1

Colour Vision

1.1 INTRODUCTION

Ten million! That is the number of different colours that we can distinguish, according to one reliable estimate (Judd and Wyszecki, 1975). It is, therefore, no wonder that we cannot remember colours well enough to identify a particular shade. People are thus well advised to take samples of their clothing colours with them when purchasing accessories that are intended to match. They are also usually well aware that it is not enough to examine the colour match in just one type of light in a shop, but to see it in daylight as well as in artificial light. Finally, a second opinion about the match, expressed by a friend or a shop assistant, is often wisely sought.

The above activity involves the three basic components of colour: sources of light, objects illuminated by them and observers. Colour, therefore, involves not only material sciences, such as physics and chemistry, but also biological sciences, such as physiology and psychology; and, in its applications, colour involves various applied sciences, such as architecture, dyeing, paint technology, and illuminating engineering. Measuring colour is, therefore, a subject that has to be broadly based and widely applied.

Without observers possessing the faculty of sight, there would be no colour. Hence it is appropriate to start by considering the nature of the colour vision provided by the human eye and brain. Before doing this, however, a brief description must be given of the way in which it is necessary to characterise the nature of the light which stimulates the visual system.

1.2 THE SPECTRUM

It is fair to say that understanding colour finds its foundations in the famous experiments performed by Isaac Newton in 1666. Before this date, opinions on the nature of colours and the relationships between them were most vague and of very little scientific use, but, after Newton's work became known, a way was open for progress based on experimental facts.

The historic experiments were performed in Trinity College, Cambridge, when Newton made a small hole, a third of an inch in diameter, in the shutter of an otherwise entirely

Measuring Colour, Fourth Edition. R.W.G. Hunt and M.R. Pointer.

dark room; through this hole, the direct rays of the sun could shine and form an image of the sun's disc on the opposite wall of the room, like a pin-hole camera. Then, taking a prism of glass, and placing it close to the hole, he observed that the light was spread out fan-wise into what he was the first to call a *spectrum*: a strip of light, in this case about ten inches long, and coloured red, orange, yellow, green, blue, indigo, and violet, along its length. The natural conclusion, which Newton was quick to draw, was that white light was not the simple homogeneous entity which it was natural to expect it to be, but was composed of a mixture of all the colours of the spectrum.

The next question which arose was whether these spectral colours themselves, red, green, etc., were also mixtures and could be spread out into further constituent colours. A further experiment was performed to test this suggestion. A card with a slit in it was used to obscure all the light of the spectrum, except for one narrow band. This band of light, say a yellow or a green, was then made to pass through a second prism, but the light was then seen not to be spread out any further, remaining exactly the same colour as when it emerged from the slit in the card. It was, therefore, established that the spectral colours were in fact the basic components of white light.

The inclusion by Newton of indigo in the list of spectral colours is rather puzzling since, to most people, there appears to be a gradual transition between blue and violet with no distinct colour between them, as there is in the case of orange between red and yellow. Several explanations of the inclusion of indigo have been suggested, but the most likely is that Newton tried to fit the colours into a scale of tones in a way analogous to the eight-tone musical scale; to do this he needed seven different colours to correspond to the seven different notes of the scale (McLaren, 1985).

In Figure 1.1, the main bands of colour in the spectrum are shown against a scale of the *wavelength* of the light. Light is a form of electro-magnetic radiation, as is also the case for x-rays, radar, and radio waves, for instance, and the property of this radiation that gives it particular characteristics is its wavelength. Radio waves have quite long wavelengths, typically in the range from about a metre to several kilometres, whereas x-rays

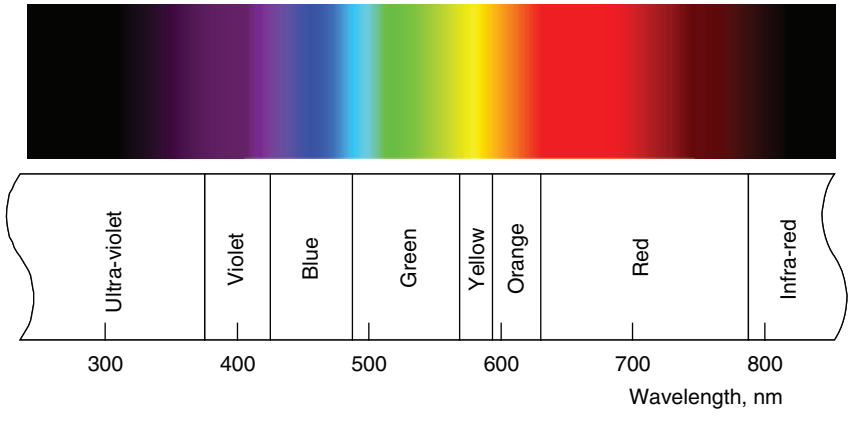

Figure 1.1 The colour names usually given to the main regions of the spectrum. Because of the limitations of printing, the colours of the spectrum cannot be shown accurately, but their general layout is displayed adequately

