Chapter 6
Form Perception

We could say that we live in a world where our retinas are constantly assaulted from everywhere. Thousands of potential stimuli in the immediate environment may reach our eyes at any moment. These various stimuli result from the interactions between the surface properties and those coming from light sources (intensities and wavelengths). Moreover, our environment is sometimes stable, sometimes not; sometimes, things are moving and sometimes, we are moving. Therefore, there is constantly an incredible variety of stimuli on the retina. Nevertheless, we extract from all this information something intelligible; moreover, this task is completed without effort. This remarkable efficiency is made possible by the functioning of some basic mechanisms described below.

6.1 Perception of Contours

We can extract a shape in the environment because it provides brightness variations. These variations are such that there are boundaries between objects. We know that there is somewhere a given object because we perceive delimitation, or an edge, between this object and its surroundings. We call this edge a contour. This contour could be considered the elementary unit of form perception.

To fully realize the importance of a contour, just think about what happens during a snowstorm. When there is too much blowing snow, it becomes no longer possible to see anything, even when you try to keep your eyes open, because the field is evenly lit (in German, this phenomenon is referred to as a *ganzfeld*—i.e., complete field). If you want to experience a ganzfeld without waiting for the next snowstorm, simply try the following activity. Take two white plastic spoons or even the two halves of a white ping-pong ball, and draw a small but clearly visible colored line on the inside of the spoons or half balls. Then, just make sure to completely cover the eyes with spoons or half balls so that no light can enter. Keep your eyes open while fixating the inside line and avoid any eye movements.

You need to maintain this fixation activity for several seconds so that the line remains at the same place on the retina.

What happens after a few seconds (less than a minute)? If there was no movement of your eyes, the line disappears. For perceiving form, even just a single line, it takes brightness variations between this form and its environment. We clearly see the line at first, but eventually lose sight if we prevent the visual system from restoring the perception of a contour. In fact, the image never remains stable for long on the retina. The image on the retina keeps moving because there are always small eye movements called microsaccades. These small involuntary eye movements create variations in time on the receptors of the retina. What the experiment with the spoons or half balls teaches us is not only that it takes contour perception to see but also that it is necessary, for avoiding the disappearance of the contour, that the image does not stabilize on the retina.

6.1.1 Edges and Subjective Contours

The contour depends mostly on the presence of an edge. The latter can be defined as a change in luminance or spectral composition occurring somewhere in the environment. Most often in the environment, contrast or texture changes will create edges. In other words, the contours are typically due to a physical phenomenon, namely, the presence of boundaries. In this case, we sometimes refer to it as the first-degree contours (or first-level contours).

As shown in Fig. 6.1, the presence of an edge is not always necessary for the formation of a contour. It is possible to perceive contours without any physical changes. Such conditions are referred to as subjective contours or second-level contours. We also refer to the term emerging contours to describe these cases where a contour is perceived although there is absolutely no physical variation producing it.

In short, perceiving a form requires to perceiving contours. The detection of these contours depends mostly on the presence of an edge caused by the heterogeneity in the stimulation. The perception of these edges also requires that variations of this stimulation occur on the retina. These variations are made possible by eye

Fig. 6.1 Examples of subjective contours. You can observe a *triangle* and a *square* on the *left* and a *horizontal line* on the *right*

microsaccades. Finally, even in the absence of edges, there may be contours, called subjective, but their presence still requires the presence of special conditions in the environment as shown in Fig. 6.1.

6.1.2 Lateral Inhibition

There is a basic mechanism in the visual system for increasing brightness variations that generate a border. When light reaches the retina at a given point, what is located just next to this point undergoes inhibition. The mechanism by which the activity of certain nerve cells affects that of its neighbors is called *lateral inhibition*. This phenomenon, first reported by Keffer Hartline and Floyd H. Ratliff, is a fundamental notion of retinal physiology.

Hartline and Ratliff (1957) worked on the visual system of horseshoe crab. This animal has the distinction of having a series of small elementary eyes, called ommatidia, rather than a dense neural network. This feature makes it easier to stimulate each eye when the effect of lateral inhibition is demonstrated. Each ommatidium can somehow be compared to the ganglion cell of the human visual system.

In the work of Hartline and Ratliff, the electrical activity of a nerve fiber, say A, is collected by means of an electrode. When the receptor corresponding to fiber A receives light stimulation, the electrical activity increases, indicating that the activity is linked to the stimulation (Fig. 6.2). When only one receptor corresponding to a neighboring fiber, B, is stimulated, the electrical activity collected from fiber A is
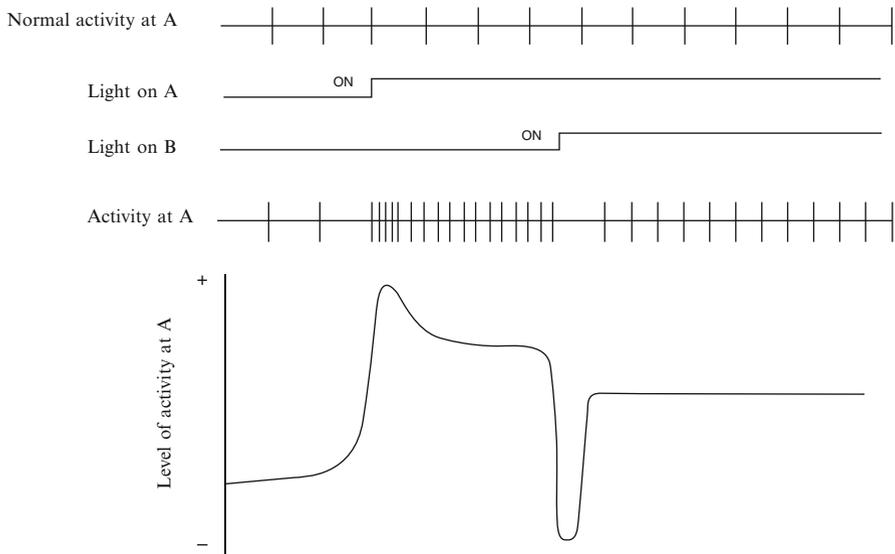

