CHAPTER 3

Human Perception and Information Processing

This chapter deals with human perception and the different ways in which graphics and images are seen and interpreted. The early approach to the study of perception focused on the vision system and its capabilities. Later approaches looked at cognitive issues and recognition. We discuss each approach in turn and provide details. Significant parts of this chapter, including many of the figures, are based on the work of Christopher G. Healey (http://www.csc.ncsu.edu/faculty/healey/PP/index.html) [150], who has kindly granted permission for their reuse in this book.

3.1 What Is Perception?

We know that humans perceive data, but we are not as sure of how we perceive. We know that visualizations present data that is then perceived, but how are these visualizations perceived? How do we know that our visual representations are not interpreted differently by different viewers? How can we be sure that the data we present is understood? We study perception to better control the presentation of data, and eventually to harness human perception.

There are many definitions and theories of perception. Most define perception as the process of recognizing (being aware of), organizing (gathering and storing), and interpreting (binding to knowledge) sensory information. Perception deals with the human senses that generate signals from the environment through sight, hearing, touch, smell and taste. Vision and audition

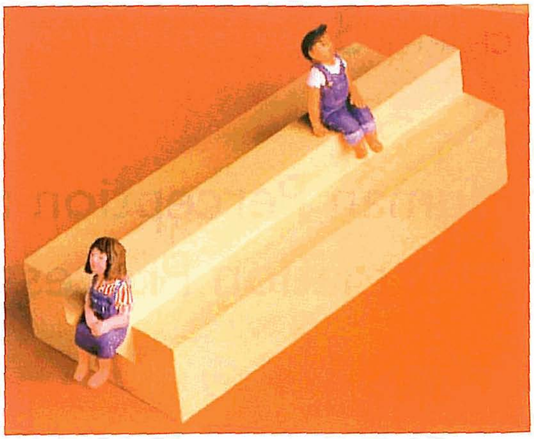


Figure 3.1. Two seated figures, making sense at a higher, more abstract level, but still disturbing. On closer inspection, these seats are not realizable. (Image courtesy N. Yoshigahara.)

are the most well understood. Simply put, perception is the process by which we interpret the world around us, forming a mental representation of the environment. This representation is not isomorphic to the world, but it's subject to many correspondence differences and errors. The brain makes assumptions about the world to overcome the inherent ambiguity in all sensory data, and in response to the task at hand.

Visual representations of objects are often misinterpreted, either because they do not match our perceptual system, or they were intended to be misinterpreted. Illusions are a primary source of such misinterpretations. Figures 3.1 and 3.2 highlight our inability to notice visual problems except on more detailed perusal. The drawings are those of physically nonrealizable objects.

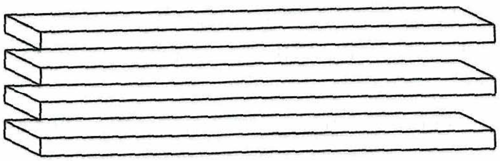


Figure 3.2. Four $\neq$ three. As in Figure **3.1**, this object would have a problem being built (there are four boards on the left and three on the right).

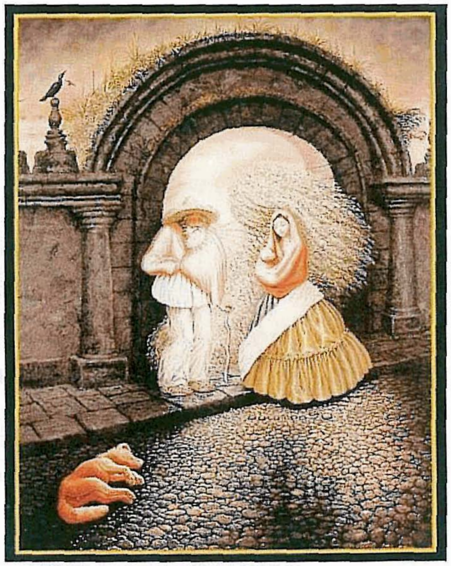


Figure 3.3. A more coinplex illusion: there are two people drawn as part of the face.

Sometimes the ambiguity presented is easily seen, but more difficult to explain. Sometimes it is not even perceived.

Figure 3.3 highlights that on first glance an image may represent a primary object, one that is perceived more obviously than the secondary others that may require more effort or time. There are many such illusions, and these are easy to construct. In effect, the artist puts together a primary image out of secondary images. There may even be tertiary ones. Tools have been developed to support such imagery. For example, Rob Silvers uses a computational technique to form an image composed of a mosaic of smaller given iinages (see Figure 3.4 and Figure 3.5, which contains a detailed view).

Our visual machinery also performs similar computations, but perhaps not as we would expect. Figure 3.6 highlight that our vision system is, foremost, not static, and secondly, often not under our full control. It is clear that there appear to be black squares being generated between the white spaces in Figure 3.6(a) and black circles in Figure 3.6(b). Why? If we forcibly stare at an intersection of the spaces between the black squares, we can actually stop the "spots" from appearing. This is akin to our stopping breathing. When we visualize data, we need to make sure that no such interferences are present that would impede the understanding of what we are trying to convey in the visualizations.

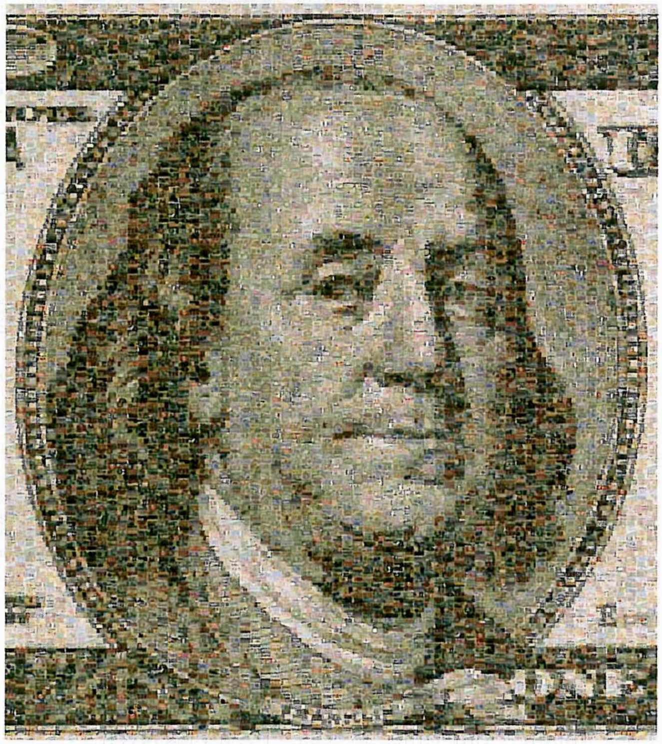
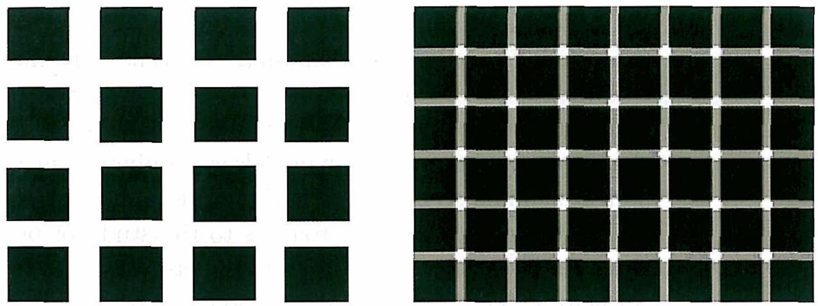


Figure 3.4. Photomosaic of Benjamin Franklin using images of international paper money or bank notes. (Photomosaic® by Robert Silvers, http://www.photomosaic.com.)

Figure 3.5. Close-up view of the eye in Figure 3.4. (Photomosaic® by Robert Silvers, http://www.photomosaic.com.)



(a) (b)

Figure 3.6. The Hermann grid illusion: (a) illusionary black squares appear over the complete image as you gaze at it; (b) similar to (a) but even more dynamic and engaging.

Similarly, Figure 3.7(a) and (b) highlight that there is more to our visual system than meets the eye (pun intended). In both of these images, we seem to have machinery forcing the interpretation of the objects we see in a specific manner. The study of perception is to identify not just this machinery, but the whole process of perception, from sensation to knowledge. What is causing the lines not to appear perfectly straight, or the triangle to stand out? More generally can we explain the causes of these and other illusions we see? These are the important questions we need to answer in order to be able to generate synthetic images that will represent data unambiguously and not pop out an artifact.

(a) (b)

Figure 3.7. (a) The Hering illusion: red lines are straight. (Use a straight edge to verify.) (b) The Kanizsa illusion: a triangle seems to pop out of the image even though no such triangle is drawn.

These illusions are due to our perceptual system's structure, and the assumptions it makes about an image or scene. The interpretations are due to a variety of reasons and are the result of how the process works. To understand this process and identify its structure, we first need to measure what we see and then develop models explaining the measured results. These models should also help explain the illusions.

There are two main approaches to the study of perception. One deals with measures, and the other with models. Both are linked. Measurements can help in the development of a model, and in turn, a model should help predict future outcomes, which can then be measured to validate the model. We can measure low-level sensory perception (which line is longer) or higher-level perception (can you recognize the bird in this scene?). Each requires a different set of tools and approaches. This approach, however, still does not explain why we see these differences, or why we recognize objects. That requires a model of the process.

Not paying attention to perception will lead to problems in visualization. For example, Figure 3.6 clearly shows how visual patterns can impact a display. We need to understand, at least rudimentarily, what aspects of visualization cannot be violated. Some of these involve color (perceived differently by individuals) and three-dimensional perception (forced interpretations by inherent perceptual assumptions, such as where a light source is typically placed). We will see several more examples later in this chapter.

3.2 Physiology

The main sensory component of vision involves the gathering and recording of light scattered from objects in the surrounding scene, and the forming of a two-dimensional function on the photoreceptors [187, 302]. Photoreceptors are generally very small sensory devices that respond in the presence of photons that make up light waves.

3.2.1 Visible Spectrum

Visible light, the light waves that are capable of being perceived by human eyes, actually represents a very small section on the electromagnetic spectrum (see Figure 3.8). This sub-spectrum ranges from about 380nm (nanometers) near ultraviolet, up through to about 700nm towards the infrared. This range is very much dependent on the individual and generally shrinks in size after the age of twenty [233]. Color blindness and total blind-

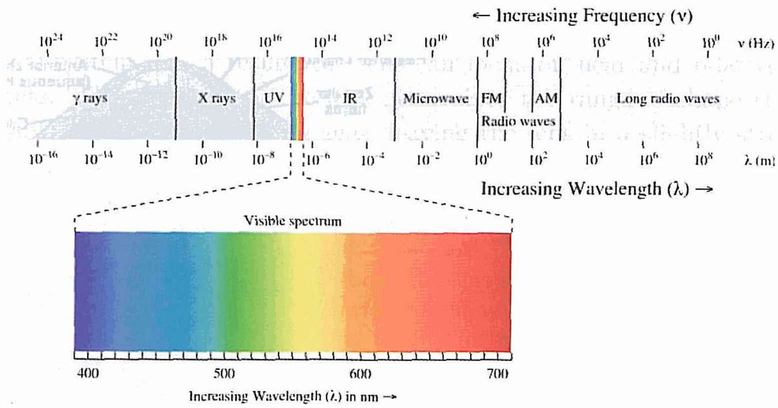


Figure **3.8.** The electromagnetic spectrum with an expanded visible light spectrum [233]. (Image courtesy Wikimedia Commons.)

ness in humans are the result of an individual not responding to certain wavelengths.

Beyond just the consideration of light is the importance of physical object properties. It is through the visual system that information concerning the external objects in the surrounding environment is captured. This exchange of information between the environment and the observer is presented to the eyes as variations of wavelengths. These variations are a result of object properties that include object geometry, scene illumination, reflectance properties, and sensor photoreceptor characteristics.

3.2.2 Anatomy of the Visual System

The human eye is a marvelous organ, yet its construction is quite simple. Figure 3.9 shows a horizontal cross-section of the right eye, viewed from above. This diagram provides names to most of the fundamental macrostructures that provide humans with the ability to see their surrounding environment. The major parts that directly involve the path taken by light waves include the cornea, iris, pupil, lens, and retina. Overall, the eye is a fluid-filled sphere of light-sensitive cells with one section open to the outside via a basic lens system, and connected to the head and brain by six motion-control muscles and one optic nerve.

Lens System and Muscles. First, the six muscles are generally considered as motion controllers, providing the ability to look at objects in the scene. The action of looking at specific areas in the environment involves orienting the

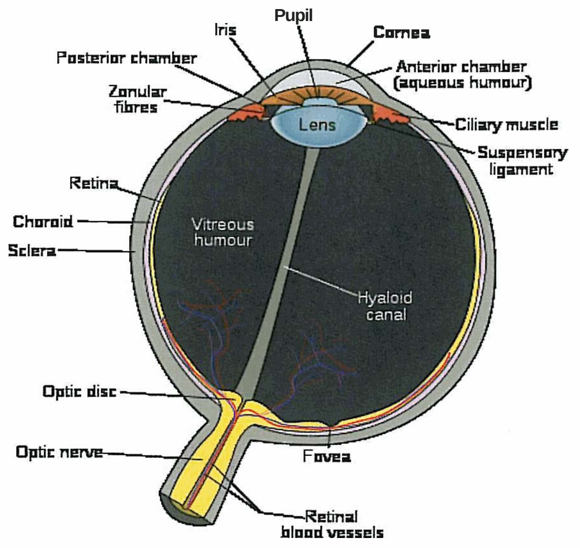


Figure 3.9. Horizontal cross-section of the human eye, viewed from above. (Image courtesy Wikimedia Commons.)

eye's optical system to the regions of interest through muscle contractions and relaxations. Also, the muscles tend to maintain the eye-level with the horizon when the head is not perfectly vertical. These muscles also play another important role in the stabilization of images. Continually making minor adjustments, eyes are never at rest, although we do not perceive these actions visually. In an engineered system, such motions are usually considered as imperfections, yet they have been found to improve the performance of the human visual system [269].

