Chapter 5
Color Perception

Colors are everywhere in our lives and we could not imagine life without them. They are useful, if only to inform us about the state of ripeness of a fruit or the status of any other food that we are about to eat. They help in basic tasks such as the detection and discrimination of objects. In addition, and perhaps most importantly, they make life enjoyable. For example, we are sensitive to the color of the walls or clothing and to their arrangement.

Yet, the study of colors long remained a mystery. We will see in this chapter that understanding the perception of colors requires the integration of the basic concepts about the nature of physical stimuli underlying visual sensation and about retinal physiology. These concepts are necessary if we want to understand what the brain must deal with for providing relevant information about what is colored in the environment.

5.1 Description of Light

Each sensory receptor is particularly sensitive to a specific form of stimulation. For example, stimuli may be chemical, as in the case of taste or smell, or mechanical, as in the case of touch. If the ear is sensitive to variations in the air pressure, the eye is for its part sensitive to electromagnetic radiation. Light, which is a particular form of this radiation, produces a visual response. Light can be described either by considering that the irradiated energy is propagated in the form of a continuous wave or by considering that it is composed of specific matter particles, the photons.

S. Grondin, *Psychology of Perception*, DOI 10.1007/978-3-319-31791-5_5

5.1.1 Intensity

Light intensity could be expressed in number of photons, but it is agreed to use different photometric units. The basic unit of photometry is called *candle*. A candle is the standard value of light intensity. For example, with a wavelength of 555 nm, a candle produces an amount of energy slightly above 0.001 W.

For understanding color perception, it is important first to identify the nature of what reaches the eye and to distinguish two types of sensory experiences, the incident light and the reflected light. The amount of energy that comes directly from a light source is the radiance, or luminous flux, whereas the amount of light emanating from that source and reaching a surface is called incident light or illuminance. Meter-candle is the term used to describe the illuminance, and this equals the illumination of a 1-m^2 surface located 1 m away from a standard candle.

The light from a source rarely reaches the eye directly, unless someone looks at this source directly. Most often, the light is reflected from various surfaces in the direction of the eye. This reflected light is called luminance. It is sometimes referred to as surface light. The luminance of a surface is expressed with a unit called candle per square meter (cd/m^2), i.e., the amount of light reflected in all directions by a surface (reflecting and diffusing light perfectly) illuminated by a meter-candle. Because the luminance was once expressed in footlambert or millilambert (mL), one can still find these units in some textbooks. To give a rough idea of the value of different luminances, snow in the sun provides 10^5 cd/m^2; an overcast sky is about 3000 cd/m^2; easy reading requires a luminance of 100 cd/m^2; and the absolute threshold is about 10^{-6} cd/m^2.

The luminance of a surface definitely depends on the incident light and also on another property called reflectance. The reflectance of a surface is its ability to reflect light. Reflectance is expressed with a coefficient. Thus, a surface which has 70 % reflectance reflects 70 % of incident light:

$$\text{reflectance} = (\text{luminance} / \text{illuminance}) \times 100$$

Sometimes, the concept of retinal illuminance could be useful. This is the amount of light that reaches the retina, and this quantity is expressed in *trolands*.

5.1.2 Wavelength and Spectral Composition

As a whole, the electromagnetic spectrum ranges from 10^{-14} to 10^8 m. This, however, is only the part of the spectrum that is visible. The eye can only perceive wavelengths that lie between 400 and 700 nm (Fig. 5.1). A nanometer is 10^{-9} m. Waves that are a little below 400 nm are called ultraviolet rays; waves above 700 nm are referred to as infrared rays. Although at the physical level the variety of waves ranging from 400 to 700 nm is a continuum, perceptually, the human observer rather

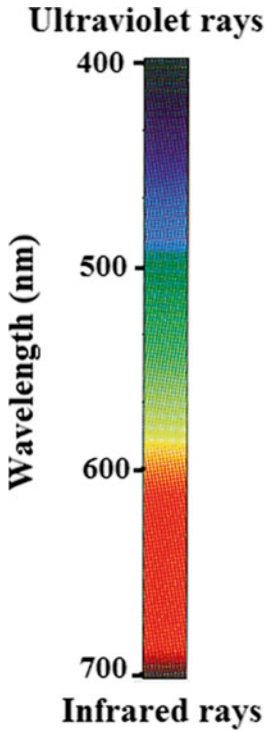

Fig. 5.1 Visible wavelength in the electromagnetic spectrum

distinguishes color categories. We can distinguish hundreds of colors, but in everyday life, we most often refer only to a few categories. In fact, we will see in the next section to what refers exactly the term color.

