24B) could be distinguished from distractors by a single feature. These disjunction trials showed the flat response time function predicted by parallel search using diffuse attention (Figure 11.2.20A). This pattern of results is just what feature integration theory predicts.

**Texture Segregation.** Another prediction of feature integration theory is that although effortless texture segregation should be possible for displays in which simple features are sufficient for the discrimination, it should not be possible for displays in which feature conjunctions define the different textures. For example, it was already known that effortless feature discrimination is possible when a texture boundary is defined by horizontal versus vertical lines (regardless of color), as in Figure 11.2.25A, or by black versus white lines (regardless of line orientation), as in Figure 11.2.25B. But what happens when the boundary is defined by feature conjunctions, such as black vertical and white horizontal elements on one side and white vertical and black horizontal elements on the other? This type of conjunction display, shown in Figure 11.2.25C, does not support effortless texture segregation, a finding that is consistent with the prediction of feature integration theory (Treisman & Gelade, 1980).

**Illusory Conjunctions.** Perhaps the most startling prediction of feature integration theory is that if attention is spread over a region including several different objects, the features may not be correctly conjoined. As a result, these “free-floating” features can form illusory conjunctions: perceptions of objects in which features are bound into objects in the wrong way. Treisman and Schmidt (1982) first tested this prediction by flashing a 200-ms display containing a red X, a blue S, and a green T between two black digits, as shown in Figure 11.2.26. It was followed by a masking display of randomly arranged colored letter fragments to eliminate any afterimage of the test display. Subjects were told first to report the two digits (to spread attention over the portion of display including the colored letters) and then to report any of the colored letter targets that they could perceive. Illusory conjunctions—such as reporting a blue X or a red T—occurred on 39% of the trials, compared with only 15% of the trials on which subjects erroneously reported a color or letter that was not present.
maps are activated by many objects. For example, a red horizontal line in a field of red vertical lines and green horizontal lines will activate many locations in the red, green, horizontal, and vertical maps. The red horizontal element can be detected only by the conjunction of both redness and horizontalness at the same location.

In feature integration theory, conjunctions are accomplished by moving the attentional spotlight to a specific location on the master attentional map. Once attention is focused on that location, all the features at corresponding locations in all the independent feature maps are properly bound into a coherent object. This object representation can then be checked for the presence of the target conjunction. If it is found, the search can be terminated. If it is not, attention must be moved to a different location to check for the presence of the target conjunction there. This process must be continued sequentially until either the target is found or all locations have been examined. Because conjunction targets require this serial search process, response time should increase approximately linearly with the number of distractors in the display, and numerous experimental results have shown that it does.

**Problems with Feature Integration Theory.** Feature integration theory has been tested many times in many ways by many investigators. For several years its predictions held up remarkably well, but systematic discrepancies eventually began to emerge. These problems have caused Treisman to make a number of modifications to her initial theory (Treisman, 1988, 1993).

One of the most important developments was the finding that conjunctions can sometimes be detected in parallel rather than requiring serial search. For example, Nakayama and Silverman (1986) reported that conjunctions of color and depth plane (e.g., a red target square in the near depth plane) produced strong pop-out against a background of distractors containing both features that were differently conjoined (e.g., green distractor squares in the near plane and red distractor squares in the far plane). In a similar experiment, they also found evidence of parallel search for conjunctions of color and direction of motion.

Such findings raise serious questions about Treisman’s claim that attention is necessary to conjoin features. There are several ways in which such results could be obtained within the processing architecture defined...
by feature integration theory. Later experiments by Treisman (1988) and by Wolfe, Cave, and Franzel (1989) supported the hypothesis that highly distinctive features allow selective access to their intersection. They found that when the features within the same pair of dimensions were dramatically different, conjunctions could be searched in parallel, whereas when they were similar, conjunctions were searched serially. Consider conjunctions of color and line length, for example. When differences in color and line length are small (e.g., pink versus beige and 3 mm versus 4 mm in length), conjunctions require serial search. When they are large (e.g., red versus green and 2 mm versus 5 mm), conjunctions pop out in parallel search.

Such findings have caused Treisman (1988; Treisman & Sato, 1990) and others (e.g., Cave & Wolfe, 1990) to revise the search mechanism of the original feature integration theory. The main change is to assume that highly distinctive features allow nontarget features to be actively inhibited.* For example, if the target is a green vertical line, inhibiting all locations containing red elements and all locations containing horizontal elements will leave the target uninhibited, regardless of the number of distractor items. Thus, it will pop out, as evidenced by fast search time functions.