Fig. 6.2 Illustration of the lateral inhibition effect exerted on a cell, *A*, previously activated by a light source, by the arrival of a light stimulus to a cell, *B*, located close to *A*

not affected. This reflects the independence of the activity of B on A when there is no direct stimulation on A. In a case where light stimulation is maintained on A, and another light stimulation excites B, then the electrical activity observed earlier on fiber A is decreased. In other words, the activity on B exerts lateral inhibition, that is to say, it reduced the activity of neighboring fiber A.

The strength of this inhibition depends essentially on two factors: the proximity between the nerve cells involved and the strength of the stimulation on the inhibitory cell. The stronger the activity of the inhibiting cell, the greater the inhibition; similarly, the closer the inhibited and inhibitory cells, the greater the inhibition effect.

Mutual effects of nerve cells or fibers on each other can be quite complicated. Suffice to say, there may be a decreased inhibitory effect in a case like the following one: Given fibers A, B, and C arranged in this order. B has some inhibitory effect on A when fiber C is not stimulated. However, when light stimulates C, the activity on C inhibits the activity of B. By having a reduced activity, B exerts in turn a less pronounced inhibitory effect on A. Thus, the electrical activity of A is higher if the fibers A, B, and C are stimulated compared to when only fibers A and B are stimulated. The activity recorded from A in the condition in which A, B, and C are stimulated remains nevertheless below that normally observed when only fiber A is excited.

6.1.3 Mach Bands

There are many fascinating perceptual effects which can be explained on the basis of lateral inhibition. A classic example of the effect of lateral inhibition is illustrated by the demonstration called Mach bands. Ernst Mach, who revealed this effect, is the same Austrian physicist and philosopher who gave his name to the unit used to express the speed of sound.

Consider the following situation where, say, black (dark gray) and white (pale gray) are separated by a gray gradient (upper part of Fig. 6.3). Although black and white are both uniform (even luminance), the insertion of a gray gradient changes the perceived brightness: the black and white are not perceived anymore as uniform. Most people perceive a particularly small dark line (very black) on the side where the luminance is low and a particularly pale line (very white) on the side where the luminance is high. At these locations, on each side of the gray, thus appear Mach bands. The main interest of this demonstration is therefore located at the two places where luminance changes occur.

It is possible to explain this Mach band effect with the lateral inhibition exerted by nerve cells on each other. For example, if we take two points close to each other, near to the midpoint where the luminance is uniformly black, these two points are subjected to similar levels of inhibition caused by neighboring cells to the left and to the right. Thus, their brightness is the same. If we rather take a point X where the luminance change begins, then the inhibition exerted by the left and by the right

Fig. 6.3 On the *lower panel*, changes in luminance (*black line*) and brightness (*green line*) corresponding to the *black and white image* (*upper panel*). *Arrows* indicate Mach bands. It is a little darker under the *left arrow* and a little brighter under the *right arrow*
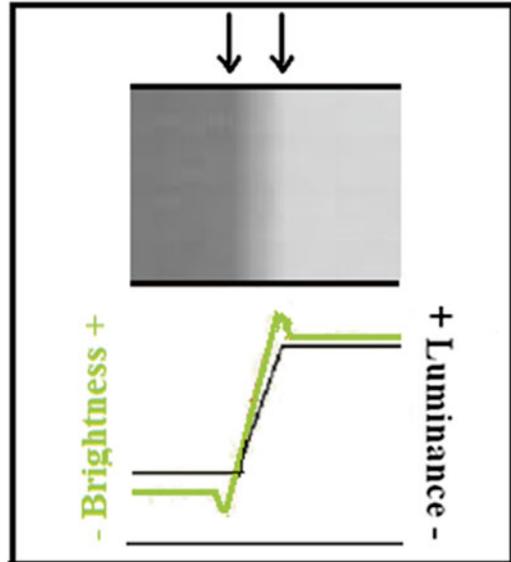
cells is not the same. If the transition initiates a luminance increase, the proportion of white of the physical stimulus increases, and, thus, the inhibition caused by the cells on this side on point X is larger than that exerted by the cells on the other side. So this transition line appears darker as it undergoes more inhibition. The same reasoning can be applied to the reverse situation: where a decrease in luminance begins as the inhibitory effect decreases. At this point of transition appears a brighter thin band.

One can also observe this Mach band phenomenon in Fig. 6.4 on which are illustrated a series of bands with uniform brightness. These bands do not appear uniform when viewed as a whole. If we do look at only one band, hiding somehow the others, then its brightness is uniform because the luminance of a given band is uniform. It is the activity exerted by the ones on the others that determines the level of neural activity of each cell and, consequently, the brightness. In brief, the figure shows a series of lateral inhibition effects.

