Perception

FIFTH EDITION

Randolph Blake
Vanderbilt University

Robert Sekuler
Brandeis University

(2006)

Boston Burr Ridge, IL Dubuque, IA Madison, WI New York San Francisco St. Louis
Bangkok Bogotá Caracas Kuala Lumpur Lisbon London Madrid Mexico City
Milan Montreal New Delhi Santiago Seoul Singapore Sydney Taipei Toronto

Chapter 5

Spatial Vision and Form Perception

OUTLINE

We rely on vision for just about everything we do: reading a menu, looking out for vehicles when crossing a busy street, kicking a football, using a mirror to put on makeup, selecting ripe apples at a roadside stand, spotting a friend in a crowded airport, and all the countless other activities we engage in every day. As these examples suggest, visually guided activities nearly always revolve around *objects*. Some tasks entail picking out an object from its surroundings, such as spotting the evening's first star in the twilight sky. We can call this process **detection.** Other tasks require the additional step of distinguishing one object from another, such as selecting a quarter from the pile of coins on your desk. We can call this second, more refined process **discrimination.** Or you may need to identify with precision a particular object or a specific person. Finding your picture in the college yearbook requires this level of specificity. This third process we can call **identification.** To deal with the environment successfully, people must be able to accomplish all three tasks rapidly and accurately: detect, discriminate, and identify. Detection reveals the presence of an object or objects; discrimination and identification serve more refined, complicated purposes. For example, visual discrimination allows an organism to sort important objects from unimportant ones. Of course, importance varies with one's current needs.

Overwhelming hunger means you'd better be able to distinguish the edible from the inedible; when preparing a research paper, you must be able to distinguish potentially relevant books from irrelevant ones. For some animals, it is imperative to discriminate potential mates who are in heat from those who are not. Identification demands one more step—not only must one object be distinguished (discriminated) from others, the precise identity of that object must be specified. To find *your* red Miata in a crowded parking lot, you must discriminate the red cars from all the rest *and* then identify yours based on other familiar, more specific cues such as the small dent in the hood.

These examples suggest a hierarchy of processes in which discrimination builds on detection, and identification, in turn, builds on discrimination. Of course, you do not consciously step through this sequence. The perceptual act of identifying some object in a complex scene often occurs effortlessly, in the blink of an eye, and mostly without your being aware. Nonetheless, the stimulus information required for the three processes differs, with each successive process requiring more refined information.

Even the seemingly simple act of detection has prerequisites. To be detected, an object must differ from its surroundings in at least one of several possible ways. The differences can involve color (a ripe, red berry against the green leaves of its bush), movement (an ant crawling on

your ice cream), shape (a chocolate-covered cherry in a box of mixed chocolates), or depth (an aspirin that has fallen on a white tile floor). Usually, several of these sources of information are available at the same time, which facilitates detection. For example, a breeze causes some red berries to move relative to the green leaves, making the berries even more conspicuous. But once the berries are detected, ripe ones must be discriminated from the unripe, which can require judging subtle color differences. And, of course, to guard against poisoning, it's a very good idea to identify the type of berry you're about to eat. Identification entails comparing the current environmental stimulus with stored knowledge about categories of berries.

This chapter and the four that follow deal with visual information that makes detection and discrimination possible. These chapters also go into the details of how the nervous system registers and processes that information. The present chapter examines how an object's spatial properties influence detection and discrimination. We use the term *spatial properties* to refer to those attributes of an object that determine the perceived size and shape of the object. Defined in this way, the term *spatial properties* is more or less synonymous with the often-used terms *form* and *visual pattern*. In this chapter, we'll be focusing on what some experts would characterize as "early vision," a term used to denote the initial stages of image analysis that, in turn, are used for more refined description of a visual scene. In the following chapter, we'll move to the rules governing that more refined description. Of course, objects have color and three dimensionality, and some characteristically move about in the environment—these are aspects of vision that are discussed in Chapters 7, 8, and 9, respectively.

We should start by defining what is meant by an "object." To do that, let's start with a real-world problem.

What Defines an Object?

On January 4, 2004, the planet Mars had its second human-made visitor, a sophisticated robotic vehicle named Spirit. (Spirit's robotic forbearer was a compact little gadget named Sojourner, the first human-made vehicle to move across Mars's surface back in 1997.) Spirit was deposited in the middle of the Gusev crater, a region of the Martian landscape of particular interest to NASA scientists because of its diverse geological composition. From high-altitude photographs, scientists suspected that Gusev's rocks and sediment held important clues to the origin and evolution of Mars, including the possible existence of water on Mars eons ago. Spirit's assignment was to navigate Gusev's treacherous terrain, collecting invaluable images and geological samples. To accomplish this daunting task, Spirit was outfitted with an array of sensitive instruments including a stereo camera that captured detailed color images of the Martian surface in vivid 3D. Electronic signals from this pair of cameras were transmitted back to earth, where Spirit's human colleagues issued commands guiding Spirit's explorations.[1]

Information from Spirit's "eyes" arrived on earth as a torrent of numbers: numerical snapshots of what Spirit encountered from moment to moment. In any snapshot, a single number represented how much light fell on a particular place on the robot's cameras during that snapshot. A large number signified a lot of light, a small number meant very little light: in this respect, Spirit's camera is like the photoreceptor matrix in the human eye (recall Figure 3.1). The numbers held the secrets to locating objects within the Martian environment—the hills, valleys, rocks, and boulders of the red planet. But how could those secrets be unlocked so that scientists, back here on earth, can reconstruct the objects that Spirit encountered?

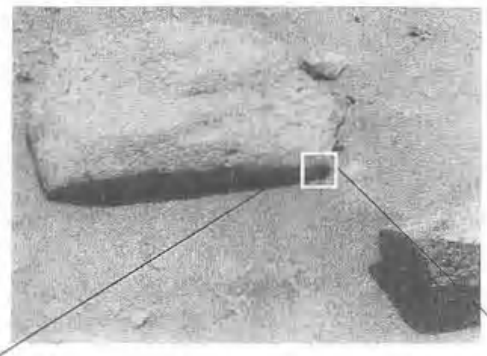
Figure 5.1 shows the collection of numbers produced when Spirit had its camera-eyes fixed on the distinctive rock formation that NASA scientist fancifully named Sashimi.[2] To transform the numbers into meaningful descriptions of Martian objects, researchers perform operations much like ones the human visual system performs when our eyes gather information from some complex scene here on earth. Of course, operations on data from Mars are carried out numerically, by powerful computers, but our visual system performs analgous operations neurally, using even higher-powered biological brains.

Spirit's numerical snapshots convey to us something about Martian objects. But how can we sift through the numbers to find signs that an object was present? For the answer to this question, we must exploit physical properties of matter. As described in Chapter 2, the patterns of light reflected from objects depend on the surface properties of those objects. Suppose that an object, let's say a rock, is composed of different material from its surroundings, say the dust-covered surface of Mars. Now, different materials (e.g., rocks and dust) usually differ in

[1] There are websites devoted to Spirit's activities on Mars, and these sites include some remarkable photos taken by the device's camera. Some of those websites are available through links at www.mhhe.com/blakeSen.

[2] Sashimi, as you probably know, is a Japanese delicacy consisting of slices of raw fish. Geological structures encountered on Mars are typically named on the basis of their resemblance to familiar objects on earth. Imagination plays a large role in this naming process.

FIGURE 5.1 | At the upper part of the figure is an image of a rock (named Sashimi) encountered at Gusev crater site on the Martian surface; the photograph was taken by one of the onboard cameras of the Martian Rover Spirit. Light levels of the pixels in a small square section of the image (outlined in white) were measured and digitized (converted into numerical values). The results of digitization are shown in the array of numbers at the bottom of the figure. High numerical values represent brighter pixels; low values represent dimmer pixels. You should have no trouble seeing the correspondence between clusters of numbers of regions in the image. This photograph, of course, is black and white, so the numbers just refer to light levels. For color photographs, each cell in the matrix of numbers would have three values specifying the amount of light at different bands of wavelength.



174	168	160	159	168	171	187	193	192	172	161	97	50	108	129	120	130
175	173	171	171	170	158	166	178	189	181	172	127	101	121	130	137	139
154	161	168	169	171	155	149	152	159	161	158	144	131	128	138	148	137
144	153	161	158	154	149	146	143	141	146	149	158	140	129	133	155	153
123	125	139	143	135	120	112	102	120	128	131	156	148	152	148	157	163
155	156	140	121	92	100	123	134	129	109	124	150	142	149	155	163	161
130	119	102	94	94	109	111	104	111	136	169	156	145	154	164	169	162
70	58	49	49	47	58	44	38	47	69	116	136	152	165	167	168	168
20	17	7	6	7	5	11	12	7	6	57	130	151	167	163	165	178
3	8	9	8	16	15	13	3	2	4	62	136	148	166	162	170	191
0	10	6	0	9	4	3	9	11	6	73	140	148	162	161	170	186
2	2	6	6	7	14	5	8	2	17	92	139	148	156	156	161	169
0	6	2	5	3	5	2	3	12	39	97	137	152	146	148	161	173
7	7	8	9	0	13	16	4	10	66	139	159	152	145	147	159	164
2	3	3	9	16	12	21	46	77	120	148	139	153	153	152	158	165
7	22	32	42	69	90	105	125	144	141	145	164	159	169	165	158	167
81	98	111	122	138	146	145	146	164	168	159	164	152	169	168	158	171
139	139	136	137	130	136	148	148	152	164	164	159	152	164	167	167	175
132	133	134	140	158	149	157	158	158	163	161	160	167	167	167	168	169
143	149	152	160	156	157	157	152	160	160	144	148	160	154	152	152	157

the amounts of light they reflect and in the wavelengths contained in that reflected light; those differences in light reflection thus can distinguish an object from its surroundings based on color, texture, lightness, and shape. A properly programmed computer can find regions in Spirit's Martian snapshots where clusters of numbers differ systematically from surrounding numbers. The computer, in other words, can "detect" an object (or parts of an object).

So, one of vision's basic aims is to identify regions where clusters of numbers (if done by computer image processing) or neural signals (if done by biological hardware) differ appreciably from surrounding numbers or neural signals. There are several ways that light reflected from an object can differ from the light reflected by the object's immediate surroundings. For example, the two might differ in intensity (bright versus dim light), spectral content (long wavelengths versus short wavelengths), or in the way that the reflected light is patterned (uniform versus uneven). These three kinds of differences produce perceptual qualities that are called lightness, color, and glossiness (recall Figure 2.16). Success in detecting an object or a part of an object depends on the *difference* between the light reflected by the object and the light reflected from its surroundings.[1] As Chapters 3 and 4 noted, neurons at multiple stages of the visual system are well suited to register such differences. The numerical array in the bottom part of Figure 5.1 summarizes the light levels within the tiny portion of the image of Sashimi outlined by the white square. Within this array of numbers you can easily see regions where the light levels change abruptly, suggesting the presence of different surfaces and, perhaps, different objects. So, identifying distinctive areas within an image is a crucial first step in machine vision and in biological vision, but of course it's not the only step. After all, distinctive areas do not by themselves tell us what we're looking at. Spirit's clusters of numbers are meaningless without further analysis, and the same is true for the cluster of intensity and wavelength values registered by the eye. The next step is to translate information about distinctive regions into information about form.

[1] As the intensity of an object's background increases, light reflected from that object must be more intense in order for the object to be seen. This is why you cannot see the stars in the daytime sky even though they are present and give off just as much light as they do at night. This is an example of the general principle known as Weber's law, which is discussed in the appendix.

What Defines "Form"?

You probably have some intuition about what is meant by an object's "form." The term typically refers to the attributes of size and shape. All objects have form. For nebulous ones such as clouds, the form may be changeable. For most biological objects, such as a person or an orange, however, stable form is more long-lasting. And for many nonbiological objects, such as the Martian rock named Sashimi, form is really long-lasting. Are there universal laws governing the perception of form? Is it useful to consider a given visual form as some composite of more elementary features? And if so, what are those features? How does the visual system extract form information from the retinal image? These questions are explored in the next several sections.

From the outset, keep in mind an important caveat about the study of form perception. Like other aspects of visual perception, form is the culmination of many processing steps, most of which transpire without your effort or awareness. Just looking at an object tells you nothing about the steps involved in deriving your impression of that object's form (Snyder and Barlow, 1988). To make this point concrete, consider an analogy. The end product of cooking can be a sophisticated dish, but just tasting the dish does not disclose the steps involved in preparing the dish. Complete understanding of the cooking process requires knowing the ingredients as well as the recipe by which those ingredients were prepared and combined. The same requirement applies to understanding form perception.

As we move through this and the next chapter, it is useful to distinguish between processing carried out within local spatial neighborhoods and processing carried out on a more widespread, global basis. This local/global distinction will become clearer as we move through this and the following chapter, but for now an analogy should help. The text on this page consists of words composed of letters constructed from simple lines and curves. Obviously to read the text requires that vision register the simple features comprising individual letters. We can construe that registration process as local since it deals with simple features in local regions of the retinal image of the page. But reading also involves clustering of letters into larger, coherent groups: words and then sentences. These grouping processes assemble local features into more global patterns that are meaningful to our task.

Our exploration of spatial vision and form perception begins with short-range, local analysis of the retinal

image, and then proceeds to longer-range, global analysis. For starters, we'll consider the question of form and its components from an historic perspective.

What Are the Components of Visual Form?

The structuralist tradition, which germinated in Europe during the late nineteenth century and flowered after being transplanted to the United States, had a simple guiding notion, namely, that ideas and perceptions are created by combining fundamental components. According to this view, simple sensations constitute the building blocks of perceived form. Seems reasonable, but how do we identify these simple sensations? Edward Bradford Titchener, a leader of the structuralist movement, championed a very simple method, called **analytic introspection.** He trained people to ignore all aspects of an object except its immediate, primary qualities. By attending only to an object's appearance and nothing else, people tried to enumerate the various sensations evoked by that object. For example, a person looking at an apple might report seeing four different colors and 13 different levels of lightness, with both color and lightness distributed over the apple's surface. According to the structuralists, these ingredients constituted the sensations from which perception of the apple was built. After analytic introspection had been in use for a while, the number of approved "elementary" sensations ballooned out of control, to more than 40,000 (Boring, 1942). Compared to the relatively modest number of chemical elements in the periodic table (currently, a few shy of 120), a set of 40,000 perceptual elements seemed excessive. To make matters worse, one person's reported elementary sensations often differed from those of another person. These disagreements were particularly striking because they involved people working in different laboratories. This suggests that the technique of introspection is susceptible to nonperceptual influences such as motivation and instruction. As you might imagine, what people *say* they see depends partly on what they *expect* to see ("Aren't apples typically red?").

In addition to the drawbacks just mentioned, introspection suffers from a logical flaw as well. Even if someone can learn to recognize a particular sensation when it occurs all by itself, there's no guarantee that the person will be able to recognize it when that sensation occurs in combination with other sensations. Is an apple's particular shade of red really independent of the glossiness of its skin or the color of the surface upon which the apple is resting? These difficulties seriously undermine the usefulness of analytic introspection as a tool for identifying "components" of perception. Nonetheless, one can embrace the structuralists' idea that perceptions are built from elementary sensations without subscribing to introspection as the method for isolating those sensations. But how can we identify the units of sensation with minimal contribution from our prejudices and expectations about what those units *ought* to be (Harris, 1998) and thus be confident we've got the right answer? One approach is to use physiological and psychophysical experiments that force the visual system itself to tell us the answer, even if those experiments produce answers that go against our intuitions.

A Local Analysis: Differences between Neighbors

Figure 5.2a is a photograph of downtown Nashville, Tennessee. Like other environments, natural or human-made, this scene comprises a range of different levels of spatial information, from coarse to very fine (Field, 1987; Tolhurst and Tadmor, 1997). Each level of spatial information defines a **scale.** To illustrate this idea, we have used a computer to break down the scene into three different sets of spatial scales, extracting and then displaying each scale separately.[4] The photograph in panel b captures the content of the scene on the broadest scale, the proverbial "big picture." Panel d contains the smallest scale of structural information from the scene—the scene's details. Panel c shows the scene's information on a scale intermediate to the other two. In panel e, you can see what happens when the three isolated scales, in panels b, c, and d, are put back together. The result is identical to the original. The differences in scale of the three images emphasize different aspects of the scene: the top image (b) emphasizes the broad layout of the scene, with skyline distinct from the river in the foreground; the middle (c) highlights individual buildings; and the bottom (d) image spotlights architectural details.

The objects we encounter every day contain spatial information on multiple spatial scales, and the features represented on those different scales harmonize with one another. But to highlight the concept of spatial scale, hybrid images can be created in which the various scales are in conflict with one another. An example of such an

[4] Any scene, this one included, actually contains a continuum of scales, not just three. We are working with three for ease of illustration.

FIGURE 5.2 | Photograph of Nashville, Tennessee, skyline (a) decomposed, or filtered, into component parts (b–d) depicting different spatial scales present in the original. Recombination of those filtered images recreates the original (e).