The optical system of the eye is similar in characteristic to a double-lens camera system. The first component is the cornea, the exterior cover to the front of the eye. Acting as a protective mechanism against physical damage to the internal structure, it also serves as one lens focusing the light from the surrounding scene onto the main lens [128]. From the cornea, light passes through the *pupil*, a circular hole in the iris, similar in function to an aperture stop on a photographic camera [233]. The iris is a colored annulus containing radial muscles for changing the size of the pupil opening. Thus, the pupil determines how much light will enter the rest of the internal chamber of the eye. The third major component is the lens, whose crystalline structure is similar to onion skin. Surrounded by the ***ciliary*** body, a set of muscles, the lens can be stretched and compressed, changing the thickness and curvature of the lens and consequently adjusting the focal length of the optical system. As a result, the lens can focus on near and relatively far objects. The elasticity of the lens determines the range of shape changes possible, which is lost as one ages, leaving the lens in a slightly stretched state [128]. Once the light has passed through this lens system, the final light rays are projected onto the photoreceptive layer, called the *retina*. The process is not as precise as camera optics, however. As Overington states:

> An important point to note about the lens system is that is has very little facility built-in for correction of many of the aberrations which are normally corrected in good quality instrumental optical systems. This inevitably means that the image produced is far from perfect. Yet the apparent image perceived appears very sharp, whilst quite phenomenally fine subtleties in the image can be observed. [269, p. 7]

The **Retina** The *retina* of the human eye contains the photoreceptors responsible for the visual perception of our external world. It consists of two types of photosensitive cells: *rods* and *cones* (see Figure 3.10) [128, 233]. These two types of cells respond differently to light stimulation. Rods are primarily responsible for intensity perception, and cones for color perception. Rods are typically ten times more sensitive to light than cones. There is a small region at the center of the visual axis known as the *fovea* that subtends 1 to 2 degrees of visual angle. The structure of the retina is roughly radially symmetric around the fovea. The fovea contains only cones, and linearly,

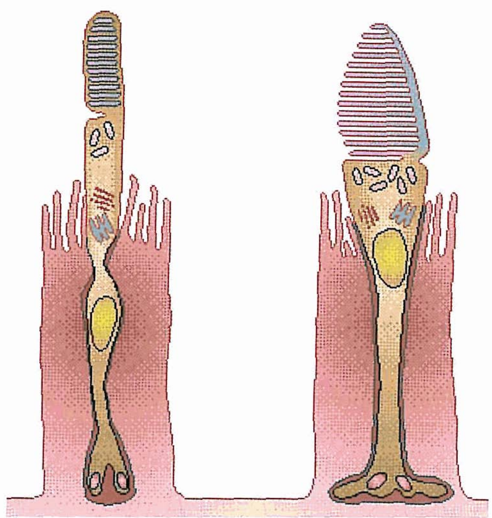


Figure 3.10. Human rod (left) and cone (right). (Image © Colour4Free.)

there are about 147,000 cones per milliineter [128]. The fovea is the region of sharpest vision. Because the human eye contains a limited number of rods and cones (about 120 million rods and G million cones), it can only manage a certain amount of visual information over a given time frame. Additionally, the information transferred from these two types of cells is not equivalent.

Another interesting fact is that the optic nerve only contains about one million fibers; thus the eye must perform a significant amount of visual processing before transmitting information to the brain. What makes the retina an unusual layer for light stimulation is the orientation of the photoreceptive cells. The whole layer of cells that maltes up the retina is actually backwards; the light rays must pass through the output neurons and optic nerve fibers first, before reaching the photosensitive cells, which are also facing away from the light source. The reason suggested for this arrangement in all vertebrates is that "eyes are actually part of the brain and represent an outgrowth from it," and that "the cells of the retina are formed during development from the same cells that generate the central nervous system" ([128] p. 18).

The eye contains separate systems for encoding spatial properties (e.g., size, location, and orientation), and object properties (e.g., color, shape, and texture). These spatial and object properties are important features that have been successfully used by researchers in psychology for simple tasks such as target detection, boundary detection, and counting. These properties have also been used extensively by researchers in visualization to represent high-dimensional data collections [384].

Rods. Rods are the most sensitive type of photoreceptor cells available in the retina; consequently, they are associated with *scotopic* vision, night vision, operating in clusters for increased sensitivity in very low light conditions. As these cells are thought to be achromatic we tend to see objects at night in shades of gray. Rods do operate, however, within the visible spectrum between approximately 400 and 700 nm [233]. It has been noted that during daylight levels of illumination, rods become *hyperpolarized*, or completely saturated, and thus do not contribute to vision [128].

Cones. On the other hand, cones provide photopic vision, i.e., are responsible for day vision. Also, they perform with a high degree of acuity, since they generally operate individually. There are three types of cones in the human eye: S (short), M (medium), and L (long) wavelengths [128]. These three types (see Figure 3.11) have been associated with color combinations using R (red), G (green), and B (blue). The long wavelength cones exhibit a spectrum peak at 560 nm, the medium wavelength cones peak at 530 nm, and the short wavelength cones peak at around 420 nm. However, it must

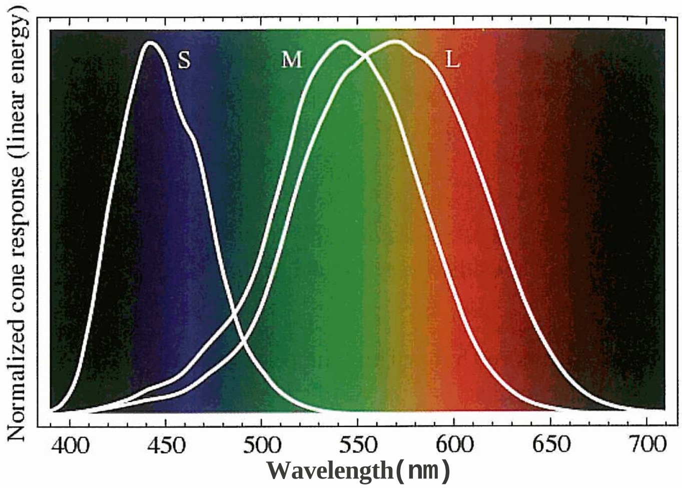


Figure 3.1 l. The retina layer contains the three types of cones (short, medium, and long) (Image courtesy Wikimedia Commons.)

be noted that there are considerably fewer short cones, compared to the number of medium and long wavelength cones [269]. In spite of this, humans can visually perceive all the colors within the standard visible spectrum. Unlike rods, cones are not sensitive over a large fixed wavelength range, but rather over a small moving-window-based range. Cones tend to adapt to the average wavelength where there is sensitivity above and below their peaks, and a shift in their response curve occurs when the average background wavelength changes [128].

Blind Spot. Given that humans have two types of photoreceptors with three types of cones, how are these cells distributed on the retina? First, there is an overall distribution of all cells across the retina, with the highest concentration occurring at the center of our visual field in the fovea and reducing in coverage towards the edges [128]. Where the optic nerve meets the retina, a blind spot occurs, due to the lack of photoreceptive cells. Second, there is a striking separation between the locations of rods and cones. The fovea consists of only cone receptors, and no rods, for highly detailed and exact vision [233]. Surrounding the fovea are three concentric bands: *parafovea* with an outer ring of 2.5-mm diameter, *perifovea* with an outer ring of 5.5-mm diameter, and the peripheral *retina*, covering approximately 97.25% of the total retinal surface and consisting largely of rods [233]. Each of these areas is marked by a dramatic reduction in cones, and it is significant to

Figure 3.12. Blind spot discovery by identifying disappearance of target.

note here that within the parafovea there already are significantly more rods than cones.

The identification (more really verification) of one's blind spot can be done simply with this Optic Disk experiment (see Figure 3.12). Close your right eye and look directly at the number **3.** You should see the yellow spot in your peripheral vision. Now, slowly move toward, or away from the screen or paper image. At some point, the yellow spot will disappear, as its sensory reflection hits the blind spot.

There are some very interesting outcomes resulting from the physiology of human eyes. First, the photoreceptive cells are packed into the retina parallel to each other, and are not directed toward the pupil [128]. Thus, the eye obtains its best stimulation from light entering straight on through the pupil.

Next, the rods and cones are packed in a hexagonal structure for optimized coverage. Such a packing scheme, in conjunction with an initially blurred image, resulting from cell sampling, has been demonstrated to provide near-optimal information transfers [269]. Another fascinating fact about the retina concerns the sampling rate of the photoreceptive cells. Through the effects of temporal smoothing, where receptors only respond every few milliseconds, humans perceive flickering lights up to a certain frequency, beyond which the eye only registers a constant light source [128].

It has been said that the United States Air Force tested pilots' ability to respond to changes in light by flashing a picture of an aircraft on a screen in a dark room for 1/220th of a second. According to these anecdotal reports, pilots were consistently able to detect the afterimage of the flash, and were also able to identify the aircraft type.

Finally, it has been shown that the human eye responds to ratios of intensities and not absolute values [128]. These ratios play an important part in adaptation and contrast sensitivity and the eye will adapt to changes in wavelength ranges.

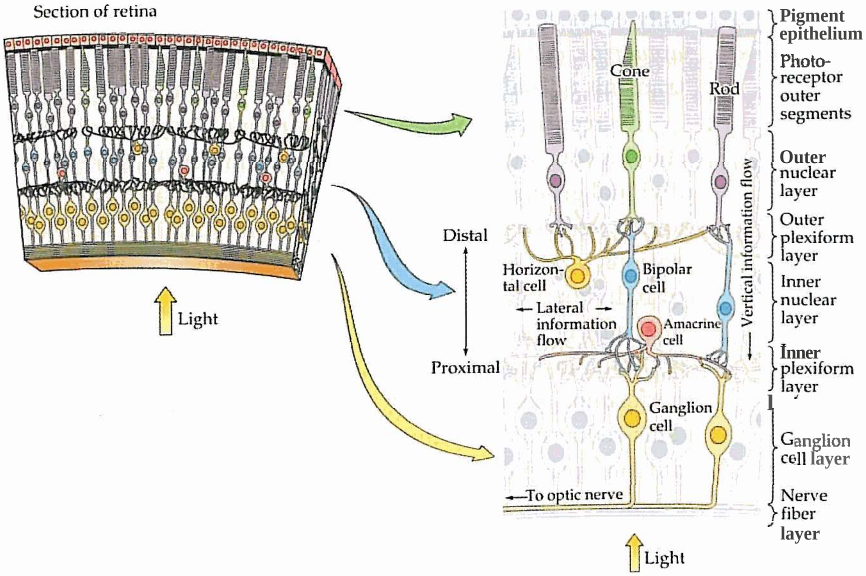


Figure 3.13. A representation of a retinal cross-section. (Image © The Brain from Top to Bottom.)

3.2.3 Visual Processing

Signal processing in humans is performed by neurons, the elementary biological components that make up the nervous system. This system operates on sequences of frequency-modulated pulses sent between two neurons in communication. Through chemical actions, each neuron stimulates other neurons — possibly hundreds to thousands of other nervous system cells — causing information to travel.

Retinal Processing. The retina of the eye is actually a complex layer of many neurons and photoreceptive cells, as depicted in Figure **3.13.** This illustration has the photoreceptors pointing up; thus, the front of the eye is pointing down, so that light first hits the bottom layer and progresses through the various layers, until it stimulates the rods and cones. The relatively large black bulbs represent the nucleus of each neuron.

There are four neuron layers within the retina that perform initial image processing on the stimulations resulting from the individual photoreceptors, the cones and rods. Figure **3.13** is a highly stylized diagram of the human retina, showing the four layers plus the top layer of receptors; again, the light enters from the bottom. These four layers are composed of individual types of neuron cells, based on their connectivity properties: horizontal cells only connect sets of receptors, *bipolar cells* connect receptors to other layers, *amacrine cells* join numerous bipolar and ganglion cells, and *ganglion cells* transmit retinal stimulation from the eye to the brain is the *optic nerve* [233].

As mentioned previously, the retina develops directly from brain cells; thus the obvious ability for preprocessing of the stimulus image. Like the individual groups of photoreceptive cells, there are also various types of bipolar and ganglion cells that have very distinct properties dealing with the combinations of rods and cones [269]. Some cones within the fovea are connected to individual ganglia via a single bipolar link. Rods on the outer periphery of the retina are grouped together and joined with bipolar cells, where several bipolar groups output to a single ganglion. Hence, the retina is already performing some kinds of image compression, and possibly segmentation. This reduction of retinal stimulation is required, as there are only about a million optic nerve fibers relaying image information to the brain, which is a hundred times less than the total number of rods and cones [4]. There is also other valuable information formed during this compression. Individual rods and cones by themselves do not provide much information, due to the limitations of the optic nerve. Furthermore, individual cones only respond to fixed wavelength ranges; thus one cell cannot provide color information. Consequently, it is through the combinations of photoreceptor stimuli that intensity and color descriptions can be obtained, which is believed to happen at a very early stage in visual processing [128].