6.1.4   Factors Influencing the Perception of Contours

Many factors may influence the creation of contours. For example, contours are perceived more easily when the visual acuity is greater. The acuteness being greater at the fovea, the contours appear more clearly in this region. The further away an image is moved from the fovea, the less clear the contours are. Similarly, subjective contours shown in Fig. 6.1 are examples of the influence of the spatial context on the creation of contours.
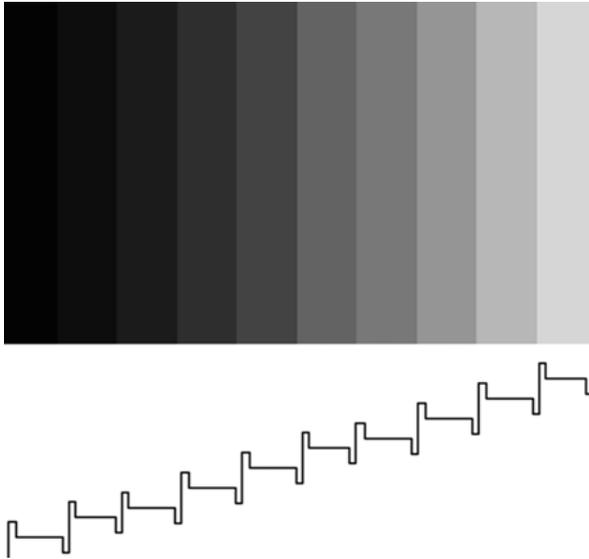
Fig. 6.4 Another illustration of the Mach bands where a series of stimuli, each of uniform luminance, appear *brighter* on the *left side* and *darker* on the *right side*

The formation of a contour takes a minimum of stimulation intensity. The intensity depends on the number of photons absorbed by the photoreceptors. Indeed, the ability of these photons to produce an effect depends on how long the eye has been stimulated. It takes a minimum of exposure time for a stimulus to be detected. Photons can benefit from a temporal summation effect. If their arrival is not sufficiently close in time, they lose that benefit. This is essentially what Bloch's law refers to. It can be summarized as follows:

$$I \times T = C$$

where the interaction between the intensity, I, and the exposure time, T, results in a constant visual effect, C. If a stimulus is very intense, it can be detected even if presented for a very short period, whereas a weaker stimulus will be detected only if presented for a longer period.

In fact, the interaction between time and intensity applies only for very short exposure times, i.e., of less than 100 ms. Beyond this period, the only crucial factor is the fact that the intensity is sufficient or not for perceiving a stimulus. This 100-ms value holds for rods; for cones, this value would rather be 50 ms. Note that there is also a law, Ricco's law, which applies only to the fovea. According to this law, the detectability of stimuli is a combination of the intensity and of the stimulated area.

There are different demonstrations which illustrate the importance of the exposure duration in the perception of contours. In this regard, an old experience of Werner (1935) is most relevant. In this experiment, a black circle and a black ring are presented alternately to a participant. The outer contour of the disk corresponds

Fig. 6.5 *Disk* and *ring* used in the experiment reported by Werner (1935). See text for explanation
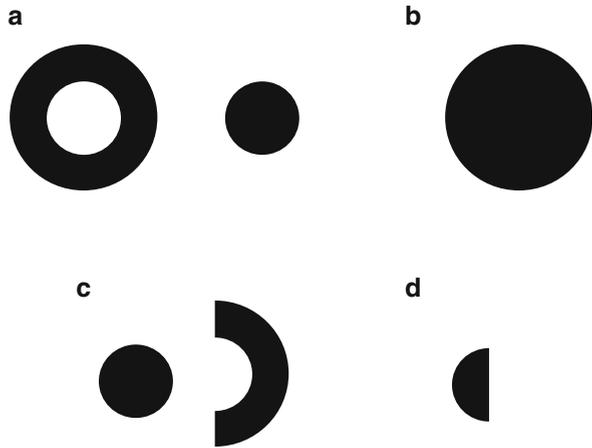
exactly to the inner contour of the ring (Fig. 6.5a). The experimenter varies the exposure time of stimuli and the pause time between exposures. When the time interval between stimuli is smaller than 100 ms, the participant perceives a full circle (Fig. 6.5b). If the interval between stimuli is more than 200 ms, the subject sees alternately the disk and the ring. Pauses with a duration ranging from 100 to 200 ms lead the participant to perceive only the ring. Thus, depending on a temporal factor, a masking effect may occur. The formation of the inner contour of the ring prevents the contour of the disk to be seen. And if it is the disk and only a half-ring (Fig. 6.5c) that are presented for 100–200 ms, only the half-disk can be detected (Fig. 6.5d).

6.2 Gestalt: Perceptual Organization

A century ago, Max Wertheimer developed a fine and influential way of approaching the study of form perception. The school of thought, known as Gestalt, which means form (or "whole form"), also received contributions from Wolfgang Kohler and Kurt Koffka, other German specialists of the psychology of sensation and perception, and collaborators of Wertheimer. The aim of the Gestalt psychology, which encompasses all the work on perceptual structuring, was to explain how the visual system combines the various elements available in the visual field.

There is in this notion of Gestalt the idea that perceiving is more than the summation of the sensations produced by stimuli. There is an organization of these stimuli. A person organizes the elements of a visual scene for extracting meaning. The organization of these elements includes two aspects that will be described in the following paragraphs. There is firstly the distinction between figure and ground and secondly the grouping of elements according to some characteristics sometimes called the laws of Gestalt.

6.2.1 Figure/Ground Distinction

When looking at a visual field, some parts are different from others. We look in a particular way in order to highlight some parts of this field. In a task as simple as looking at a piece of art on a wall, there is a way of looking. Our gaze is focused on the dominant object of our visual field, the piece of art, and the nearby field, the wall, serves as a background. So there is a fundamental distinction, the figure as opposed to the ground, in our way of looking. These two parts of the field have their own characteristics.

In a visual scene, the contour seems to belong to the figure rather than to the ground. The figure looks like something and appears to be closer than the ground. There may sometimes be an ambiguity in the figure, as shown in Fig. 6.6, which can be solved according to the way of looking. In Fig. 6.6, on the left, the black part is perceived as the ground, and, consequently, white diamonds are perceived spontaneously. In Fig. 6.6, on the right, white diamonds are much less likely to be perceived spontaneously. Indeed, it is much easier to imagine that the white part can be the ground in the right than in the left figure. Consequently, we perceive much more easily that there are black diamonds in the right figure.