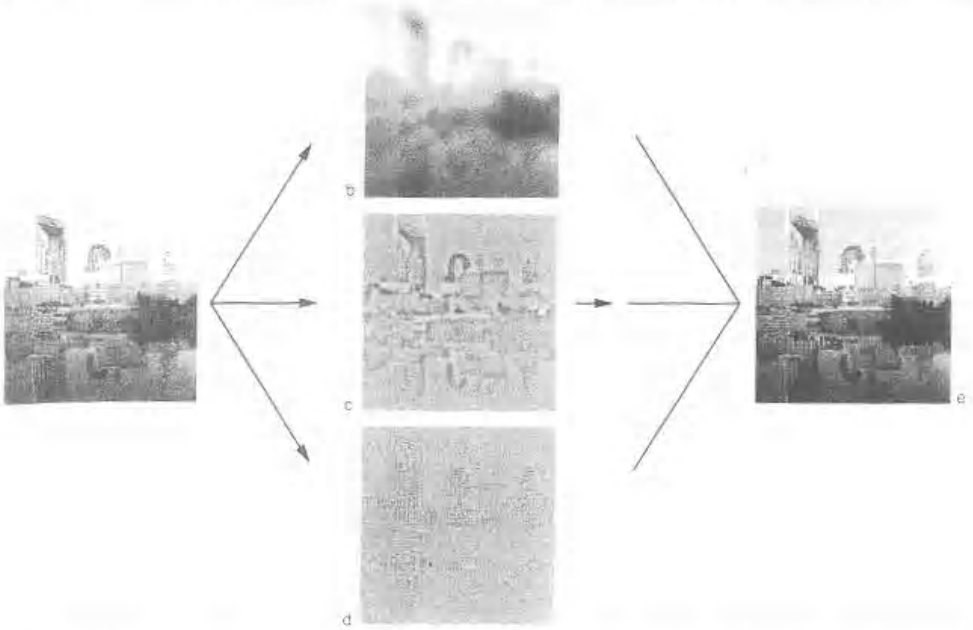


image is shown in Figure 5.3. Upon looking at this photograph, most people see the unsmiling face of a man. But this indicates that they're actually seeing only half of the picture. To create this photograph, Philippe Schyns and Aude Oliva blended two different faces—one an unsmiling male (carried by fine scale information), the other a smiling female (carried by low scale information). Although the image contains both faces, the unsmiling male is what we see because his face has been computer-processed to make it dominate the blend. Want to see the other member of the composite? Try squinting or defocusing your eyes as you look at the figure—you should see the other face, a smiling female. If you wear glasses, simply removing them may do the trick. If none of these tricks works for you, try dimming the lights or moving farther away from the picture until the smiling face appears. All of these maneuvers—squinting, dimming the lights, and so forth—alter the balance of information in the two composites, reducing the effectiveness of the small scale (detailed) information and allowing the form carried by the lower scale information to emerge. Ordinarily, of course, information registered at low spatial scale is concordant with information registered at high spatial scale, so you'd never see this kind of perceptual competition.

As we noted before, objects and scenes in the natural environment typically contain information at multiple spatial scales, from very fine to very coarse. A viewer's goals determine what components of this blend are paramount. Sometimes it's important to see the forest, (the overall, "global" view), while at other times it's the trees or even the leaves that interest us (the detailed, "local" view). Ideally, then, your brain's neural representation of a scene would contain information from all scales simultaneously.

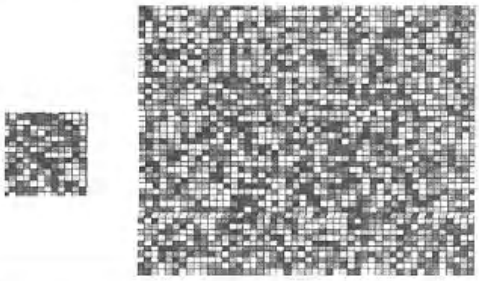
To probe vision's flexibility in using information on different scales, Schyns and Oliva (1997) employed hy-

FIGURE 5.3 | Two superimposed faces, one portrayed at a coarse spatial scale (smiling woman) and the other at a fine spatial scale (unsmiling man). Under normal viewing conditions, you will probably see that the male's face dominates the percept. The text explains how you can tip the balance in the direction of the other, female face.

brid images like the one in Figure 5.3. After each hybrid was briefly presented, observers reported what they saw. From these reports Schyns and Oliva inferred which spatial scale observers were relying on. For example, if you look at Figure 5.3 and report seeing a smiling woman, you must be accessing the figure's coarse scale of spatial information.

In the first phase of Schyns and Oliva's experiment, observers were sensitized to one scale of information—for some people this was the coarse scale, for the others, it was the fine scale. During this sensitization phase, observers saw hybrids containing meaningful information on only one scale; the other scale constituted a meaningless pattern. The aim of this sensitization phase was to "tune" or "prime" each observer's visual system to one scale or the other. Then, without warning to the observer, the composition of the hybrids changed. In this phase of the experiment, each hybrid contained meaningful images on both scales, coarse and fine. Nonetheless, observers continued relying on the scale they had been sensitized to ("coarse" for those sensitized to coarse, and "fine" for those sensitized to fine). So the sensitization experience dramatically influenced how observers saw the ambiguous test hybrids. Moreover, after the experiment, virtually all observers were surprised to learn that the test hybrids actually contained two different meaningful images since they were aware of seeing just one.

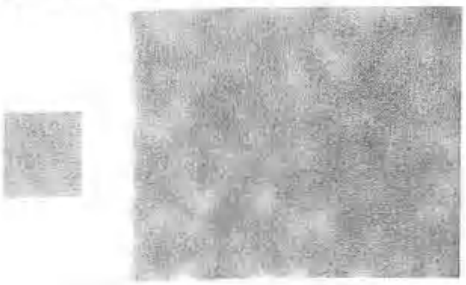
FIGURE 5.4 | The puzzle piece on the right fits somewhere in the complete puzzle on the left. Can you find the piece's location?



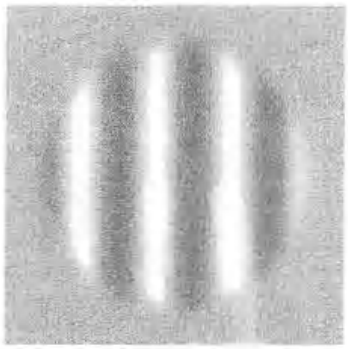
Access to representations of scenes and objects at different spatial scales allows the visual system to solve problems that would be intractable using just a single representation (Hildreth, 1986). You will learn about some of those potentially intractable problems in later chapters. For now, a simple analogy may illustrate the utility of creating neural representations of a visual scene on several different scales simultaneously. Think about the optimum strategy for assembling a complex jigsaw puzzle (a thought experiment we introduced in Chapter 4, when speculating about the lateral geniculate nucleus's function). Before attempting to fit together small, individual pieces of a puzzle, most people sort them into larger clusters based on common features. For example, we gather all the light blue pieces in a pile, knowing that they probably all go together to represent the sky; the dark brown pieces are placed in a separate pile because they'll make up the large tree trunk in the foreground. The large dark regions (foreground) and the large light regions (background) define the image's coarse spatial information. In effect, you're exploiting spatial scale information to get a head start on the assembly process. If you started by trying to assemble the pieces based on the picture's fine detail, the task would be overwhelming.

To illustrate spatial scale's utility, we've created a special jigsaw puzzle that's unusually hard to assemble. The assembled puzzle is shown in the right-hand part of Figure 5.4, and as you can see the puzzle portrays an irregular array of black, white, and gray checks. Notice that

This is the same puzzle and piece as in Figure 5.4, only now portrayed at a coarse spatial scale.



A sinusoidal grating pattern.



there is no meaningful structure in the layout of these checks. Shown on the left of Figure 5.4 is one small, square piece of this puzzle, and this piece appears somewhere within the completed puzzle on the left. Can you figure out where the piece originated? Taking a hint from the visual system, we've recreated the puzzle and the piece as they might be represented at some coarser spatial scale—these modified images are shown in Figure 5.5. Note that operating on coarse scale information, the task becomes easier—you probably have less trouble locating where this piece came from.

But this analogy is a little misleading. Assembling a puzzle involves figuring out which pieces interlink with one another. You begin by randomly shuffling the pieces to scramble their spatial arrangement. In vision, however, the two-dimensional spatial layout of the image is preserved; no scrambling is involved. But here's a key: There are neural operations in vision where the brain must compare the spatial arrangement of "pieces" from one image and match them to the spatial arrangement of "pieces" in another image. This is true for binocular vision (where the separate images are formed in the two eyes), and it's true for motion perception (where the successive images are formed over time); you'll learn about both in Chapters 8 and 9, respectively. For now, you should have a clearer understanding of the concept of spatial scale and its potential utility when it comes to image analysis.

But how is this multiscale analysis actually implemented in human vision? In 1968, two vision scientists, Fergus Campbell and John Robson, hypothesized that the visual system uses neurons with different-sized receptive fields to create a series of neural representations on different scales. Their theory was inspired by physiological studies of neurons in the visual system. At different levels of the visual system, responses of various neurons form a neural representation of the light that is falling on the retina. As discussed in Chapters 3 and 4, because receptive fields vary in size among neurons, the responses of different subsets of neurons constitute a neural representation at different spatial scales. Each subset of neurons conveys a characteristic kind of information defined by the spatial scale, so each set comprises an information *channel*. Because Campbell and Robson's theory relates vision of spatial form to activity within many such channels, it is known as the **multichannel model.**

To test their hypothesis, Campbell and Robson (1968) needed visual patterns whose properties could be varied systematically in ways that would have a varying, but predictable, impact on neurons in the visual cortex. Because cortical neurons are tuned to stimulus orientation and to spatial scale (size), the researchers needed patterns whose orientation and scale could be varied systematically. In addition, knowing that neurons' responses vary with image contrast (the difference between adjacent regions' light levels), they needed to specify and vary contrast as well. The visual pattern shown in Figure 5.6 satisfies these requirements. Such patterns, called **gratings,** can be generated on a computer monitor or television display. The one in Figure 5.6 is called a **sinusoidal grating** because the intensity of its light and dark bars varies gradually across the grating's width, in a sinusoidal fashion.

Gratings as Tools for Exploring Form Perception

What Are Gratings?

Gratings have four defining characteristics: **spatial frequency, contrast, orientation,** and **spatial phase.** These characteristics can vary independently of one another, which means that one can be changed without affecting the others. Let's examine these characteristics one by one.

Spatial frequency refers to the number of light and dark regions imaged within a given distance on the retina. One third of a millimeter is a convenient unit of retinal distance because an image that size subtends one degree of visual angle (recall the discussion of visual angle at the end of Chapter 2). To give you an idea of this angular unit of size, your index fingernail forms an image one degree in width when your finger is viewed at arm's length; a typical human thumb, not just the nail, but the entire width at the widest point, forms an image about twice as big, two degrees of visual angle (O'Shea, 1991). Recall from Chapter 2 that the size (or visual angle) of the retinal image cast by an object depends on the distance of that object from the eye. As the distance between the eye and the object decreases, the object's image subtends a greater visual angle.

The unit employed to express spatial frequency is the number of cycles that fall within one degree of visual angle (one cycle consists of one dark and one light bar). A grating of high spatial frequency—many cycles within each degree of visual angle—is made up of narrow bars. A grating of low spatial frequency—fewer cycles within each degree of visual angle—contains wide bars. Because spatial frequency is defined in terms of visual angle, a grating's spatial frequency changes with viewing distance. As this distance decreases, each bar casts a larger image. As a result, the grating's spatial frequency decreases as the distance decreases. To give you an example, when held at arm's length, the grating in Figure 5.6 has a spatial frequency of about 1.0 cycle per degree of visual angle. Doubling the viewing distance doubles the grating's spatial frequency to 2.0 cycles per degree.

Contrast is related to the intensity difference between the light and dark bars of the grating. If this difference is great, the grating's contrast is high; a small difference means the contrast is low. If the contrast is low enough, the bars of the grating may not even be visible. At sufficiently low contrast, the computer monitor or television screen would appear uniform and unpatterned; the grating contrast would be "below the threshold for visibility." At higher contrasts, the pattern would be visible. In Figure 5.3, the image of the unsmiling male face tends to predominate because the contrast of that image exceeds that of its companion female face image.

Quantitatively, contrast runs from 0 percent (when there is no difference at all between the intensity of the light and dark bars) to 100 percent (when the difference between light and dark bars is at its maximum).[5] The contrast of the print you are reading is about 90 percent; the contrast in the grating shown in Figure 5.6 is about 40 percent.

Orientation refers to the axis of the grating's bars. The grating in Figure 5.6 is a vertical grating. If it were rotated through 90 degrees, it would become a horizontal grating.

Spatial phase refers to a grating's position relative to some landmark. A convenient landmark is the left edge of the display. Looking at that edge, we can say that a grating "begins" with a dark bar, a light bar, or something in between. The gratings at the top and bottom of Figure 5.7 are in opposite phase; one begins with a light bar, the other with a dark one. The phase of the middle grating is midway between those at the top and bottom.

Phase specifies how components of an image relate to one another, which is a crucial aspect of form perception. In fact, when we create hybrid gratings composed of two components, the appearance of the hybrid depends on the relative spatial phase of the two. Figure 5.8 shows a pair of hybrid gratings consisting of a 1 cycle/degree grating added to a 3 cycle/degree grating (these spatial frequencies assume viewing at arm's length). In the right-hand hybrid, the two components were added so that their bright bars in the middle of the figure were superimposed. In the left-hand hybrid, the middle bright bar of one coincides with a dark bar of the other. Notice that these two hybrids are noticeably different even though they are composed of identical components; phase affects the appearance of the combination.

Manipulating these four properties of gratings—spatial frequency, contrast, orientation, and phase—we can construct any visual pattern, including a human face. To start, let's see how far we can get using just two orientations. Our plan is to start with a pair of gratings, one with bars oriented diagonally clockwise (CW) and the other with bars oriented diagonally counterclockwise (CCW), and add pairs of higher-frequency grat-

[5] There are several formulae for computing contrast. For gratings, though, contrast is equal to the difference between the light intensity of the lightest part of the grating and the light intensity of the dimmest part, divided by the sum of these two quantities.

FIGURE 5.7 | These three gratings differ in phase only.

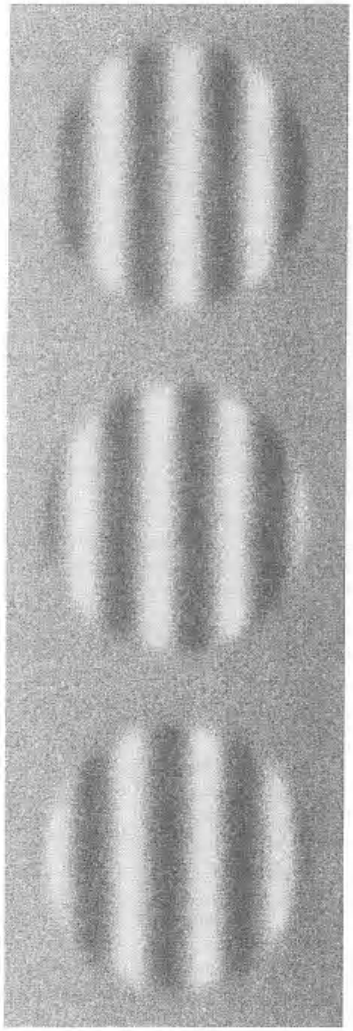
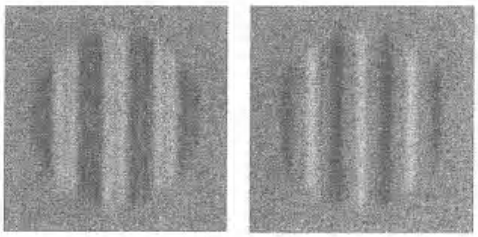


FIGURE 5.8 | The same components added in different phases yield distinctly different compound gratings.



ings, each additional pair consisting of a CW and a CCW component. Looking at Figure 5.9, in A you see the first two gratings superimposed. Adding in the next two components (1) results in B. Adding to this the next two components (2) produces the pattern shown in C. In each case, the added components are nothing more than a pair of obliquely oriented gratings of a particular frequency. Note two things about C's evolution. First, from A through C, the combination begins to resemble a plaid. Here, interactions between components are creating visual structures (diamonds) that are not associated with the individual components themselves. Second, from A through C, the diamonds become more sharply defined, even though only sinusoidal components with no sharp edges were used. Repeating a point made previously when discussing introspection, you cannot always tell what elementary components make up a pattern.

Now let's try to synthesize an even more complex figure, a photograph of a natural scene. This synthesis requires that we use more than just two orientations. Because it would take a great many frequency components to synthesize the natural scene, it would be tedious to show each individual step of the process. Instead, at each stage we'll add whole clusters of frequencies. Looking at Figure 5.10, A shows the frequency and orientation cluster that we'll start with. From this point, we follow the same procedure used in Figure 5.9, successively adding the clusters shown in 1 and 2 to create B and C. Note how each cluster makes its own unique contribution to the final product, C. Bear in mind that this complex scene results from adding various simple components that differ in frequency, contrast, orientation and phase. Notice, incidentally, that A, 1, and 2 represent the low, medium, and high spatial frequency information of the final product.

As just shown, spatial frequency components can be used to create a visual scene. Because they provide a vocabulary rich enough to express important aspects of visual form, gratings became a popular tool for studying

FIGURE 5.9 | Steps involved in generating a plaid.

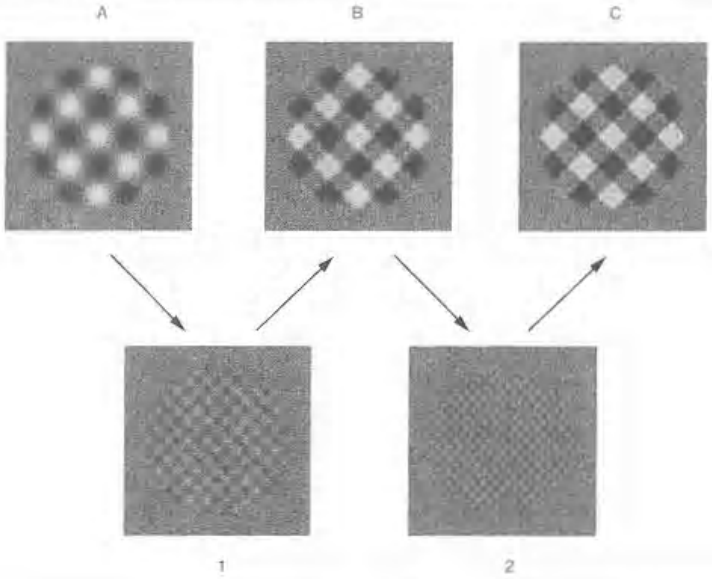


FIGURE 5.10 | Steps involved in generating a natural scene, in this instance, a photograph of flowers.