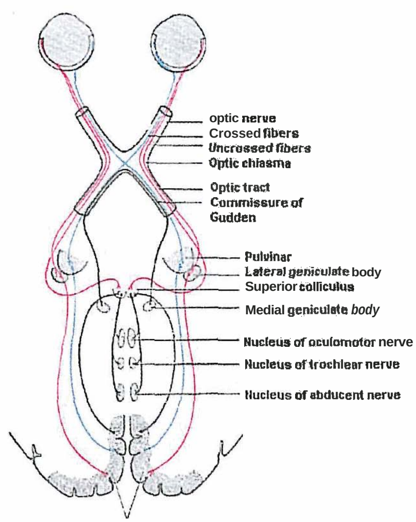


Figure 3.14. The anatomy of the visual system. (Image courtesy Wikimedia Commons.)

The Brain. The brain is the center of all bodily functions and is composed of the majority of neurons found in the human nervous system. The overall structure of the brain is divided into two hemispheres, left and right, with the addition of a few smaller structures located under these hemispheres. Of importance is the fact these hemispheres have relative functional regions, one of which is designed for processing visual stimulation [233]. Before the optic nerves from each eye reach the inner regions of the brain they partially cross at the optic chiasma—half the fibers from each eye cross to the opposite side of the corresponding brain region (see Figure 3.14). Thus, each hemisphere receives visual information from both eyes, possibly to help with the perception of depth. As there is so much visual processing performed at both the eyes and within the brain, these linked organs form an integral visual system [4].

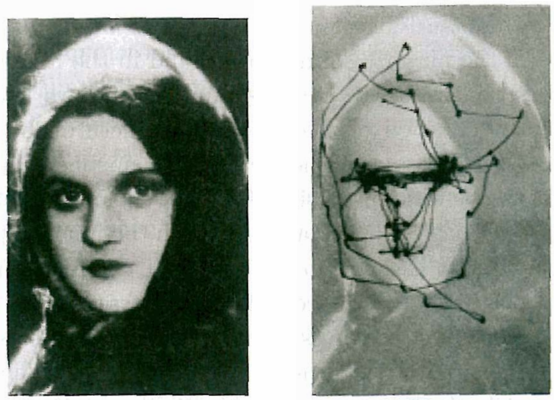
3.2.4 Eye Movement

Perhaps the most critical aspect of perception is the importance of eye movement in our understanding of scenes, and therefore images. It explains, for example, the illusionary black dots in the earlier figures [123,309,334]. There are a variety of eye movements performed for scene interpretation.

Smooth pursuit movements. These are just as their name implies. The eyes move smoothly instead of in jumps. They are called pursuit because this type of eye movement is made when the eyes follow an object. For example, to make a pursuit movement, look at your forefinger at arms' length and then move your arm left and right while fixating on your fingertip. Such movements are also called conjugate eye *movements* or coordinated eye movements. The angles from the normal to the face are equal (left and right as well as up and down).

Vergence eye movements. These result from nonconjugate movement and yield different angles to the face normal. Moving a finger closer to the face and staring at it will force the eyes inward, resulting in vergence movement. Defocusing to merge depths in illusions is another example.

Saccadic eye movements. These result from multiple targets of interest (not necessarily conscious). The eye moves as much as 1000 degrees per second, bringing the gaze on those targets within 25 msec. It holds its position once on target. Selected targets are determined in the frontal part of the cerebral cortex. The selection is discriminatory, dependent on a variety of parameters, and somewhat random.



(a) (b)

Figure 3.15. (a) The face used to study eye tracking. (b) The results of the tracking of the gaze.

Saccadic masking. Saccadic masking or suppression occurs during two states between saccadic views. The gap produced is ignored (some say blocked). A continuous flow of information is interpreted, one that makes sense. The higher level visual system filters out the blurred images acquired by the low level one, and only the two saccadic stop views are seen.

Marketing research has helped identify how to set up advertisements to force the visual focus on objects of interest. For example, when looking at the face in Figure 3.15(a), we find that the eye moves as in Figure 3.15(b). Note how the concentration of vertices highlights the targets to which the eye is attracted. The same tracking for the left image is shown on the right one in Figure 3.16. Note the role of the boundaries and the key focal points of faces.

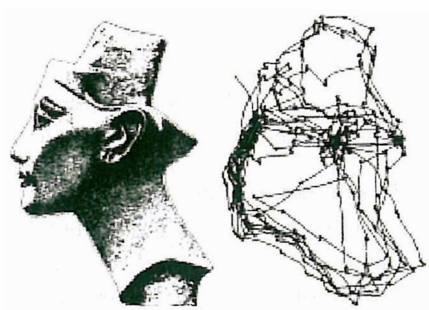


Figure 3.16. The right image shows the path followed by the eye in looking at the image on the left. Note the targets, which can easily be identified from the concentration of vertices of the path, and note the role of the boundary.

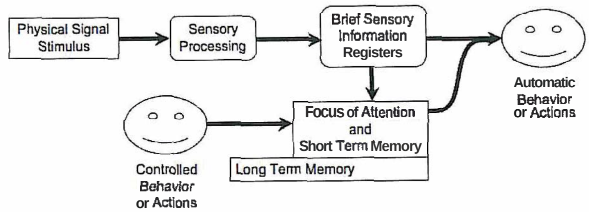


Figure 3.17. Classic model of the flow of sensory data for cognition (based on [72]).

3.3 Perceptual Processing

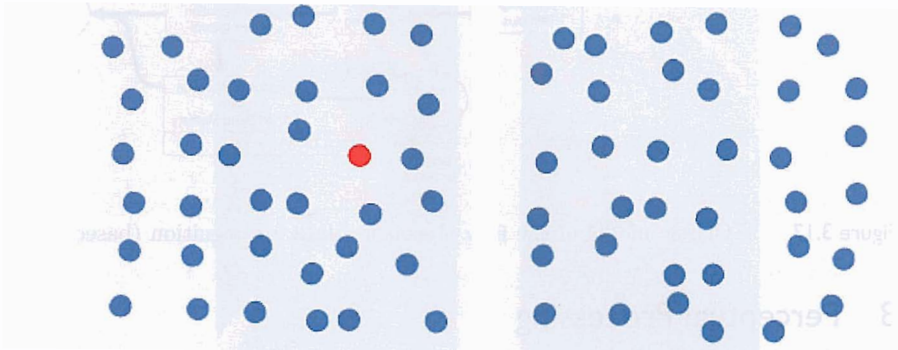
We use the classic model of information processing [72] for understanding the flow of sensory information, from the low level pre-attentive to the higher cognitive levels (Figure 3.17). This model highlights that memory is involved in post processing, but this is known to be only partially correct. Perception can be intrinsic and uncontrolled (preattentive) or controlled (attentive).

Automatic or preattentive perception is fast and is performed in parallel, often within 250ms. Some effects pop out and are the result of preconscious visual processes. Attentive processes (or perception) transform these early vision effects into structured objects. Attentive perception is slower and uses short-term memory. It is selective and often represents aggregates of what is in the scene. Low-level attributes are rapidly perceived and then converted to higher-level structured ones for performing various tasks, such as finding a door in an emergency. We first focus on low-level attributes, then turn to higher level ones, and finally put it all together with memory models.

3.3.1 Preattentive Processing

For many years vision researchers have been investigating how the human visual system analyzes images. An important initial result was the discovery of a limited set of visual properties that are detected very rapidly and accurately by the low-level visual system. These properties were initially called preattentive, since their detection seemed to precede focused attention. We now know that attention plays a critical role in what we see, even at this early stage of vision. The term preattentive continues to be used, however, since it conveys an intuitive notion of the speed and ease with which these properties are identified.

Typically, tasks that can be performed on large multielement displays in less than 200 to 250 inilliseconds (msec) are considered preattentive. Eye movements take at least 200 msec to initiate, and random locations of the



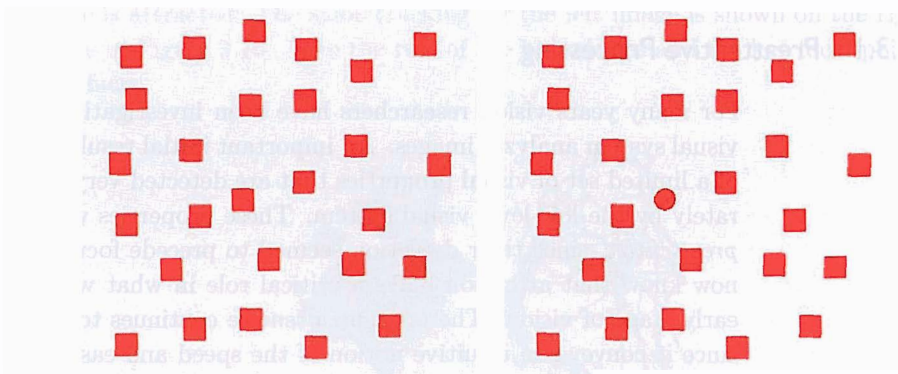
(a) Target is present in a sea of blue circle distractors.

(b) Target is absent.

Figure 3.18. An example of searching for a target red circle based on a difference in hue.

elements in the display ensure that attention cannot be prefocused on any particular location; yet viewers report that these tasks can be completed with very little effort. This suggests that certain information in the display is processed in parallel by the low-level visual system.

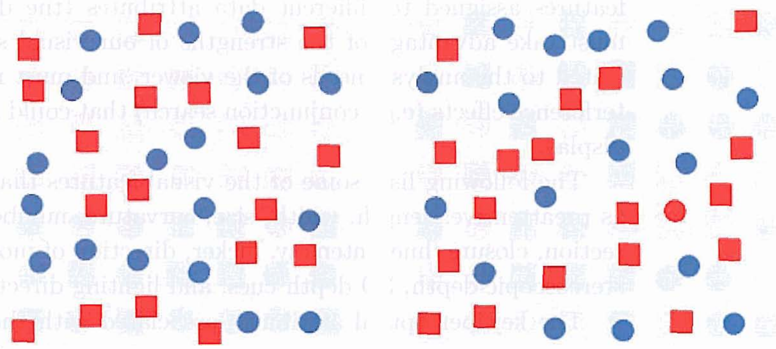
A simple example of a preattentive task is the detection of a red circle in a group of blue circles (Figure 3.18). The target object has a visual property "red" that the blue distractor objects do not (all nontarget objects are



(a) Target is absent in a sea of red square distractors.

(b) Target is present.

Figure 3.19. An example of searching for a target red circle based on a difference in curvature.



(a) Target is absent in a sea of red square and blue circle distractors.

(b) Target is present.

Figure **3.20.** An example of a conjunction search for a target red circle.

considered distractors). A viewer can tell at a glance whether the target is present or absent. In Figure 3.18 the visual system identifies the target through a difference in hue, specifically, a red target in a sea of blue distractors. Hue is not the only visual feature that is preattentive. In Figure 3.19 the target is again a red circle, while the distractors are red squares. As before, a viewer can rapidly and accurately determine whether the target is present or absent. Here, the visual system identifies the target through a difference in curvature (or form).

A unique visual property in the target (e.g., a red hue in in Figure 3.19(a) or a curved form in Figure 3.19(b)) allows it to "pop out" of a display. A target made up of a combination of nonunique features (a *conjunction* target) normally cannot be detected preattentively. Figure 3.20 shows an example of conjunction search. The red circle target is made up of two features: red and circular. One of these features is present in each of the distractor objects (red squares and blue circles). This means the visual system has no unique visual property to search for when trying to locate the target. If a viewer searches for red items, the visual system always returns true, because there are red squares in each display. Similarly, a search for circular items always sees blue circles. Numerous studies have shown that this target cannot be detected preattentively. Viewers must perform a time-consuming serial search through the displays to confirm its presence or absence.

If the low-level visual system can be harnessed during visualization, it can be used to draw attention to areas of potential interest in a display. This cannot be accomplished in an ad-hoc fashion, however. The visual features assigned to different data attributes (the data-feature mapping) must take advantage of the strengths of our visual system, must be well-suited to the analysis needs of the viewer, and must not produce visual interference effects (e.g., conjunction search) that could mask information in a display.

The following lists some of the visual features that have been identified as preattentive: length, width, size, curvature, number, terminators, intersection, closure, hue, intensity, flicker, direction of motion, binocular luster, stereoscopic depth, 3D depth cues, and lighting direction.

The key perceptual attributes associated with the above include luminance and brightness, color, texture, and shape. Luminance is the measured amount of light coining from some place. Brightness is the perceived amount of light coming from a source. Perceived brightness is a nonlinear function of the amount of light emitted by the source, typically a power function $S = a^i$, where S = sensation and i = intensity. Note that these look very different on a screen versus on paper. Texture is the characteristic appearance of an area or surface. Whereas texture applies to multiple sensory objects (the texture of a music, the texture of a fabric), shape is strictly a geometric attribute.

Experiments in psychology have used these features to perform the following preattentive visual tasks:

Target detection. Users rapidly and accurately detect the presence or absence of a "target" element with a unique visual feature within a field of distractor elements (Figures 3.18, 3.19, and 3.20);

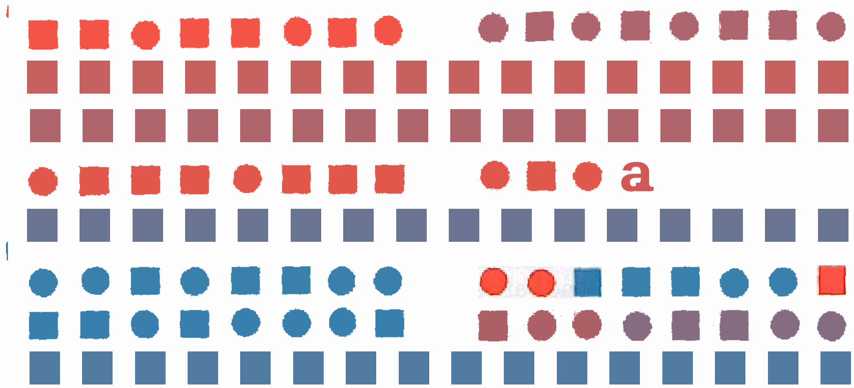
Boundary detection. Users rapidly and accurately detect a texture boundary between two groups of elements, where all of the elements in each group have a common visual property;

Region tracking. Users track one or more elements with a unique visual feature as they move in time and space; and

Counting and estimation. Users count or estimate the number of elements with a unique visual feature.