have extremely short wavelengths, typically about a millionth of a millimetre or shorter. Light waves have wavelengths in between, ranging from slightly above to slightly below a half a millionth of a metre. To obtain convenient numbers for the wavelengths of light, the unit used for expressing them is the *nanometre* (abbreviation, *nm*), which is a millionth of a millimetre, or 10^{-9} of a metre; this is the unit used in Figure 1.1. It must be emphasised that the colour names and wavelength boundaries given in Figure 1.1 are only intended as a rough guide; each colour gradually merges into the next so that there is really no exact boundary; moreover, the colour appearance of light of a given wavelength depends on the viewing conditions, and is also liable to be slightly different from one observer to another. Even so, the names given in Figure 1.1 are useful to bear in mind when considering data that are presented as functions of wavelength. Radiation having wavelengths longer than those of the visible spectrum and less than about 1 mm is called *infrared*; and that having wavelengths shorter than those of the visible spectrum and longer than about 100 nm is called *ultraviolet*. These radiations can provide radiant energy that tans the skin or warms the body, for instance, but they cannot normally be seen as light. In colour science, although it is the long-established practice to identify different parts of the spectrum by using wavelength, it would be more fundamental to use *frequency*. This is because, for light from any part of the spectrum, as it passes through a medium, its wavelength decreases by being divided by the refractive index of that medium; however, the velocity also decreases in the same proportion, so that the frequency (the velocity divided by the wavelength) remains constant. The values of wavelength quoted are usually as measured in air, and, although those measured in vacuum would be more fundamental, they differ by only about 3 parts in 10 000. (The velocity of light in vacuum is about 2.998×10^8 metres per second.)

1.3 CONSTRUCTION OF THE EYE

A diagrammatic representation of a cross-section of the human eye is given in Figure 1.2. Most of the optical power is provided by the curved surface of the *cornea*, and the main function of the *lens* is to alter that power by changing its shape, being thinner for viewing

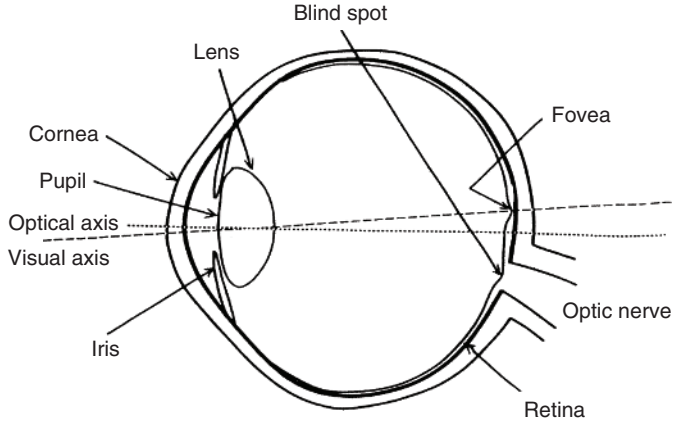

Figure 1.2 Cross-sectional diagram of the human eye

distant objects and thicker for near objects. The cornea and lens acting together form a small inverted image of the outside world on the *retina*, the light-sensitive layer of the eye. The *iris*, the annular-shaped coloured part of the eye that we see from the outside, changes its shape, having a central aperture that is only about 2 mm in diameter in bright light, but which is larger in dim light, having a maximum diameter of about 8 mm. The aperture referred to is called the *pupil*, and is the area through which the light passes. The iris, by changing its diameter, provides some compensation for changes in the level of illumination under which objects are seen; however, this compensation only amounts to a factor of about 8 to 1, rather than the 16 to 1 to be expected from the ratio of the squares of the diameters, because rays that pass through the edge of the pupil are less effective in stimulating the retina than those that pass through the centre, a property known as the *Stiles-Crawford effect*.