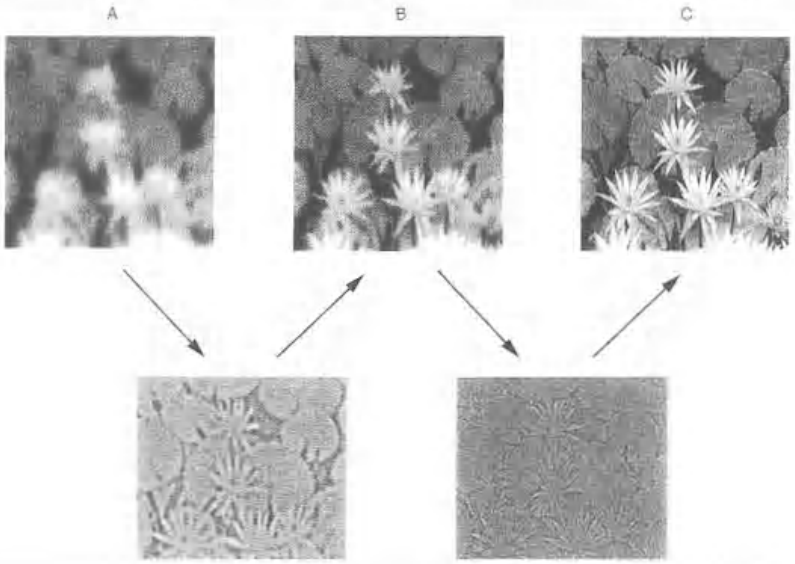


FIGURE 5.11 | Which of these three gratings appears highest in contrast and which appears lowest in contrast?

form perception. However, just because the photographic image of a scene can be synthesized from spatial frequency components does not mean that the visual system analyzes the scene into those components. Later in this chapter, we will consider evidence for scale-dependent analysis in human vision. Eventually, we'll show you how gratings are used to explore human pattern perception. However, since *human* vision is rather complicated, we begin with a much simpler system.

As you already know from Chapters 2 and 3, there are many parallels between the eye and a camera. As an entrée to human form perception, then, let's see how the performance of a camera's lens might be measured. A lens, as you know, is the optical component responsible for producing a focused image on the plane of the film or, with a digital camera, on the plane of the charge-coupled device (CCD). The quality of that image is only as good as the optical quality of the lens. The following section explains the technique we can use to measure the optical quality of a lens, a technique that can be generalized to measurement of the quality of human vision.

Using Gratings to Measure Performance

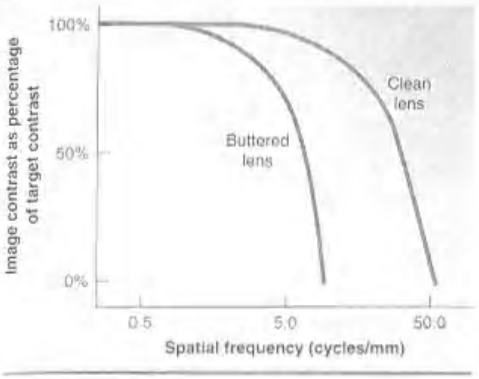
Measuring a lens's performance requires two steps. First, we use the lens to create an image of an object, preferably one like a grating whose spatial content can be quantified. Then we compare the image produced by the lens with the actual object. For example, using a lens, we would create an image of a grating of specified spatial frequency and contrast. We could then determine how good an image the lens had created. But "good" is an extremely vague term. How can we quantify it?

One approach is simply to judge the appearance of the image. But subjective impressions can be misleading. To illustrate, look at the three gratings in Figure 5.11 and rank them in terms of their apparent contrast (these gratings differ in spatial frequency). Most people would rank them in the order shown, with the leftmost grating deemed lowest in contrast. But this is wrong, for all three gratings have precisely the same physical contrast. In a moment, we'll consider why the perceived contrast differs from the actual contrast. But first, let's explore in greater detail this problem of evaluating image quality. A better, more objective way to assess the quality of images such as the one in Figure 5.11 is by means of a photometer, an instrument that measures light levels. The following example shows how this can be accomplished.

Suppose we use an expensive, high-quality lens to cast an image of a "target" grating on a clean, white paper. We can use the photometer to determine the intensity of the light and dark portions in the image and, hence, the contrast of the image of the grating produced by the lens. We'll repeat these measurements for different spatial frequencies, always using target gratings of the same contrast. We can graph the results in the following way. The horizontal axis of the graph will show spatial frequency; the vertical axis will show the image's contrast (as a percentage of the target's contrast). The resulting plot is termed a **transfer function,** because it specifies how contrast is transferred through the lens. Typical graphs of this kind are shown in Figure 5.12.

Look first at the curve labeled "clean lens" in the graph. Note that up to a certain spatial frequency the contrast in the image is identical to that of the target. For these frequencies, the lens faithfully reproduces the target. However, for still higher spatial frequencies, the contrast in the image is reduced even though the contrast in the target is constant. For these spatial frequencies, the lens reproduces the target less faithfully. The frequency

FIGURE 5.12 | Two transfer functions for a lens. The curves specify how contrast in the image formed by the lens is related to contrast in the object.



at which the image contrast falls to zero is called the **cutoff frequency.** Once the frequency in an actual target exceeds this value, the image will no longer contain any contrast whatsoever—the target itself might as well have zero contrast.

Now look at the curve labeled "buttered lens" in Figure 5.12. This curve connects the points we would observe if we repeated the experiment after having made one messy modification: smearing the lens by running a buttery finger over its surface. At very low spatial frequencies, the smear makes little difference in the performance of the lens. However, at intermediate and high spatial frequencies, butter on the lens degrades the contrast in the image. This is shown by the difference between the curves for the lens in its buttered and unbuttered states. Note also that the cutoff frequency for the buttered lens is lower than that for the clean lens. This difference between the curves makes intuitive sense: a high-quality lens excels at imaging fine spatial detail *and* coarse spatial detail, whereas a low-quality lens images only the latter. Incidentally, rubbing butter on a lens isn't the only way to degrade the lens's quality—smudges from fingerprints and accumulations of dust will have the same effect. (For those of you who wear glasses, think about how much sharper the world looks when you clean the lenses on your spectacles.)

But few people dedicate their cameras exclusively to taking photographs of gratings. To see how transfer functions such as those of Figure 5.12 can be applied to photographs of more natural scenes, the information in the scene must be related to gratings.

One method for doing this comes from the work of Jean Baptiste Fourier, a nineteenth-century French mathematician who happened also to be a friend and an associate of Napoleon. As part of a prestigious mathematics contest, Fourier had to develop equations expressing how heat is transferred from one body to another. He recognized that extremely complex equations would be needed and that those equations would have to be general enough to apply to a wide variety of different bodies. To satisfy these requirements, Fourier developed a powerful simplification. He showed that if some quantity (such as heat) changed in a complex manner over time, that complex function of time could be approximated by a series of simple sinusoidal functions. This simplification was an enormous breakthrough because it allowed Fourier to simplify a mathematically difficult function down into less complex, more tractable components. From here, the problem could be solved by working with the simple components (incidentally, Fourier won the contest). In recognition of his accomplishment, his technique is known as **Fourier analysis** (Bracewell, 1989; see also the websites at www.mhhe.com/blake5).

But how does Fourier's solution enable us to relate simple sinusoidal functions to a photograph of some scene taken through some lens? First, we treat the scene as the sum of a series of simple sinusoidal components. Then, using the lens's transfer function, we evaluate how the lens would image each of those components individually.

Consider the lens whose transfer function is given by the lighter line in Figure 5.12 (the lens smeared with butter). If we used that lens to photograph a scene containing many *very* fine details, the resulting image would be low in contrast and would appear *very* washed out. This is because fine detail is equivalent to high spatial frequency. As the transfer function shows, the buttered lens does a poor job of transferring high spatial frequencies. It reduces the contrast of any high spatial frequencies contained in a scene. Although this lens could faithfully represent the general shape of a large target (such as a tree that is near the camera), it would not be adequate for fine details (such as the wrinkles in the tree's bark). This illustrates that sinusoidal targets can predict the quality of a photograph produced by a lens.

To reiterate, several steps are involved. First, we determine the transfer function of the lens. Second, we analyze the visual scene into its spatial frequency components. With these pieces of information in hand, we determine which spatial frequency components will be preserved in the image of that scene and which will not. The first step, measuring the transfer function, is straightforward in the camera. But how easy is it to measure a transfer function for a visual system such as your own? If we did know your transfer function, we could predict the visibility of scenes you might look at. As we'll describe later, these predictions confer some practical benefits. Our next goal, then, is to derive a transfer function for human vision comparable to the one we derived for a lens.

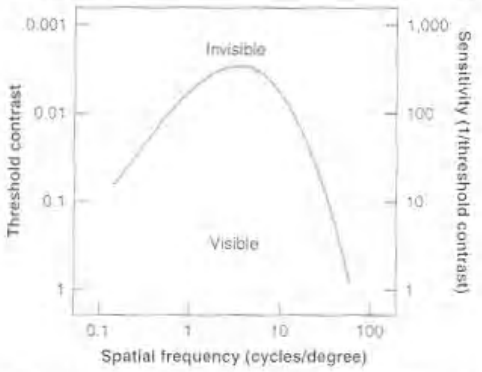
The Contrast Sensitivity Function as a Window of Visibility

The Human Contrast Sensitivity Function

There's one major stumbling block to measuring a transfer function for human vision: We cannot duplicate with humans the procedure employed with a lens. While we can produce sinusoidal gratings of known contrast, it's difficult to measure the image such gratings produce because that image is inside the eye. Besides, measuring this image would give only part of the visual system's complete transfer function. While describing the eye's *optical* components, this transfer function would not reflect the *neural* components of the visual system. And since we are interested in visual perception, not just the image formed in the eye, we must be concerned with the *perceptual* transfer function, which depends both on the optical transfer function and on the neural transfer function.

How, then, can we measure the perceptual transfer function? If your visual system (both its optical and its neural components) did a good job of transferring some particular spatial frequency, it stands to reason that you'd need little contrast to see a grating of that frequency. In other words, you'd be relatively sensitive to that frequency. However, if your visual system did a poor job of transferring that spatial frequency, you'd need more contrast to see it. You'd be relatively *insensitive* to that frequency. In general, the sensitivity of the visual system determines the threshold contrast needed to detect a given spatial frequency. By measuring **contrast thresholds** for different spatial frequencies, we

FIGURE 5.13 | A contrast sensitivity function for an adult human. The horizontal axis is scaled in units specifying the number of pairs of light and dark bars of the grating falling within 1 degree of visual angle on the retina. The vertical axis on the left specifies the contrast necessary for just detecting a grating; the vertical axis on the right plots sensitivity, the reciprocal of contrast threshold.



can derive a curve that describes the entire visual system's sensitivity to contrast. We call this curve the **contrast sensitivity function (CSF),** to distinguish it from the transfer function of a lens. The term *sensitivity* is a reminder that we are dealing with a property of the visual system, not just a property of the stimulus. As you'd expect from everyday usage of the term *sensitivity*, someone is said to have high sensitivity if that person requires little contrast to see a pattern. By the same token, someone is said to have low sensitivity if that person requires considerable contrast to see a pattern. Defined in this way, sensitivity is inversely related to threshold contrast.

Figure 5.13 shows a CSF for a human adult. This curve defines the adult's **window of visibility.** Before explaining the importance and usefulness of the CSF, let's describe how it can be measured. A test grating is created electronically on a specially designed and calibrated television screen. The screen displays a grating of fixed spatial frequency. Using a knob (like the contrast control on a television set), the observer adjusts the contrast until the bars of the grating are just barely visible—turning

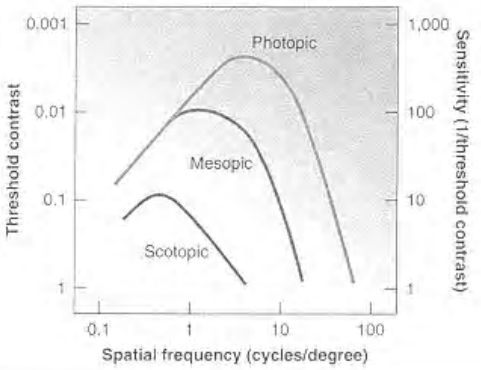
down the contrast any further would reduce the grating to invisibility.[6] This barely visible contrast value is then recorded, and the procedure is repeated several more times to produce multiple estimates of the visibility threshold for that spatial frequency; the average of these settings is defined as the "contrast threshold" for that test grating. Now we repeat the entire procedure for other test spatial frequencies, generating contrast threshold values spanning a range of spatial frequencies. The resulting set of threshold values are plotted in the graphical format shown in Figure 5.13. The horizontal axis specifies the grating's spatial frequency, plotted as the number of cycles within a degree of visual angle. The vertical axes plot the minimum contrast required to see the grating, with the left-hand axis plotting these values in units of contrast and the right-hand axis plotting values as the inverse of this contrast value (defined as sensitivity). To see your own CSF, use the web resources listed for Chapter 5 at www.mhhe.com/blake5.

This curve defines a window of visibility: The region underneath the curve represents combinations of contrast and spatial frequency that can be seen, while the region above the curve represents combinations that cannot be seen. To clarify this idea, pick any point on the CSF curve. Because this point is the threshold contrast for seeing that pattern, decreasing the pattern's contrast (moving upward from the curve) renders the pattern invisible. Conversely, increasing the pattern's contrast (moving downward from the curve) makes the pattern more visible.

Note that in one respect the *shape* of the human CSF resembles the shape of the transfer function of a lens (Figure 5.12): Both curves display a high-frequency cutoff. However, in another respect the two are different. In particular, the CSF drops at low frequencies, whereas the function for the lens does not. The visual system, in other words, is less sensitive to very low spatial frequencies than it is to intermediate ones. As a result, there is a range of spatial frequencies, toward the center of the horizontal axis in Figure 5.13, where humans are maximally sensitive. Gratings are less visible if they lie on either side of this optimum spatial frequency; a person requires higher contrast in order to see them. The same line of reasoning can be applied to a visual scene or photograph of that scene. If the objects in a scene have most of their spatial frequency information

[6] This method of measuring visibility thresholds is called the *adjustment method*, and it is described in the appendix. Also in the appendix are other techniques for estimating visibility thresholds.

FIGURE 5.14 | Contrast sensitivity functions measured at three light levels.



around the optimum point on the CSF, those objects will be clearly visible even when they are low in contrast. If those objects contain only very low spatial frequencies (very large objects) or only very high spatial frequencies (very small objects or fine details), they will be less visible and their contrast will have to be high in order for those objects to be seen. This also explains why the gratings in Figure 5.11 *appear* different in contrast: their apparent contrast varies with the sensitivity to different spatial frequencies.

You know from experience that you are able to see better under some conditions than others. If the CSF and your ability to see are importantly related, conditions that change one should also change the other. In fact, this is precisely what happens. Let's consider one such condition.

As discussed in Chapter 3, resolution is poor under scotopic conditions. That is why it's hard to read in dim light. Because resolution involves seeing fine detail, we'd expect decreased light to affect particularly that portion of the CSF corresponding to fine detail. Indeed this happens, as the curves in Figure 5.14 illustrate. The upper curve shows the CSF measured under photopic (daytime) light levels; the middle curve shows a CSF measured under mesopic (twilight) conditions; the lowest curve depicts the CSF associated with scotopic (dim light) viewing conditions. As the level of light decreases from daylight to twilight, visual sensitivity drops primarily at high spatial

BOX 5.1 Putting the CSF to Work

The contrast sensitivity function captures aspects of human vision that escapes measures such as visual acuity. Visual acuity, as described in Chapter 3, is a measure of the finest spatial detail that the visual system can resolve. Visual acuity measures only the spatial (size) factors that limit vision, so other factors are optimized. For instance, when an eye chart is printed with very light gray ink on a gray card stock rather than with very black ink on a white card stock, the letters are harder to see (Regan, 1988). In this case, the reduced contrast of the letters limits visual acuity, preventing one from assessing performance on the basis of size alone. When measuring visual acuity, then, one tries to optimize contrast and illumination so they do not limit performance.

When measuring visual acuity, one is interested in how size *alone* limits vision. When measuring the CSF, one is interested in how *both* contrast and size limit vision. In fact, two people can have exactly the same visual acuity but have different contrast sensitivities, as the following demonstrates.

Arthur Ginsburg and his associates used the CSF to predict how well pilots would be able to see objects in the air and on the ground. At least under conditions of reduced visibility (twilight or fog, for example), visual acuity gives a poor account of a pilot's visually guided performance. In fact, very fine details that might normally be seen are invisible at twilight or in fog. Ginsburg found that a pilot's CSF was a good predictor of the pilot's ability to see targets on the ground from the air.

Ginsburg tested pilots in a sophisticated aircraft simulator that provided a panoramic view through the plane's windscreen. Pilots flew simulated missions and then landed. On half of their landings, an obstacle (another plane) blocked the runway, requiring the landing to be aborted. Ginsburg determined how close each pilot came to the obstacle before aborting the landing. Even though all were experienced jet pilots, they varied in the distance at which they could spot the obstacle. The best pilots saw the obstacle three times farther away than did the worst. Significantly, pilots with the highest contrast sensitivities were able to see the obstacle from the greatest distances. Pilots with the lowest contrast sensitivities had to get closer to the obstacle before it became visible. Visual acuity was unrelated to the performance of the pilots on this test (Ginsburg, Evans, Sekuler, and Harp, 1982).