In general, a figure has a shape and some meaning, whereas the background is rather disorganized. Indeed, several objective factors determine this figure/ground distinction. These factors are illustrated in Fig. 6.7. These factors are reported to be objective because they are determined by the stimuli. An image placed in a vertical horizontal orientation will be more readily perceived as a figure than if it is placed in diagonal directions. Thus, it should be easier to perceive a white cross in the left portion of Fig. 6.7a than it is in the right portion. In this latter case, due to *orientation*, we perceive more spontaneously the gray cross than the white cross. Similarly, a smaller (or thinner) image is more easily perceived as a figure than a larger image. This factor is called the *relative size*. Thus, in Fig. 6.7b, we perceived more easily a
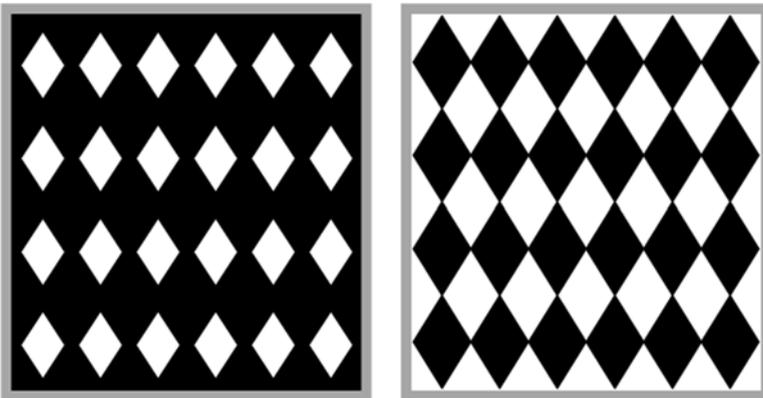

Fig. 6.6 Illustration of the propensity to see, on the *left*, *white diamonds* on a *black background* and on the *right*, *black diamonds* on a *white background*
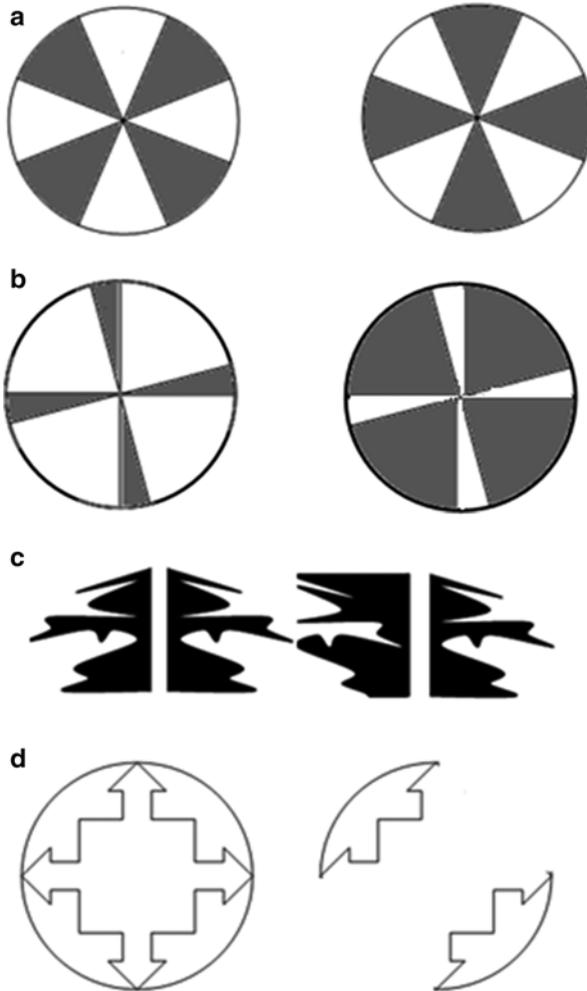
Fig. 6.7 Objective characteristics of the figure/ground segregation: (**a**) orientation, (**b**) size, (**c**) symmetry, and (**d**) inclusion (see text)

gray cross than a white cross on the left, but the reverse on the right. In fact, the thin crosses on Fig. 6.7b are more spontaneously perceived than the crosses on Fig. 6.7a.

Another very powerful factor is *symmetry* (or regularity). When objects or figures are symmetrical, they are more likely to be perceived as figures. Thus, because the four black pieces of Fig. 6.7c are on a white background, they tend to be perceived as figures. However, it would have been possible to perceive a white figure between the two black parts in the center, but being very irregular, this white figure cannot be spontaneously perceived. In addition, the two black parts on the left being symmetrical, they are easily perceived, more than are the two rightmost black portions of the figure.

When an image is inside another, chances are that it will also be recognized as a figure rather than as ground. This factor is called *inclusion* (or *surroundedness*). Thus, the square in the middle of Fig. 6.7d, on the left, does not act like ground but is part of a complex figure in a circle. That said, different portions of the circle could have been perceived as figure, as illustrated in the right part of Fig. 6.7d.

Note that there are various other objective factors that may contribute to figure/ground differentiation. For example, the patterns within an image can be crucial for perceiving a figure; this factor is called the *internal articulation*. Also, various subjective factors are likely to influence this differentiation. Among these factors, there is the previous experience of the person perceiving, as well as the elements toward which attention is directed. That individual traits exert influence on what is extracted from a given visual scene is hardly surprising for clinical psychologists using projective tests.