It is extremely rare that a light beam contains only one wavelength. Should this happen, it would be called a monochromatic light. Most often, a light beam comprises several wavelengths and thus composes a so-called polychromatic light. All the light energy, however, will not necessarily be distributed equally among all wavelengths. Indeed, different lights vary according to their different spectral compositions. The relative importance of the different waves therefore varies from one light to another.

Between a monochromatic light and a polychromatic light extending over a wide range of waves, there are many possible variations. If a light is monochromatic, it will be reported as being pure. Indeed, the more light is concentrated in a narrow band, the purer it is. In contrast to the purity of the monochromatic light, there may be a case where, for a given beam, all the light energy of all visible wavelengths is distributed into equal proportions. In such a case, we will refer to a white light, and the purity of this light will be null (zero).

To end this section, it is relevant to note that the composition of the light that reaches the eye depends on two factors. Of course, it depends on the spectral composition of the light emitted by a source. It also depends on the properties of a given surface. We refer to reflective properties, in the case of reflected light, or to transmission properties, when light is transmitted through something. In short, two factors determine what reaches to the eye: the emitted light and the properties of a given surface.

5.2 Perceptual Dimensions of Color

What is normally called color most often refers to one of the three basic dimensions that make up the experience of color. This dimension often called "color" is indeed hue. There are chromatic hues (green, yellow, etc.) and achromatic hues. Chromatic hues are determined by the wavelength, but the achromatic hues rather range from white to black, passing through the different shades of gray. In the latter case, their hue is neutral (we can say that there is no hue).

If the different shades of gray do not differ in their hue, how can we distinguish them? The eye can discriminate these grays, and the black and the white, on the basis of the different degrees of lightness. The continuum extends from zero lightness (the case of black) to maximum lightness or almost (the case of white). In between, there is a whole continuum of gray. In the same way that there are different degrees of lightness for distinguishing achromatic stimuli, there are different degrees of lightness for chromatic stimuli. For either chromatic or achromatic hue, it is indeed the term brightness that is used to refer to this concept of lightness or lightness of stimuli. More specifically, brightness will be qualified as light or dark when describing a surface, but when dealing with a light source, the description will be in terms of more or less intense.

In addition to hue and brightness, there is a third perceptual dimension for describing a visual stimulation. This third dimension is called saturation and refers to the degree of purity of light. For example, one can have the impression that a particular green seems to contain more or less green or, in other words, seems to contain more or less gray. When an impression of gray is larger, it is that the light has lost purity. A light that is losing in purity is said to be less and less saturated. On the contrary, if a green, for instance, seems very accentuated or highly concentrated, it means that the saturation level is high.

If a color would contain a lot of gray to the point of losing the impression that there is any color, this would mean that its saturation is null (zero). What would be perceived then would be located somewhere between white and black. Figure 5.2 synthesizes the three fundamental dimensions to be understood to fully grasp what can be experienced with respect to colors.

5.3 Color Mixtures

In order to efficiently describe the experience of color perception, we must integrate the information above about the physical bases of light stimulation, as well as other principles. Thus, it is necessary to understand the concept of primary colors and to distinguish between additive and subtractive color mixtures.