Other findings that require an important modification in feature integration theory concern the level of the visual system at which features are defined. In Treisman’s early writings, feature maps were discussed as early image-based representations, much like the features of Marr’s retinotopically organized primal sketch (see Section 4.3.1). But such representations are not consistent with some findings in the visual search paradigm for at least two reasons. One is that subjects move their eyes rapidly around the array during search, and this would cause the locations of the elements to shift every time subjects made an eye movement. Keeping track of which items had been examined in a serial search would be extremely difficult. It therefore seems likely that the master location map (at least) is coded in terms of perceived locations in the environment rather than retinal locations.

*The same effect can also be produced by facilitation of target features rather than inhibition of nontarget features. This is actually the mechanism suggested by Wolfe et al. (1989) and Cave and Wolfe (1990). Unfortunately, there is no easy way to distinguish between these two alternatives empirically.

![Figure 11.2.27](Image) Pop-out of high-level features. The odd element pops out from distractors when they depict coherent 3-D objects that differ in spatial orientation (A) but not when they are similarly colored 2-D patterns (B). (After Enns & Rensink, 1990.)

Another problem is that recent research has shown that pop-out can occur for high-level, postconsoncy properties. For example, Enns and Rensink (1990) found pop-out among objects that appear to be 3-D prisms with differently colored (or shaded) surfaces (Figure 11.2.27A) but serial search for 2-D patterns that are matched with the prisms in terms of complexity (Figure 11.2.27B). If the feature maps were based on image-level properties, the prisms would be processed in much the same way as the 2-D patterns. Other recent findings also show that pop-out occurs after stereoscopic depth perception and visual completion have been achieved (He & Nakayama, 1992). Therefore, we conclude that the features involved in visual search can come from relatively late perceptual representations.

The original story of visual search in feature integration theory has obviously been complicated by these new findings. Nevertheless, feature integration theory has
provided the basic framework for understanding the integrative function of visual attention for many years and is still viable, albeit in a more complex form. Future research will undoubtedly require further modifications in its structure, but the insights it has provided have revolutionized the understanding of visual attention.

**Object Files.** Spatially focused attention appears to bind independently registered features into the representation of a unitary multifeatured object. In many experimental situations, this may be sufficient to characterize the brief perception of a stationary display. But in the real world, as well as in more complex experimental tasks, perception is dynamic and temporally extended.

As we discussed earlier in this chapter, eye movements are used to sample information in different locations via a sequence of fixations. An object whose gross features were perceived initially in the periphery may later be fixated, adding more detailed information about its specific shape, color, texture, and so on. Sometimes the new information that is perceived may substantially modify earlier information, although the perceiver almost always realizes that it is the same object. For example, I just saw a dark blob in my peripheral vision that I initially thought was my cat. But when I made a saccade to that object, it turned out to be a paper bag of roughly the same size and shape as a sitting cat. Such new information must be integrated with old to keep knowledge of the environment coherent and current.

Some objects in the environment change their visible properties over time. People, animals, and even inanimate objects move in a variety of complex ways, and such changes must be tracked by the visual system to maintain an accurate, up-to-date representation of the environment. If an object changes its location, for example, we see it not as a new object in its new location, but as the same one, even if some new features are visible after it has moved. How are we to understand the visual system’s ability to cope with these dynamic changes in visual information?

Kahneman and Treisman (1984) have suggested that the results of perceptual analyses are integrated into temporary representations of objects and events called **object files.** They view object files as being analogous to the case files maintained in a police station to keep track of incidents currently being monitored. Initial in-

**Figure 11.2.28** An object file. Object files are memory representations of objects that are used to keep track of basic perceptual information—in this case, about a cat named “Motor.” As this information changes, the object file is updated.

formation about a theft or shooting is used to open a new file. As further information and evidence become available, they are entered into the same file, updating its previous contents by either adding new information or modifying existing information.

An object file would likewise be opened for each new object that is perceived. It initially holds the conjunctions of features provided by focused attention to the object’s location, more or less as hypothesized by Treisman’s feature integration theory. At this point, the file might include information about the time, the object’s location, its gross shape, and its approximate color (see Figure 11.2.28). As further information accumulates about it, as when it is fixated following a saccade, the newly visible features are integrated into the existing object file. At some point the object may be identified as an instance of a known category (e.g., a cat) or even a known exemplar (e.g., my cat Motor). If the object moves, its new location and direction of motion would be entered into its existing object file. The object file thus maintains the coherence of an object’s representa-
tion over time, even though the contents of the file may change rather dramatically as a result of updating.