The retina lines most of the interior of the approximately spherically-shaped eye-ball, and this provides the eye with a very wide field of view. However, the retina is far from being uniform in sensitivity over its area. Colour vision is very limited for stimuli seen beyond about 40° off the visual axis (Hurvich, 1981), and in this area vision is used mainly for the detection of movement. Within the 40° on either side of the eye's axis, the ability to see both colour and fine detail gradually increases as the eye's axis is approached, the area of sharpest vision being termed the *fovea*, which comprises approximately the central 1.5° diameter of the visual field. An area within this, termed the *foveola*, corresponds to a field of about 1°. A curious feature of the fovea and foveola is that they are not centred on the *optical axis* of the eye, but lie about 4° to one side as shown in Figure 1.2, thus resulting in the visual axis being offset by this amount. About 10° to the other side of the optical axis (equivalent to about 14° from the fovea) is the *blind spot*, where the nerve fibres connecting the retina to the brain pass through the surface of the eye-ball, and this area has no sensitivity to light at all. There is also an area covering part of the fovea, called the *yellow spot* or *macula lutea*, containing a yellowish pigment. In addition to these spatial variations in the retina, there are changes in the types of light receptors present in different areas. In the foveola, the receptors are all of one type, called *cones*; outside this area, there is, in addition, another type, called *rods*. The ratio of cones to rods varies continuously from all cones and no rods in the foveola to nearly all rods and very few cones beyond about 40° from the visual axis. Finally, the individual cones and rods are connected to the brain by nerve fibres in very different ways, depending on their position: in the foveola, there are about the same number of nerve fibres as cones; but, as the angle from the visual axis increases, the number of nerve fibres decreases continuously until as many as several hundred rods and cones may be served by each nerve fibre.

1.4 THE RETINAL RECEPTORS

The function of the rods in the retina is to give monochromatic vision under low levels of illumination, such as moonlight and starlight. This *scotopic* form of vision operates when the stimuli have luminances of less than some hundredths of a candela per square metre (cd m^{-2}; for a summary of photometric terms and units, see Appendix 1).

The function of the cones in the retina is to give colour vision at normal levels of illumination, such as daylight and typical indoor artificial light. This *photopic* form of vision operates when stimuli have luminances of several cd m^{-2} or more.

There is a gradual change from photopic to scotopic vision as the illumination level is lowered, and, for stimuli having luminances between several cd m^{-2} to some hundredths of a cd m^{-2}, both cones and rods make significant contributions to the visual response, and this is called *mesopic* vision. The wavelengths of the light to which the rods are most sensitive are shorter than is the case for most of the cones, and, as a result, as the illumination level falls through the mesopic range, the relative brightnesses of red and blue colours change. This can often be seen in a garden at the end of the day; red flowers that look lighter than blue flowers in full daylight look darker than the blue ones as the light fades. This is known as the *Purkinje phenomenon*.

The rods and cones are so named because of their shapes, but they are all very small, being typically about a five-hundredth of a millimetre in diameter, with a length of around a twenty-fifth of a millimetre. They are packed parallel to one another and face end-on towards the pupil of the eye so that the light is absorbed by them as it travels along their length. They are connected to nerve fibres via an extremely complicated network of cells situated immediately on the pupil-side of their ends. The nerve fibres then travel across the pupil-side of the retina to the blind spot where they are collected together to form the *optic nerve* which connects the eye to the brain. Hence, before the light reaches the receptors, it has to pass through the cells and nerve fibres, which are largely transparent. In each eye, there are about 6 million cones, 100 million rods, and 1 million nerve fibres.