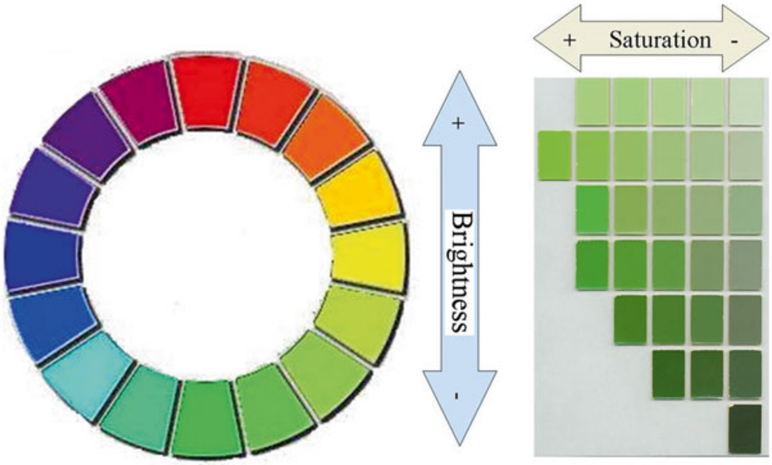

Fig. 5.2 The three basic dimensions at the basis of different shades of colors. The different hues are on the *left*, and the *green squares* on the *right* have different brightness and different saturations

5.3.1 Primary Colors

Despite the adaptive significance of colors in many animal species, and human life in general, it was not until the seventeenth century that new ideas allowed some understanding of the perception of light and color. Until then, the perception of what appeared white to people was interpreted as an absence of color. Intuitively, this naïve interpretation was quite appropriate.

Supported by a simple empirical demonstration, Isaac Newton reported this important idea: the white rather consists of a summation of all colors. His experience consisted of passing beams of white light (sun rays) through a small opening and then through a prism (Fig. 5.3). Beyond the prism, these rays reached a screen. On the screen, these rays did not appeared white anymore, but rather showed the entire color spectrum, the diffraction of the different rays being linked to their wavelength. Newton completed his argument by adding, reversed, a second prism which had the effect of recomposing white light. This demonstration led Newton to conclude that all colors, that is to say, all the wavelengths, were contained in the white light. Newton also advanced another great principle of color perception: to any color corresponds a second color which, mixed with the first, leads to white. These colors are called complementary colors.

Another great idea would later advance our understanding of color perception: there are primary colors, and there are three such primary colors. Primary colors are colors whose combination allows the production of white and the whole range of other colors. Many combinations of colors may constitute the three primary colors. The key point is to select three colors where the mixture of two of them cannot produce the third. On the basis of an arbitrary decision of the *Commission internationale de l'éclairage* (CIE), the three primary colors are defined as blue (435.8 nm), green (546.1 nm), and red (700 nm).

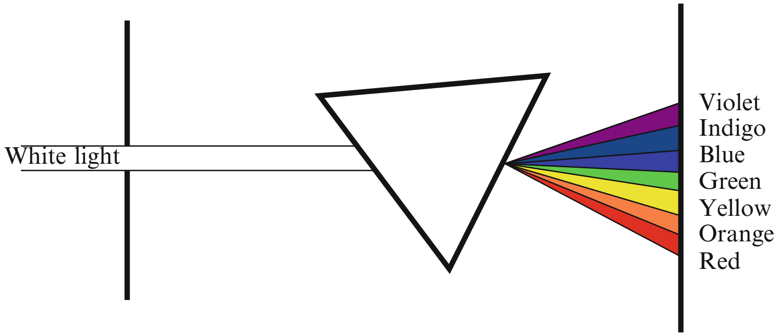

Fig. 5.3 Newton's experiment demonstrating that *white light* contains all colors of the spectrum

5.3.2 Addition and Subtraction

These different concepts relative to light and color are somewhat counterintuitive in that they fail to explain certain phenomena observed in everyday life. For example, working with crayons, each child has experienced the emergence of the green when blue and yellow were mixed. This observation leads some children to believe, wrongly, that yellow, but not green, is a primary color because green would result from a mixture. However, the mixture of two light beams projected on a same location, one that would previously been passed through a yellow filter and the other through a blue filter, will not permit to obtain green. Understanding the difference in the results obtained with crayons and with light beams requires distinguishing between the following two basic concepts: the additive mixtures and the subtractive mixtures.