3.3.2 Theories of Preattentive Processing

A number of theories have been proposed to explain how preattentive processing occurs within the visual system. We describe four well-known models: feature integration theory, *texton* theory, similarity theory, and guided search theory. We also discuss briefly the phenomenon of postattentive *vi-*



(a) A boundary defined by a unique feature hue (red circles and red squares on the top, blue circles and blue squares on the bottom) is preattentively classified as horizontal.

(b) A boundary defined by a conjunction of features (red circles and blue squares on the left, blue circles and red squares on the right) cannot be preattentively classified as vertical.

Figure 3.21. All example of a boundary detection, from Treisman's experiments.

sion, which shows that prior exposure to a scene does not help a viewer answer questions about the content of the scene.

Feature Integration Theory. Anne Treisman was one of the original researchers to document the area of preattentive processing. She provided important insights into this phenomenon by studying two important problems. First, she tried to determine which visual properties are detected preattentively [359,360,362]. She called these properties *"preattentive features"* [361]. Second, she formulated a hypothesis about how the human visual system performs preattentive processing [358].

Treisman ran experiments using target and boundary detection to classify preattentive features. For target detection, subjects had to determine whether a target element was present or absent in a field of background distractor elements (Figures *3.18* amd *3.20)*. Boundary detection involved placing a group of target elements with a unique visual feature within a set of distractors to see if the boundary could be preattentively detected (Figure *3.21)*.

Treisman and other researchers measured preattentive task performance in two different ways: by response time and by accuracy. In the response time model viewers are asked to complete the task (e.g., target detection) as quickly as possible while still maintaining a high level of accuracy. The number of distractors in a scene is repeatedly increased. If task completion time is relatively constant and below some chosen threshold, independent of the number of distractors, the task is said to be preattentive. If the task were not preattentive, viewers would need to search serially through each display to confirm a target's presence or absence. Increasing the number of elements in the display would therefore produce a corresponding increase in the time required to report on the target.

In the accuracy model, the display is shown for a small, fixed exposure duration, then removed from the screen. Again, the number of distractors in the scene varies (i.e., increases) across trials. If viewers can coinplete the task accurately, regardless of the number of distractors, the feature used to define the target is assumed to be preattentive. A common exposure duration threshold is 200 to 250 msec, since this allows subjects only "one look" at the scene. The human visual system cannot decide to change where the eye is looking within this time frame.

Treisman and others have used their experiments to compile a list of visual features that are detected preattentively, as mentioned above. It is important to note that some of these features are asymmetric. For example, a sloped line in a sea of vertical lines can be detected preattentively. However, a vertical line in a sea of sloped lines cannot be detected preattentively. Another important consideration is the effect of different types of background distractors on the target feature. These factors must often be addressed when trying to design display techniques that rely on preattentive processing.

To explain the phenomenon of preattentive processing, Treisman proposed a model of low-level human vision made up of a set of feature maps and a master map of locations. Each feature map registers activity in response to a specific visual feature. Treisman suggested a manageable number of feature maps, including one for each of the opponent color primaries (green, red, yellow, and blue), as well as separate maps for orientation, shape, texture, and other preattentive features.

When the human visual system first sees an image, all the features are encoded in parallel into their respective maps. A viewer can access a particular map to check for activity, and perhaps to determine the amount of activity. However, the individual feature maps give no information about location, spatial arrangement, or relationships to activity in other maps.

This framework provides a general hypothesis that explains how preattentive processing occurs. If the target has a unique feature, one can simply access the given feature map to see if any activity is occurring. Feature maps are encoded in parallel, so feature detection is almost instantaneous. A conjunction target cannot be detected by accessing an individual feature map. Activity there may be caused by the target, or by distractors that share the

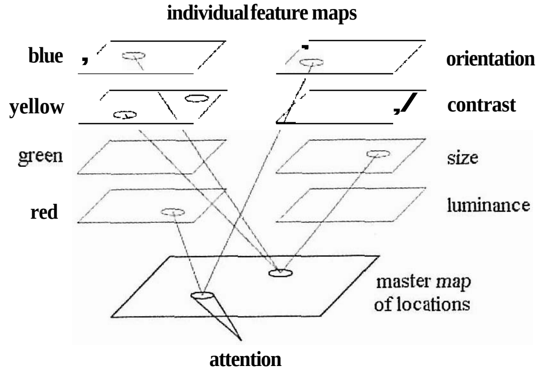


Figure 3.22. Treisman's feature integration model for early vision; individual maps can be accessed to detect feature activity; focused attention acts through a serial scan of the master map of locations

given preattentive feature. To locate the target, one must search serially through the master map of locations, looking for an object with the correct combination of features. This use of focused attention requires a relatively large amount of time and effort.

In later work, Treisman has expanded her strict dichotomy of features being detected as being either in parallel or in serial [359, 362]. She now believes that *parallel* and *serial* represent two ends of a spectrum. "More" and "less" are also encoded on this spectrum, not just "present" and "absent." The amount of differentiation between the target and the distractors for a given feature will affect search time. For example, a long vertical line can be detected immediately among a group of short vertical lines. As the length of the target shrinks, the search time increases, because the target is harder to distinguish from its distractors. At some point, the target line becomes shorter than the distractors. If the length of the target continues to decrease, search time decreases, because the degree of similarity between the target and the distractors is now decreasing.

Treisman has also extended feature integration to explain certain cases where conjunction search is preattentive. In particular, conjunction search tasks involving motion, depth, color, and orientation have been shown to be preattentive by Nakayama and Silverman [263], Driver et al. [86], and Wolfe et al. [399]. Treisman hypothesizes that a significant target-nontarget feature difference would allow individual feature maps to ignore nontarget information contained in the master map. For example, consider a search for a green horizontal bar within a set of red horizontal bars and green vertical bars. This should result in a conjunction search, since horizontal and green occur within each of the distractors. In spite of this, Wolfe et al. [399] showed that search times are independent of display size. If color constituted a significant feature difference, the red color map could inhibit information about red horizontal bars. Thus, the search reduces to finding a green horizontal bar in a sea of green vertical bars, which can be done preattentively.

Texton Theory. Bela Julesz was also instrumental in expanding our understanding of what we "see" in an image. Julesz's initial investigations focused on statistical analysis of texture patterns [180–183, 185]. His goal was to determine whether variations in a particular order statistic were seen (or not seen) by the low-level visual system. Examples of variations in order statistics include contrast (a variation in a texture's first-order statistic), orientation and regularity (a variation of the second-order statistic), and curvature (a variation of the third-order statistic). Unfortunately, Julsz's results were inconclusive. First-order variations were detected preattentively. In addition, some (but not all) second-order variations were also preattentive, as were an even smaller set of third-order variations.

Based on these findings, Julsz modified his theory of how preattentive processing occurs. He suggested that the early visual system detects a group of features called textons [179, 184, 185]. Textons can be classified into three general categories:

- elongated blobs (e.g., line segments, rectangles, ellipses) with specific properties such as hue, orientation, and width;
- terminators (ends of line segments);
- crossings of line segments.

Julesz believed that only a difference in textons or in their density can be detected preattentively. No positional information about neighboring textons is available without focused attention. Like Treisman, Julesz suggested that preattentive processing occurs in parallel and focused attention occurs in serial. Figure 3.23 provides an example of textons that appear different in isolation, but have the same size, number of terminators, and join points. This shows that even when each appear very different in isolation, it may be difficult, if not impossible, to differentiate any pattern when in a texture or grid.

Julesz used texture segregation, the task of locating groups of similar objects and the boundaries that separate them, to demonstrate his theory

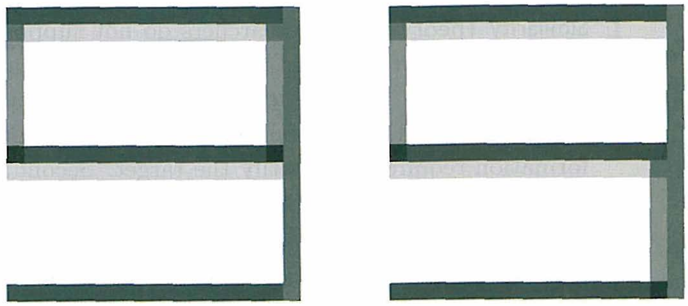


Figure **3.23.** Two simple textons, easily differentiable.

(other researchers, including Treisman, also used this type of task, for example, identifying the orientation of the boundary between groups of common elements in Figure 3.21). Figure 3.24 shows an example of an image that supports the texton hypothesis. Although the two objects look very different in isolation, they are actually the same texton. Both are blobs with the same height and width. Both are made up of the same set of line segments, and each has two terminators. When both are oriented randomly in an image, one cannot preattentively detect the texture boundary between the target group and the background distractors.

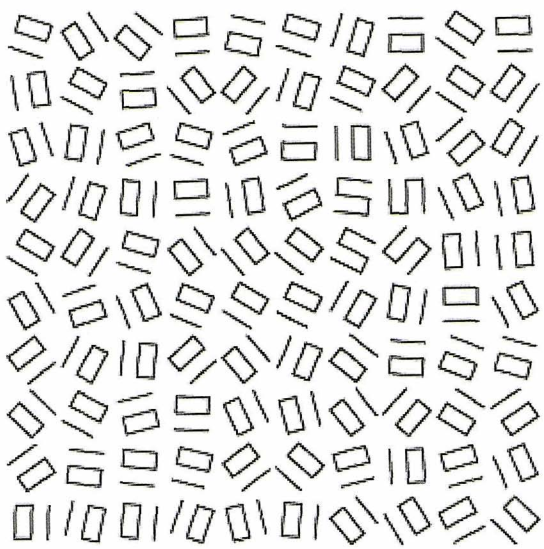


Figure **3.24.** A target group of b-textons is difficult to detect in a background of a-textons when a random rotation is applied.

Similarity Theory. Some researchers do not support the dichotomy of serial and parallel search modes. Initial work in this area was done by Quinlan and Humphreys [281]. They investigated conjunction searches by focusing on two factors. First, search time may depend on the number of items of information required to identify the target. Second, search tiine may depend on how easily a target can be distinguished from its distractors, regardless of the presence of unique preattentive features. Treisman addressed this second factor in her later work [359]. Quinlan and Humphreys found that Treisman's feature integration theory was unable to explain the results they obtained from their experiments. Duncan and Humphreys developed their own explanation of preattentive processing. Their model assumes that search ability varies continuously, depending on both the type of task and the display conditions [88,89,258]. Search tiine is based on two criteria: T-N similarity and N-N similarity. T-N *similarity* is the amount of similarity between the targets and nontargets. N-N *similarity* is the amount of similarity within the nontargets themselves. These two factors affect search time as follows:

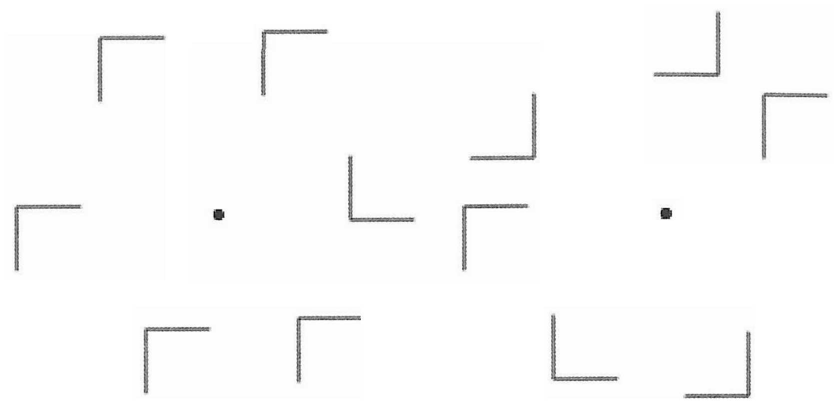
1. As T-N similarity increases, search efficiency decreases and search time increases.

2. As N-N similarity decreases, search efficiency decreases and search time increases.

3. T-N similarity and N-N similarity are related (see Figure 3.25); decreasing N-N similarity has little effect if T-N similarity is low; increasing T-N similarity has little effect if N-N similarity is high.

Treisman's feature integration theory has difficulty explaining the results of Figure 3.25. In both cases, the distractors seem to use exactly the same features as the target, namely oriented, connected lines of a fixed length. Yet experimental results show displays similar to Figure 3.25 on the left produce an average search time increase of 4.5 msec per additional distractor, while displays similar to Figure 3.25 on the right produce an average search time increase of 54.5 msec per additional distractor.

In order to explain the above and other search phenomena, Duncan and Humphreys proposed a three-step theory of visual selection. The visual field is segmented into structural units. Individual structural units share some common property (e.g., spatial proximity, hue, shape, motion). Each structural unit may again be segmented into smaller units. This produces a hierarchical representation of the visual field. Within the hierarchy, each



(a) High N-N (nontarget-nontarget) similarity allows easy detection of target L.

(b) Low N-N similarity increases the difficulty of detecting the target L.