1.5 SPECTRAL SENSITIVITIES OF THE RETINAL RECEPTORS

The rods and the cones are not equally sensitive to light of all wavelengths. In the case of the rods, the initial step in the visual process is the absorption of light in a photosensitive pigment called *rhodopsin*. This pigment absorbs light most strongly in the blue-green part of the spectrum, and decreasingly as the wavelength of the light becomes either longer or shorter. As a result, the spectral sensitivity of the scotopic vision of the eye is as shown by the broken curve of Figure 1.3. This curve is obtained by having observers adjust the strength of a beam of light of one wavelength until the perception it produces has the same intensity as that produced by a beam of fixed strength of a reference wavelength. If the strength of the variable beam had to be, for example, twice that of the fixed beam, then the scotopic sensitivity at the wavelength of the variable beam would be regarded as a half of that at the wavelength of the fixed beam. These relative sensitivities are then plotted against wavelength to obtain the broken curve of Figure 1.3, the maximum value being made equal to 1.0 by convention. To obtain a sensitivity curve representing scotopic vision, it is necessary to use beams of sufficiently low intensity to be entirely in the scotopic range, and the curve of Figure 1.3 was obtained in this way. It is based on results obtained from about 70 observers (22 in a study by Wald, 1945; and 50 in a study by Crawford, 1949) and represents scotopic vision of observers under 30 years of age; above this age, progressive yellowing of the lens of the eye makes the results rather variable. The curve represents the scotopic sensitivity for light incident on the cornea, and thus the effects of any absorption in the ocular media are included. The strengths of the beams can be evaluated in various ways, but the convention has been adopted to use the amount of power (energy per unit time) per small constant-width wavelength interval. If the beams used have the same small width of wavelength throughout the spectrum, then all that is required is to know the relative power

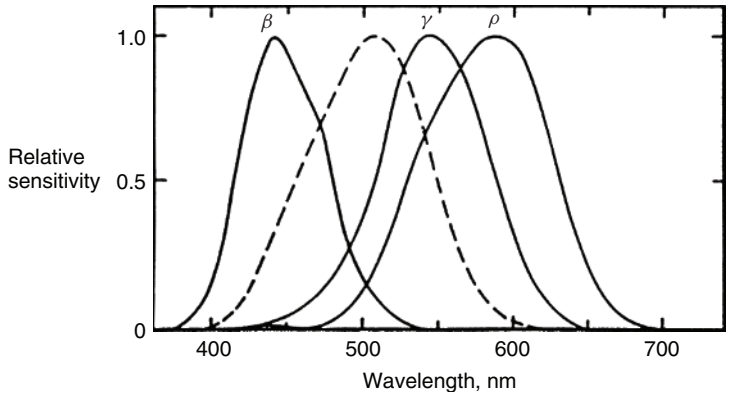

Figure 1.3 Broken line: the spectral sensitivity of the eye for scotopic (rod) vision. Full lines: spectral sensitivity curves representative of those believed to be typical of the three different types of cones, ρ, γ, and β, of the retina that provide the basis of photopic vision. The sensitivities are for equal power per small constant-width wavelength interval

in each beam. However, if the beams have different widths of wavelength, then the values of the relative power per unit wavelength interval have to be determined for each beam.

A system having a single spectral sensitivity function, such as that shown by the broken line in Figure 1.3, cannot, on its own, provide a basis for colour vision. Thus, although, for example, light of wavelength 500 nm would result in a response about 30 times as great as the same strength of light of wavelength 600 nm, the two responses could be made equal simply by increasing the strength of the 600 nm beam by a factor of about 30 times. The system is thus not able to distinguish between changes in wavelength and changes in intensity, and this is what is needed to provide a basis for colour vision. Scotopic vision therefore provides only shades of whites, greys, and blacks, as occur in moonlight.

In the case of the cones of the human retina, it has proved difficult to extract the photosensitive pigments, and our knowledge of them has had to be obtained largely by indirect means. These include very careful measurements of the light absorbed at each wavelength of the spectrum by individual cones removed from eyes that have become available for study (Dartnall, Bowmaker, and Mollon, 1983), and deductions from experiments on colour matching together with data (Estévez, 1979) on colour defective vision (to be discussed in Section 1.10). Also the genes for the pigments have been expressed in tissue cultures, enabling the pigments to be produced for study (Nathans, Merbs, Sung, Weitz, and Wang, 1992). As a result of these studies, sets of curves typified by those shown by the full lines of Figure 1.3 have been obtained. (See also Stockman, Sharpe and Fach, 1999; Stockman and Sharpe, 2000; Stockman, 2008.)

The exact shapes of the curves that best typify the spectral sensitivities of the cones are still a matter of some debate, and in some sets the right-hand curve peaks at about 565 nm (Smith and Pokorny, 1972) instead of at about 585 nm as in Figure 1.3; but the set shown in Figure 1.3 shows all the important features of any reasonably plausible set, and is adequate for our present descriptive purposes. These curves represent the spectral

sensitivities for light incident on the cornea, and allowance has thus again been made for any absorptions in the ocular media; these types of curve are often referred to as *action spectra*. For convenience, they have been plotted so that their maxima are all equal to 1.0.