In another application of contrast sensitivity, Mary Jo Nissen measured the CSFs of patients with Alzheimer's disease (Nissen et al., 1985). Although this progressive disease of the brain is best known as a destroyer of memory, in many patients, Alzheimer's disease brings other symptoms, including diminished vision. Nissen's patients showed substantial but varying degrees of contrast sensitivity loss, affecting both coarse (low frequency) and fine (high frequency) patterns. These findings raise the possibility that some portion of the cognitive losses shown by Alzheimer's patients results from diminished vision (Sekuler and Sekuler, 2000).

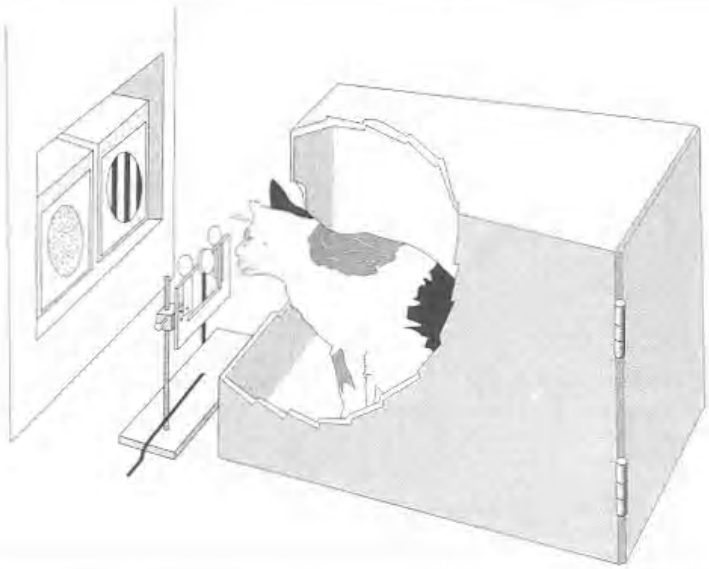
These are just two examples of CSF's usefulness in predicting visual performance in everyday settings. Other uses include predicting how well various visually impaired people can get around in their environments (Marron and Bailey, 1982), gauging the disabling effects of glare from various types of lighting sources (Carlsson et al., 1984), and enhancing printed materials for use by the visually impaired (Peli and Peli, 1984).

frequencies; lower frequencies are little affected. But when light levels drop even further to nighttime values, sensitivity decreases at low frequencies, too.

Think about what these curves imply for your vision as conditions of illumination change. Driving at night, you may be unable to see the fine details (high frequencies) of the shrubs alongside the road. At the same time, you probably will be able to see larger objects (low frequencies), such as another car, just about as well as you do under daylight conditions. If you park your car in an unlit place and turn off your headlights, you will be operating under light conditions like those producing the scotopic curve. Under such conditions, even large objects will be difficult to see. Box 5.1 gives several examples of practical and medical uses of the CSF.

In summary, the CSF characterizes the ease with which people are able to detect objects of various sizes and perceive the structural detail, such as texture, of those objects. Conditions that alter the CSF, such as light level, change the visibility and appearance of objects. In a sense, these conditions thrust one into a different visual world. From this it follows that creatures with different CSFs must actually perceive different visual worlds. The next two sections develop this intriguing possibility.

FIGURE 5.15 | Apparatus for testing the cat's contrast sensitivity.



CSFs of Other Species

Cats are notorious for seeming to watch things that their human owners can't see, as if there were ghosts present. Such inexplicable behavior enhances the cat's reputation for spookiness. While we don't deny that cats can be spooky, it may well be that under certain conditions they actually do see things that are real, though invisible to their owners (Blake, 1988a). We'll explain what we mean by this, and in so doing, show how the CSF allows you to compare your vision to that of other animals.

Just as it does for humans, the CSF defines a window of visibility for other species. Provided they have sufficient contrast, objects producing retinal images composed of spatial frequencies falling within the range of a creature's CSF will be visible to that creature. Objects producing images composed of frequencies outside that CSF will be invisible, regardless of their contrast. Thus, if you know some animal's CSF, you can predict what that animal will be able to see and what it won't be able to see. But how do you measure the CSF in a nonverbal animal such as the cat? The basic problem is to determine how little contrast a cat needs to distinguish a grating from a uniform field of the same brightness. One setup for doing this is illustrated in Figure 5.15.

The cat faces two adjacent television screens, one displaying a grating, the second a uniformly bright field. The cat is rewarded with a morsel of food if it pushes against a small plastic nose-key located just in front of the television displaying the grating. The experimenter randomly presents the grating on the left or right television screen. When the contrast of the grating is above the cat's threshold, a hungry cat will respond correctly on virtually every trial. When the contrast falls below the cat's threshold, the animal will simply guess and will be correct only half the time. The cat's threshold, then, is defined as the contrast that allows it to respond correctly on 75 percent of the trials, a level of performance midway between chance and perfection. This threshold contrast can be measured for different spatial frequencies, and the cat's CSF can then be plotted in the same way as it is for humans. (This procedure, incidentally, constitutes a forced-choice threshold test, and the details of this method are covered in the appendix.)

FIGURE 5.17 Contrast sensitivity functions for cats and human beings.

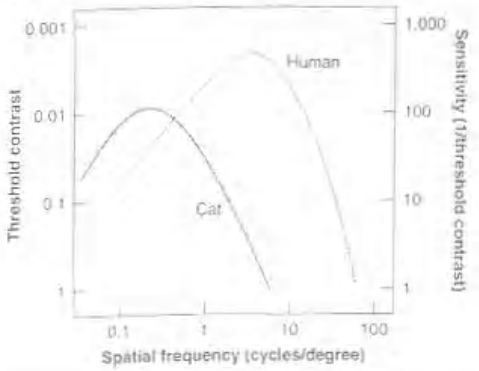


FIGURE 5.18 Contrast sensitivity functions for five species.

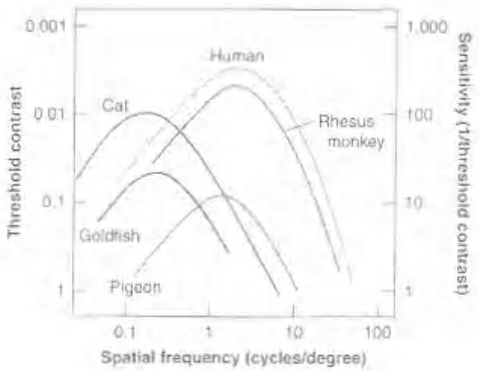


Figure 5.16 shows a typical CSF for a cat (grey line) and for a human (colored line) tested under comparable conditions. Note first the area common to the two CSFs. This overlapping region defines combinations of spatial frequencies and contrasts that both you and a cat can see. Next, note the regions where the two CSFs do not overlap. Within these two regions, one creature—you or the cat, depending on which one has the higher sensitivity—can see patterns that are invisible to the other. At high spatial frequencies, your sensitivity is better than the cat's; at low spatial frequencies, the reverse is true. Now suppose that a cat is sitting on your lap. If you are watching television, you will be able to see fine details in the picture that your cat cannot see. This is because those fine details are composed of high spatial frequencies that fall outside the cat's window of visibility. At the same time, if something large and very low-contrast appears in the room—say, an indistinct shadow on a wall—the cat may see it even though you cannot. In this instance, the shadow falls outside your window of visibility. These large, low-contrast objects could be the invisible "ghosts" that enhance the cat's reputation for spookiness.

CSFs have been measured for more than a dozen different species, and it's instructive to compare them. Figure 5.17 shows several representative curves. Look first at the curve labeled "Human"; this is the human adult CSF that you've seen in several previous figures. Next note the curve labeled "Rhesus monkey." Like several other nonhuman primates that have been studied, the rhesus monkey's CSF is highly similar to that of a human, suggesting that the world would probably appear very similar to you and to a monkey sitting on your lap. It is certainly unlikely that you'd see things that the monkey could not, or vice versa. Next note that the goldfish's CSF is displaced toward lower spatial frequencies, a fact that makes sense considering where the goldfish lives. Its aquatic environment prevents high spatial frequencies from ever reaching the fish's eye (Lythgoe, 1979). Like the cat, the goldfish is equipped for seeing either very large objects or smaller ones that are quite nearby. In general, there seems to be a good fit between what an animal uses its eyes for and where its CSF lies along the spatial frequency scale.

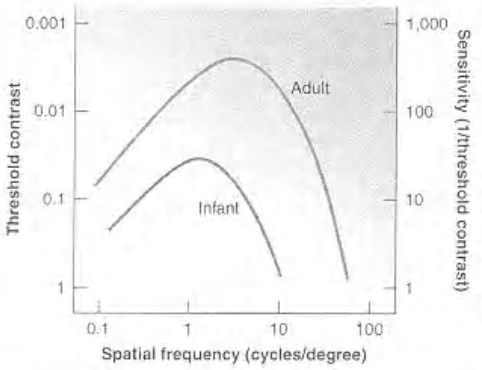
Age and the CSF

On the basis of differences between the CSFs of various species, we've pictured what the world might look like to the owners of these CSFs. In the same way, we can consider how the world might appear to human individuals at various points in their lifetimes. To begin, suppose while you are reading this book, someone puts a human infant on your lap (where previously you've held a cat and a monkey). How does your visual world compare to that of the infant? This question, incidentally, has intrigued philosophers and parents for centuries.

It is hard to know what very young, preverbal infants see. Obviously, you can't use the same methods to study

FIGURE 5.18 | An infant's preference for looking at a pattern can be used to measure the infant's contrast sensitivity.

FIGURE 5.19 | Contrast sensitivity functions for an infant and for an adult.



infant vision that you use with cooperative, attentive adults. To get around this limitation, researchers have exploited a naturally occurring tendency exhibited by infants: they prefer to look at complex rather than dull scenes, and this preference can be exploited to measure the infant's ability to see gratings (Atkinson and Braddick, 1998). Here's how it works.

Confronted with a patch of grating and a patch of uniform brightness, an infant prefers to look at the grating, presumably because the presence of the contours makes the grating more interesting than the blank patch (see Figure 5.18). During testing, the bars of the grating can be made sufficiently low in contrast that the infant cannot detect the contours and, therefore, cannot tell the difference between the patch of grating and the blank patch. In this case, the infant will show no preference for the grating over the uniform field. So, we can estimate the infant's contrast threshold for a given spatial frequency by fixing the grating spatial frequency and varying its contrast over a series of trials. By doing this, we can discover the minimum contrast at which the infant exhibits a preference for looking at the grating—this value defines the infant's contrast threshold at that spatial frequency. And by repeating this procedure at different spatial frequencies, we can derive the entire CSF for the infant.

The basic findings are summarized in Figure 5.19, which shows CSFs for an infant somewhere between 3 to 6 months old and for a typical adult. Note that the infant's window of visibility is very different from the adult's. An infant held on your lap will not be able to see fine spatial details visible to you. In this respect, the infant more closely resembles a cat. But unlike a cat, the infant does not have an advantage over you at low frequencies: You should be able to see everything that the infant can see. Also, even for spatial frequencies visible to both of you, the infant will require more contrast than you do.

The CSF in Figure 5.19 delimits the range of visible spatial frequencies visible to the infant, but it doesn't portray what the infant sees when looking about the natural environment. To get some idea of what the world looks like through an infant's eyes, we have simulated infants' quality of vision at several different ages. The simulations, shown in Figure 5.20, are grounded in measurements of contrast sensitivity (Teller, 1997).

In a sense, these simulations confirm what some parents have noticed: Their very young infants seem oblivious to everything except very large, high-contrast objects (Banks, 1982). Incidentally, the lack of sensitivity to high frequencies does not stem from optical causes but from the fact that the infant's immature visual nervous system fails to encode high frequencies. In effect, infants are best suited for seeing objects located close to them (recall that spatial frequency is distance dependent), which makes sense from a behavioral standpoint.

The infant's CSF improves steadily over the first year or so of its life. However, this improvement stalls at an immature level if the infant does not receive normal visual experiences. Several visual disorders can alter the quality

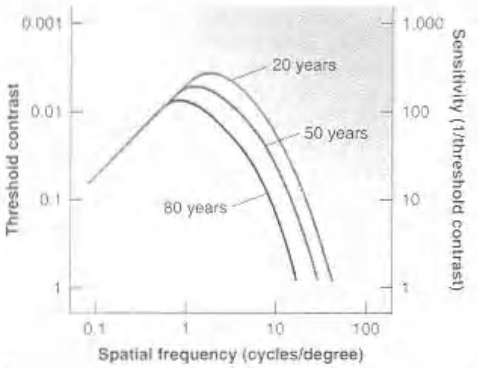
FIGURE 5.20 | This series of photographs shows what the world would look like through the eyes of infants at different stages of development. As acuity improves with age, the infant is able to see finer and finer spatial detail

of visual experiences received by an infant and hence keep spatial vision from its normal course of development. First of all, any condition that chronically blurs the images reaching the infant's retina will limit the information available to the visual system. Optical blur of this sort can result from myopia or hyperopia, or from congenital cataracts or corneal scars (recall our discussion of optics in Chapter 2). Misalignment of an infant's two eyes can also retard the development of good spatial vision. When its eyes are not properly aligned, an infant must suppress or ignore visual input from one eye in order to avoid seeing double (a condition we will discuss in Chapter 8). For reasons still to be learned, continuous suppression of one eye can lead to a loss in spatial vision, a condition called *amblyopia* (a visual disorder introduced in Chapter 4).

Fortunately, infants afflicted with any of these disorders can recover normal spatial vision, providing the disorder is corrected sometime during the first few years of life (von Noorden, 1981; Maurer et al., 1999). But if correction is postponed until the child reaches school age, the prognosis for full recovery is much poorer. Apparently there is a critical period early in life when the visual nervous system requires normal input to mature properly. During this period, neural connections are still being formed. This critical period of neural development ends by the time a child reaches 3 or 4 years of age. If the visual nervous system arrives at this stage incompletely developed because of inadequate visual experience, any neural abnormalities are irrevocably preserved throughout the remainder of life.

Because early visual experience is crucial to the brain's normal development, visual disorders in infants must be detected and then corrected as early as possible. The realization that early detection and intervention are vital has revolutionized the way that pediatric ophthalmology and pediatric optometry treat human infants. For example, if congenital cataracts are left untreated for several years, the development of normal connections in a child's brain is retarded, and even if surgery is done later in life, the quality of vision is unlikely to be optimal.

FIGURE 5.21 | Contrast sensitivity functions for three age groups.



So far our discussion has focused on spatial vision in infants and young children. Consider now what happens to the CSF during the remainder of the life span. The CSF remains more or less stable through young adulthood, but after age 30, systematic changes in the CSF begin reappearing. Figure 5.21 shows how the CSF changes from age 20 to age 80 (Owsley, Sekuler, and Siemsen, 1983).

By now you should have no trouble interpreting what a CSF implies about a person's ability to see. Suppose an elderly aunt takes the place of that infant on your lap; you should be able to predict from Figure 5.21 how her visual world might differ from yours.

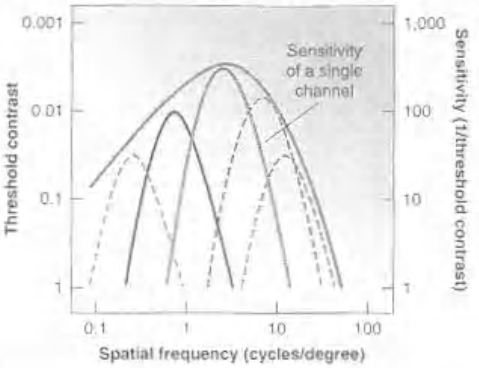
At least some of your aunt's diminished sensitivity to high frequencies results from optical changes in her eyes. For example, as she has grown older, her pupil has become smaller, which means that her retina receives considerably less light than yours (Weale, 1982). This reduced illumination of her retina mimics changes that would be seen in your CSF as you went from photopic to mesopic conditions (see Figure 5.14). Still, there remains some degree of loss in contrast sensitivity that cannot be accounted for by optical factors, implicating age-related neural changes (Bennett, Sekuler, and Ozin, 1999). Whatever its origins, changes in the CSF as well as in other measures of vision (Haegerstrom-Portnoy, Schneck, and Brabyn, 1999) document that people experience very different visual worlds at different stages of their lives. Society needs to be aware of these visual changes, so that older people can be helped to adjust to their changing visual world. To give just a couple of simple suggestions, providing large-print menus in restaurants can be very helpful to an older person whose reduced acuity makes it more difficult to read fine print. And providing extra illumination for reading that menu is a good idea, too, since the older person's smaller pupil is effectively reducing the illumination for that individual.

The Structural Basis of the CSF

As we mentioned previously in this chapter, Campbell and Robson (1968) proposed that the human visual system contains sets of neurons, each capable of responding to targets over only a restricted range of spatial frequencies. This range of "preferred" frequencies varies from one set of neurons to another. According to this hypothesis, the sensitivities of these frequency-tuned neurons, or channels, together determine the overall CSF (see Figure 5.22).

One reason for believing that the CSF depends on several different channels is the fact that certain conditions can alter one portion of the CSF without affecting others. As Box 5.2 explains, diseases attacking the visual system can produce this frequency-selective change in the CSF. (Evidently, subsets of neurons are differentially affected by disease.) But we don't have to wait for disease to change the CSF. We can change it intentionally, using a technique called **selective adaptation.** This procedure, mentioned in Chapter 4 in the context of orientation, produces a temporary loss of sensitivity to particular spatial frequencies.

FIGURE 5.22 | The overall contrast sensitivity function is determined by the individual sensitivities of neurons responsive to limited ranges of spatial frequency.