Although Kahneman and Treisman (1984) were not specific about precisely what kind of representations are contained within an object file, it is reasonable to think of them as dynamically changing structural descriptions of some sort, because they presumably support object categorization. Once the object has been identified, appropriate categorical information from memory can be added to the object file. Such information can provide expectations about what might occur in the future: what new parts might come into view as the viewpoint changes, what sort of motion it might exhibit, and so forth. Object files are thus a representational structure that can mediate between incoming low-level sensory information and internally generated high-level expectations. Once formed, object files may allow top-down unconscious monitoring of changes. That is, as long as the features of an object stay constant or change only in expected ways, the object file may be able to keep track of such changes without requiring the deployment of focused attention. If unexpected changes occur, however, attention may be directed to the object in question to update the new information.

What sort of evidence is there for the existence of object files and the role of attention in setting them up? One interesting experimental prediction is that once an object file has been established for an object, there will be some processing benefit from reperceiving it as having the same properties rather than new ones. Kahneman, Treisman, and Gibbs (1992) have examined this possibility using a priming technique called the reviewing paradigm. The structure of trials in the reviewing paradigm is illustrated in Figure 11.2.29 for an experiment involving motion. Subjects are initially presented with two simple geometrical objects in a static preview display. Letters are flashed briefly within these objects and then turned off. In the linking display, the two objects move continuously from their initial positions to their final positions in the target display. When the objects stop moving, a single letter is presented in one of the objects. The subjects' task is to name this target letter as quickly as possible, and response time is measured.

The crucial factor is the relationship between the target letter and the preview letters. On same-object trials the target letter is the one that was initially presented in the same object of the preview display. On different-object trials the target is the letter that was initially presented in the other object. And on no-match trials, the target is a new letter, one that does not match either of the initially presented letters. Kahneman, Treisman, and Gibbs (1992) reasoned that if there were separate object files for each object that tracked them over time, subjects would be faster at naming the target in the same-object condition than in the different-object condition because the target letter would be contained in the object file for the same object. This prediction was confirmed for both slow- and fast-moving objects. Subjects were 23 ms faster at naming the target in the same-object condition than in the different-object condition when the objects moved quickly and 41 ms faster when they moved slowly.

Comparing performance on the different-object trials with that on the no-match trials provides a measure of how specific the letter priming effect is. If merely presenting the target letter in the initial preview display...
produced priming, subjects should be faster at naming the target in the different-object trials. In fact, however, performance was almost exactly the same on different-object and no-match trials. Consistent with the predictions based on object files, the facilitating effect of previewing the target letter in the initial display is quite specific to the object in which it was presented.

Kahneman, Treisman, and Gibbs (1992) reported the results of many experiments similar to this. In each case, they found significant object-specific facilitation in naming the target letter. It occurred both when the preview objects did not move and when they did, when they were seen in apparent motion and in continuous motion, and when as many as four (but not eight) objects were contained in the preview display. The object-specific reviewing effect is therefore robust experimental evidence for the existence of temporary object representations that track attended objects over time and space. This is precisely the rationale for postulating object files in the first place.

11.2.7 The Physiology of Attention

We have surveyed a sampling of what is known about human visual attention from behavioral methods. We now turn to what is known about the physiology of attention: How is visual attention accomplished by neural mechanisms in the brain?

Until recently, the only source of evidence about the physiology of attention came from studies of brain lesions due to strokes, tumors, gunshot wounds, and other neurological traumas. Patients with functional deficits of attention from such injuries are still the main source of information about the gross location of attentional mechanisms in the brain, but recent technological advances have brought other methods to bear, such as brain imaging and single-cell recording. Functional brain imaging techniques (including PET, and fMRI; see Section 2.2.3) not only have enriched the study of neurological patients, but also have brought physiological studies to the domain of normal perceivers. And although single-cell recording techniques have been available for many years, they have only recently been used to study awake, behaving animals. Almost all previous studies were performed on anaesthetized animals that were essentially comatose and therefore not attending at all.