6.2.2   Perceptual Grouping

The visual perceptual system tends to group automatically, that is to say, without cognitive effort, certain elements present in the visual field. This grouping is based on basic principles identified by the Gestaltists. These organizational principles are sometimes referred to as Gestalt laws or Gestalt grouping rules.

We tend to group together elements that are close to each other. This tendency is called the law of *proximity*. Thus, we perceive spontaneously, in Fig. 6.8a, four groups of three elements rather than 12 elements. A series of elements may be equidistant from each other, but some of them can be grouped together because of their resemblance. This is what the law of *similarity* stipulates (Fig. 6.8b). A third
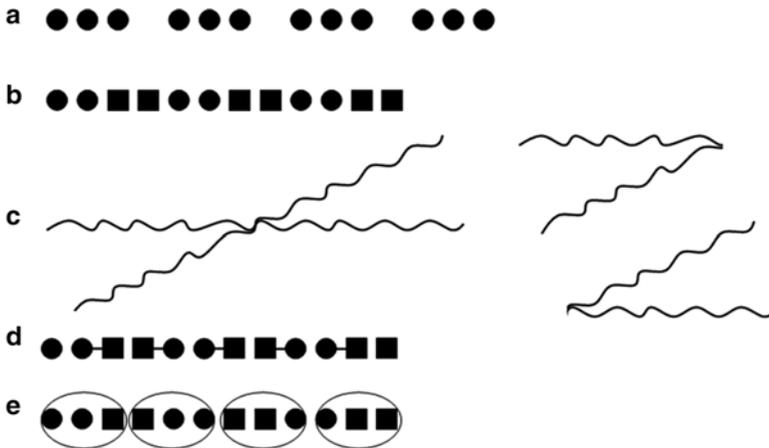

Fig. 6.8 Illustration of the Gestalt laws: (**a**) proximity, (**b**) similarity, (**c**) good continuation, (**d**) connectedness, (**e**) common region

law, known as the *good continuation*, reveals that the elements of a set forming a continuous series, or are part of the continuity relative to each other, tend to be seen as if they were one unit. Two intersecting lines are perceived in Fig. 6.8c, rather than the two items shown just to the right of these lines.

More recently, in what might be called a modern Gestalt, other perceptual organization principles were reported. These new principles are indeed very powerful. The first is called *connectedness* (or uniform connectedness). The fact of connecting elements together, as is shown in Fig. 6.8d, has more impact than the principles described earlier. Similarly, items that are part of the same region are seen as if they belong to the same entity. This principle, known as *common region*, is illustrated in Fig. 6.8e.

Furthermore, there are other laws of Gestalt, for example, the law of *closure*. According to this law, the visual system tends to see figures or objects as if they were complete, be it fully or in part; if the figure is not complete, the visual system manages to reach closure (see subjective contours, Fig. 6.1). Also, the more regular and symmetrical a shape is, the more it imposes itself to the perceptual system. This is known as the law of the *pragnanz*, also called the *law of good form* (and sometimes law of symmetry). Finally, another very powerful factor that organizes the visual perception is related to the fact that some elements might be in motion. If elements of a visual scene move in the same direction, they are perceived as being grouped together. This is called the law of *common fate* (or of *common motion*).

6.3   Theory of Multiple Spatial Channels

A very original way to address the issue of form perception was proposed in the late 1960s. This approach, developed by F. W. Campbell and J. G. Robson, is based on spatial frequency analysis and is sometimes referred to as the multiple spatial channels theory (Campbell & Robson, 1968).

6.3.1   Basic Concepts

The multiple spatial channels theory is based on a simple and clever idea: each image can be decomposed into a series of cyclical variations in luminance. The reader already familiar with the physical bases of auditory perception knows that sound can be interpreted as pressure variations over time. Similarly, a visual scene can be described as luminance variations, but instead of being described as variations as a function of time, they are described as a function of space.

A full understanding of this theory requires knowing that the size of the retinal image depends on the distance from which an object is viewed. For an image of a given size and for a given distance, the size on the retina is twice as small if the image is twice as far. When we look at an image, the spatial frequency thus depends

on variations in luminance ("light/dark") and on the distance from which the image is perceived. For a given visual angle, there are a number of these variations. For example, an object with a diameter of 175 mm and located 10 m from the person subtends a visual angle of about 1°. A variation of periods alternating between light luminance and dark luminance is called a cycle. This makes it possible to express what is viewed on the following terms: the number of cycles per degree of visual angle. That is called *spatial frequency*, which is one of the four characteristics allowing the understanding of Campbell and Robson's idea.

A visual scene—a grating—like the one shown in Fig. 6.9 can be described by means of a sine wave. The *spatial frequency* is higher in C than in A or B. The difference between A and B is due to a second characteristic: the *contrast*. For a given cycle ranging from a light band to a dark band, the intensity variation is not the same. The light band is brighter in B than A. When the differences between light and dark bands are large, the contrast is high. If the contrast is too low for perceiving a difference between the two areas, it means that the contrast is below the visibility threshold. The contrast level can be quantified by means of a percentage scale ranging from 0 to 100 %, i.e., from the weakest (incapacity to perceive) contrast to the highest contrast.

Two more features complete the description of a visual scene. The gratings like those in Fig. 6.10 (left column) are identical, but their position is not the same. It is their *spatial phase* that distinguishes them. Finally, the bars of the gratings can be more or less inclined. Those on the left column and those on the right column differ on the basis of a fundamental characteristic called *orientation*.