The full curves of Figure 1.3 have been labelled ρ, γ, and β, to distinguish them. If Figure 1.3 is compared to Figure 1.1, it is clear that the ρ curve has a maximum sensitivity in the yellow-orange part of the spectrum, the γ curve in the green part, and the β curve in the blue-violet part. Various designations have been used by different authors for the three types of cone to which the three curves refer, including L, M, S (Long, Medium, and Short wavelength), π_5, π_2, π_1 (after the work of Stiles), and R, G, B (sensitive mainly to the Reddish, Greenish, and Bluish thirds of the spectrum). The L, M, S and R, G, B designation are perhaps the most widely used, but *L* is used for luminance, *M* for correlates of colourfulness, and *s* for correlates of saturation, and it is convenient to keep R, G, B to represent red, green, and blue lights and colours; hence the similar, but distinctive, Greek symbols, ρ, γ, β, have been adopted instead.

It is clear that there are *three* different curves for the cones, and they correspond to three different types of cone, each containing a different photosensitive pigment. We now have a basis for colour vision. For, if we consider again lights of wavelengths 500 and 600 nm, it is clear that the 500 nm light will produce about twice as much γ response as ρ response, whereas the 600 nm light will produce about twice as much ρ response as γ response. If the strengths of the beams are altered, the ρ and γ responses will also alter, but their ratios will always remain typical of those for their respective wavelengths. Hence, in this case, the strengths of the signals can indicate the intensities of the lights, and the ratios the wavelengths of the lights. We can thus distinguish between changes in the intensity, and changes in the wavelength, of the lights, and hence a basis for colour vision exists. Most colours consist of a mixture of many wavelengths of the spectrum, not just a single wavelength as considered so far; but the above argument is quite general, and changes in spectral composition of such spectrally complex colours will cause changes in the ratios of the cone responses, and changes in the amount of light will cause changes in the strengths of the responses. Of course, in general, there will also be changes in the ratios of the β to γ and β to ρ responses as the spectral composition is changed, and these further assist in the discrimination of colours.

The different types of cone, ρ, γ, and β are distributed more or less randomly in the retinal mosaic of receptors on which the light falls (Mollon and Bowmaker, 1992; Hofer, Carroll, Neitz, Neitz, and Williams, 2005; Williams, 2006). The relative abundance of the three types of cone varies considerably from one observer to another (Williams, 2006), but it is always found that there are many fewer β cones than ρ and γ; one estimate of their relative abundances is that they are, on average, in the ratios of 40 to 20 to 1 for the ρ, γ, and β cones, respectively (Walraven and Bouman, 1966). This rather asymmetrical arrangement is, in fact, very understandable. Because the eye is not corrected for chromatic aberration, it cannot simultaneously focus sharply the three regions of the spectrum in which the ρ, γ, and β cones are most sensitive, that is, wavelengths of around 580 nm, 540 nm, and 440 nm, respectively. The eye focuses light of wavelength about 560 nm, so both the ρ and γ responses correspond to images that are reasonably sharp; the β cones then have to receive an image that is much less sharp, and hence it is unnecessary to provide such a fine network of β cones to detect it.

1.6 VISUAL SIGNAL TRANSMISSION

When light is absorbed in a receptor in the retina, the molecules of its photosensitive pigment are excited, and, as a result, a change in electrical potential is produced. This change then travels through a series of relay cells, and eventually results in a series of voltage pulses being transmitted along a nerve fibre to the brain. The rates at which these pulses are produced provide the signal modulation, a higher rate indicating a stronger signal, and a lower rate a weaker signal. However, zero signal may be indicated by a *resting rate*, and rates lower than this can then indicate an opposite signal. The pulses themselves are all of the same amplitude, and it is only their frequency that carries information to the brain. The frequencies involved are typically from a few per second to around 400 per second.

It might be thought that, as there are four different types of receptor, the rods and the three different types of cone, there would be four different types of signal, transmitted along four different types of nerve fibre, each indicating the strength of the response from one of the four receptor types. However, there is overwhelming evidence that this is not what happens (Mollon, 1982). While much still remains unknown about the way in which the signals are encoded for transmission, the simple scheme shown in Figure 1.4 can be regarded as a plausible framework for incorporating some of the salient features of what is believed to take place. The strengths of the signals from the cones are represented by the symbols, ρ, γ, and β. These strengths will depend on the amount of radiation usefully absorbed by the three different types of cone, and on various other factors (as will be discussed in more detail in Chapter 15).