Unilateral Neglect. The most common neurological condition in which attention is impaired is called unilateral neglect (or hemineglect). Neglect is a name for a complex, clinically defined constellation of neurological symptoms. Whether it has a single underlying neurological cause is as yet unknown. However, the symptoms of neglect are commonly associated with brain injuries in certain locations, principally in the parietal lobe of the right hemisphere. It can also happen with parietal damage in the left hemisphere, although this condition is less severe and therefore less frequently reported. Acute neglect is often manifest only during the first few weeks following a stroke or other injury, when brain tissues swell in the vicinity of the damage. In the following weeks, it usually diminishes and may even disappear entirely as the swelling subsides. In some cases a permanent but less severe attentional disability remains, depending on the site and size of the irreversible portion of neurological damage.

The primary symptom of neglect is that patients systematically fail to notice objects on the side of the world opposite (contralateral) to their brain injury. For example, typical neglect patients with right parietal damage will orient toward the right side of the world by keeping their head turned noticeably to the right, fail to look at people on the left side of their hospital bed, and fail to eat food on the left side of their plate, even when they are hungry. A classical perceptual finding in patients with unilateral neglect is that when they are asked to draw a scene or copy a picture, they will draw only the right side of it or only the right sides of each object, as illustrated in Figure 11.2.30. This also happens when they draw something from memory, as exemplified by their drawings of a clock. A patient with right hemisphere damage either will draw only the right half of the clock or will try to squeeze all 12 numbers into the right side (see Figure 11.2.30). This clock-drawing task is a standard bedside diagnostic test for the presence of

---

3 Unilateral neglect is sometimes called inattention, but we will reserve that term for the nonpathological conditions in which someone is not attending...
neglect. Another standard clinical test is to ask patients to cross out all the lines drawn on a page. They do fine with the lines on the same side of the page as their lesion (the ipsilateral lines) but fail to cross out the ones on the opposite side (the contralateral lines). Interestingly, if they are asked to erase all the lines, rather than cross them out, they will begin on the right but eventually erase them all.

Such findings indicate that neglect patients have severely impaired perception of objects or parts of objects on the side contralateral to their lesion, as though they were blind in that region of the visual field. However, more refined tests show that the problem is not sensory in nature. Rather, neglect patients appear not to be able to attend to objects on the contralateral side or appear to be able to do so only with great difficulty.

One intriguing piece of evidence that the problem in unilateral neglect is not sensory is that the same problems occur in visual imagery of scenes that were well known to the patients before their brain injury (Bisiach & Luzzati, 1978). When two neglect patients with right parietal damage were asked to imagine the buildings in the Piazza del Duomo in Milan from a vantage point on one side of the plaza, they systematically failed to report the ones that would have been on the left side of their image. When they were then asked to imagine the same scene from the opposite side of the plaza, they named the buildings they had failed to report previously (because these now fell in the intact right half of their image) and failed to report the buildings they had successfully reported previously (because these now fell in the neglected left half of their image). Therefore, neglect appears to operate at a level high enough to affect imagery as well as perception and therefore cannot be solely a problem of low-level sensory input.

Further evidence of the nonsensory nature of neglect is that the neglected half can be defined relative to an object rather than to the field of view (Driver & Helligan, 1991). An object can be presented entirely within their “good” (ipsilateral) field, yet neglect patients do not perceive aspects of the object on the side of the object contralateral to their lesion. This finding is particularly interesting because it implicates object-based frames of reference in neglect. It appears that patients neglect not only the contralateral portion of the frame defined by the whole visual field, but also the contralateral portion of frames centered on individual objects. This is consistent with the so-called object-based theories of attention we discussed earlier in this chapter (see Section 11.2.3).

Posner and his colleagues have studied neglect patients using the attentional cuing paradigm (see Section 11.2.2) to find out how attention is impaired. Through careful analysis of the results from a series of experiments, they determined that neglect patients are able to move attention normally and engage attention to a new object but are impaired in disengaging attention from the currently attended object (Posner, Walker, Friedrich & Rafal, 1987). To explain the directionality of this syndrome, the disengage deficit must be specific to the direction of the required movement. That is, it must be difficult to disengage attention from objects on the uninjured side to detect objects on the impaired side. This appears to be the case. It explains, among other
things, why patients act as though they see only the lines on the good side of the page in the cross-out test yet appear to see all of them in the erasure test. The crossed-out lines are still present on the good side of the page, and so patients have great difficulty disengaging attention from them to cross out the lines on the bad side. But in the erasure test, the lines on the good side disappear from the page once they are erased. Eliminating them thus provides a kind of physical disengagement—since they are not there any more—allowing the patient eventually to see and erase all the lines on the page.