In everyday life, visual scenes are rarely that simple or as clear-cut as described in Figs. 6.9 and 6.10. Figure 6.11e, for example, is more complex. However, it contains a series of simpler elements. Using a mathematical procedure known as Fourier analysis, it is possible to decompose a complex scene on the basis of simpler elements, in this case a series of sine waves. The gratings in Fig. 6.11a, b are used to form the grating illustrated in Fig. 6.11d. If we add the grating in Fig. 6.11c to the ones in Fig. 6.11a, b (or to the grating of Fig. 6.11d), we obtain the complex figure
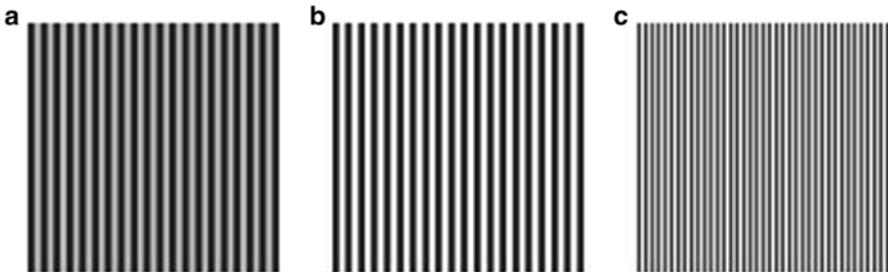

Fig. 6.9 Spatial frequency is much higher in (**c**) than in (**a**) or (**b**), but (**a**) and (**b**) differ because the contrast is higher in (**b**) than in (**a**)

Fig. 6.10 Gratings in the *left column* have different phases, whereas those in the *right column* have different orientations
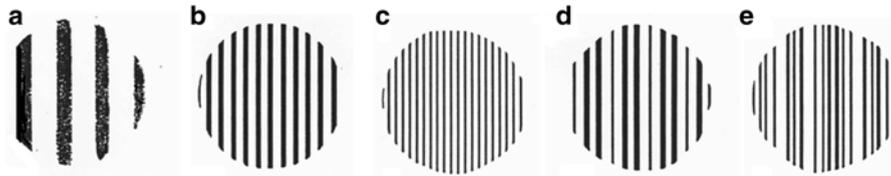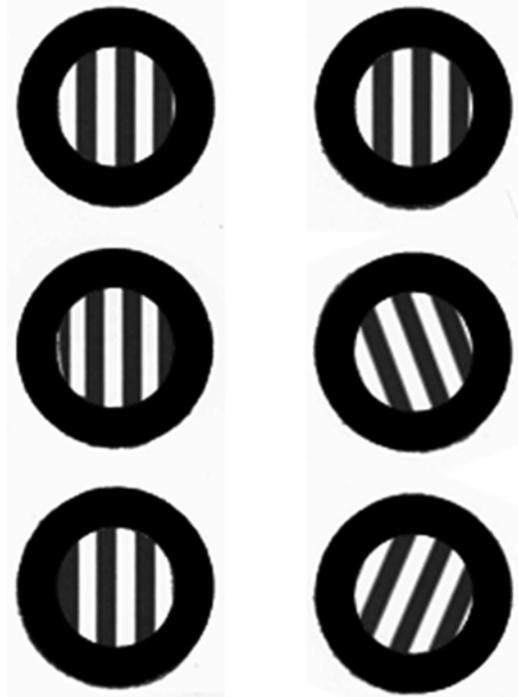

Fig. 6.11 Grating (**e**) is complex, but is made in the end of the mixture of gratings (**a–c**); and grating (**d**) results from the mixture of gratings (**a**, **b**)

reported in Fig. 6.11e. Note that the spatial frequency in Fig. 6.11c is much higher than the one in Fig. 6.11b, which is itself much higher than that in Fig. 6.11a.

Although we do not consciously have the impression that they are there, each of the components of a grating like the one shown in Fig. 6.11e acts on the brain. These components excite different sets of neurons. For each component, there is therefore in the visual cortex a set of neurons that are specific to it. At the cortical level, for viewing a form, it is necessary to synchronize the activity of a series of specialized neurons.

In the context of Campbell and Robson's explanation based on spatial frequencies, such a set of neurons is called a channel. This channel essentially acts like a frequency detector. Each channel is sensitive to the spatial frequencies which extend over a narrowband. Also, because multiple channels are often activated at the same

Fig. 6.12 *Block image* of
Cosette, blurred when seen
from up close, but clearer
as we move away. Cosette
is a character from Victor
Hugo's *Les Misérables*,
drawn originally by
illustrator Émile Bayard
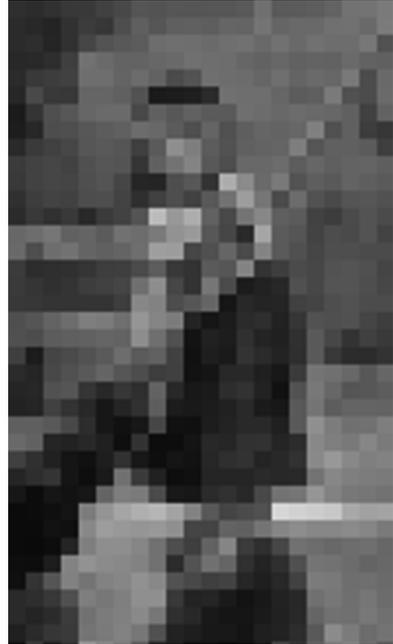
time, we talk about the theory of multiple channels, and more specifically about the
multiple spatial channels theory, since it refers to spatial frequencies.

When using images made of squares like that reported in Fig. 6.12, it is possible
to realize that perceiving a form means filtering what these images are in terms of
the spatial frequencies that they contain. Depending on the distance between the
observer and the image, it is not the same spatial frequencies that are involved, and,
therefore, different specialized channels are activated. The squares add noise to the
image. When we are at a normal reading distance, the image is not clear: a series of
small squares are perceived. When we move away, the spatial frequency is changed
and a clearer vision is restored.