Common experiences are examples of subtractive mixtures. They are based on the mixture of pigments, that is to say, on the fact that different objects contain a substance which absorbs certain wavelengths and reflects others. Thus, the color of objects does not depend on the properties of light, but rather on how pigments respond to light. In other words, an additive mixture is based on the addition of wavelengths, while a subtractive mixture prevents certain wavelengths to contribute in the color of an object. This impediment is caused by the presence, in this object, of pigments which absorb certain wavelengths. These absorbed waves cannot be reflected and, by extension, will not reach the eye and will not be perceived. With an additive mixture of colors, the resulting color will be brighter than each of the colors used in the mixture; in contrast, a subtractive mixture will result in a decrease of the brightness compared to each of the colors used. Figure 5.4 illustrates the concepts of subtractive and additive mixtures.

It is possible to predict the addition of certain colors on the basis of certain rules. The understanding of these rules is facilitated by observing the color circle shown in Fig. 5.5. This circle illustrates two subjective dimensions of color: (1) the circumference means the hue and (2) the radius designates the saturation. The circumference covers all wavelengths of the visible spectrum from violet (about 400 nm) to

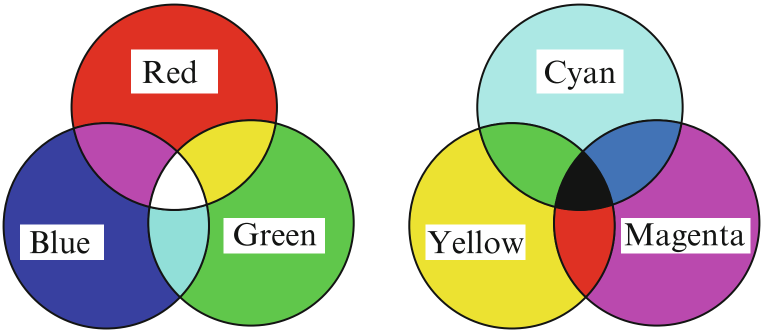

Fig. 5.4 Illustration of the resulting color from an additive (*left*) or subtractive (*right*) mixture

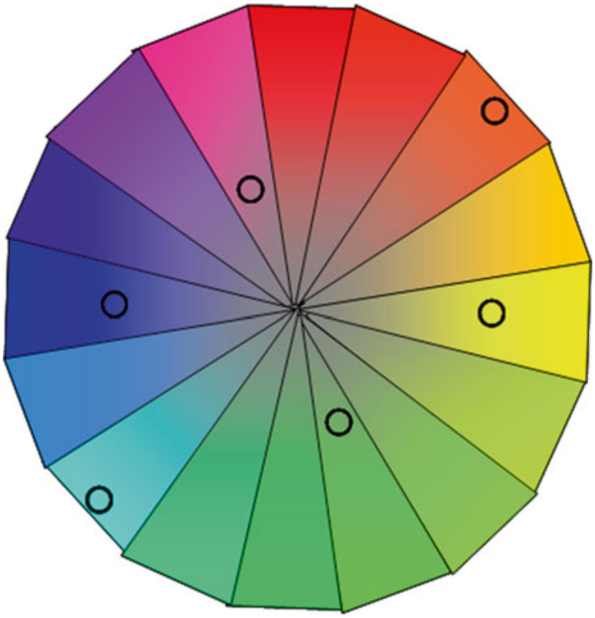

Fig. 5.5 Illustration with the *color wheel* of the resulting additive mixture. For a given pair of points diametrically in opposition, if one provides the same amount of intensity on each side, the resulting mixture is a point in the middle, i.e., some *gray* or *white*

red (about 700 nm). Also, the further away we get from the center of the circle, the greater the saturation is. The center corresponds to a zero degree of saturation, i.e., gray or even to white if the brightness is high.

On this circle, the complementary colors are diametrically opposed. Also, the closer we get to the center, the less saturated they are. If we take two equal amounts of light energy associated with two complementary colors which are located at equal distances from the center, then the resulting mixture gives white (or gray).

However, if one of these two complementary colors is less saturated than the other, it is necessary to increase the intensity of the light flux on the less saturated so that no color persists. Also, by choosing two colors on this circle that are not complementary, it will not be possible to obtain an achromatic mixture. In addition, all colors are not present on the color circle. These are called nonspectral colors and can only be obtained by mixing at least two colors. Purple is an example of nonspectral color.