Figure **3.25.** Example of N-N similarity affecting search efficiency for a target shaped like the letter L.

structural unit is described by a set of properties (e.g., spatial location, hue, texture, size). This segmentation process occurs in parallel.

Because access to visual short-term inemory is limited, Duncan and Humphreys assume that there exists a limited resource that is allocated among structural units. Because vision is being directed to search for particular information, a template of the information being sought is available. Each structural unit is coinpared to this template. The better the match, the more resources are allocated to the given structural unit, relative to other units with a poorer match.

Because units are grouped in a hierarchy, a poor match between the template and a structural unit allows efficient rejection of other units that are strongly grouped to the rejected unit. Structural units with a relatively large number of resources have the highest probability of access to the visual short-term memory. Thus, structural units that most closely match the template of information being sought are presented to the visual short-term memory first. Search speed is a function of the speed of resource allocation and the amount of competition for access to the visual short-term memory.

Given these three steps, we can see how T-N and N-N similarity affect search efficiency. Increased T-N similarity ineans that more structural units match the template, so competition for visual short-term memory access increases. Decreased N-N similarity ineans that we cannot efficiently reject large numbers of strongly grouped structural units, so resource allocation time and search time increase.

Guided Search Theory. More recently, Jeremy Wolfe has suggested a visual search theory that he calls "guzded search" [397, 399, 403]. He hypothesized that an activation map based on both bottom-up and top-down information is constructed during visual search. Attention is drawn to peaks in the activation map that represent areas in the image with the largest combination of bottom-up and top-down influence.

Like Treisman, Wolfe believes that early vision divides an image into individual feature maps (see Figure 3.26). In his theory, there is one map for each feature type (e.g., one map for color, one map for orientation, and so on). Within each map, a feature is filtered into multiple categories. For example, in the color map there might be independent representations for red, green, blue, and yellow. Wolfe had already found evidence to suggest that orientation is categorized into steep, shallow, right, and left [401]. The relationship between values within a feature map is different than the relationship between values from different maps (the relationship between "red" and "blue" is different than the relationship between "blue" and "shallow").

Bottom-up activation follows feature categorization. It measures how different an element is from its neighbors. Differences for each relevant feature map are computed and combined (e.g., how different are the elements in terms of color, how different are they in terms of orientation?) The "metrics" used to measure differences in each feature map are still being investigated.

Top-down actzvation is a user-driven attempt to find items with a specific property or set of properties. For example, visual search for a blue element would generate a top-down request that activates "blue" locations. Previous work suggests subjects must specify requests in terms of the categories provided by each feature map [398, 401]. Thus, subjects could search for "steep" or "shallow" elements, but not for elements rotated by a specific angle. Obviously, subjects should pick the category that best differentiates the target from its distractors. Finding the "best" category is often non-intuitive, however. Wolfe suggests this might explain cases where subjects' performance for a task improves over time.

The activation map is a combination of bottom-up and top-down activation. The weights assigned to these two values are task dependent. A conjunction search would place priority on top-down information, since bottom-up results are, in essence, useless. A search for a target with a unique feature would assign a high weight to bottom-up activation. Hills in the activation map mark regions that generated a relatively large amount of bottom-up or top-down influence. There is no information in the activation map about

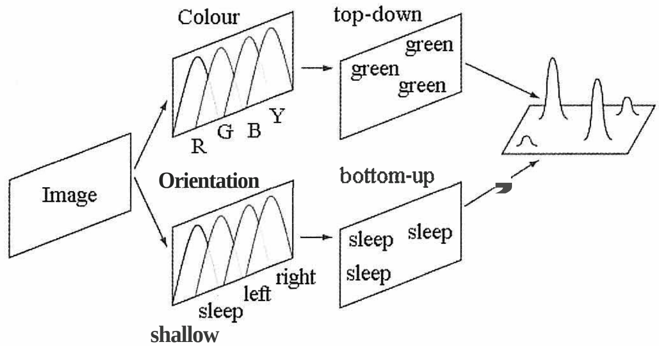


Figure **3.26.** Framework for guided search, the user wants to find a green steep target; image is filtered into categories for each feature map. Bottom-up and top-down activation "mark" regions of the image; an activation map is built by combining bottom-up and top-down information, attention is draw to the highest "hills" in the map [150].

the source of a hill. High activation from a color map looks exactly the same as high activation from an orientation map. A subject's attention is drawn from hill to hill in order of decreasing activation.

Wolfe's theory easily explains traditional "parallel" visual search. Target elements produce the highest level of activation, regardless of the number of distractor elements. This causes the target to "pop-out" of the scene in time independent of the number of distractors. This also explains Duncan and Humphreys' similarity theory results. Low N-N similarity causes distractors to report higher bottom-up activation, since they now differ from their neighbors. High T-N similarity causes a reduction in the target elements' bottom-up activation. Moreover, guided search also provides a possible explanation for situations where conjunction search can be performed preattentively [263, 399, 400]. User-driven top-down activation may permit efficient searching for conjunction targets.

Postattentive Vision. Preattentive processing asks in part: "What visual properties draw our eyes, and therefore our focus of attention, to a particular object in a scene?" An equally interesting question is: "What happens to the visual representation of an object when we stop attending to it and look at something else?" Jeremy Wolfe addressed this question in his work on postattentive vision [402]. The intuitive belief that a rich visual representation accuinulates as we look at more and more of a scene appears not to be true. This provides important insight into why the low-level visual system performs the way it does. The results also act as a bridge between preattentive processing and the new area of change blindness, which shows that people are often "blind" to significant variations that occur between glances at a scene.

Attention to different objects may allow a viewer to learn what is in a scene (if the objects are familiar and recognizable), but it does not allow the viewer to see the scene in a different manner. In other words, the preattentive visual representation of an object after a viewer studies it and looks at something else appears to be identical to its representation before the viewer studied it. No additional information is "saved" in the visual system after the focus of attention shifts to a new location.

Wolfe argues that if multiple objects are recognized siinultaneously in the low-level visual system, it would involve a search for links between the objects and their representation in long-term memory (LTM). LTM can be queried nearly instantaneously, compared to the 40–50 msec per item required to search a visual scene. Preattentive processing can help to rapidly draw the focus of attention to a target with a unique visual feature (e.g., little or no searching is required in the preattentive case). To remove this assistance, Wolfe designed targets with two critical properties (Figure 3.27):

- The targets were formed from a conjunction of features (e.g., they could not be detected preattentively).
- The targets were arbitrary combinations of colors and shapes (e.g., they were not objects that could be semantically recognized and remembered on the basis of familiarity).

Wolfe initially tested two search types. In both cases, viewers were asked to answer as quickly as possible while maintaining a high level of accuracy (e.g., a response-time search):

Traditional search. Text on a blank screen was shown to identify the target. This was followed by a display containing 4, 5, 6, 7, or 8 potential target objects in a 3×3 array (formed by combinations of seven colors and five shapes (Figure 3.27 (top)).

Postattentive search. The display to be searched was shown to the user for a specific duration (up to 300 msec). Text identifying the target was then inserted into the scene (Figure 3.27(bottom)). Results showed that the postattentive search was as slow (or slower) than the traditional search, with approximately 25–40 msec per object required for the target present trials. This implies that previewing the scene provides

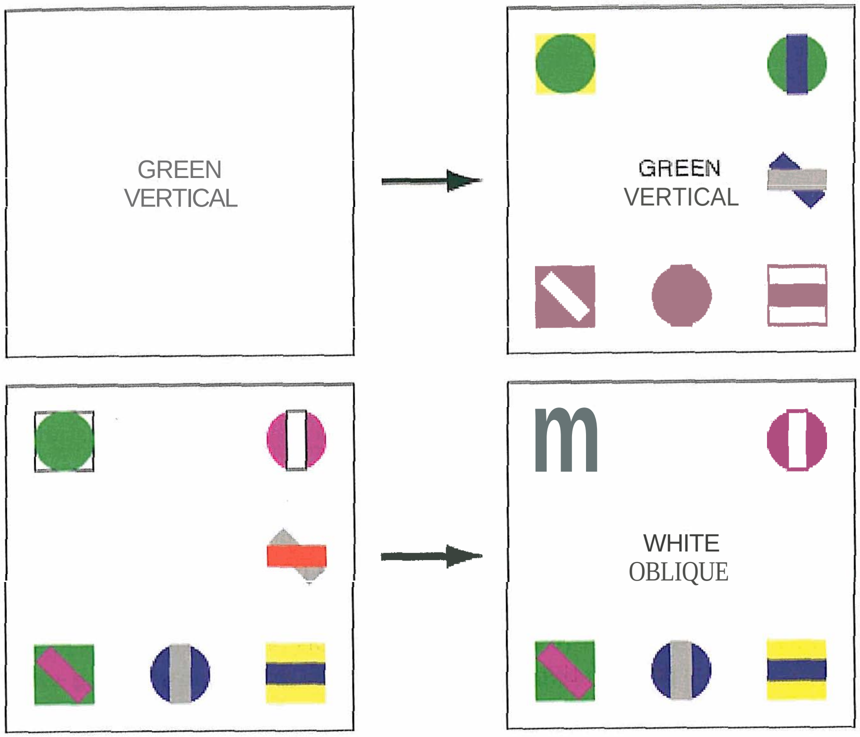


Figure 3.27. Examples of search for color-and-shape conjunction targets, both with and without a preview of the scene: (top) no preview of the scene is shown (although text identifying the target is shown prior to the search)—in this case, the green vertical target is present; (bottom) a preview of the scene is shown, followed by text identifying the target; in this case, a white oblique target is not present.

no advantage to the viewer for finding a conjunction target. In order to explore further, Wolfe studied a number of different search scenarios to test for any benefit from previewing the scene.

These scenarios include:

Repeated search. Viewers were asked to search the same display five times for five different targets. The display was shown with target text, and after an answer was provided (target present or absent), the target text changed to identify a new target. This experiment tested whether additional exposure to the display improved search performance.

Repeated search with letters. Viewers searched in a manner identical to repeated search, but with displays containing letters rather than combinations of colors and shapes. This experiment tested whether the type of target used affected search performance.

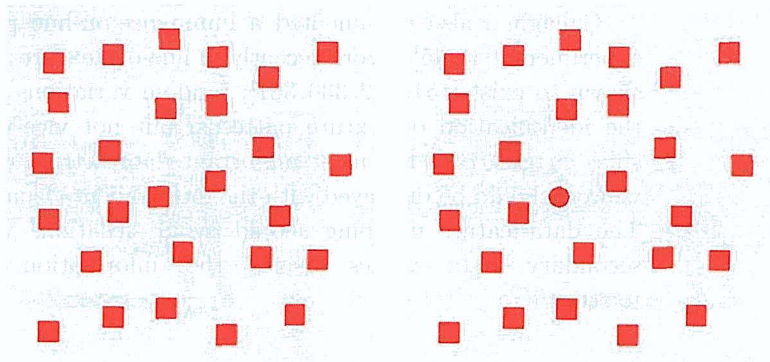
Repeated search versus memory search. Viewers were asked to search a group of five letters 350 times for a target letter. Half the viewers were shown the five letters. The other half were required to memorize the five letters prior to the target queries. This experiment tested whether a prolonged exposure to a set of objects improved search performance. It also tested to see how visual search and short-term memory search performance differed.

In each case, viewers continued to require 20-50 msec per object to complete the search. Wolfe's conclusion was that sustained attention to the objects tested in his experiments did not make visual search more efficient. This has a significant potential impact for visualization design. In most cases, visualization displays are novel, and their contents cannot be committed to long-term memory. This means that studying a display may offer no assistance in searching for specific data values. In this scenario, methods that draw attention to areas of potential interest within a display (i.e., preattentive methods) would be critical in allowing viewers to rapidly and accurate explore their data.

3.3.3 Feature Hierarchy

Based on our understanding of low-level human vision, one promising strategy for multidimensional visualization is to assign different visual features to different data attributes (e.g., building a data-feature mapping that maps data to a visual representation). This allows multiple data values to be shown simultaneously in a single image. One key requirement of this method is a data-feature mapping that does not produce visual interference. Interactions between different visual features hide or mask information in a display. Obviously, we want to avoid this situation during visualization. One simple example of visual interference is the conjunction search shown in Figure 3.20. If we want to search rapidly for combinations of data values, care must be taken to ensure that the resulting combinations contain at least one unique feature for the visual system to cue on.

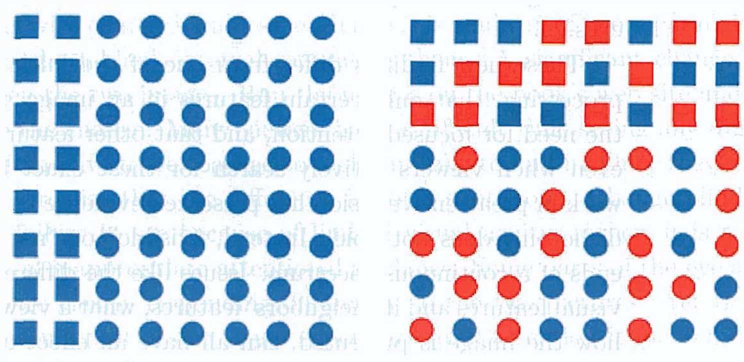
Other types of visual interference can also occur. An important type of interference results from a feature hierarchy that appears to exist in the visual system. For certain tasks, the visual system seems to favor one type of visual feature over another. For example, during boundary detection,



(a) (b)

Figure 3.28. An example of hue-on-form feature hierarchy: (a) a horizontal hue boundary is preattentive identified when form is held constant; (b) a vertical hue boundary is preattentively identified when form varies randomly in the background.

researchers have shown that the visual system favors color over shape (Figures 3.28 and 3.29). Background variations in color interfere with a viewer's ability to identify the presence of individual shapes and the spatial patterns they form [46]. If color is held constant across the display, these same shape patterns are immediately visible. The interference is asymmetric: random variations in shape have no effect on a viewer's ability to see color patterns.