Measurement of selective adaptation entails several steps. First, you assess the person's CSF. Then, the person views a high-contrast grating for a minute or so, thereby adapting to one spatial frequency at a given orientation. Steady viewing of this adaptation grating decreases responses from visual neurons that respond to the grating (Movshon and Lennie, 1979; Albrecht, Farrar, and Hamilton, 1984). After the person has adapted to this grating for a minute or two, you redetermine the CSF (interspersing threshold measurements with additional adaptation to keep the level of adaptation high). The typical outcome of such experiments (for example, Blakemore and Campbell, 1969) is shown in Figure 5.23. In each panel, the arrow on the horizontal axis indicates the frequency to which the person adapted. Note that the notch carved in each CSF is centered about the adaptation frequency. So adapting to one spatial frequency diminishes sensitivity to that frequency and neighboring ones, leaving more remote frequencies unaffected. This process of selective adaptation works even when the adaptation frequency is contained in a complex pattern.

To understand how selective adaptation works, look again at Figure 5.22. If just one of the channels is fatigued by adaptation, that channel will respond more weakly when its preferred spatial frequency is presented. This means that a grating of that preferred spatial frequency would look somewhat washed out compared to its normal appearance (Blakemore, Muncey, and Ridley, 1973). A low-contrast grating that is normally visible will stimulate the adapted channel so weakly that the

BOX 5.2 When Things Go Wrong with Pattern Vision

The way in which diseases affect vision can provide clues about the structural basis of the CSF. Some neurological cases can produce a notch in the CSF—a loss of contrast sensitivity limited to a certain range of target sizes. Moreover, the location and severity of sensitivity loss can vary from one patient to the next.

In a study by Ivan Bodis-Wollner (1972), patients' notch losses resulted from stroke damage to the visual cortex. (A stroke is caused by an interruption of the blood supply to a portion of the brain.) Other diseases can produce similar losses. The most common of these diseases is multiple sclerosis, a disease that attacks the myelin on nerve fibers, including fibers that make up the optic nerve. Even though their visual acuity is good, some multiple sclerosis patients complain that the world appears "washed out." Presumably, this washed-out appearance is related to the nervous system's diminished capacity to code contrast.

In addition, about 30 percent of all people who have multiple sclerosis experience Uhthoff's symptom, a condition first described by Wilhelm Uhthoff, an eminent late nineteenth century ophthalmologist. For individuals with Uhthoff's symptom, exercise or emotional strain heightens their visual problems for several minutes. No one yet understands how emotional or physical strain or exercise causes these effects. But these conditions offer a unique opportunity to study the visual system under conditions of transient impairment. In one study (Sekuler, Owsley, and Berenberg, 1986), a 30-year-old accountant with Uhthoff's symptom reported that when he was emotionally upset his vision grew hazy and objects lost much of their apparent contrast. For example, during a confrontation at work, his boss's face appeared totally washed out and featureless. Surprisingly, during these episodes the patient retained his ability to read columns of small numbers, an important part of his job. Thus, he seemed to suffer a size-selective loss, retaining the ability to see fine details while losing the ability to see larger objects.

This strange set of symptoms was confirmed by comparing the patient's contrast sensitivity before and immediately after he exercised. Contrast sensitivity for high spatial frequencies was unchanged by exercise, but sensitivity for intermediate frequencies dropped. Longer periods of exercise produced an even greater loss of vision: Not only was contrast sensitivity drastically reduced at all frequencies, but visual acuity was seriously impaired as well.

We do not yet understand the specific physiological basis for these restricted losses in the CSF. But such losses do imply that the CSF—and the ability to see patterns of different sizes—depends on the coordinated responses of different sets of visual cells, any one of which may be attacked and damaged by disease.

FIGURE 5.23 | Adapting to a particular spatial frequency produces a temporary depression in the contrast sensitivity function at and near the adapting spatial frequency.

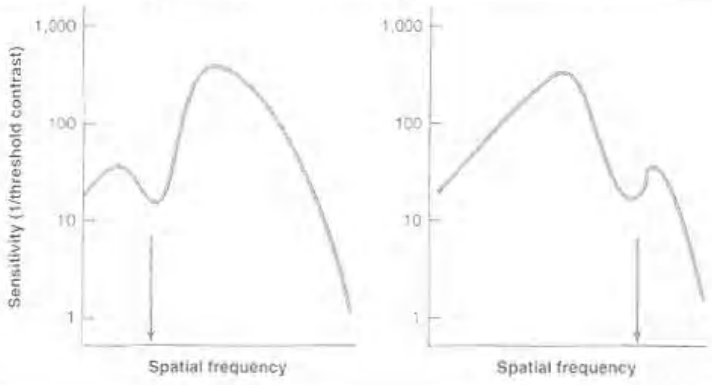
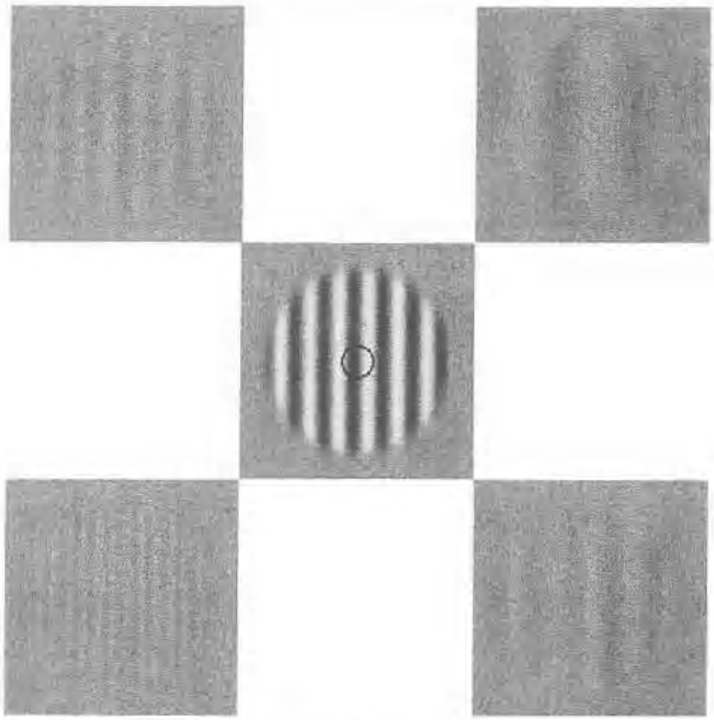


FIGURE 5.24 | By following the instructions in the text, you can experience the consequences of selective adaptation.



grating will be below threshold. In order to attain visibility, this grating's contrast will have to be boosted, resulting in a dip in the CSF. The width of this dip depends on the selectivity of the fatigued channel. Presumably, unaffected regions of the CSF depend on channels that were not fatigued by the adaptation grating.

You can demonstrate selective adaptation for yourself. Figure 5.24 shows several different low-contrast test gratings and one high-contrast adaptation grating. Before looking at the adaptation grating (in the center of the figure), note the apparent contrast of the test gratings surrounding it. Now inspect the high-contrast grating for about a minute. Don't stare fixedly, but allow your eyes to roam around the circle in the grating's center. Then, without delay, look at one test grating and note its apparent contrast. After looking back at the adaptation grating for another minute, note the apparent contrast of another test grating. By adapting and then examining each test grating in turn, you'll find that adapting alters the appearance of only some test gratings. Now you've seen for yourself that spatial frequency adaptation is selective.

Selective adaptation probably occurs in visual cortical cells (Albrecht, Farrar, and Hamilton, 1984). Such cells are selective for spatial frequency (size), but as Chapter 4 showed, they are also selective for orientation. As a result, you might expect grating adaptation to be selective for both orientation *and* spatial frequency. The logic behind this assertion resembles that used to explain spatial frequency selective adaptation.

You can demonstrate for yourself that this adaptation effect really is selective for orientation. Rotate this book by 90 degrees (turn it sideways) and then adapt to the high-contrast grating in Figure 5.24. This maneuver will orient the contours in that grating horizontally. After

adapting for a minute, quickly return the book to its upright position and look at the vertically oriented test gratings. Unlike before, *all* of the test gratings should remain visible. In other words, adapting to horizontal has no effect on the contrast threshold for detecting vertical, regardless of similarities in spatial frequency. This implies that the cells adapted by the horizontal grating are not at all involved in detecting vertical contours.

This demonstration is a reminder of one of the properties exhibited by visual cortical cells: orientation selectivity. Recall that another characteristic of cortical cells is their binocularity: Most of them receive input from *both* eyes. This implies that if you adapt just one eye to a grating, you will be able to observe the consequences of adaptation while looking through either the adapted eye or the unadapted eye (Blake and Fox, 1972). You can use Figure 5.24 to verify this fact also.

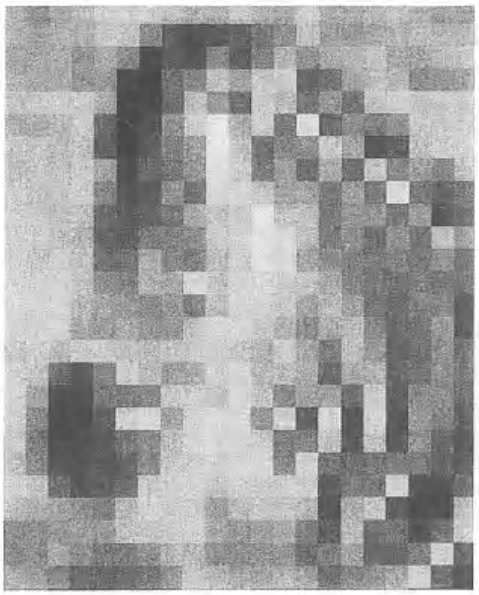
These adaptation effects imply that different neural channels are used to detect different spatial frequencies, since it is possible to affect the sensitivity of one without altering the others. And this same conclusion—different channels for different spatial frequencies—is also implied by results obtained using other techniques (see Wilson and Wilkinson, 2004, for a succinct review of those results). We'll not go into that additional evidence and, instead, we'll turn to evidence showing that these distinct channels strongly interact with one another. We'll start with a demonstration of this interaction and then consider why these interactions are a good idea in terms of visual object perception.

Interaction among Channels and Contrast Normalization

Look at Figure 5.25 and see if you can perceive a familiar object. Need some help? Try squinting your eyes, a maneuver that blurs your vision and thus reduces the high-frequency details in the retinal image of this picture (if you wear glasses, try viewing the photo with glasses off). Once those details are attenuated, the remaining low frequency information becomes relatively stronger and the object becomes conspicuous. In effect, the high frequency details contained in the edges of the blocks were camouflaging the lower frequency pattern that is present in Figure 5.25. (To understand how this block portrait was created, see Figure 5.26.)

This demonstration is just one piece of evidence pointing to the existence of strong interactions among the multiple channels revealed by selective adaptation. Results using the adaptation procedure point to the same conclusion. Consider, for example, the apparent contrast of a grating. Adaptation to, say, a set of horizontal contours *reduces* the apparent contrast of gratings of all orientations, not just horizontal, when those gratings are high in contrast; the same is true when the adaptation and test gratings differ in spatial frequency (Snowden and Hammett, 1996). So the perception of patterns at clearly visible levels of contrast, unlike their appearance at the contrast threshold of visibility, seems to depend on activity across multiple channels tuned to different orientations and to different spatial frequencies. This cross-channel interaction occurs, however, only among spatial frequencies and orientations imaged at the same location within the visual field. We know this because perceived contrast of a pattern at a given location is not altered by adaptation occurring within another portion of the visual field.

FIGURE 5.25 | Can you tell what is pictured here?



But what purpose do these interactions serve? One idea is that the responses of channels tuned to different spatial frequencies and orientations within a given, local region of an image are "normalized" based on the overall contrast among these components (Heeger, 1992). According to this idea of **contrast normalization,** visual re-

FIGURE 5.26 | The picture in the previous figure was constructed by setting the light level of each block so that it was the average intensity of the region covered by the block. Squinting (thus blurring the high frequencies) should make the detail in the two photographs more nearly equivalent.

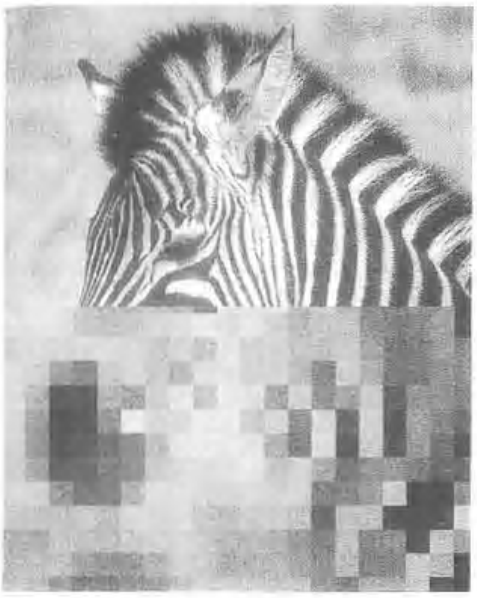
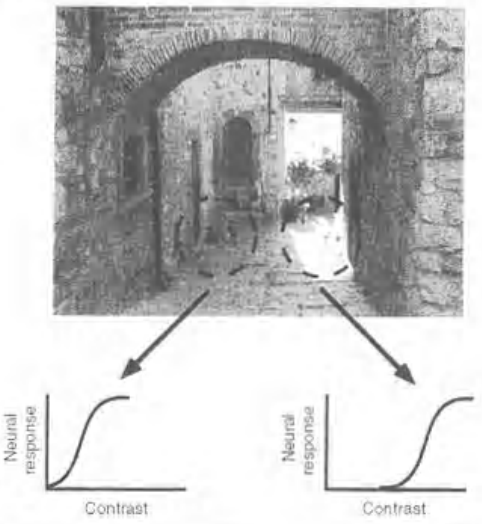


FIGURE 5.27 | Many natural scenes contain some regions where the local contrast among contours is relatively low and other regions where contrast is high. Visual neurons can respond to only a limited range of contrast variations, but through contrast normalization that range can be dynamically shifted to operate efficiently within the prevailing levels of contrast. The two graphs in the lower part of this figure show the contrast-response curves for visual neurons with receptive fields stimulated by relatively low contrast images (left-hand curve) and by relatively high contrast images (right-hand curve).



sponses are combined from all active neurons whose receptive fields fall within a local neighborhood, and this combined signal, in turn, modifies the individual responses of those active neurons; this process can occur very quickly. But what is the advantage of normalizing neural responses based on some combined level of responsiveness? There are two reasons for doing this (summarized in Schwartz and Simoncelli, 2001). The first reason has to do with maximizing the ability of neurons to respond over a wide range of contrast values. The response of a visual neuron increases with contrast, but a neuron can respond only so vigorously after which its response levels off at its maximum value. Because of the limited response range of the neuron, the range of contrasts over which the neuron's activity level varies is relatively narrow compared to the complete range of contrast values encountered during ordinary vision. It is as if the neuron has a limited range of numbers with which to specify any value within a broad range of numbers. Through contrast normalization, however, the neuron can adjust its range of responsiveness to cover the prevailing level of contrast within a local region of the image. This normalization process is illustrated schematically in Figure 5.27. In effect monitoring a given, restricted region of an ima their responsiveness to ensure that they can rep contrast levels contained within that region of t It is important for normalization to be spatially and to not extend over the entire image, because field of view is very likely to contain a range o values much too wide to be effectively accomm global normalization applied to all regions of t (see the image in the top part of Figure 5.27).

That's the first reason for pooling responses across multiple channels. The second reason has to do with maintaining the stimulus selectivity of visual neurons. Recall that a given neuron selectively responds only to a narrow range of orientations and a limited range of sizes (or spatial frequencies). In Chapter 4 you saw examples of orientation selectivity summarized in the form of a "tuning curve" (recall Figure 4.10), where the peak of the curve signaled the preferred orientation and the breadth of the curve represented the range of orientations capable of stimulating that neuron. Moreover, we saw in Figure 4.25 that the pattern of activity among these "tuned" neurons could signal the orientations of contours. Now, if the shapes of the neurons' tuning curves were to vary depending on the contrast of the contours activating those neurons, the patterns of activity produced by a given contour would change depending on the contour's contrast. And that would not be visually desirable, for then we could mistake changes in contrast for changes in orientation. Fortunately we don't make that mistake, and that's because tuning curves *retain* their shapes despite variations in contrast. And they retain their shapes because the responses of those neurons are normalized for contrast. In effect, the visual system is willing to tolerate slight errors in perceived contrast to avoid confusing errors in perceived orientation.

Contrast normalization also offers one possible explanation for why the object in Figure 5.25 is difficult to see unless you blur the image. The sharp edges of the blocks in Figure 5.25 contain strong contrast signals within the high spatial frequency range of the picture. These strong signals bias the normalization process toward higher contrasts and, thus, effectively weaken the contrast signals within the low spatial frequency components contained in the image. By blurring the image, the high spatial frequencies are removed, their contribution to contrast normalization is weakened, and the low spatial frequency contrast signal is boosted. (For other possible factors involved in this phenomenon, see Hess, 2004.)

To sum up, recent studies of detection and pattern visibility converge on a common view of the human visual system. In this view, detection of any spatial target depends on responses generated in a set of visual neurons tuned to contours of particular sizes and orientations. Each set of neurons is responsible for signaling the presence of targets on a particular scale and orientation. Moreover, the responses of those neurons are adjusted based on the level of contrast contained within the region of the image being stimulated by those neurons. In the next sections, we'll see how this model of spatial vision may account for other aspects of form perception as well.

Form Discrimination

You've now learned how the CSF defines a window of visibility within which objects can be seen. Our emphasis was on form detection as determined by an observer's sensitivity to contrast and spatial frequency. Now we are ready to extend our analysis of spatial vision to form discrimination, which is the ability to tell one object from another. Before getting into details, though, it will be useful to specify what must go on within the visual system to make discrimination possible.