On a more practical level, this means that if an image with squares is presented
during the news on television to hide the face of a criminal or accused, you now
know that you have better chances to identify the person if you move away from the
TV! You also know that when you change the angle with which you are looking at
something, you see things differently. This information will be precious the next
time you visit a visual art gallery or museum where paintings are exhibits. Your
impression on a piece of art might change if you adopt different perspectives. The
effect is even more striking in a gallery like the Orsay Museum in Paris, for exam-
ple, which exhibits the works of great Impressionist masters. Different angles and
different distances allow to increase the appreciation of the works, for instance, of
Van Gogh, Renoir, or Monet or of artists like Georges Seurat and Paul Signac who
used pointillism to create impressions.

6.3.2   *Contrast Sensitivity Function*

The multiple spatial channels theory offers a new interpretation of the perception of form and, by the same occasion, a new way of approaching the study or the measurement of visual abilities. We are able to perceive images at different distances with the involvement of different spatial frequencies. However, for the different spatial frequencies, we do not have the same efficiency for perceiving. As is the case for the range of audible frequencies in the field of audition (Chap. 2), or visible wavelengths as we saw in the previous chapter, we are not sensitive to all spatial frequencies.

More specifically, just like it is possible to compensate some deficit in the perception of some auditory frequencies by increasing the loudness of the sound, it is possible to perceive an image, for certain spatial frequencies, only by increasing the contrast. In other words, the perception threshold has to be increased. This link between the spatial frequency and the perception threshold is described by what is called the *contrast sensibility function* (CSF). In brief, the contrast sensitivity is described as a function of the spatial frequency.

For humans, the sensitivity is at its maximum at about 3 cycles/degree. It is at this frequency that the threshold is the lowest. The extent of the sensitivity of the visual system varies from one animal species to another and depends on the light level. Given the demands of their environment, it is not surprising to learn that goldfish have a maximum sensitivity for images with a spatial frequency of about 0.3 cycle/degree, as opposed to the hawks that are crisscrossing the sky, searching for prey on the ground, who have a maximum sensitivity for spatial frequencies of about 30 cycles/degree.

This CSF concept has interesting practical implications. In fact, it measures the visual abilities more completely than does the traditional visual acuity test, the Snellen chart (Fig. 6.13). With the latter, the visual ability is only tested in an optimal condition, i.e., in a condition where contrast is high. Also, tests are executed

Fig. 6.13   A few lines from the Snellen chart

only with high frequencies. The visibility conditions we face in everyday life are not always optimal. For example, it may well happen that you have to drive a car in a more or less thick fog or when it is snowing or raining heavily. In such occasions, contrast is not at all at maximum. It may well be that people with the best visual acuity, as measured by the Snellen chart, do not have the greatest contrast sensitivity for low spatial frequencies. This issue also applies to aircraft pilots who must deal with all kinds of weather conditions, including flying through the clouds (Ginsburg, Evans, Sekuler, & Harp, 1982).

For historical purpose, note that the Snellen chart (or test) is a traditional eye examination tool for quantifying visual acuity. Developed by the Dutch Herman Snellen in the mid-nineteenth century, this tool is still used today. These charts are calibrated in different ways (different letter sizes). A conventional manner of using consists in reading from a distance of 20 ft (6 m in Europe). Letters have to be read with of a single eye, one letter at a time, down to the smallest letters that can be read. The goal is to determine whether a person can read at 20 ft what is normally read from that distance. When we say that a person has a 20/15 read, we say that that person reads at 20 ft what a person normally reads at 15 ft. This person has a good view.

6.4   Form Recognition

It is difficult to evoke form perception without mentioning the ability to recognize it. For recognizing a form, we have to have a representation of it. It therefore becomes necessary, for understanding the mechanisms of form perception, to refer to higher level concepts. This part of the chapter will not be just about what stimulates the retina but rather about what is kept from these stimuli and from the different visual scenes, objects, or faces.

6.4.1   Templates or Characteristics?

To fully understand the theoretical interest of the study of recognition, it is first necessary to understand the requirements of the task. All readers of this book know the letter A and can easily recognize it, whether it is a, **a**, *a*, A, **A**, or *A*. Yet, it is likely that you have never seen one or some of these As. Now imagine all versions of A you wrote by hand or even better and all versions of A that all humans have written in the past year. Even having seen only a very small percentage of these As, you would be able to recognize most of them. This indicates that we do not have to have seen everything a first time to make it possible to recognize some forms. Despite all the possible transformations of the same object, we do recognize it. This capability to recognize a visual stimulus in spite of the multiple changes it has undergone, or despite the new perspective for observing it, is called invariance.

A first model to account for this ability to recognize is called "template match-ing." According to this model, a template is kept in memory and superimposed on a form in order to verify to what extent this template and the form are similar. Once the learning of letter A is completed, it becomes possible to attempt to match the template that was learned with the one that is perceived. Such a theoretical perspec-tive is based on the need to store a vast repertoire of images and templates. This idea has the advantage of being simple, but does not allow to really explaining invari-ance. We should learn everything a first time, which does not seem very economical when considering the space in memory such learning would require for storing information.

Rather than learning templates, perhaps we learn features. This perspective states that the stimuli are rather defined as combinations of basic features. If we take the previous example, learning letters, instead of making comparison with a template, it would rather be a comparison with the definition of what is retained. For letters, relevant questions would be, for example, the following: "Are there any lines with a vertical orientation? Are there any intersections? Are there any curves?" Given the specificity of certain cells for processing depending on the orientation (Chap. 4), this model has a certain plausibility from a physiological point of view (Hubel & Wiesel, 1968). Also, in an experiment in which the task is to determine whether the two letters presented are identical or not (Gibson, Schapiro, & Yonas, 1968, in Reed, 1982), the response time will be longer if the letters are alike (e.g., P and R), than if they are not (e.g., G and W). In other words, the processing time is longer and must be more complete if several features are in common.