(a) (b)

Figure 3.29. Another example of hue-on-form feature hierarchy: (a) a vertical form boundary is preattentively identified when hue is held constant; (b) a horizontal form boundary cannot be preattentively identified when hue varies randomly in the background.

Callaghan also documented a luminance-on-hue preference during her experiments [44, 45]. More recently, a hue-on-texture interference has been shown to exist [151, 152, 330, 361]; random variations in hue interfere with the identification of texture patterns, but not vice-versa. These hierarchies suggest that the most important data attributes (as defined by the viewer) should be displayed with the most salient visual features, if possible. The data-feature mapping should avoid situations where the display of secondary data values masks the information the viewer wants to see.

3.3.4 Change Blindness

Recent research in visualization has explored ways to apply rules of perception to produce images that are visually salient [384]. This work is based in large part on psychophysical studies of the low-level human visual system. One of the most important lessons of the past twenty-five years is that human vision does not resemble the relatively faithful and largely passive process of modern photography [278, 359, 361, 402, 403]. The goal of human vision is not to create a replica or image of the seen world in our heads. A much better metaphor for vision is that of a dynamic and ongoing construction project, where the products being built are short-lived models of the external world that are specifically designed for the current visually guided tasks of the viewer [92, 244, 288, 325]. There does not appear to be any general-purpose vision. What we "see" when confronted with a new scene depends as much on our goals and expectations as it does on the array of light that enters our eyes.

These new findings differ from one of the initial ideas of preattentive processing, that only certain features in an image are recognized without the need for focused attention, and that other features cannot be detected, even when viewers actively search for these exact features. More recent work in preattentive vision has presented evidence to suggest that this strict dichotomy does not hold. Instead, "visible" or "not visible" represent two ends of a continuous spectrum. Issues like the difference between a target's visual features and its neighbors' features, what a viewer is searching for, and how the image is presented, can all have an effect on search performance. For example, Wolfe's guided search theory assumes both bottom-up (e.g., preattentive) and top-down (e.g., attention-based) activation of features in an image [397, 399, 403]. Other researchers have also studied the dual effects of preattentive and attention-driven demands on what the visual system sees [360, 362]. Wolfe's discussion of postattentive vision also points to the

Figure 3.30. Only one image of many examples of change blindness, each image shows a frame from a sequence which contains a significant variation from the other frame; the animations are available on the book's web site. All sequences courtesy of Ron Rensink; see his discussion of change blindness for additional resources [287]. See also the famous basketball example.

fact that details of an image cannot be remembered across separate scenes, except in areas where viewers have focused their attention [402].

New research in psychophysics has shown that an interruption in what is being seen (i.e., a blink, an eye saccade, or a blank screen) renders us "blind" to significant changes that occur in the scene during the interruption. This change blindness phenomenon can be illustrated using a task similar to a game that has amused children reading the comic strips for many years [244, 288,3251. Figure 3.30 shows a pair of images from a series of movies dealing with change blindness; each movie is made up of two separate images, with a short blank interval separating them. A significant change occurs between the two images. Run the movies on the book's web site and try to locate the change. Many viewers have a difficult time seeing any difference and often have to be coached to look carefully to find it. Once they discover it, they realize that the difference is not a subtle one. Change blindness is not a failure to see because of limited visual acuity; rather, it is a failure based on inappropriate attentional guidance. Some parts of the eye and the brain are clearly responding differently to the two pictures. Yet this does not become part of our visual experience until attention is focused directly on the objects that vary.

The presence of change blindness in our visual system has important implications for visualization. The images we produce are normally novel for our viewers, so prior expectations cannot be used to guide their analyses. Instead, we strive to direct the eye, and therefore the mind, to areas of interest or importance within a visualization. This ability forms the first step towards enabling a viewer to abstract details that will persist over subsequent images.

Dan Simons offers a wonderful overview of change blindness in his introduction to the **Visual** *Cognition* special issue on change blindness and visual memory [325]. We provide a brief summary of his list of possible explanations for why change blindness occurs in our visual system. Interestingly, none of these explanations by themselves can account for all of the change blindness effects that have been identified. This suggests that some combination of these ideas (or some completely different hypothesis) is needed to properly model this phenomenon.

Overwriting. One intuitive suggestion is that the current image is overwritten, either by the blank between images, or by the image seen after the blank. Information that was not abstracted from the first image is lost. In this scenario, detailed change can only be detected for objects the viewer focuses on, and even then, only abstract differences may be recognized.

First Impression. A second hypothesis is that only the initial view of a scene is abstracted. This is plausible, since the purpose of perception is to rapidly understand our surroundings. Once this is done, if the scene is not perceived to have changed, features of the scene should not need to be re-encoded. This means that change will not be detected, except for objects in the focus of attention. One example of this phenomenon is an experiment conducted by Simons and Levin [323, 324]. Subjects were asked to view a short movie. During a cut scene in the movie, the central character was switched to a completely different actor. Subjects were not told to search for any unexpected change in the movie (i.e., they were naive to the presence of the change). After viewing the movie, subjects were asked if they noticed anything odd. Nearly two-thirds of the subjects failed to report that the main actor was replaced. When queried, 70% of the subjects who failed to see the change described the central character using details from the initial actor, and not the replacement. This suggests that their first impression of the actors was the lasting one.

Nothing Is Stored. A third explanation is that after a scene has been viewed and information has been abstracted, no details are represented internally. This model suggests that the world itself acts as a memory store; if we need to obtain specific details from the scene, we simply look at it again. A somewhat weaker form of this model suggests that some detail is preserved between scenes (e.g., the details of the objects in the viewer's focus of attention). In this way, we are blind to change unless it affects our abstracted knowledge of the scene, or unless it occurs where we are looking in the scene.

Everything Is Stored, Nothing Is Compared. Another intriguing possibility is that details about each new scene are stored, but cannot be accessed until an external stimulus forces the access. For example, if a man suddenly becomes a woman during a sequence of images, this discontinuity in abstracted knowledge might allow us to access the details of past scenes to detect the change. Alternatively, being queried about particular details in a past scene might also produce the stimulus needed to access this image history. In one study, an experimenter stops a pedestrian on the street to ask for directions [325]. During this interaction, a group of students walks between the experimenter and the pedestrian. As they do this, one of the students takes a basketball that the experimenter is holding. After providing the directions, the pedestrian is asked if anything odd or unusual changed about the experimenter's appearance. Only a very few pedestrians reported that the basketball had gone missing. When asked specifically about a basketball, however, more than half of the remaining subjects reported it missing, and many provided a detailed description. For example, one pedestrian reported, "Oh yeah, he did have a ball, it was red and white." Not only was the pedestrian able to recall the presence of the basketball when prompted; he was also able to provide specific details about its unique appearance.

Feature Combination. A final hypothesis is that details from an initial view might be combined with new features from a second view to form a combined representation of the scene. Presumably, viewers would not be aware of which parts of their mental image come from the first scene, and which come from the second. The details being combined must make sense, and must be consistent with the viewer's abstract understanding of the scene; otherwise, the change will be recognized as "impossible" or "out of place."

3.4 Perception in Visualization

Figure 3.31 shows several examples of perceptually motivated multidimensional visualizations:

1. A visualization of intelligent agents competing in simulated e-commerce auctions: the x-axis is mapped to time, the y-axis is mapped to auction (each row represents a separate auction), the towers represent bids by different agents (with color mapped to agent ID), height is mapped to bid price, and width is mapped to bid quantity.

(a) (b) (c)

Figure 3.31. Examples of perceptually motivated multidimensional visualizations: (a) a visualization of intelligent agents competing in simulated e-commerce auctions; (b) a visualization of a CT scan of an abdominal aortic aneurism; (c) a painter-like visualization of weather conditions over the Rocky Mountains.

2. A visualization of a CT scan of an abdominal aortic aneurism: yellow represents the artery, purple represents the aneurism, and red represents metal tines in a set of stents inserted into the artery to support its wall within the aneurism.

3. A painter-like visualization of weather conditions over the Rocky Mountains across Utah, Wyoming, and Colorado: temperature is mapped to color (dark blues for cold, to bright pinks for hot), precipitation is mapped to orientation (tilting right for heavier rainfall), wind speed is mapped to coverage (less background showing through for stronger winds), and pressure is mapped to size (larger strokes for higher pressure).

We briefly describe how perceptual properties of color, texture, motion, and nonphotorealism have been used in visualization.

3.4.1 Color

Color is a common feature used in many visualization designs. Examples of simple color scales include the rainbow spectrum, red-blue or red-green ramps, and the grey-red saturation scale [383]. More sophisticated techniques attempt to control the difference viewers perceive between different colors, as opposed to the distance between their positions in RGB space.

This improvement allows:

Perceptual balance. A unit step anywhere along the color scale produces a perceptually uniform difference in color.

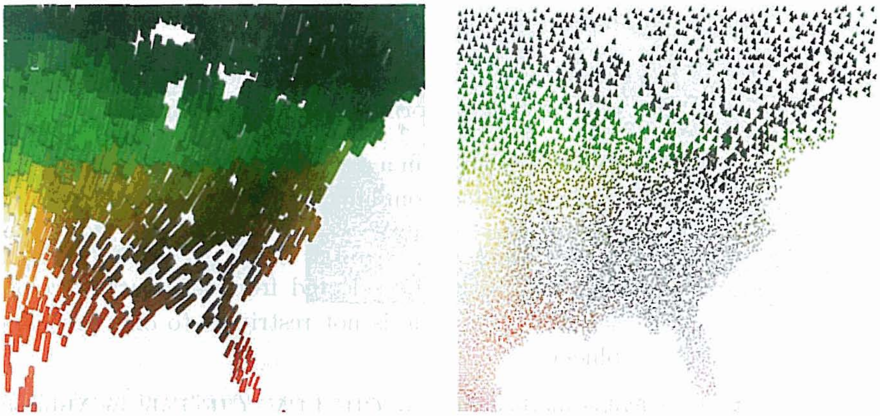
Distinguishability. Within a discrete collection of colors, every color is equally distinguishable from all the others (i.e., no specific color is "easier" or "harder" to identify).

Flexibility. Colors can be selected from any part of color space (e.g., the selection technique is not restricted to only greens, or only reds and blues).

Color models such as CIE LUV, CIE Lab, or Munsell can be used to provide a rough measure of perceptual balance [30, 65, 260]. Within these models, Euclidean distance is used to estimate perceived color difference. More complex techniques refine this basic idea. Rheingans and Tebbs plotted a path through a perceptually balanced color model, then asked viewers to define how attribute values map to positions along the path [290]. Nonlinear mappings emphasize differences in specific parts of an attribute's domain (e.g., in the lower end with a logarithmic mapping, or in the higher end with an exponential mapping). Other researchers have constructed rules to automatically select a colormap for a target data attribute [24,296]. Properties of the attribute, such as its spatial frequency, its continuous or discrete nature, and the type of analysis to be performed, are used to choose an appropriate color representation. Ware constructed a color scale that spirals up around the luminance axis to maintain a uniform simultaneous contrast error along its length [383]. His solution matched or outperformed traditional color scales for metric and form identification tasks. Healey and Enns showed that color distance, linear separation, and color category must all be controlled to select discrete collections of equally distinguishable colors [152, 153].

Figure 3.32 shows historical weather conditions over the eastern United States for Marcli, with color mapped to temperature (blue and green for cold, to red and pink for hot), luminance mapped to wind speed (brighter for stronger winds), orientation mapped to precipitation (more tilted for heavier rainfall), size mapped to cloud coverage (larger for more cloudy), and frost frequency mapped to density (denser for higher frost).

Healey's color selection technique combines different aspects of each of these methods. A single loop spiraling up around the Laxis (the luminance pole) is plotted near the boundary of our monitor's gamut of displayable colors in CIE LUV space. The path is subdivided into r nained color regions (i.e., a blue region, a green region, and so on). Here, n colors can then be



(a) (b)

Figure 3.32. Example of color representations for weather maps: (a) a nonphotorealistic visualization using simulated brush strokes to display the underlying data; (b) a traditional visualization of the same data using triangular glyphs.

selected by choosing n/r colors uniformly spaced along each of the r color regions. The result is a set of colors selected from a perceptually balanced color model, each with a roughly constant simultaneous contrast error, and chosen such that color distance and linear separation are constant within each named color region (Figure 3.32).

3.4.2 Texture

Texture is often viewed as a single visual feature. Like color, however, it can be decomposed into a collection of fundamental perceptual dimensions. Researchers in computer vision have used properties such as regularity, directionality, contrast, size, and coarseness to perform automatic texture segmentation and classification [145, 285, 286, 345]. These texture features were derived both from statistical analysis, and through experimental study. Results from psychophysics have shown that many of these properties are also detected by the low-level visual system, although not always in ways that are identical to computer-based algorithms [3, 76, 179–181, 183, 330, 362, 403].

One promising approach in visualization has been to use perceptual texture dimensions to represent multiple data attributes. Individual values of an attribute control its corresponding texture dimension. The result is a texture pattern that changes its visual appearance based on data in the underlying data set. Grinstein et al. visualized multidimensional data with "stick-figure" icons whose limbs encode attribute values stored in a data element [140]; when the stick-men are arrayed across a display, they form texture patterns whose spatial groupings and boundaries identify attribute correspondence. Ware and Knight designed Gabor filters that modified their orientation, size, and contrast, based on the values of three independent data attributes [382]. Healey and Enns constructed perceptual texture elements (or *pexels)* that varied in size, density, and regularity [151,152]; results showed that size and density are perceptually salient, but variations in regularity are much more difficult to identify. More recent work found that 2D orientation can also be used to encode information [391]; a difference of 15 degrees is sufficient to rapidly distinguish elements from one another. A follow-on to these studies showed that certain 3D orientation properties can also be detected by the low-level visual system [237].