When acute neglect subsides, it sometimes leaves a permanent disability called extinction. In extinction, unlike full-blown neglect, a patient can see objects on either side of the visual field without difficulty. However, if patients are shown two objects at once, one on the good side and one on the bad side, they will report seeing only the one on the good side. This appears to be a less debilitating residue of neglect, but it can be explained by the same kind of disengagement problems.

Balint’s Syndrome. An even more severe neurological condition related to attentional deficits is Balint’s syndrome (Balint, 1909; Holmes & Horax, 1919; Rafal, in press). It results in what seems to be an almost complete inability to see anything except a single fixated visual object. Such patients take little interest in events occurring around them, staring instead at inconsequential objects for extended periods of time. The condition is often so debilitating that those who suffer from it are functionally blind. They must use conscious strategies such as closing their eyes to break fixation from one object so that they can look at another.

Patients with Balint’s syndrome typically suffer from four main symptoms:

1. Ocular apraxia: the inability to change fixation from one object to another, as though the gaze were stuck on the currently fixated object. This renders impossible any visual task requiring more than one fixation. Because almost all real-world activities require multiple fixations, Balint’s patients are severely impaired in everyday life.

2. Simultagnosia: the inability to perceive more than one object at a time during a single fixation. Even in a complex field of many objects, patients suffering from Balint’s syndrome perceive only the object they are currently fixating. This is true even when two objects occupy the same location, as when a patient can see a person’s face but cannot tell whether that person is wearing glasses.

3. Spatial disorientation: the inability to orient and localize objects correctly, including both their egocentric direction and their depth. This results in profound inaccuracy not only in perceiving the visible environment, but also in comprehending visual memories of places that were well known before the brain injury.

4. Optic ataxia: the inability to reach out and touch an object in space. This is perhaps not too surprising, given that patients are unable to localize objects correctly in space, but it is one of the most profoundly debilitating symptoms of the condition.

The first two symptoms may result from an underlying inability to disengage attention from the fixated object (Farah, 1990; Posner et al., 1987). If attentional shifts precede and direct eye movements, as some researchers believe (e.g., Rizzolatti, Riggio, Dascalu, & Umiltà, 1987), then a complete inability to disengage attention from its current object would produce ocular apraxia. And if attention is required for conscious perception of objects, as Mack and Rock (1998) suggest, restricting attention narrowly to the fixated object would also explain simultagnosia, the inability to perceive more than one object at once.

An alternative hypothesis to explain the nature of Balint’s syndrome is that these problems can be understood as resulting from the loss of the master spatial map within Treisman’s feature integration framework. If attention binds features into individuated object tokens by selecting locations in this map, and if Balint’s patients cannot differentiate locations because there is no master map, it follows that they should form many illusory conjunctions in the one object that they do see. The Balint’s syndrome patient whom Friedman-Hill, Robertson, and Treisman (1995) studied was found to do so quite noticeably.

Balint’s syndrome is important from a theoretical standpoint because it provides powerful neurological evidence for the existence of an object-based attentional system. Its most significant perceptual symptom is that only one object is perceived at any time, even when two or more objects spatially overlap. For example, when
Balint’s patients view a comb and a spoon in an overlapping cross configuration, they see only the comb or only the spoon, never both (Rafal, in press). This is hard to understand in terms of space-based theories based on metaphors of spotlights or zoom lenses but easy to understand in terms of object-based theories, for it is exactly what would be predicted. Another intriguing and relevant fact is that although one Balint’s syndrome patient was completely unable to determine which of two separate lines was longer (because he could see only one of them at any one time), he could always report whether a rectangle was square or elongated (Holmes & Horax, 1919). That is, he had no trouble attending to the square holistically as a unitary figure even though it was composed of four line segments, but he was unable to attend to just two line segments when they were presented as separate objects.

Another finding that supports the object-based attention hypothesis concerns the performance of two Balint’s syndrome patients on a task requiring them to determine whether a display contained circles of just one color (either all red or all green) or two colors (half red circles and half green circles). If the circles were separate and intermixed with randomly positioned, disconnected black lines, their performance was essentially at chance. However, if each line in the two-color display connected a red circle and a green circle, they performed much better. The connecting lines apparently unified the two circles into a single dumbbell-shaped object whose color could then be processed as a single object (Humphreys & Riddoch, 1992). This finding supports the powerful role of element connectedness in object formation (Palmer & Rock, 1994a), as discussed in Section 6.1.2.