So it seems that form recognition is based on features. But how does this process work? Does it work in sequence, where each element is processed successively, or is there some simultaneous processing, i.e., a parallel processing? According to Selfridge (1959), who developed a theory called the pandemonium, this processing is done in parallel and involves three steps. The different features (curve, angles, etc.) are first recorded, and then, specialized units (feature demons) take care of identifying them. Units representing letters (cognitive demons) then handle the need to reveal the level of agreement between the letter they represent and the recorded features. At a third level, units (decision demons) would be assigned to the identifi-cation of cognitive demons having demonstrated the highest level of agreement.

6.4.2   A Computational Approach

In the pandemonium theory, there is a first sign of a computational theory of form recognition. The goal within this approach is to develop programs (series of cal-culations) used to make connection between what occurs on the retina and the representation of objects and of the physical world. If neurophysiology provides information about the hard drive, it does not inform us about the dynamics (the processes involved) allowing to perceive and recognize form.

For Marr (1982), the perceptual representation is a construction involving different steps. There is a first filtering step which allows the extraction of the main features of an image. According to the filter properties that can be associated with receptor fields which have different sizes and allow to accentuating contours, it becomes possible to extract more or less rough idea of the image. For instance, narrower filters are more sensitive to higher spatial frequencies.

The information derived from this filtering operation thus results in a primitive sketch in two dimensions (2D). This is a fundamental first step in the computational theory of Marr. The different variations of light intensity reaching the retina are translated into features such as curves, intersections, etc. In short, the contours are detected and the main features of the image are drawn. This step can be compared to that of the draft in pencil performed by a painter. Next is a 2.5-D representation where the features are rather arranged according to the direction, depth, shadows, or texture. At this stage, the object is not yet a structured whole. All tridimensional information is not fully grasped. At this stage of processing, the sketch depends on the perspective of the observer, and, consequently, a change of perspective might prevent the recognition. The third stage is that of the 3-D model. It is centered on the object rather than on the observer's perspective. The surfaces are structured in volumetric components.

6.4.3 A Structural Model

Another model to account for the tremendous ability to recognize form was proposed by Biederman (1987). According to this author, this recognition is based on structural components. Somewhat along the line reported earlier for describing letters according to structural features, one could describe the objects based on a set of basic structures. One could compare this viewpoint to the idea that a few dozen phonemes allow to produce and recognize the thousands of words in a language (see Chap. 3). Thus, the description of all objects might be reduced to a series of basic components. Objects in memory would therefore be represented in the form of a spatial arrangement of geometric components. These components act in some way as phonemes in language. They are called geons, for "geometric ions." According to Biederman, there are 36 geons; Fig. 6.14 illustrates some of these. Note that these geons resemble cylinders that Marr and Nishihara (1978) were using to describe different forms.
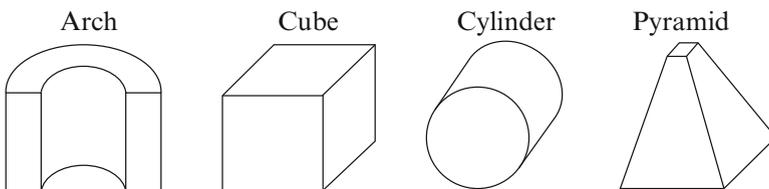

Fig. 6.14 A few examples of geons, the basic structures of the recognition-by-components model of Biederman (1987)

Geons can be extracted directly from 2D features. It is these geons and their relative position that determine the object. If we have a cylinder with a bow on the side, we have the representation of a cup; but if the bow is on top, it will be rather a bucket. Whatever an observer standpoint is, this description in terms of structures and their relationship does not change. With this model, it becomes possible to account for this crucial property, that is, spatial invariance. This model is interesting because it offers some resistance to constraints that sometimes accompany the perception of objects. For recognizing, the key point is to avoid the degradation of geons.

6.4.4   Agnosia

There exists a deficit specific to the identification or recognition of objects. This problem is called agnosia and is diagnosed as such when it is neither an intellectual disability nor a sensory disorder or a disorder of language. In general, it is said that agnosia is caused by perceptual problems or problems of representations in memory.

A first type of apperceptive agnosia may depend on the difficulty of extracting basic features like corners or edges. In other words, this is a very serious problem. This case is referred to as form agnosia. Another type could occur even when the features can be perceived, but in such cases, it is not possible to extract a whole configuration. This is called integrative agnosia. Another type of agnosia is called transformational. In this case, the agnosia is caused by the difficulty to recognize objects presented from a new angle.

Moreover, there are two categories of agnosia related to a problem of mnemonic representations. There is agnosia caused by the loss of structural representations, which entrains imaging trouble and the loss of a sense of familiarity with the object. The other kind of agnosia is the so-called associative agnosia, characterized by the inability to find the meaning of the object. This agnosia occurs because the semantic representation in memory is deteriorated or because it is not possible to access this representation.

Finally, prosopagnosia is the name given to disorder consisting of an inability to recognize faces, even one's own face. In such a case, the view of the face does not activate a sense of familiarity or biographical elements. In some cases, the affected person presented with a face is unable to identify whether the presented face is that of a young or of an old person, or that of a man or woman, or what facial emotion is expressed. Note in conclusion that face recognition is in itself a specific and fascinating subfield of study of form recognition (Tsao & Livingstone, 2008). One of the debated issues is related to the fact that the face would first be perceived as a whole (holistic model) as opposed to a viewpoint where the features and their spatial organization would be analyzed before face recognition.