The addition of color also obeys another law. If we mix equal amounts of different colors, the resulting brightness is greater than the average brightness of the colors used in the mixture. Also, if the mixed quantities are unequal, the resulting brightness is closest to that of the color presented in highest quantity.

Finally, note that there are other types of color mixtures. These other color mixtures are reported below in a section about color effects and illusions.

5.4 Theories of Color Vision

Two major views have long been opposed when attempting to explain color vision. A first view point, supported by Thomas Young in the early nineteenth century and also by Hermann von Helmholtz a few decades later, is known as trichromatic theory of Young-Helmholtz. Essentially, this theory states that color vision depends on the presence of three types of receptors in the eye. It is postulated that these receptors are sensitive to all wavelengths, with a maximal sensitivity for a given length. These receptor types are more sensitive to blue, green, and red. In fact, Young and Helmholtz knew that, for a person having no color vision deficit, an additive mixture of red and green gives yellow. So they explained the vision of yellow by the excitation of the receptors of red and receptors of green. Indeed, according to them, any color could be explained by different excitation levels of the three receptor types.

Later in the nineteenth century, various observations not compatible with the trichromatic theory led Ewald Hering to develop another theory of color vision. In particular, Hering observed that people asked to choose colors that do not seem to be a mixture tend to discern four, and not three, primary colors: blue, green, red, and yellow. He also observed that people never report perceiving a greenish red or a yellowish blue. Moreover, the fact that people perceiving neither red nor green can perceive yellow was also a major objection to the trichromatic theory of Young-Helmholtz. Finally, Hering also knew that prolonged exposure to a color can create a strange effect, as discussed below.

Thus, Hering rather proposed the *opponent process* theory to account for the wide range of perceived colors. This theory states that color perception is based on the operation of pairs of opponent colors. These pairs are red and green, blue and yellow, and white and black to reflect brightness perception. In this way, if a neuron is excited by the presence of a color, it will be inhibited by the presence of the opposite color.

Interestingly, contemporary data from physiology provide support for both theories. With a technique called microspectrophotometry, it is possible to quantify the proportion of light, for a given wavelength, absorbed by the photoreceptors. Thus, it was possible to observe that there are actually three types of cone, each having a maximum light absorption for different wavelengths, as suggested by the theory of Young and Helmholtz. The exact value of these wavelengths varies somewhat depending on the study. For instance, maximum absorptions were reported at 420, 530, and 560 nm in macaques (Bowmaker, Dartnell, & Mollon, 1980) and 425, 534, and 564 nm in humans (Bowmaker & Dartnell, 1980). Since these values loosely correspond to red, green, and blue, respectively, some authors use the terms the red cones, green cones, and blue cones (sometimes also called γ, α, and β fibers). Although it may be simpler to adopt these terms, especially in the context of the trichromatic theory, it is more accurate to call them S, M, and L to respectively designate the cones having maximum light absorption at short, medium, and long wavelengths. Indeed, as the values reported above indicate, the values are closer to the long wavelengths than to the short ones.

Other physiological data rather allow to support the other theory of color vision, that of Hering. However, contrary to the contention of Hering, these opponent processes are not located at the receptor level. An investigation of the functions of nerve cells beyond photoreceptors reveals that some cells actually work according to an opponent principle. This investigation was conducted at different levels between the photoreceptors and the striate cortex, particularly at the level of the ganglion cells and of the lateral geniculate nucleus. In both cases, the opponent responses are comparable. Based in particular on the wavelength at which a cell becomes inhibited rather than excited, DeValois, Abramovet, and Jacobs (1966) grouped the opponent cells of the lateral geniculate nucleus into four categories (see also DeValois & DeValois, 1988):

$$R+G- \quad R-G+ \quad B+Y- \quad B-Y+$$

where R = red, G = green, B = blue, and Y = yellow and where + means that cells are excited by the presence of the designated color and − means that they are inhibited (Fig. 5.7). We also find two types of non-opponent cells in the lateral geniculate nucleus. These cells respond to all stimulations, either by increasing their activity (white+/black−) or by decreasing it (black+/white−).