Unlike unilateral neglect, Balint’s syndrome typically involves bilateral lesions in the parietal and/or nearby occipital cortex. This fact suggests the possibility that Balint’s syndrome may be the result of a bilateral deficit in disengaging attention (e.g., Farah, 1990). This idea fits with the hypothesis that unilateral neglect results from an inability to disengage attention from objects in the good side to objects in the bad side: If the brain damage were bilateral (as it is in Balint’s syndrome), the patient would be unable to disengage attention to move to either side, thus freezing it on the currently attended object. Other indications suggest that Balint’s syndrome may be different from a bilateral version of neglect, however. The locations of the bilateral lesions that produce Balint’s syndrome are slightly different from the unilateral lesions of neglect, being more likely to occur in the parieto-occipital junction than in the temporo-parietal junction, where neglect lesions are typically found (Rafal, 1996).

**Brain Imaging Studies.** The physiological basis of attention has also been studied in normal perceivers. Using sophisticated neuroimaging techniques, researchers have been able to determine which areas of the brain are active when people engage in activities involving attention. In general, they have corroborated conclusions previously drawn from studies of focal brain damage: The parietal lobe, particularly the right parietal lobe, is especially important in moving attention from one object to another.

One imaging study provides evidence of why damage to the right parietal lobe is so much more likely to produce unilateral neglect than damage to the left. PET images were used to examine what parts of the brain were active when normal subjects performed a continuous version of Posner’s cuing task among several locations in the right or the left half of the visual field (Corbetta, Meizlin, Dobmeyer, Shulman, & Petersen, 1991). As the subjects fixated their gaze overtly on a single box in the center of the display, they covertly attended to targets that were presented in a sequence of different boxes in the right visual field for one PET scan and in the left visual field for another. As Figure 11.2.31 shows, when attention was directed to the left visual field, the right parietal lobe showed greatly increased activity, as would be predicted from studies of neglect. When attention was directed to the right visual field, however, both the left and right parietal lobes were highly active. Thus, the right parietal lobe appears to be involved in attending to both halves of the visual field, whereas the left parietal lobe is involved only in attending to the right visual field. This helps to explain why the symptoms of neglect are so much more pronounced following damage to the right parietal lobe. Right parietal damage destroys the ability to attend to the left visual field but leaves the right visual field intact (owing to the unharmed left parietal lobe), whereas left parietal damage allows attention to both the right and left visual
Figure 11.2.31 Brain activity while attending to the right versus left visual field. PET scans show activity in the right parietal lobe when subjects were attending to the left visual field but activity in both left and right parietal lobes when they were attending to the right visual field. (From Posner & Raichle, 1994.)

fields (owing to the unharmed right parietal lobe), as indicated in Figure 11.2.32.

Electrophysiological Studies. At a much smaller scale of analysis, single-cell physiologists have recently begun to study attentional effects by applying their techniques to awake behaving animals. The results have begun to provide a glimpse of the effects of attention on the responses of individual neurons. Only a little is known yet, but there are already clear indications that selective attention to a given spatial location or object can restrict the functional size of a cell’s receptive field and increase its resolution for responding to specific features.

The reader may recall from Section 1.3.3 that the size of cells’ receptive fields increases by a factor of 100 or more as one traces the flow of information from LGN to primary visual cortex (area V1) to inferotemporal cortex (area IT) where pattern recognition is thought to occur. A cell that is sensitive to a monkey’s face in IT, for example.
ample, may have a receptive field as large as 25°, firing rapidly when a monkey face appears anywhere within that area. But how is the monkey to know where this face is located? And how is it to be combined with other features that are coded by other cells in other visual areas? This is the binding problem of feature conjunction that we considered in Section 11.2.6. Moran and Desimone (1985) have reported physiological evidence that the response of an individual neuron is gated by the locus of the animal's attention—that is, that the cell's functional receptive field is restricted to the region to which the animal is attending.

Moran and Desimone found that the responses of cells in areas V4 and IT were strongly affected by the locus of the monkey's attention. The firing rate to the same stimulus was only one-third as great when the monkeys did not attend to it as when they did. Because the sensory conditions in the experiment were carefully controlled to be identical in these two cases, the difference in the cell's response can be due only to the locus of the animal's attention. Therefore, cells in V4 and IT are strongly affected by attention, almost as though their effective receptive field were restricted to the attended location. This suggests a suppressive function of attention in which it blocks processing of non-attended information. Cells in areas V1 and V2 were not affected by attention, however. Such results suggest that V4 may be the first visual area in the occipitotemporal pathway where attention has this suppressive effect.