Interrante, Kim, and Hagh-Shenas have studied the use of different texture types and orientations for showing the shape of an underlying 3D object. Initial experiments investigated textures that varied in luminance (e.g., greyscale patterns) [172, 173, 208]. More recent work has studied the use of relief textures. The textures were arrayed over the surface using orientations that were either *isotropic* (e.g., all following a common direction), or *anisotropic* (e.g., following different directions based on a property at that point on the surface). Preliminary results suggest that anisotropic textures that follow both the first or second principal curvature directions produce surface perception that is as good or better than either principal direction alone, or than other orientation rules [207].

3.4.3 Motion

Motion is a third visual feature that is known to be perceptually salient. The use of motion is common in certain areas of visualization, for example, the animation of particles, dye, or glyphs to represent the direction and magnitude of a vector field (e.g., fluid flow visualization). Motion transients are also used to highlight changes in a data set across a user-selected data axis (e.g., over time for a temporal data set, or along the scanning axis for a set of CT or MRI slices). As with color and texture, our interest is in identifying the perceptual dimensions of motion and applying them in an effective manner. Three motion properties have been studied extensively by researchers in psychophysics: flicker, direction of motion, and velocity of motion.

For visualization purposes, our interest is in flicker frequencies F (the frequency of repetition measured in cycles per second) that are perceived as discrete flashes by the viewer. Brown noted that frequency must vary from 2–5% to produce a distinguishable difference in flicker at the center of focus ($1.02 \leq \Delta F \leq 1.05$), and at 100% or more for distinguishable difference in flicker in the periphery ($\Delta F \geq 2.0$) [39, 122, 257].

Tynan and Sekuler showed that a decrease in a target object's velocity or an increase in its eccentricity increased identification time [367], although in all cases viewers responded rapidly (200–350 msec for targets in the periphery, 200–310 msec for targets in the center of focus). In addition, van Doorn and Koenderink confirmed that higher initial velocities produce a faster response to a change in the velocity [163, 250, 372, 373]. They claim this is due to the need for the target to traverse a "critical distance" before it can be detected. For a baseline velocity V_1 and a target velocity $V_2 = 2V_1$, approxiinately 100 msec is needed to see the velocity change from V_1 to V_2 for slow V_1 (1° per second) and approximately 50 msec for faster V_1 (2° per second or higher).

Researchers in psychology have used properties of motion to extend a viewer's ability to perform basic exploration tasks. Nakayama and Silverman showed that coherent motion or stereoscopic depth can be used to separate elements into coherent groups, allowing viewers to search each group independently [263]. For example, consider searching for a red circle in a background of red squares and blue circles, a situation that normally produces a time-consuming serial search for the target. If the red elements are animated to move up and the blue elements are animated to move down, however, the target is immediately visible. Applying different motion patterns to the red and blue groups allows a viewer's visual system to separate them and search them independently, producing the rapid search for a curved element (a red circle) in a background of linear elements (red squares). Similar results can be achieved by displaying the red and blue elements on different stereoscopic planes. Driver et al. showed that oscillation can also be used to separate elements into independent visual groups, but only if the oscillation pattern is coherent [86]. For example, a viewer could identify a red circle in a set of red squares and blue circles if all the red items oscillate up and down in lock step, and all the blue elements oscillate left and right in lock step. If the elements oscillate "out of phase," however (i.e., some red elements start moving down while others are still moving up), viewers are forced to revert to serial search. More sophisticated motion patterns have also been analyzed, although with less success in terms of achieving high-speed search performance. Braddicli and Holliday studied both divergence (e.g., squares increase or decrease in size over a period of time, then snap baclc to their original size) and deformation (e.g., rectangles deform from tall and skinny to short and wide, then snap baclc to their original shape) [33]. Although the basic motion properties being shown can be rapidly identified in isolation, the coinbinations that form deformation and divergence were not detected by the low-level visual system. See also the iconic extensions to pixels called *moxels* [414].

Properties of motion have been extended to visualization design. Animated motion is used in flow visualization to show the direction and speed of different flow patterns (e.g., by Kirby [210]). Kerlick proposed the use of animated glyphs to visualize 2D and 3D multidimensional data sets [206]. He designed a set of "*boids*" to encode attribute values at specific locations in the data set, for example, a sphere boid to query data values at a user-selected location, or pyramid and dart boids that animate over a vector field to visualize its shape. Bartram et al. studied the use of variations in color, shape, and motion to "notify" viewers while they were engaged in a separate, attention-demanding task [18]. Results showed that applying motion to a static glyph significantly eased recognition, coinpared to changing the glyph's color or shape. This finding held both when the glyph was near the center of focus and when it was located on the periphery of the viewer's gaze. The authors also studied how distracting a secondary motion cue was judged to be. Flicker was the least distracting, followed by oscillating motion, then divergence, and finally movement over long distances. Related work by Bartram et al. confirmed that different motion paths can be used to perceptually group glyphs in a manner similar to the worlc of Nalcayama and Silverman [263] or Driver et al. [19]. The groups can then be searched independently for a target feature.

3.4.4 Memory Issues

Three types of memory are relevant to our study of perception in visualization:

Sensory memory. Sensory memory is high capacity information storage. It is effectively preattentive eye filters. Large quantities of information are processed very fast (less than 200 msec). Such learning is physical and can be harnessed by repeated actions. This explains the importance, for example, of positional learning in typing or playing piano (it feels almost as if the memory is in the hand and fingers).

Short-term memory. Short-term memory analyzes information from both sensory and long-term storage. It has limited information capacity. It occurs at a high level of processing, but the time span is limited typically to less than 30 seconds. It represents the beginning of thinking. It can be harnessed by grouping and repetition, by not requiring users to remember too many things, and by chunking. The chunks are grouped objects remembered as a unit, with the number limited to 5 to 9 (see Section **3.5**).

Long-term memory. Long-term memory is complex and theoretically limitless, much like a data warehouse. This storage is multicoded, redundantly stored, and organized in a complex network structure. Information retrieval is a key problem and access is unreliable and slow. It can be harnessed by using association mnemonics and chunking.

The following was distributed as an example highlighting how memory supported the quick scanning of words in a document, showing that not all letters are needed.

> *Rinadeg Oedrr*
>
> *Aoccdrnig to a rscarhee at Cigdmabre Uinervtisy, it deosn't mtetar in waht oredr the ltteers in a wrod are, the olny iprmoatnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a taotl mses and you can sitll raed it wouthit porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe.*

In the next section we will see how memory can play a part in studying a visualization.

3.5 Metrics

How many distinct line lengths and orientations can humans accurately perceive? How many different sound pitches or volumes can we distinguish without error? What is our "channel capacity" when dealing with color, taste, smell, or any other of our senses? How are humans capable of recognizing hundreds of faces and thousands of spoken words? These and related issues are important in the study of data and information visualization. When designing a visualization, it is important to factor in human limitations to avoid generating images with ambiguous, misleading, or difficult-to-interpret information. Many efforts have been made to try and ascertain these limits, using experiments that test human performance on measuring and detecting a wide assortment of sensed phenomena. This section presents an overview of some early seminal work on measuring perceptual capabilities and relate it to current work in data visualization. The sorts of questions we would like to be able to answer include:

- What graphical entities can be accurately measured by humans?
- How many distinct entities can be used in a visualization without confusion?
- With what level of accuracy do we perceive various primitives?
- How do we combine primitives to recognize complex phenomena?
- How should color be used to present information?

The answers to these and other questions will enable us to design more effective visualizations, and to have a better understanding of how accurately we are cominunicating information with the visualization.

3.5.1 Resource Model of Human Information Processing

To be able to measure and compare human perceptual performance on various phenomena, one needs a metric, a gauge or yardstick that can reliably evaluate performance and associate numbers with the results of testing a group of subjects. George Miller, in 1956 [255], borrowed the concept of channel capacity from the field of information theory. Suppose that we assume the human being is a communication channel, taking input (perceiving some phenomena) and generating output (reporting on the phenomena). The overlap between input and output is the information about the phenomena that has been perceived correctly, and is thus the amount of transmitted information.

For each primitive stimulus, whether it be visual, auditory, taste, touch, or smell, we measure the number of distinct levels of this stimulus that the average participant can identify with a high degree of accuracy. The results will follow an asymptotic behavior, e.g., at a certain point, increasing the number of levels being used causes an increase in the error rate, and no additional information will be extracted from the source stimulus. Miller called this level the "channel capacity" for information transfer by the human. He measured it in bits (borrowing again from information theory), depending on the number of levels that the average human could measure with high accuracy. Thus if errors routinely begin when more than 8 levels of a phenomenon are tested, the channel capacity for this phenomenon is **3** bits.

As in all experiments involving human subjects, it is important to establish controls, so that only the single isolated phenomenon is being tested. Training time must therefore be limited, as some individuals can fine-tune their perceptual abilities much faster than others. For the same reason, we need to avoid including the results from *specialists.* Clearly, a musician will likely be more able to accurately perceive sound pitches than the average subject, and a cartographer or navigator will be able to identify spatial features more readily than someone who does spatial analysis less frequently. Related to this is the aspect of *context;* it is very important to design perceptual experiments to be as context-free as possible, as we don't want to bias the results via associations and factors that have little to do with perception. Finally, the experimental data should be free of error and noise; while real data and phenomena are rarely noise-free, it is difficult to obtain accurate measurements from data of variable quality.

There are many other guidelines for the design of perceptual experiments. This section, and the contents of the rest of the chapter, are merely meant to illustrate the general procedure for conducting this sort of analysis. Those wishing to understand the process in more detail are directed to the literature in perceptual psychology, social sciences, and human factors analysis.

3.5.2 Absolute Judgmentof 1D Stimuli

A large number of experiments have been performed over the years to ascertain the ability of humans to judge absolute levels of different stimuli. In this section, we summarize a number of these experiments (from [255]) in terms of the number of bits in the channel capacity of humans, as defined earlier. For each, we provide the name of the researcher, the experimental set-up, and the number of levels that could, on average, be accurately measured.

1. *Sound pitches (Pollack)*: Subjects were exposed to sets of pitches at equal logarithmic steps (from 100–8000 cps). The result was that the average listener could reliably distinguish 6 pitches. Varying the range didn't change the results appreciably; subjects who correctly classified 5 high pitches or 5 low pitches could not accurately classify 10 when combined. This is a channel capacity of 2.5 bits.

2. *Sound loudness (Gardner):* In another auditory experiment, the loudness of a sound was varied between 15–110 dbs. On average, 5 levels were accurately discerned, for a capacity of 2.3 bits.

3. *Salinity (Beebe-Center):* Taste perception had similar results. By varying salt concentrations from 0.3 to 34.7 gm per 100 cc water, subjects

were found to be able to distinguish just under 4 levels, on average, corresponding to a capacity of 1.9 bits.

4. *Position on a line (Hake/Gardner)*: In an experiment much more relevant to data visualization, this experiment varied the position of a pointer located between two markers. Participants attempted to classify its position either from a list of possibilities or on a scale of 0 to 100. Most subjects were able to correctly label between 10 and 15 levels, though this increased with longer exposure. This corresponds to a channel capacity of 3.25 bits.

5. *Sizes of squares (Eriksen/Hake)*: In another graphics-related experiment, the size of squares was varied. Surprisingly, the capabilities of humans to accurately classify the sizes was only between 4 and 5 levels, or 2.2 bits.

G. *Color (Eriksen):* As color is often used to convey information in visualization~,it is important to understand how well this attribute is perceived. In experiments that varied single color parameters, it was found that users could correctly classify 10 levels of hue and 5 levels of brightness, or 3.1 and 2.3 bits, respectively.

7. *Touch (Gelard):* In this unusual experiment, vibrators were placed at different locations on the chest area. Several parameters were varied individually, including location, intensity, and duration. The results estimated the capacity at 4 intensities, 5 durations, and 7 locations.

8. *Line geometry (Pollack):* Lines have many attributes that can be used to convey information. In this experiment, line length, orientation, and curvature were tested. The results were: 2.6–3 bits for line length (depending on duration), 2.8-3.3 bits for orientation, and 2.2 bits for curvature with constant arc length (while only 1.6 bits for constant chord length).

To summarize these experiments, there appears to be some built-in limit on our capability to perceive and accurately measure 1D signals. The average from these experiments was 2.6 bits, with a standard deviation of .6 bits. This means that if we want users of our visualization systems to be able to extract more than G or 7 levels of a data value with accuracy, we must look at other means.

3.5.3 Absolute Judgmentof Multidimensional Stimuli

One solution to the dilemma regarding this limitation on the number of levels of a data value that can be accurately ineasured is to use more than one stimulus simultaneously. A logical assumption would be that if we combine stimulus A, with a channel capacity of C_A bits (or 2^{C_A} levels), and stimulus B, with a channel capacity of C_B bits (or 2^{C_B} levels), we should get a resulting capacity of approximately $C_A + C_B$, or the product of the two numbers of levels. Unfortunately, experiments have shown otherwise:

1. *Dot in a square (Klemmer/Frick)*: Given that a dot in a square is actually two position measurements (vertically and horizontally) we should get a capacity that is twice that of gauging the position of a marker on a line (6.5 bits), but it was measured at 4.6 bits.