The rod and cone receptors are shown in Figure 1.4 to be connected to *neurons* (nerve cells) that eventually result in just three, not four, different types of signal in the nerve fibres. One of these signals is usually referred to as an *achromatic signal*; its neurons collect inputs from both rods and all three types of cone. Because of the different abundances of the ρ, γ, and β cones, the cone part of its signal is represented as:

$$2\rho + \gamma + (1/20)\beta$$

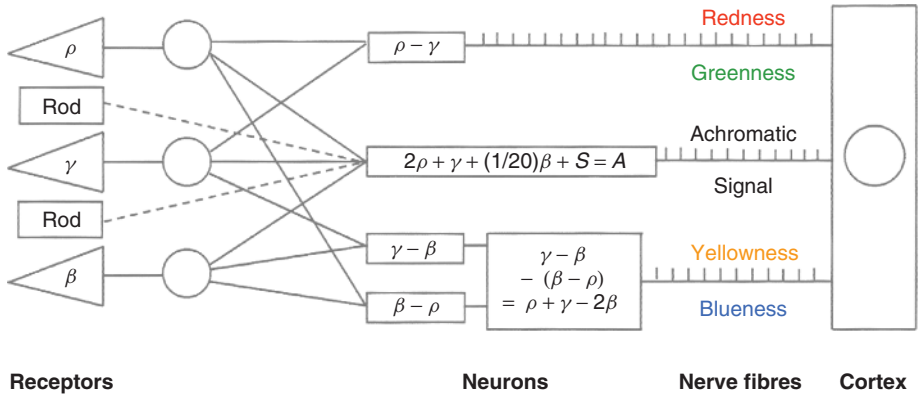

Figure 1.4 Greatly over-simplified and hypothetical diagrammatic representation of possible types of connections between some retinal receptors and some nerve fibres

(The 1/20 factor in the above expression represents a very small contribution from the β cones to the achromatic signal; some studies suggest that there is no such contribution.)

If the scotopic contribution from the rods is represented by S, then the total achromatic signal is:

$$2\rho + \gamma + (1/20)\beta + S = A$$

The other two signals in the nerve fibres are usually referred to as *colour-difference* signals. Three basic difference signals are possible:

$$\rho - \gamma = C_1$$
$$\gamma - \beta = C_2$$
$$\beta - \rho = C_3$$

To transmit all three of these signals would be to include redundancy, since, if two of them are known, the third can be deduced from the fact that $C_1 + C_2 + C_3 = 0$. In fact, there is various evidence to suggest that the signals actually transmitted resemble:

$$C_1 = \rho - \gamma$$

and

$$C_2 - C_3 = \gamma - \beta - (\beta - \rho) = \rho + \gamma - 2\beta$$

Behavioural studies have shown the presence of well-developed colour vision in various non-human species; in many cases, physiological experiments on these species have revealed signals of three general types that are broadly similar to the signals, A, C_1, and $C_2 - C_3$, proposed above.

1.7 BASIC PERCEPTUAL ATTRIBUTES OF COLOUR

We shall now consider some of the perceptual attributes of colour, in the context of these visual signals. There are three basic attributes, brightness, hue and colourfulness; they are defined as follows:

Brightness

Attribute of a visual perception according to which an area appears to exhibit more or less light. (Adjectives: *bright* and *dim*.)

Hue

Attribute of a visual perception according to which an area appears to be similar to one, or to proportions of two, of the perceived colours red, yellow, green, and blue.

Colourfulness

Attribute of a visual perception according to which an area appears to exhibit more or less of its hue.

The achromatic channel is largely responsible for providing a basis for the attribute of brightness: every colour has a brightness, and, as this channel collects responses from all types of receptor, it could indicate an overall magnitude of response for all colours. Hence we could have:

A	large	Bright colours
A	small	Dim colours

There is some evidence that the magnitude of the colour-difference signals may also make a contribution to brightness; if this is so, then the stimulation of the β cones could contribute to brightness even if they did not contribute to the achromatic signal.