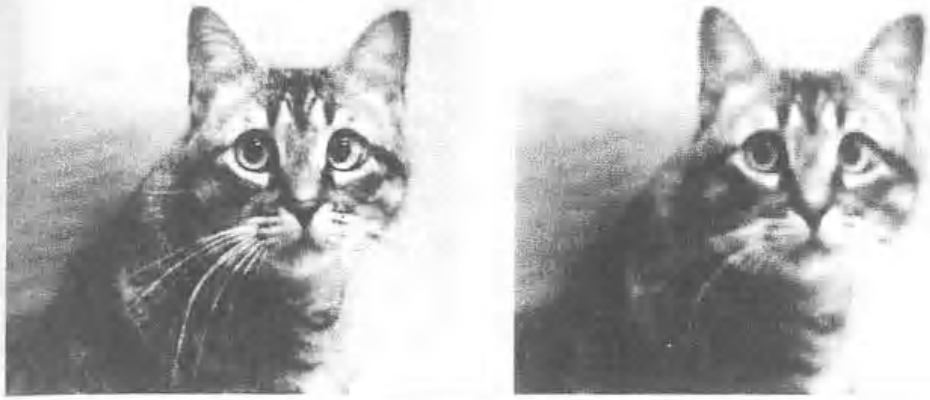
The perception of pattern, or spatial structure, depends on the responses of cells in the visual system. This dependence sets limits on what stimuli will appear different and what stimuli will look alike. It stands to reason that all stimuli producing *identical* patterns of neural activity within the visual system will be indistinguishable from one another.[7] But if stimuli produce different activity patterns, potentially you have some basis for telling them apart. Whether you actually can tell them apart depends on how different the neural responses are. In Chapter 4 you read that the response of any one cell, say a simple cell in the visual cortex, provides ambiguous messages about the characteristics of visual features of objects in the visual world. In all likelihood, the visual appearance of objects depends on the pattern of activity within an *ensemble* of cells, not just one cell alone.

Metamers

If you look at two exact duplicates of the same object, each produces precisely the same effects on your visual system, and they will look identical. But two objects that are physically different can also have identical neural effects and, therefore, appear identical. In other words, things need not *be* identical to *look* identical. Two objects that are perceptually indistinguishable from one another, despite their physical differences, are called **metamers.**

[7] In principle, "identical patterns" of neural activity means that the very same neurons produce the very same sequence of action potentials in response to two or more stimuli. Of course, it is highly unlikely that precisely the same neurons produce precisely the same pattern of activity upon repeated presentation of the *same* stimulus. So practically speaking, "identical" really means equivalent in a statistical sense, that is, with insignificant variability.

FIGURE 5.28 | The two photographs are identical except that frequencies falling above the cat's high-frequency cutoff have been eliminated from the right-hand photograph.



We can learn about neural processing in human vision by determining which visual patterns are metamers and which are not. Metamers exist because two or more stimuli, although different along some dimension (e.g., size) are insufficiently different to produce distinctive responses within the visual nervous sytem—whatever dimension distinguishes those stimuli is too subtle for the neurons responsive to those stimuli. Consequently, we are "blind" to the differences between the stimuli. When your visual system generates identical responses to two objects, the objects will appear identical, even if they are physically different; those two objects are said to be metameric. The following simple exercise will introduce you to one example of metamers, in this case a pair of pictures.

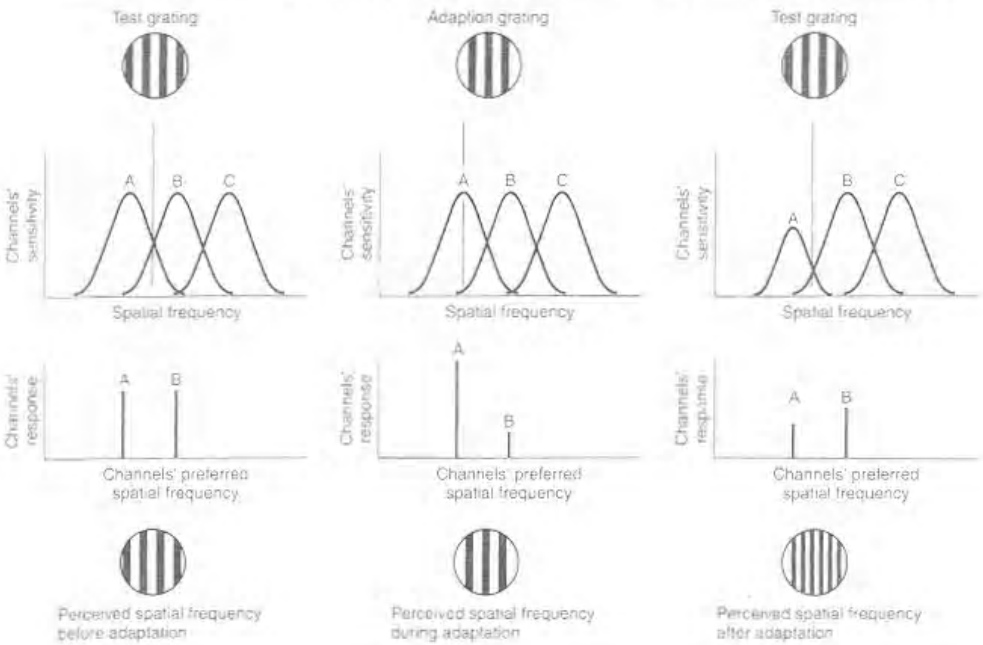
In Figure 5.28, the two photographs are identical except for the presence of high spatial frequency information in the left-hand photograph. You are probably able to see that the two photos are slightly different because, from this reading distance, these high spatial frequencies are visible to you; in other words, the two photographs are *not* metameric. But now prop the book up on a table and view these two photographs from a distance of about 15 feet. Now the high frequencies in the left-hand photograph will fall beyond the limits of your acuity, rendering the two photographs metameric: they appear identical even though the two are physically different. At the greater viewing distance, both photographs produce equivalent patterns of activity in your visual system. Viewing these two photographs under dim light illumination will also render them metameric even at ordinary reading distance. Looking back at Figure 5.14, you should understand why.

Incidentally, if a cat were to view this same pair of photographs from your normal reading distance under good illumination (conditions where you can tell the photos apart), the high frequencies in the left-hand photograph would be outside of the cat's window of visibility. Thus, the two photographs would be indistinguishable to a cat, meaning the pair would be metamers for the cat at both viewing distances.

But it's not only light level or distance that determines whether two stimuli are metameric. Any condition that alters the response of the nervous system can influence whether stimuli are metameric. You've already seen one example of this principle at work, when you experienced the temporary invisibility of a low-contrast grating following adaptation (Figure 5.24). Adaptation momentarily reduced the responsiveness of neurons tuned to that spatial frequency and orientation, to the point where those neurons were no more active than they are when confronted with an uncontoured, uniform gray field. The low-contrast grating, as a consequence of adaptation, was temporarily metameric with an uncontoured field.

Let's consider another visual phenomenon where we use adaptation to alter the responsiveness of the visual nervous system and, thereby, to change stimuli from metameric to nonmetameric. To do this, we turn to the

A neural account of how contour size is encoded in the visual system. (See text for details.)



size aftereffect, which is robust and easy to induce. Before experiencing the size aftereffect, let's preview why the effect happens. We start with the widely accepted theory that the visual system represents size by the patterns of neural activity among channels tuned to spatial frequencies (Blakemore, Nachmias, and Sutton, 1970). Figure 5.29 shows how a simplified version of this theory applies to the perceived size (i.e., spatial frequency) of a grating. In the upper two graphs, the horizontal axes represent spatial frequency and the vertical axes represent sensitivity. These curves, you may notice, correspond to several of the multiple channels we talked about previously (recall Figure 5.22). To make the argument easier to follow, only three channels are shown. As the upper left panel indicates, channel A prefers the lowest frequency of the three, B prefers an intermediate frequency, and C prefers the highest frequency.

Suppose we present a test grating whose frequency is halfway between the preferred frequencies of channels A and B. (This condition is not crucial for the argument; it merely simplifies the description.) Note that the test grating produces equal responses within channels A and B (bottom left panel). According to the theory, this neural code—equal activity in A and B—determines the way the grating looks.

But if we did something to change the distribution of activity produced by that test grating, the grating's appearance would be altered. Producing a change in activity can be accomplished using an adaptation procedure much like the one already described. First, a person looks at a grating and judges its spatial frequency. We'll refer to this as the "test grating." Next, the person adapts to a high-contrast grating whose spatial frequency is lower than that of the test grating. We'll call this the "adaptation grating." Finally, after a minute or so of adaptation, the person looks back at the test grating and again judges its spatial frequency. Does the test grating still look the same? To answer this question, consider how adaptation would have affected the neural code previously described.

As we have pointed out, the test grating produces equal responses in channels A and B prior to adaptation. Suppose also that the adaptation grating is one to which

FIGURE 5.30 | These samples cause the viewer to experience the size-shift aftereffect. (Follow the instructions given in the text.)

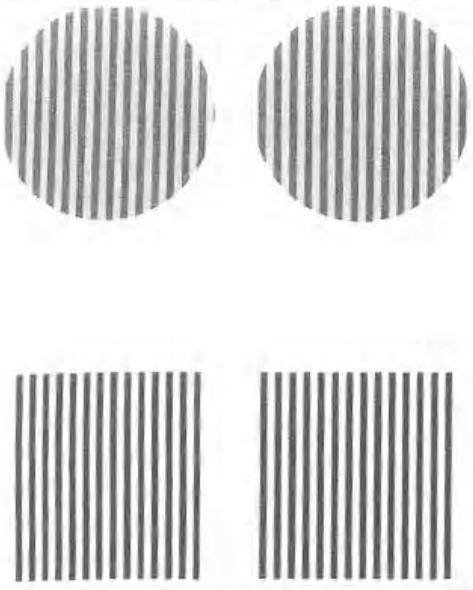
A is highly sensitive and to which B is relatively less sensitive (as illustrated in the middle panel of Figure 5.29). Prolonged adaptation will thereby reduce the sensitivity of A considerably but B very little—this selective change in sensitivity is illustrated in the upper right panel. Now when presented following adaptation, the test grating will evoke a nearly normal response from B but a weakened response from A (lower right panel). The test grating does not usually produce this pattern of responses—a greater response in B than in A. Instead, this pattern of responses is usually evoked by a spatial frequency higher than that of the test grating. As a result, after adaptation, the spatial frequency of the test grating should appear higher than usual. And this is precisely what happens (Blakemore, Nachmias, and Sutton, 1970). On the basis of this theory, can you predict what would happen if a person were to adapt to a spatial frequency higher than that of the test grating?

Using the test displays in Figure 5.30, you can experience one version of this size aftereffect yourself. To reassure you that this aftereffect is not peculiar to gratings, we've borrowed from Stuart Anstis a demonstration using more familiar patterns: alphabetic text. Still, the outcome is the same: a temporary change in apparent size. To make the effect easier to see, this demonstration simultaneously produces two size aftereffects in opposite directions. Here's how to use Figure 5.30. Look at the black dot between the upper and lower panels of the left side of the figure. With your eyes fixed on the black dot, verify that the letters in the text above the line look the same size as the letters below.

We'll now explain what you must do to alter the apparent size of these letters. Because timing is important, finish reading the instructions before you begin to execute them. First, adapt, using the right side of the figure. The patch of small, tightly packed letters corresponds to high spatial frequencies, while the patch of larger, more spread out letters corresponds to low spatial frequencies. To adapt, slowly move your eyes back and forth along the bar between the two patches of letters for about 90 seconds. *Caution:* Do not look directly at either patch; keep your gaze on the horizontal bar between the two. Otherwise, you'll intermix the two types of adaptation. Finally, at the end of adaptation, look back at the dot between the patches in the left side of the figure. Keeping your eyes on the dot, note the sizes of the letters above and below. You will see that letters in the upper patch look larger than those in the lower patch. Here, then, is a situation where altering your visual system's sensitivity makes ordinarily identical stimuli temporarily nonmetameric.

The point to be remembered from this discussion of metamers is that the ability to discriminate among objects depends on the extent to which those objects produce different neural responses. Carrying this idea one step further, the ability to make very fine discriminations among objects indicates that the underlying neural events are very sensitive to variations along whatever dimension(s) distinguish those objects. For example, compare the two patterns in the upper part of Figure 5.31. The bars in the circular patch on the left differ very, very slightly in orientation from the bars in the patch on the right. If you can see that these two are slightly different, this must mean that the two orientations are producing reliably different patterns of activity within your visual system. Next, compare the pair of patterns in the bottom part of the figure. Can you see that the bars in the figure on the left are ever so slightly narrower than the ones on the right? Again, if your answer is yes, then evidently your visual nervous system is producing discriminably different patterns of activity when you view one pattern and then the other. These tiny yet discernible differences between members of each of these pairs of patterns reveal that the visual nervous system is exquisitely sensitive to variations in size and in orientation.

FIGURE 5.31 | Visual figures that are discriminable, even if just barely, must be producing different patterns of neural activity within the visual system. Looking at the top pair of patterns, the orientation of the bars on the right is slightly different from the orientation of those on the left. The angular difference in orientation is only about 2 degrees, a value right at the threshold for discrimination. In the bottom pair of patterns, the bars on the right are about 3 percent larger than those on the left, again a difference close to the limits of discrimination.



Form Defined by Texture Differences

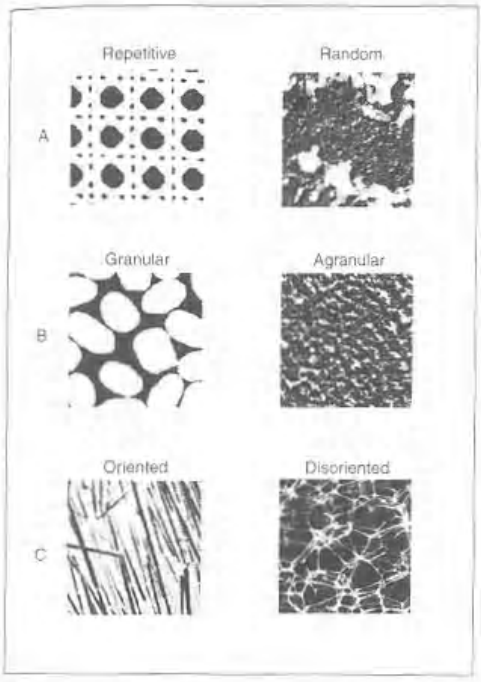
So far we have focused on detection and discrimination of forms defined by intensity differences that create spatial contrast at different spatial scales. Thinking back to the robotic vehicle Spirit as it explored the Martian surface, its camera would register the presence of a rock if that rock reflected a different amount of light than did the background upon which it was resting. But what if the rock and its background were the same color and reflected equivalent amounts of light? Would Spirit miss it altogether? Not necessarily. Rocks, after all, typically have surface markings—scratches, streaks, and other irregularities—that differ from the more regular, grainy texture of the dusty Martian ground upon which the rock rests. If NASA scientists had programmed Spirit and its image processing computers to register spatial discontinuities in texture (**texture contrast,** we could call it), the rock could be detected even though its overall lightness was no different from the ground. Texture contrast provides another, complementary, means for detecting and discriminating form. Let's spend a few moments considering texture.

To begin, think about familiar objects, whose textures are distinctive—for example, a pebbled path, the skin of an alligator, the bark of a tree, or a brick wall. Think about what makes each of these surfaces distinctive, putting color aside. As a first approximation to a definition, these examples suggest that texture depends upon the sizes and shapes of the surface's markings as well as the distribution of those markings over the surface. Can we be more precise about the key stimulus features that determine perception of texture in natural and human-made surfaces?

To answer that question, Rao and Lohse (1996) utilized as test materials photographs of 56 different surfaces taken at various magnifications and conditions of illumination. Based on similarity judgments made by people who viewed these images, Rao and Lohse identified three main physical dimensions that governed texture perception. One of the dimensions captures the repetitiveness, or regularity, of texture elements. Figure 5.32A shows examples of textures at opposing extremes of this dimension. The second dimension reflects the granularity of the texture. An extremely granular texture is composed of randomly distributed, large elements of roughly uniform size. Examples are shown in Figure 5.32B. The last prominent dimension reflects the presence of a strong local orientation within various regions of a surface. Examples of two extremes along this dimension are shown in Figure 5.32C. To make sure you understand their scheme, look around you, note three or four different textured surfaces, and try to place them into Rao and Lohse's framework.

So, Rao and Lohse's work provides a way to describe textures. Work by other investigators (for example, Nothdurft, 1985; Wolfson and Landy, 1998) shows that human observers are reasonably good at judging shapes defined solely by texture differences (a skill we'd like to embody in Spirit's circuitry). Two examples of shape defined by texture appear in Figure 5.33. In the right-hand figure, the "grain" of the central textured region sets it off from its surroundings; in the left-hand figure, the regularity of the texture distinguishes center from surround. It is important to note that neither of these rectangular shapes would be detectable if you had to rely on luminance differences alone, because the average light intensity in local regions throughout both figures is constant.

FIGURE 5.32 | Examples of textures representing ends of the perceptual dimensions identified by Rao and Lohse (1996). Row A: repetitive versus random; row B: granular versus agranular; row C: oriented versus disoriented.



Still, the central rectangular region can been seen immediately and with no effort—the shape appears to "pop out" from the background.