Later studies have shown that attention can also have an enhancing effect when the amount of attention is manipulated (Spitzer, Desimone, & Moran, 1988). The amount of attention that was paid to the stimuli was manipulated by having one block of trials require an "easy" discrimination between dissimilar stimuli (e.g., red versus aqua and orange versus blue) and another block require a "difficult" discrimination between similar stimuli (e.g., red versus orange and blue versus aqua). Recordings from V4 cells showed responses that were larger and more tightly tuned to the same stimulus in the difficult task than in the easy one. The same results were found for similar versus dissimilar orientations as for colors.

These two studies demonstrate some of the ways in which attention can influence the firing of individual cells in the visual system, but they do not identify the source of these effects. Presumably, there is an attentional subsystem somewhere in the brain that sends signals to these regions that modulate their responses depending on how attention is deployed, in both direction and amount. Where might this subsystem be located?

Desimone, Wesinger, Thomas, and Schneider (1990) reasoned that the lateral pulvinar nucleus (Figure 11.2.12) might be a good bet because it projects directly to visual cortex and because humans with lesions in this area have difficulty engaging attention (Rafal & Posner, 1987). They tested their conjecture by deactivating this region chemically in monkeys to see what effect it would have on selective attention. After injecting a drug (muscimol) into the lateral pulvinar that stopped all neural activity there, they found that the monkey was impaired in attending selectively to one stimulus in the presence of a distractor, but not to the same stimulus without a distractor. When the effects of the drug dissipated, the monkey was again able to selectively attend to the relevant stimulus in the presence of another. These results suggest that deactivating the lateral pulvinar nucleus eliminates the gating effect of attention that Moran and Desimone (1985) had found on the response of IT cells. This is just what they predicted would happen if this region were controlling selective attention to different regions of space. Similar results were found when portions of the superior colliculus were deactivated, suggesting that this region is also involved in controlling selective attention (Desimone et al., 1990).

11.2.8 Attention and Eye Movements

We began this chapter by considering the nature of overt selection through eye movements that change the optical content of the retinal image. We then discussed what is known about covert selection through atten-

Chapter 11

568
The similarities between eye movements and attention are so compelling that we briefly considered the idea that attention is like an internal eye that can be moved around to sample the visual field much as the eye can be moved around to sample the visual world. Although this metaphor has limited utility for reasons we discussed, it underscores the intuitively close relation between the two. This relation is strengthened by the fact that shifts of attention normally occur more quickly than saccadic eye movements (Posner, 1980). Benefits of valid cues and costs of invalid ones in Posner’s cuing paradigm begin to occur in less than 100 ms, long before a saccadic eye movement can be initiated. It also appears that the same kinds of stimulus conditions that lead to involuntary eye movements to an event also lead to involuntary capture of attention via so-called pull cues (Jonides, 1981).

Despite these connections, insight into the precise relation between attention and eye movements is a recent development. The dominant view, called the pre-motor theory, suggests that overt orienting through eye movements and covert orienting through attentional movements are controlled by closely related mechanisms and that eye movements normally follow attentional movements. This relation can be broken when you attend to something you are not looking at directly, but that is an unusual occurrence outside of attention experiments. Such covert orienting of just the attentional system would occur when the eye movement that normally would follow an attentional shift is inhibited, leaving only the attentional shift (Rizzolatti, Riggio, Dascola, & Umlita, 1987). According to this theory, attention normally drives the saccadic eye movement system, directing it to appropriate objects given the stimulus conditions and the task at hand. The hypothesis is that visual attention is first either summoned to a salient event (by the equivalent of a pull cue) or internally directed to an important location (by something akin to an internally generated push cue), and then the eyes follow attention.

This pre-motor theory of the relation between attention and eye movements makes several predictions. Behaviorally, it suggests that when an observer executes a saccade to a particular location, attention will enhance perception of events occurring there before the eye movement is actually executed. It also suggests that if an eye movement is made to a particular location, attention will necessarily precede that eye movement and therefore cannot be sent in some other direction. Unequivocal support for such hypotheses proved difficult to find for many years (e.g., Klein & Pontefract, 1994), but recent studies have found good evidence for both of these predictions using a cuing paradigm similar to Posner’s (see Section 11.2.2).