2. *Salinity and sweetness (Beebe-Center):* In an experiment that combined sucrose and salt solutions, the total capacity should have been twice that of measuring salinity alone, or **3.8** bits. However, it was measured at **2.3** bits.

3. *Loudness and pitch (Pollack):* The combination of two auditory channels should have produced a capacity equal to the sum of the results for pitch and loudness in isolation, or 4.8 bits, but it was ineasured at **3.1** bits.

4. *Hue and saturation (Halsey/Chapanis)*: Combining hue and saturation should have resulted in a capacity of **5.3** bits, but it was measured at only **3.6** bits.

5. *Size, brightness, and hue (Eriksen)*: In an experiment combining geometry and color, the size, hue, and brightness of shapes were varied. The sum of the individual capacities is 7.6 bits, but a capacity of only 4.1 bits was observed.

6. *Multiple sound parameters (Pollack/Ficks)*: In a very ambitious experiment, G auditory variables (frequency, intensity, rate of interruption, on-time fraction, duration, and location) were varied. As individual stimuli, each had a capacity of 5 values, so the results should have been 15,600 combinations that could be accurately discerned. However, the results were only 7.2 bits of channel capacity, or 150 different combinations.

To summarize, combining different stimuli does enable us to increase the amount of information being communicated, but not at the levels we might hope. The added stimuli resulted in the reduction of the discernibility of the individual attributes. With that said, however, having a little information about a large number of parameters seems to be the way we do things. This agrees with linguistic theory, which identifies 8 to 10 dimensions, where each can only be classified in two or three categories.

We now look at strategies for improving the the information content of data visualizations by taking advantage of alternative perceptual skills.

3.5.4 Relative Judgment

Williain Cleveland and his colleagues have performed a number of experiments in graphical perception to better understand the ways information can be communicated via images [67]. Their emphasis, rather than on *absolute measurement* (classification), was on *relative judgment*. Thus, the task they were interested in was the detection of differences, rather than extracting a numeric value. In Figure 3.33, it is much easier to detect and gauge the change in heights when the bars are surrounded by a box (a relative change).

They studied how well humans gauge differences using the following 10 graphical attributes (shown in Figure 3.34):

1. angle;
2. area;
3. color hue;
4. color saturation;

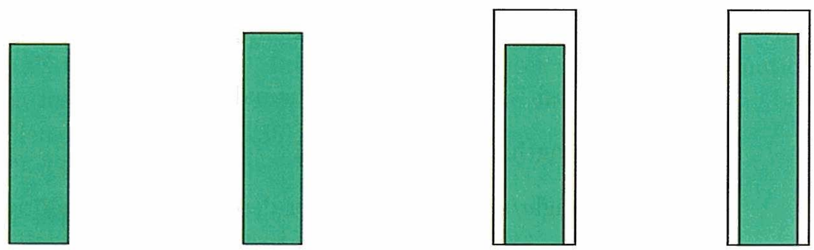


Figure 3.33. The boxes on the left are not the same size, but it is difficult to estimate the magnitude of the difference. The same boxes are shown on the right. The encapsulating frame makes it easier to gauge the relative difference between them.

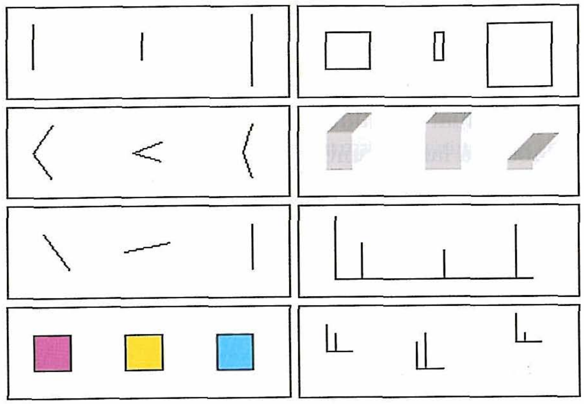


Figure 3.34. Examples of graphical attributes used in perceptual experiments. Left column (from top): length, angle, orientation, hue. Right column: area, volume, position along a common scale, position along identical, nonaligned scales.

5. density (amount of black);
6. length (distance);
7. position along a common scale;
8. position along identical, nonaligned scales;
9. slope;
10. volume.

Their experiments showed errors in perception ordered as follows (increasing error):

1. position along a common scale;
2. position along identical, nonaligned scales;
3. length;
4. angle/slope (though error depends greatly on orientation and type);
5. area;
6. volume;
7. color hue, saturation, density (although this was only informal testing).

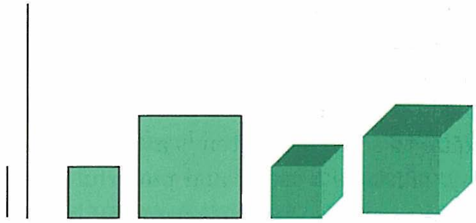


Figure **3.35.** Illustration of Stevens' Law. The size ratio for each pair is 1:4. This magnitude is readily apparent in the lines, but it is easily underestimated in the squares and cubes.

This seems to support the idea that bar charts and scatterplots are effective tools for communicating quantitative data, as they both depend on position along a common scale. It also suggests that pie charts are probably not as effective a mechanism, as one is either judging area or angles.

Two important principles came into play with these experiments. The first, named Weber's Law, states that the likelihood of detecting a change is proportional to the relative change, not the absolute change, of a graphical attribute. Thus, the difference between a 25-centimeter line and a 26-centiineter line should be no easier to perceive than the difference between a 2.5- and a 2.6-centimeter line. This means that simply enlarging an object or otherwise changing the range of one of its attributes will not, in general, increase its effectiveness at communicating information.

A second useful principle, known as Stevens' Law, states that the perceived scale in absolute measurements is the actual scale raised to a power. For linear features, this power is between 0.9 and **1.1;** for area features, it is between 0.6 and 0.9, and for volume features it is between 0.5 and 0.8. This means that as the dimensionality of an attribute increases, so increases the degree at which we underestimate it. This implies that using attributes such as the volume of a three-dimensional object to convey information is much less effective and much more error-prone than using area or, better yet, length (see Figure **3.35).**

3.5.5 Expanding Capabilities

The experiments described in the previous three sections indicate that our abilities to perceive various stimuli, and graphical phenomena in particular, is fairly limited. If we need to communicate information with a higher capacity, we must investigate strategies for expanding our capabilities. One way, as illustrated in the previous section, is to try and reconfigure the communication task to require relative, rather than absolute, judgment. Thus, in many cases, we can supplement a visualization so that the viewer just needs to gauge whether an item's attribute is greater than, less than, or equal to some other item's attribute. This is why adding grid lines and axis tick marks is a useful and powerful addition to a visualization.

We can also increase capacity by increasing the dimensionality, as seen in the experiments on multiple stimuli. In most cases, adding another stimulus will lead to larger bit rates. However, there is likely to be a limit to the number of dimensions that can be reasonably managed. This *span of perceptual dimensionality*, according to Miller [255], is hypothesized to be about 10. Another problem with this solution is that in graphics there are a limited number of parameters that we can use (color, size, position, orientation, line/fill style, and so on), although when we discuss *glyphs* in Chapter 7 we will examine efforts to pack many more dimensions into the components of a composite graphical entity.

Another potential strategy is to reconfigure the problem to be a sequence of different absolute judgments, rather than simultaneous stimuli. In this manner, we might be able to overcome some of the loss of capacity that was shown in the experiments on measuring multiple stimuli. If the viewer is directed to examine a sequence of visualizations and compose the measurements from each, we may be able to achieve an improved communication rate. This leads to the analysis of immediate memory.

3.5.6 The Relationship to Immediate Memory

Many studies have examined human memory performance. *Immediate (short-term) memory* is used for very short-term recall, often immediately after a stimulus has been received. Many games have been devised that are based on one's immediate memory skills. Studies have shown the span of immediate memory to be approximately 7 items. In other words, people, in general, can remember with accuracy a sequence of 7 or so stimuli. One question that might arise is whether this is related to our span of absolute judgment, as the capacities are similar.

The answer is that they are unrelated. Absolute judgment is limited by the amount of information, while immediate memory is limited by the number of items, no matter how complex. Thus, they are measurements at different granularities; absolute judgment is measured in *bits* corresponding to distinct levels, while immediate memory involves *chunks* of varying size or complexity. Several experiments involving binary digits, decimal digits, letters, syllables, words, and mixtures have shown that the number of chunks that can be remembered is relatively constant. An interesting observation is that we can remember 6 or so monosyllabic words, but also G or so multisyllabic words.

It is conjectured that we "chunk" things at the largest logical unit. But what is that logical unit? Can we increase its coinplexity to increase our capacity? This process is known as *recoding*.

3.5.7 The Role of Recoding

Recoding is the process of reorganizing information into fewer chunks, with more bits of information per chunk. For example, in the process of learning Morse code, one starts by connecting patterns of dots and dashes into letters, and then longer patterns into words. This process is also found in other avenues of learning, including music and dance. A similar concept, known as compilation, can be found in the artificial intelligence field as a form of machine learning.

Many experiments have been designed to study the ability of humans to recode information in this manner. Experiments in recalling long strings of binary digits show nearly linear improvement with chunk size. In other words, remembering a sequence of N individual binary digits is coinparable to the effort of remembering a sequence of N binary digit chunks of length 2 or **3.**

One problem is that the way we perform recoding differs from person to person. We remember events by creating a verbal recoding of what we saw, and then elaborate from this coded version. This accounts for variations in witness testimonies to a crime or accident; in the recoding process, different aspects are chunked together, and depending on the complexity of the chunks, it may be difficult to recall exact details (we are convinced that our particular decoding is a very accurate depiction of what took place). It also explains how people can change their appearance fairly dramatically (make a major change in hair style, switch from glasses to contacts, gain or lose significant weight) and have it go unnoticed by friends and colleagues. As long as the new attributes fit within the decoded memories, the change may not be detected.

3.5.8 The Role of Focus and Expectation

Related to the use of multiple data coding attributes and sequences of decisions is the work reported by Chapman [55], who observed that in images with multiple attributes, but with observers only reporting on one, prior notification of focus resulted in significantly better results than post selection of focus. This may seem obvious, but it is important, as it indicates that people do better when focusing on a single attribute at a time. Recall from the experiments on judging multiple stimuli that the performance was worse (often much worse) than the combination of the capacities of the individual stimuli. Chapman's work indicates that if the user can focus attention on a small set of attributes (or one attribute), he or she can reduce the influence of the attributes that are outside of the focus group. Thus, if viewers can be trained to look for certain features in isolation and have forewarning as to what those features should be, their performance can be improved. If users are uninformed as to the features containing the most relevant information, it is less likely that they will be able to extract the desired information at the desired accuracy.

This seems directly related to change blindness, an attempt to probe the types of visual representations being built when looking at a scene. The visual system makes assumptions to fill in details outside of the focus of attention. For example, if no motion transient is seen, the visual system may assume that the scene is static. This explains why one can "miss" a big change in a location not being focused on during an eye saccade.

If this theory is accurate, pre-focusing the viewer on a particular feature or feature-value would help, as one would only need to build one Boolean map to search for and/or identify what is being looked for (the target). Without prefocusing, one would build maps with some other priority, possibly building and discarding multiple maps until one hits on the right one.

3.5.9 Summary on Metrics

Many factors are involved in communicating information via the human perceptual system. The span of absolute judgment and immediate memory limits our ability to perceive information accurately. We can expand this ability by reformatting into multiple dimensions or sequences of chunks. We can also take advantage of the fact that our ability to perform relative judgment (detection) is more powerful than our absolute (measured) judgment abilities.

In terms of the implications to data visualization, for applications where absolute judgment is required, the best we can do with a single graphical attribute is between 4 and 7 values. To get a larger range of recognizable levels, we must repose the problem in multiple dimensions, do a sequence of simple decisions, or perform some type of chunking.

Alternatively, we could redefine the problem in such a way that relative, rather than absolute, judgment could be used to focus attention, with a second, more quantitatively accurate, stage following the initial focus of attention.

3.6 Related Readings

Parts of the section on metrics came from the excellent article by George Miller, "The Magic Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information" [255].

More recent work on graphical perception is from the chapter entitled "Graphical Perception" in Williain S. Cleveland, *The Elements of Graphing Data* [67].

Work on AI and cognition include, Kurzweil, *The Age of Spiritual Machines* [226] and Looks et al., "Novainente: An Integrative Architecture for Artificial General Intelligence" [239].

3.7 Exercises

1. List the features you believe you use in recognizing a friend by sight and/or by sound. How might you use related features to communicate a data set?

2. Design an experiment that would integrate an eye tracking study with a target discovery test.

3. Design an experiment to identify which is better for visualizing a linear pattern in a large data set: a simple point plot, or a point plot where the points are circular, rectangular, colored, or vibrating. Guess at the outcome.

4. Since about 8% of males are color deficient [235] (with less than 1 % for females) mostly in the red and green ranges, how would you deal with color in the display of a scatterplot?

3.8 Projects

1. Take the scatterplot code you've written. Consider some perceptual attribute you've read about and are interested in. Generate a display

for a perceptual study, say a target (one objector or a pattern) to be identified within an area of distractors. Ask a few classmates if they can easily identify the target.

2. Write a program to reproduce one of the perceptual experiments, varying either a single graphical attribute or multiple ones. Start with two or three values for a given attribute, and increase this number until you (or a willing friend) start making errors over a short sequence of samples. Describe what feature you are testing, whether you are testing for absolute or relative judgment, and what your results are.

3. Using the VIAT Windows-based software available on the book's web site, design an experiment for some of the perceptual features described in this chapter.