Thus, color vision can be explained with a system that is somewhat of a compromise between the theories of Young-Helmholtz and Hering. Specifically, this system, shown schematically in Fig. 5.6, has two levels: the three types of cones transmit information to a more central level of processing (DeValois & DeValois, 1975). In the retina, the information is captured by three types of cones reacting optimally to their wavelength: C cones to short waves, M for medium waves, and L for long waves. At ganglion cell level, the information coming from the photoreceptors exert an activating or inhibiting effect on some of the four types of opponent cells or two types of non-opponent cells. For example, the cones sensitive to shorter wavelengths would activate the B+Y− system and inhibit the Y+B− system.

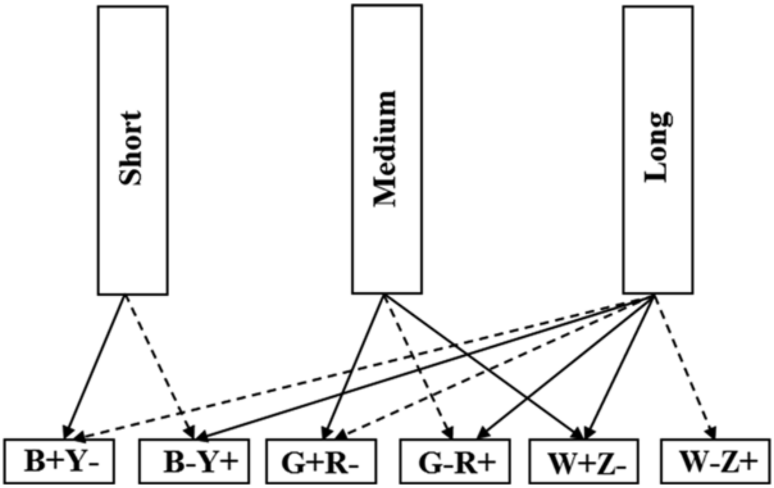

Fig. 5.6 Compromise theory by De Valois and De Valois where color vision depends on the neural activity at two levels. (1) Cones are particularly sensitive to short, medium, and long wavelengths. (2) At the next level, there are opponent (the *four rectangles* on the *left*) and non-opponent (the two on the *right*) processes. *B* blue, *Y* yellow, *G* green, *R* red, *Z* black, *W* white, + activation, − inhibition, *continuous line* activation, *dotted line* inhibition

With such a two-level system, it is possible to explain hue with the excitation of the R+G−, R−G+, B+Y− and B−Y+ opponent processes. These processes also help explain why complementary colors cannot coexist. For example, we cannot perceive greenish red, but perceiving greenish blue makes sense (if the S and M cones are excited). Brightness would be explained by the activity of non-opponent white-black and black-white cells. Finally, saturation would depend on the fact that the activity of the opponent processes would be higher than the one of the White + Black− system.

Finally, color perception probably also depends on other complex mechanisms. Researchers have identified, in the striate cortex, clusters of cells that react only to colors (Livingstone & Hubel, 1987; Michael, 1978). A property of these cells is to have double-opponent receptive fields.

5.5 Chromatic Effects

While there are only few definitive explanations of the different perceptual phenomena related to color, it remains relevant to describe some of them. Some phenomena reveal that color is not simply a matter of wavelengths or physical stimulation. It is possible to obtain one particular color mixture depending on how these colors are presented. It may indeed happen that the brain makes an average synthesis of what is presented. Colored portions of a visual field may be confused because of their density. For example, if small squares of two different colors alternate horizontally and vertically, you can distinguish them from each other if you are close to the image. You discern correctly the color of each square. However, if you sufficiently move away from the image, you will reach a point where you will no longer distinguish colors correctly. The entire image will appear in a different color, which will indeed be the synthesis of the two colors used. This phenomenon is referred to as a spatial optical mixing. In the same vein, it is possible to create conditions leading to a temporal optical mixing. This time, you might very well discriminate between two colors on a circle, but if you were turning the circle (as when spinning a top), this would lead, at a certain speed, to the inability to succinctly distinguish the two colors, and the brain would be forced to make an average synthesis of the two colors.