In one experiment, Hoffman and Subramaniam (1995) investigated whether preparing to make a saccade to a cued location facilitated perception in that location before the eye movement was actually executed. As shown in Figure 11.2.33, subjects were first shown a computer-generated display containing four empty rectangles. Immediately after they pressed a key, an arrow was presented that indicated to which box an eye movement was to be made. Then, after a randomly determined delay of 500–2000 ms, a brief tone signaled the subject to execute the cued saccade. After this cue, it took subjects at least 200 ms to actually begin the saccade. To test whether preparing for this eye movement enhanced perception in the cued location, letters were presented in the boxes at 0, 50, or 100 ms after the tone cue, well before any eye movement could be made. The subject’s task was to discriminate which target letter had been presented (T or L) in one of the boxes when E’s and F’s were presented in the other boxes as distractors. The crucial question is whether subjects were better at discriminating which target was presented if it appeared in the box toward which the eye movement was programmed than if it appeared in the other boxes.

The results showed a substantial advantage for the location for which the eye movement was intended. Discrimination performance was about 90% correct when
the target’s location matched the intended saccade’s location but about 70% in the other three locations. On control trials in which no arrow was presented and no saccade was made, performance was about 80% correct, intermediate between the two cuing conditions. Thus, Hoffman and Subramaniam’s results indicate that both benefits and costs are associated with preparing for an eye movement, much like the benefits and costs associated with purely covert attentional shifts as measured by Posner’s cuing paradigm. This finding thus supports the pre-motor theory.

In a second experiment, Hoffman and Subramaniam (1995) examined whether attention could be directed toward a location different from the one to which an eye movement was planned. Using the same series of displays and a slightly modified procedure, they decoupled eye movements from the expected target location. In this study, subjects always had to make a saccade to a given box throughout a block of trials, for example, to the box on the left. The arrow now provided information about the likely location of the target letter, as in Posner’s original attentional cuing paradigm: On 75% of the trials it was valid, and on 25% of the trials it was invalid. On control trials, when no eye movement was required, normal costs and benefits of attentional cuing were found. When eye movements were required, however, performance was always better at the location toward which an eye movement was planned than at other locations and was unaffected by where attention should have been directed on the basis of the cue. This finding shows that attention is locked into the planning of saccades and cannot be redirected to a different location even if that is the optimal strategy. Similar conclusions hold for making slow pursuit eye movements (Kowler & Zingale, 1985).

In the physiological domain, the pre-motor theory predicts close correlations between the centers that control eye movements and those that control attention. Indeed, this is the case. The superior colliculus, for example, is known to be involved in controlling movements of both attention and the eye. Two other structures thought to play a major role in high-level motor control of the eye—the frontal eye fields in the frontal cortex and the posterior parietal cortex—also appear to be involved in controlling spatial attention (Desimone, Wessinger, Thomas, & Schneider, 1990). More specifically, the pre-motor theory predicts that damage to the brain centers involved in making eye movements should lead to corresponding deficits in shifting attention covertly. In support of this prediction, patients whose superior colliculus is damaged not only have been found to have difficulty making voluntary saccades, but are also impaired in covert attentional orienting as measured by Posner’s cuing paradigm (Rafal, Posner, Friedman, Inhoff, & Bernstein, 1988). At a cellular level, evidence is also accumulating that the superior colliculus is involved in covert shifts of attention (Desimone, Wessinger, Thomas, & Schneider, 1990; Gattas & Desimone, 1992).

Although the subject is far from well understood, it is becoming increasingly clear that there is an intimate relationship between eye movements and attention as mechanisms of visual selection. Somewhat counter-intuitively, the relation appears not to be that attention follows eye movements but that it precedes eye movements as an integral step in preparing to move the eyes. On reflection, this relation makes sense. One reason is that attentional shifts are purely neural operations and can therefore be executed more rapidly than ocularmotor responses that require actually rotating an eye in its socket. Another reason is that because attention is a central (rather than peripheral) process, it is better suited to controlling where the eye should be directed. Attention can therefore gain access to information about what locations contain important information and should thus be selected for higher resolution processing. Because attentional movements appear to drive eye movements rather than the other way around, attention may legitimately be viewed as the primary mechanism of visual selection, with eye movements playing an important but supporting role.
Suggestions for Further Reading

Eye Movements


Attention


Mack, A., & Rock, I. (1998). Inattentional blindness. Cambridge, MA: MIT Press. This research monograph presents the important series of experiments by the authors on what is perceived under conditions of inattention.