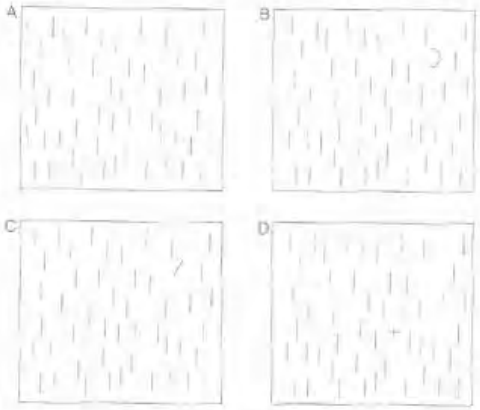
Anne Treisman (1986) and others have exploited this pop-out effect as a way to distinguish features that can be quickly and effortlessly registered from those that require time-consuming scrutiny. Some typical displays used in these kinds of pop-out experiments are shown in Figure 5.34. By presenting a "target" at randomly chosen locations among a field of "distracters," Treisman was able to measure how long observers took to locate the target. To distinguish targets that pop out from those that do not, judgment times were measured with varying numbers of nontarget, distracter elements. If a target popped out, the time needed to spot that target would be relatively unaffected by the number of embedded distracter elements. Pop-out, in other words, would imply that visual mechanisms are carrying out their searches for the target simultaneously, in parallel, throughout the entire visual field.

Many experiments show that pop-out may be produced by differences in stimulus attributes including color, line curvature, line tilt, and familiarity. And, of relevance to our present discussion, pop-out can be produced by texture differences between "target" and "background" (Bravo and Blake, 1990). There are, however, limits to pop-out from texture-defined shapes. To see this, look at Figure 5.35. The cluster of plus signs forming a horizontal shape pop out conspicuously, but, if you continue to look very hard, you may see a second region of texture difference that escaped your attention. See the cluster of Ts embedded in the Ls in the figure's right half? Only by scrutinizing the elements individually, and comparing them against one another, do you discover the existence of a group of Ts.

Ellsworth Kelly, a noted contemporary American artist, has used texture differences in paintings to create

FIGURE 5.33 | Two examples showing that shape can be defined by texture differences between figure and ground.

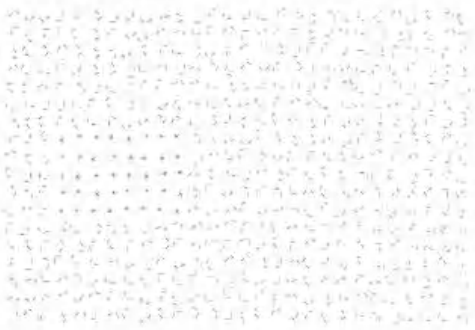
FIGURE 5.34 | Four displays typical of those in Treisman's work on target pop-out. The display in the upper left-hand panel shows "distractors" only with no "target." When displays like this appear, observers respond no as quickly as possible (by hitting one key on a computer keyboard). The other three panels show displays in which a target does appear within an array of distractors. When displays like this appear, observers respond yes as quickly as possible (by hitting another key on the computer keyboard). In a typical experiment, the number of distractors is varied, and reaction times on "yes" trials are then plotted as a function of the number of distractor elements. If the target pops out from the distracts, those reaction times will be more or less constant regardless of the number of distractors.



interesting illusory edges. A black-and-white reproduction of one of Kelly's paintings appears in Figure 5.36. You can probably see a set of sharp vertical edges that divide the painting into nine equal sections. In this piece, the artist was attempting to capture the essence of a pattern of shadows he had seen on a staircase (Shapley, 1996). Kelly stripped away brightness differences, leaving only shadow textures to define the nine steps. To get really good insight into what actually produces the illusory steps, take a card or sheet of paper and move it slowly up and down over Kelly's painting. By occluding or revealing various sections of the painting you should be able to identify minimal texture information that is needed to produce Kelly's illusory steps.

In addition to boundaries and shapes, texture information can also specify the three-dimensional curvature of a surface, particularly when the textures defining those surfaces are complex and consist of multiple orientations (Zaidi and Li, 2002; Li and Zaidi, 2000). Figure 5.37 shows a computer-generated image that simulates the appearance of a uniformly textured surface that is corrugated in depth, with the corrugations running along the horizontal axis. We will return to texture as a depth cue in Chapter 8.

FIGURE 5.35 | Some regions defined by texture differences are conspicuous whereas others require scrutiny.



This is a good time to take stock of what you've learned so far in this chapter. We have focused on visual processes that extract information about spatial scale and contour orientation. These same processes locate the edges and boundaries of objects by signaling the presence of discontinuities in light level or texture. These processes, which represent an early stage of vision, generate a map of whatever local features happen to be present throughout the visual field. Roughly speaking, these local features comprise parts of objects (often quite small). Although visual perception may begin by creating a neural inventory of parts and their locations, it certainly doesn't end there. To appreciate this point, let's go back to Spirit, the Martian explorer trying to identify objects using its onboard cameras.

When we last looked, Spirit had helped to create an inventory of Martian object parts and their locations. These were represented in the snapshots taken by Spirit's camera (Figure 5.1). The many clusters and strings of similar numbers provide some clues about locations in the image where object parts seem to be located. But how do we put these clues together to reproduce the objects in the scene that Spirit's camera caught? This next step is challenging, for it requires assembling the parts into meaningful wholes. To appreciate this challenge, look at Figure 5.38. We have no

FIGURE 5.36 | "La Combe I" by Ellsworth Kelly (painted in 1950). Note the strong illusory vertical contours along the endpoints of the thick diagonal bars. The figure shown here is a black-and-white reproduction of the original, which is red paint on white canvas. The original is 38 inches by 63.5 inches.

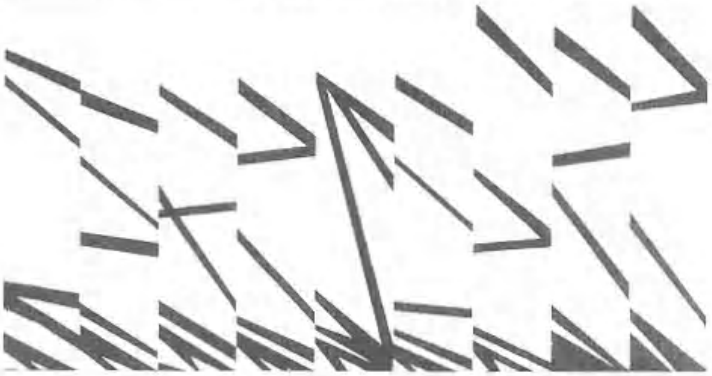


FIGURE 5.37 | Computer-generated image of a textured surface that appears to have 3D corrugations.

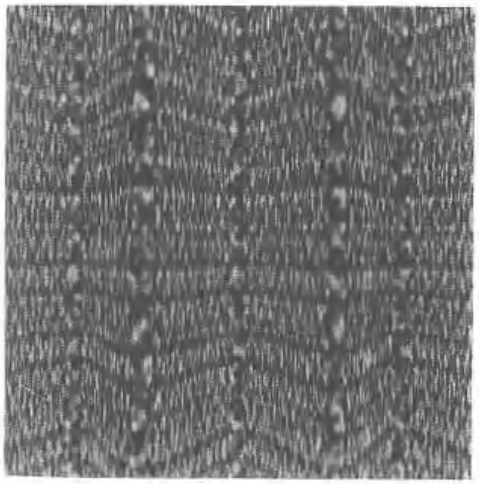
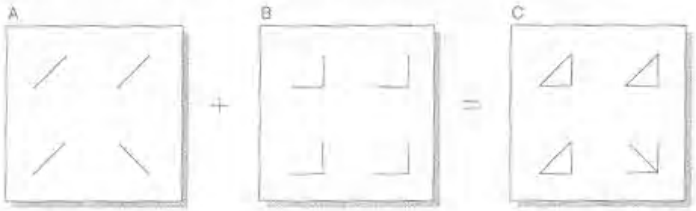


FIGURE 5.38 | Perceiving an object entails segregating contours into a meaningful whole.

FIGURE 5.39 | Displays producing the configural superiority effect.



trouble recognizing this as a young girl gazing out a window. Simple, it seems, until we stop to realize what vision had to accomplish to arrive at this conclusion. At early stages, vision must identify regions in the image that are most likely to be neighboring parts of a single surface and, hence, neighboring parts of a single object. This must be done even when parts of the image (such as the girl's cheek) are partially obscured by other objects in the scene (such as the crosspieces of the window). In effect, vision must assign different image parts to various objects even when those parts overlap. If you still think the task is easy, look back at Figure 4.1d, the seemingly random array of black-and-white splotches. Remember how difficult it was initially to identify this as a Dalmatian? This demonstration reminds us that assembling image parts into a meaningful whole is not always as easy as it seems.

So, how does Spirit's electronic visual system compute which feature parts should be grouped together to comprise a shared object? What rules must be embodied in its circuitry to accomplish this feat? For clues to the answers, we again should consider how our own, biological visual systems do it. To do that, we need to go back in time to the early part of the twentieth century, long before Spirit was even imagined.

Putting the Parts Together: Global Context and the Gestalt Principles

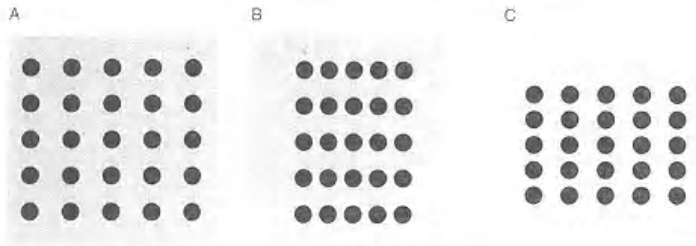
Previously in this chapter, we mentioned the structuralists' idea that perceptions could be decomposed into primary sensations. In Germany, shortly after the turn of the twentieth century, a school of thought emerged in reaction to the structuralist idea that composites could be decomposed into their parts. Called the **Gestalt** movement, this school of thought argued that objects appear as they do in virtue of the parts' *relations* to one another. It follows, therefore, that the process of decomposition—the structuralists' manifesto—destroys the essence of the object. Listen to the words of Wolfgang Köhler, one of the founders of the Gestalt movement:

> We do not see individual fractions of a thing; instead, the mode of appearance of each part depends not only upon the stimulation arising at that point but upon the conditions prevailing at other points as well. (1920/1938, p. 20)

A compelling confirmation of this statement was provided by James Pomerantz (1981). Pomerantz had people view arrays of elements like the ones shown in Figure 5.40. For each array, observers had to identify as quickly as possible the single item that was perceptually distinct from the others. For each presentation, Pomerantz measured how long it took to make this judgment. On average, observers took 1.9 seconds to spot an odd element among a cluster of four (in panel A of Figure 5.39, the odd element is the reverse diagonal at the lower right). Pomerantz also created a second array of figures, by adding to each diagonal an L-shaped figure (see panel B), producing the more complex figures shown in panel C. When required to spot the odd item in this new set, observers took only 0.75 second—65 percent faster than with the "simpler" figures. The more complex figure, in other words, acquired an emergent property not available in its individual, component features, a property that made it easier to discriminate from its neighbors. This so-called **configural superiority effect** dramatizes what became the credo for the Gestalt psychologists: "The whole is different from the sum of its individual parts." (In more contemporary parlance, one could say that a feature is known by the company it keeps.)

For the Gestalt psychologists, then, the key was identifying principles of organization that specified how parts relate to one another. In the following paragraphs we'll examine those principles and see examples of them

FIGURE 5.40 | Objects close together tend to unite perceptually into groups.

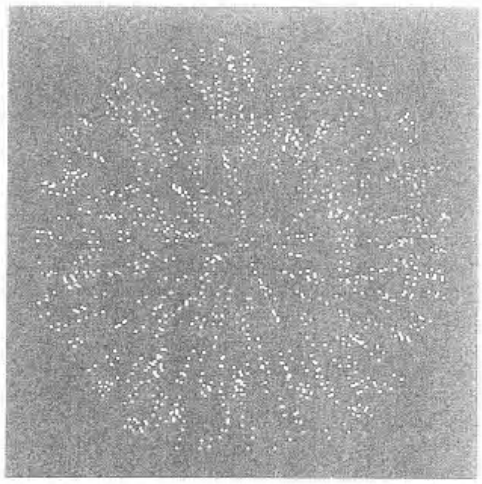


in action (for a more extensive discussion, see Palmer, 1999). The Web resources for Chapter 5, available at www.mhhe.com/blake5, has more material, including demonstrations, of these principles.

Gestalt Principles of Organization

The principle of **proximity** describes the tendency of nearby objects to group together as a single perceptual unit. As you can see in panel a of Figure 5.40, no strong, stable organization seems to predominate when the circles are equally spaced. Instead, they can be seen sometimes to cluster into rows and other times, into columns. (You may occasionally even see the circles grouping into diagonal strings.) None of these weak organizations persists, however, and none seems compelling. But notice what happens when the vertical distances among neighboring dots are made larger than the horizontal distances, as in panel b of the figure. With this simple modification, the dots form horizontal strings, or rows. Panel c illustrates what happens when the horizontal distances are made larger than the vertical: Dots group into vertical strings, or columns. So, simply varying the proximity of circles biases the perception in favor of one organization or the other. Another intriguing example of proximity's potency is illustrated in Figure 5.41. Called a **Glass pattern** (in honor of its inventor, Leon Glass, 1969), the dots in this figure appear to have an overall radial pattern to them. What does proximity have to do with this overall appearance? The global, radiating pattern comes about because throughout the figure the individual dots tend to pair up with a neighbor because of their proximity (technically each pair is called a "dipole"). And throughout the entire figure, the implied orientation of these pairs varies systematically in a manner implying a radial shape. And it is proximity that forms the basis for this global pattern.

FIGURE 5.41 | These dots convey an overall sense of a radial pattern, but this global form is defined solely by the arrangements of local pairs of dots that are grouped by their proximity.



Max Wertheimer, a famous Gestalt psychologist, called attention to another major organizational tendency, the principle of **similarity.** "Other things being equal, if several stimuli are presented together, there is a tendency to see the form in such a way that the similar items are grouped together" (1923/1958, p. 119). The dimensions of similarity that control grouping include lightness, orientation, and size. Examples of these effects are illustrated in

FIGURE 5.42 | Objects similar in shape and size tend to group together.

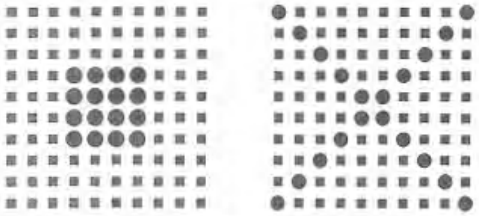


FIGURE 5.43 | A demonstration of closure.

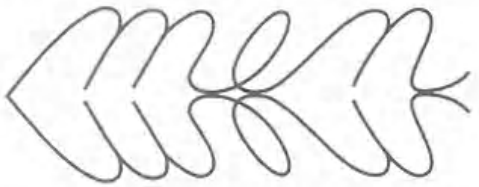


Figure 5.42. Notice in the left-hand panel how the circles form a perceptual group in the midst of the squares. Apparently the differences between circles and squares in both lightness and shape contribute to this grouping. In the right-hand panel, the disks form two diagonals crisscrossing the squares. Here, shape alone promotes grouping.

A third related grouping principle is called **closure,** and you can see it in action by looking at Figure 5.43. At first the figure may look like a meaningless row of heart-like forms. But the pattern contains a handwritten version of a simple and common word: "men." In fact, the pattern contains two copies of that word, one a mirror image of the other. Closure causes the two mirror images to cohere into a single, closed pattern, thus obscuring the characteristics of the two separate components.

Another of the Gestalt principles is **good continuation,** or more simply, continuation. This refers to the tendency to perceive clusters or strings of individual elements as forming a single contour, or path (Polat, 1999). In the preceding chapter, Figure 4.1B showed you one example of contour integration based on the alignment of individual features. A snake in the left-hand panel of the figure was easy to spot in a field of randomly oriented lines. But the snake in the other panel was hard to find. The random figures in the background were the same in both panels, but the snakes differed from one another in the amount by which their successive segments changed. The individual segments of the more visible snake differed from one another by angles of 30 degrees, whereas the less conspicuous snake's segments differed by 60 degrees. By the way, the snake would have been even more conspicuous if successive angular differences were even smaller or if the snake's "head" had grabbed onto its "tail" to form a closed figure (Kovács and Julesz, 1993; Pettet, McKee, and Grzywacz, 1998).

Good continuation is one of the Gestalt principles whose underlying neural mechanism is fairly well understood. Recall that Chapter 4 described a scheme whereby neighboring oriented line segments, like those comprising the snakes, could be "connected" by the visual system. Specifically, neural responses to nearby contour segments of similar orientation could be enhanced if neurons with neighboring receptive fields and similar preferred orientations were interconnected in a way that allowed them to reinforce one another's activity. These highly selective linkages could arise out of long-range horizontal connections that selectively associate neurons with adjacent receptive fields. These associations would selectively favor continuities between neurons whose preferred orientations were similar, and would discriminate against continuities between neurons with dissimilar preferred orientations (Field and Hayes, 2004). The presence of an appropriate stimulus would be signaled by simultaneous activation of a set of neurons whose receptive fields had the appropriate orientation tuning and spatial locations.

But what evidence do we have that these neural circuits really do promote visual grouping by good continuation? If those circuits did serve this function, we would expect that the visual features activating these circuits would be features typically present in natural images of real-world scenes—the images generated as we look around our natural, everyday world. So, are the visual stimulus conditions revealed by laboratory contour integration experiments mirrored by the visual characteristics of natural images? To answer this question, one must first ascertain the characteristics of natural images. To that end, Bill Geisler and colleagues (Geisler et al., 2001) measured the geometrical properties of a wide variety of natural images, including images of flowers, trees, a mountain, a lake and river, and some woods (see Figure 5.44). To characterize the contours in such images, these investigators displayed each image on a computer monitor and had people move a cursor to select all the oriented elements that be-

FIGURE 5.44 | Examples of natural scenes used by Geisler, Perry, Super, and Gallogly.

longed together in a single, shared contour. For example, if there were a riverbank in an image, people selected each of the oriented segments that outlined the riverbank. This operation was repeated for all the contours in every image.