The effects caused by temporal constraints are not restricted to cases involving colors. Sometimes, black and white arrangements, such as the one in Fig. 5.7, can generate different colors. If one spins such black stripes on white background, colors appear. Since these colors vary from one person to the other, this phenomenon is called subjective colors. According to Henri Piéron, a French psychologist who worked in the first half of the twentieth century, the configuration and the rotational speed of the disk would influence selectively the receptors to red, green, and blue as the receivers do not all have the same response speed. Other authors argue instead that the explanation is not located in the retina itself. The stimulation would reach the brain directly and would produce a sequence of neural events that would be interpreted, because of its resemblance to the actual effect of colored stimuli, as a chromatic stimulus.

The *simultaneous contrast* is a subjective enhancement of color differences. In other words, the perceived hue depends on the context (Fig. 5.8), and this context can accentuate differences. This could be caused, according to Helmholtz, by an unconscious inference about brightness. We will return to this concept of unconscious inference, in the context of depth perception (Chap. 7). For Hering, the effect would rather be due to lateral inhibition (which will be discussed in the next chapter). Essentially, this means that when a region of the reception system is excited by

Fig. 5.7 Arrangement in *black* and *white*—Benham's top—which allows, when spinning quickly, to create an impression of color

Fig. 5.8 Example of simultaneous contrast where the *pink square* in the *middle* appears *darker* on the *right* than on the *left*

Fig. 5.9 Is it possible to change this flag of Ivory Coast into that of France? Yes. You simply need to fixate on the flag above for a minute and then look at a *white surface*. After a few seconds, you should see new colors appearing

a chromatic stimulus, its neighboring regions remain insensitive to stimuli of the same color. What is obtained is rather the activation of the response to the complementary color.

In contrast to a simultaneous contrast, there are *assimilation or equalization effects*. This effect is a subjective attenuation of color differences or of brightness differences when stimuli are placed close to each other. In other words, this effect occurs when a color borrows somehow the color of its neighbor.

A fairly spectacular phenomenon occurs when fixating a surface, and then another surface, rather than looking at two stimuli spontaneously as was the case for simultaneous contrast or assimilation effect. This temporal phenomenon is called *afterimage*. When you fixate on a color image over a long period, say 1 min, then immediately after fixate on a white surface, you see an afterimage appearing. However, rather than seeing the initial colors, i.e., the ones you have previously been fixating, you will eventually see the complementary colors appearing on the white surface (Fig. 5.9).

According to some researchers, the prolonged exposure leading to the formation of consecutive images is due to the fatigue of receptors specialized in the perception of the presented color(s). If, after prolonged exposure, we look at a white surface, which contains all colors, there will be a greater response of the non-fatigued receptors. These receptors are indeed those responding to the complementary color(s). The existence of such a phenomenon in which the complementary colors appear after a fixation period provides support to the Hering's position described above.

Another form of color aftereffect, called the McCollough effect, is particularly fascinating (McCollough, 1965). The effect can be obtained by fixating each of the top gratings (see Fig. 5.10) for about fifteen seconds and then by looking at the other gratings below. The color of the perceived afterimage depends on the orientation of the bars (of the gratings). The color of the afterimages will tend to be red between the vertical bars but green between the horizontal bars. While it is believed that the