From this vast collection of measurements, Geisler and colleagues extracted the essential statistical properties of natural contours. Recognizing that the geometrical relationships among segments were particularly important, the research team computed the orientation and position differences among all pairs of segments that people had attributed to each shared contour. Their quantitative findings can be summarized broadly and qualitatively: Adjacent segments of any single natural contour tend to have very similar orientations (take one small stretch of riverbank and an immediately adjacent stretch), but segments of the same contour that are further apart (one stretch of riverbank versus another, more distant stretch of that same bank), tend to have orientations that are more disparate from one another.[8]

Geisler and colleagues used this information about natural images to create 216 different families of synthetic test figures—serpentine, broken contours made up of varying numbers of line segments. Within each family, all members shared the same quantitative relationship to the statistical properties of natural contours. But the families varied in how well each conformed to those natural statistics.

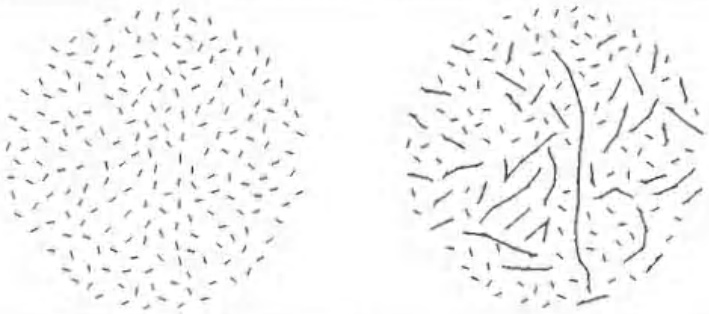
These carefully designed figures were then used in a forced-choice psychophysical experiment. In each trial, participants saw two patterns: one was a collection of many, small, randomly oriented line segments; the other pattern was a similar collection, but with a test figure embedded in it. The participants had to identify which pattern contained the serpentine contour. Geisler reasoned that if the contour integration properties of the human visual system mirrored the statistics of the natural world, the detectability of any particular test contour should depend upon how closely *its* statistics matched the average co-occurrence statistics of contours in the natural world.

Detailed predictions of contour detectability were derived from a simple computer model whose few rules mimic the way the human visual system might perform the same task. The model linked individual line segments into a shared, larger contour, but did so in accordance with the statistics of natural images. In particular, the probability that any two segments would be linked together grew as their orientations became more similar, but declined as the distance between segments increased. Finally, linkages were transitive, that is, if the computer model linked line segment a to line segment b, and it also linked segment b to segment c, then the model would link a and c, creating, in this example, an extended contour three segments long.

Figure 5.45 gives you some idea of how the model behaved. The panel at the left shows a typical visual array that Geisler used to test observers. At first glance, the short segments in the array may appear to be random in orientation. If you scrutinize the array, though, you will find a number of places where nearby segments have similar orientations. The computer model would tend to link these together into a contour. In the right-hand panel, lines are drawn through individual segments that Geisler's model would connect. As you can see, the model's rules for connecting segments generates a number of small contours, and one long one. Note, in particular, that the model discriminates against linkages between adjacent segments whose orientations differ widely from one another.

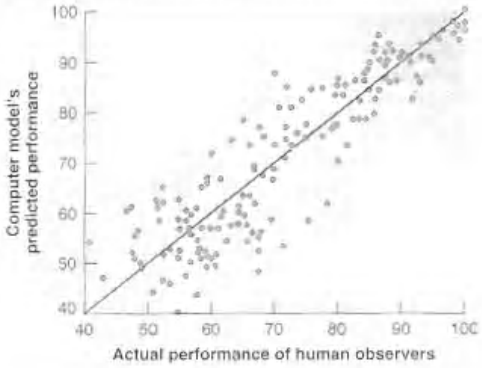
[8] Obviously, there are exceptions to these tendencies, such as instances of relatively sharp changes in orientation, as with the pointy leaves of maple trees, but overall these tendencies do capture the main statistical characteristics of natural images.

FIGURE 5.45 | Contours extracted by Geisler's model. At the left is a set of contour segments used to test both human observers and the computer model. At the right are the contours extracted by the model.



For each of their 216 families of synthetic figures, Geisler and colleagues compared the model's predictions to the actual performance of human subjects. To illustrate the power of this simple model, Figure 5.46 shows, for each condition, the model's prediction plotted against observers' actual performance. If the model were perfect, and if performance had no inherent variability whatsoever, all the data points in the figure would lie along the diagonal line. Given how elementary the model was, Figure 5.46 shows excellent agreement between the predictions of this very simple model, on one hand, and human performance, on the other. This agreement strongly supports the hypothesis that development of contour integration mechanisms, either during evolution or during the first years of any individual's life, is driven by the occurrence statistics of images encountered in the natural world. Geisler and Diehl (2002) have extended this basic idea, explaining processes by which the properties of perceptual systems are harmonized with the properties of the physical world.

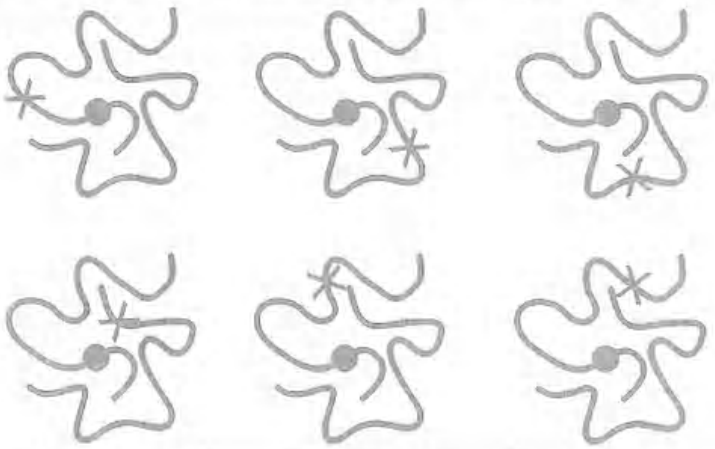
FIGURE 5.46 | Agreement between predictions of Geisler's computer model and performance of human observers. The diagonal line represents perfect agreement between the computer model and people's performance.



Given the importance of Gestalt principles of perceptual organization, we should ask at what stage in visual development these principles of organization actually begin operating. Elizabeth Spelke and her colleagues used the preferential looking technique (described previously in this chapter; see Figure 5.18) to assess infants' sensitivity to objects whose boundaries were defined by Gestalt principles such as figural similarity and good continuation (Spelke et al., 1993). Although adults describe the objects as compelling, three-month-old infants seem oblivious to the objects, implying that the individual features defining the objects were not being grouped. During the next half-year, infants slowly but surely show signs of increasing use of grouping rules. It is not until about age 10 or 11 that adult levels of performance are achieved, implying that the neural circuits underlying contour integration mature relatively slowly within the human cerebral cortex (Kovács et al, 1999).

Gestalt principles of organization are sometimes called *laws*. If they're truly laws, however, there are nonetheless circumstances where grouping tendencies are stymied, and need lots of help, effort, and extra time. Take the twisted and tangled pairs of curves in Figure 5.47. Start with the left panel by fixating on the black circle. Now decide whether the x lies on the same curve as the fixation mark. Do the same for each of the other panels.

FIGURE 5.47 | Figures used to study visual assessment of contour relationships ("curve tracing"). Each figure consists of a pair of intermingled curves. For each figure, fixate the black circle and judge whether the x lies on the same curve as the circle or not. Note how the ease of judgment varies from figure to figure.



Although you certainly can perform the task, it's not easy. In fact, the time it takes to determine whether the x and the fixation point share a line depends upon how far apart the two are on that line (Jolicoeur, Ullman, and Mackay, 1986; Pringle and Egeth, 1988). The visual system requires time to trace the curve from the fixation point by moving the focus of attention along the curve.

As you know, proximity *promotes* perceptual grouping, but here, grouping by proximity *retards* perception: the tracing time lengthens as the two curves are brought closer together. After all, performance in this task requires keeping the two curves separate, not grouping them into a single entity. This is further demonstrated by the fact that curve tracing slows dramatically if the two curves touch or cross one another (Roelfsema, Scholte, and Spekreijse, 1999). Apparently, the task requires not only differentiation of one contour from its background, but also differentiation of one contour from another contour.

So, the Gestalt "laws" operate more like suggestions than reliable, ironclad laws. But laws that operate this way are common in science. Take Boyle's law. It tells us that the volume occupied by a gas is inversely related to the gas's pressure. But pressure is not the only factor that comes into the equation. To predict what volume will be occupied, you also need to take into account the temperature, as the combined gas law does. No Gestalt principle operates in isolation, unless a stimulus has been specially designed to allow it to do so (as we did in Figures 5.42 and 5.43). In the natural world, potentially conflicting Gestalt principles are reconciled in accord with rules that are not yet fully understood.

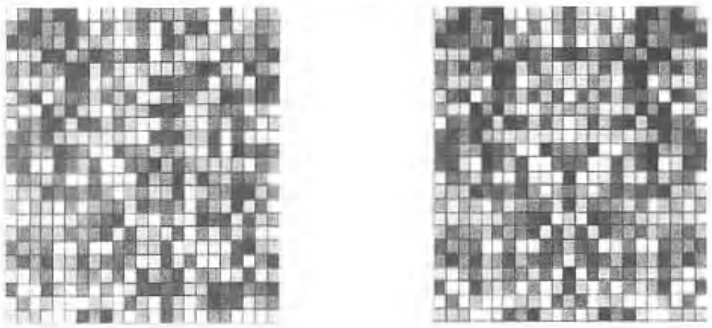
Before closing our discussion of spatial vision and grouping, we need to consider one other potent stimulus characteristic that significantly impacts the appearance of objects, the characteristic termed *symmetry*.

Global Structure and Symmetry

Some objects, including most biological ones, have an unusually conspicuous structural property that deserves special comment: that property is **symmetry,** the tendency for those parts of an object centered about a given axis to be highly similar in shape, texture, and color. Symmetry influences many of our perceptual judgments, including the quality of fruit (round apples are preferred to irregularly shaped apples), the structural integrity of a bridge, and the aesthetics of art (Weyl, 1952]). The human body is approximately bilaterally symmetric, and some evolutionary biologists even believe that bilateral symmetry of the human face significantly contributes to perceived attractiveness (Mealey, Bridgstock, and Townsend, 1999). Given the ubiquitous importance of symmetry, how does the visual system recognize this important characteristic of objects?

To come up with an answer to this question, let's start by looking at an example of symmetry. In Figure 5.48, the two panels are perceptually quite distinct from one another: The one on the left looks like a collection of random black,

FIGURE 5.48 | Demonstration of symmetry detection. Both rectangles comprise elements of varying lightness. In the rectangle at the left, those elements are distributed randomly; in the rectangle at the right, the random elements in the left-hand side of the square have been duplicated in mirror-image fashion on the right-hand side. This gives rise to bilateral symmetry that is easily seen.



gray, and white squares. The panel on the right is also a collection of black, gray, and white squares, but these squares are bilaterally symmetric; the two halves are mirror images of one another. To make this panel, we duplicated one half of the right-hand panel, flipped it around a vertical axis, and then put the two halves together, side by side. But you could detect the bilateral symmetry even before you were told how the figure was made. Detecting individual squares—regions that differ from their immediate surroundings—is an early step in the process for detecting symmetry about a vertical axis. But it's not enough. The visual system also has to register similarities in the spatial arrangement of the squares, and this requires recognizing comparable structures—such as the figure's horizontally elongated and aligned patterns (Dakin and Herbert, 1998).

The tendency to perceive symmetry must be really powerful, because we automatically "impose" an assumption of symmetry when interpreting potentially ambiguous visual scenes. What do you see when you look at the pair of shapes in the top part of Figure 5.49? No doubt you perceive a diamond-shaped object partially occluded by a circular-shaped object. But that's not the only possible perceptual interpretation: The right-hand portion of the partially occluded object could be almost any shape (including a semicircle, in which case the two objects would be butting up against one another). Our overwhelming tendency to see a diamond is caused by the strong propensity to assume that the partially occluded object is symmetrical. It is no wonder that the Gestalt psychologists considered symmetry as a fundamental principle of perception.

Symmetry is a powerful contributor to aesthetic judgments, including our reactions to representational and nonrepresentational works of art (Ramachandran and Hirstein, 1999). What's the origin of this aesthetic impact? And can experience override that preference? To explore these questions, Ingo Rentschler of the University of Münich turned to some very simple figures: compound gratings comprising two different spatial frequencies in varying phase relationships to another (Rentschler et al., 1999). You can see some of these compound gratings in Figure 5.50. Compound gratings may not be as beautiful as your own favorite painting, but using them rather than paintings makes it possible to connect known physical properties to any aesthetic responses they may evoke.

Rentschler and colleagues presented all possible pairs of 16 different compound gratings. For each pair, people chose the compound they liked best. The entire set of preference judgments was converted into a quantitative measure preference for each grating. Overall, people had the strongest dislike for compounds 2 and 4 shown in Figure 5.50, and they had the strongest preference for compounds 1 and 3. Apparently, two factors underlie these ratings (and the ratings for the other 12 gratings, which are not shown here). One factor is symmetry—compounds 1 and 3 have greater bilateral symmetry than do 2 and 4. The other factor is simplicity, as reflected in the number of easily visible bars—the preferred compounds have fewer visible bars. In a follow-up experiment, Rentschler found that experience with these compounds altered people's preferences. Bilateral

FIGURE 5.49 | In the top portion of the figure, we see the object on the left as a diamond partially occluded by a circle. In fact, the occluded portion of that object could be any shape whatsoever, including any of the three shapes shown at the bottom. Vision favors the diamond because that perceptual interpretation maintains symmetry.

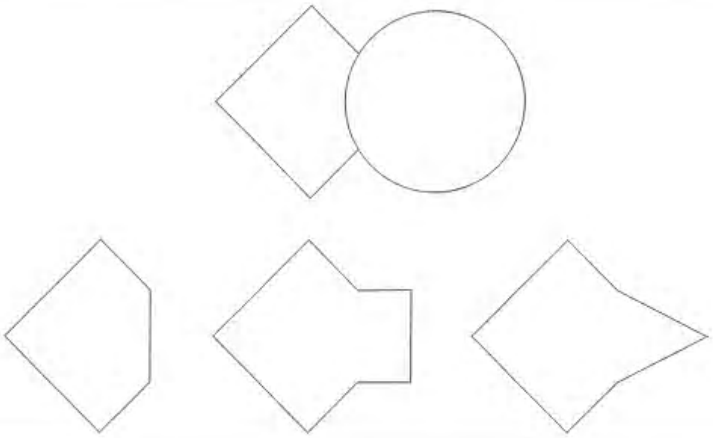
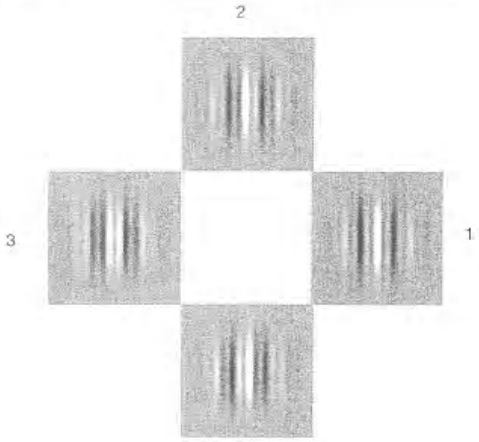


FIGURE 5.50 | Compound gratings representing the two preference dimensions Rentschler extracted from people's judgments of liking. Compounds 1 and 3 exemplify bilateral symmetry; compounds 2 and 4 exemplify visual complexity.



symmetry remained as potent as ever, but preference for complexity increased or decreased depending upon a person's prior training.

Preference for symmetry, then, depends on a relatively invariant processing of visual information, but the aesthetic pleasure derived from simplicity/complexity depends upon experience. At any moment, your preferences reflect both the enduring properties of your nervous system as well as the taste acquired through your experience.

This completes our survey of the spatial properties that define objects and their parts, and hence, determine our ability to detect and discriminate those objects. As we close, think back once more to Spirit slowly traversing the Martian landscape as it surveys whatever its cameras encounter. If Spirit's circuitry can mimic the perceptual operations described in this chapter—multiscale analysis and shape-from-texture grouping—then our little robot has taken one giant step in humankind's attempt to discover the Martian environment. But Spirit's electronic messages still must be transformed into identifiable objects. Clues about how that can be accomplished are the focus of Chapter 6.

SUMMARY AND PREVIEW

This chapter emphasized that visual information is used to distinguish objects from their backgrounds and to discriminate objects from one another. We presented various views of form perception, including structuralism, Gestalt psychology, and the multichannel model derived from measurements of the CSF. The multichannel model is an attempt to relate human form perception to what is known about the visual cortex. It can account for many aspects of human and animal vision, including detection, discrimination, and certain visual illusion, but it, too, has its limitations. In the next chapter, you'll learn how the visual system puts together the information it acquires to generate useful representations of objects.

KEY TERMS

analytic introspection
closure
configural superiority effect
contrast
contrast normalization
contrast sensitivity function (CSF)
contrast threshold
cutoff frequency
detection
discrimination
Fourier analysis
Gestalt
Glass pattern
good continuation
grating
identification
metamers
multichannel model
orientation
proximity
scale
selective adaptation
similarity
sinusoidal grating
size aftereffect
spatial frequency
spatial phase
symmetry
texture contrast
transfer function
window of visibility