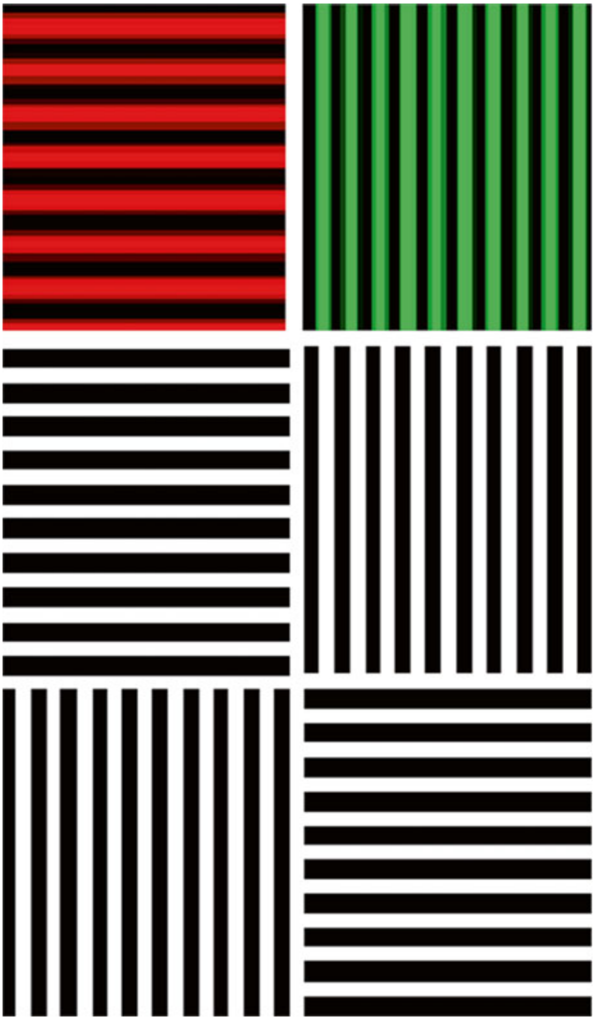

Fig. 5.10 Images required for producing the McCollough effect (see text)

consecutive images like the one described in the previous paragraph should be attributable to the neuronal adaptation at the retina level (a low level of processing), the McCollough effect would rather be caused at a higher processing level, i.e., where orientation is processed (namely, in the V1 area). And what would happen if you lean your head or if you turn the book at 90°? Try it!

Finally, just as there is a constancy phenomenon for other dimensions of visual perception, as we will see later, there is the so-called color constancy. With color constancy, it remains possible to recognize the true color of objects despite the chromatic variations of lighting, if these variations remain moderate. In other words, even if the daylight starts fading, or if an interior room is dimly lit (enough to stimulate the cones) or illuminated with light of a certain color (but not too intense), a red sweater should continue to appear red, as it is, for example, in the light of day. Thus, the visual system probably has the property of transmitting the differences in spectral composition, just like it can transmit intensity differences.

5.6 Clinical Aspects

There are several color vision disorders. The difficulty of discriminating yellow and blue affects equally men and women and touches less than 1 % of the population. The most common color vision problems are related to the discrimination of red and green and occur more frequently in men than in women (approximately 8 % against less than 1 %). This difference is caused by genes. Genes associated with these colors are located on X chromosome. Considering that women receive two X chromosomes instead of one as is the case with men, they will have this color vision disorder only if both X chromosomes are deficient. That is the reason why women are less likely to be affected by a red-green deficit.

There are three major categories of abnormal color vision. The first is called abnormal *trichromatism* and refers to a partial insensitivity to one of the three primary colors. In this category, we distinguish the protanomaly, deuteranomaly, and tritanomaly. People with protanomaly (approximately 1 % of men are affected) require a greater amount of red for perceiving as yellow the red-green mixture. With deuteranomaly, there is a need for a greater amount of green for perceiving as yellow a red-green mixture: it affects about 5 % of men. Finally, we refer to tritanomaly for describing the need for a greater amount of blue for perceiving as "blue green" a mixture of blue green.

A second major category of color vision deficit is called abnormal *dichromatism* and consists in a complete insensitivity to one of the three primary colors. Thus, a protanope, who is blind to red (affecting approximately 1 % of men), sees in yellow and blue, since red and bluish green are seen as gray. A deuteranope is blind to green (affecting about 1 % of men) and also sees in yellow and blue since the bluish red and green are seen as gray. Finally, a tritanope sees only red and green, but this deficit is very rare. Purple and yellow green are seen as gray.

The third major category is *monochromatism*. Extremely rare, this problem means that vision is summed up in shades of gray. It is caused by the lack of functioning cones, and, therefore, there is no surprise that this problem results in a decreased visual acuity.

It should also be noted that color vision disorders can be caused by damage to the V4 area of the visual cortex and not only by a problem related to the functioning of the cones. Finally, it is possible to detect color vision problems using the Ishihara test. This test consists of a series of color plates on which appear through a set of colored points, numbers, or shapes. People with color vision disorders have difficulty, for example, to correctly identify certain numbers when they are unable to perceive the colors used to illustrate the numbers.