If we assume that, for white, grey, and black colours, $\rho = \gamma = \beta$, then the colour difference signals, C_1, C_2, and C_3 would be zero for these colours. The hues of colours could then be indicated thus:

C_1	positive	Reddish colours
C_1	negative	Greenish colours
$C_2 - C_3$	positive	Yellowish colours
$C_2 - C_3$	negative	Bluish colours

The particular hue of any colour could then be indicated by the ratio of C_1 to $C_2 - C_3$, which corresponds to C_1, C_2, and C_3 being in constant ratios to one another; and the colourfulness of colours could then be indicated by the strengths of the signals C_1 and $C_2 - C_3$, zero indicating zero colourfulness (that is, white, grey, or black, the *achromatic colours*), and signals increasingly different from zero indicating the degree to which the hue is exhibited in colours possessing a hue (the *chromatic colours*). The four hues, red, green, yellow, and blue are known as *unique hues*, or *unitary hues*.

1.8 COLOUR CONSTANCY

One of the most important practical uses of colour is as an aid to the recognition of objects. Objects can, however, be illuminated under a very wide range of conditions; in particular, the level and colour of the illumination can vary very considerably. Thus bright sunlight represents a level of illumination that is typically about a thousand times that inside a living room; and electric tungsten-filament lighting is much yellower than daylight. However, the human visual system is extremely good at compensating for changes in both the level and the colour of the lighting; as a result of this *adaptation*, objects tend to be recognized as having nearly the same colour in very many conditions, a phenomenon known as *colour constancy*. Colour constancy is only approximate, and considerable changes in colour appearance can sometimes occur, as in the tendency for colours that appear purple in daylight to appear distinctly redder in tungsten light; but colour constancy is, nevertheless, an extremely powerful and important effect in colour perception.

We can, for the moment, regard colour constancy as corresponding to the ρ, γ, and β responses for whites, greys, and blacks, being approximately equal to one another and constant, no matter what the level and colour of the illuminant. (This will be discussed more fully in Sections 3.13 and 6.12, and in Chapter 15.)

1.9 RELATIVE PERCEPTUAL ATTRIBUTES OF COLOURS

Let us now consider, as an example, a white and a grey patch seen side by side on a piece of paper. If we observe the patches in bright sunlight they will look very bright, and if we take them into the shade, or indoors, they will look less bright: but the white will still look white, and the grey will still look grey. By means which are not fully understood, the eye and brain subconsciously allow for the fact that the lower brightnesses are not caused by changes in the objects, but by changes in the illumination. The same is also true for changes in the colour of the illuminant over the range of typical 'white light' illuminants. Thus, in tungsten light, the patches still look approximately white and grey, and certainly not yellow and brown. (It is to explain these phenomena that the *Retinex Theory* was produced by Edwin Land; see, for example, Land and McCann, 1971.) This is such an important phenomenon that certain relative perceptual attributes of colours are given separate names.

The term *lightness* is used to describe the brightness of objects relative to that of a similarly illuminated white. Thus, whereas brightness could depend on the magnitude of an achromatic signal such as A, lightness could depend on a signal such as A/A_n, where the subscript n indicates that the signal is for an appropriately chosen reference white. Changes in the level of illumination would tend to change A and A_n in the same proportions, thus tending to keep A/A_n constant; hence lightness would tend to remain constant for a given colour. Whites and greys could then be recognised as such by their *lightnesses*, independent of their *brightnesses*.

Just as it is possible to judge brightness relative to that of a white, so it is also possible to judge colourfulness in proportion to the brightness of a white, and the relative colourfulness then perceived is called *chroma*. It is well known that, as the illumination level falls, the colourfulness of objects decreases. Thus, in bright daylight, a scene may look very colourful, but it will look less so under dark clouds; and as the illumination level falls in the evening the colourfulness gradually reduces to zero when scotopic levels are reached. Over most of the photopic range of illumination levels however, the colours of objects are recognised as approximately constant. Let us take, as an example, a red tomato on a white plate. In bright daylight, the red tomato looks very colourful: its red hue is exhibited very strongly. If we then view it at a much lower level of illumination it will look less colourful (its hue will not be exhibited so strongly); but the white plate will also look less bright, and the visual system then subconsciously judges that the lower colourfulness in the dimmer light is caused by the lower level of illumination characterised by the lower brightness of the white. The colourfulness judged in proportion to the brightness of the white is then seen to be unchanged, and this relative colourfulness is the chroma. Thus, whereas colourfulness could depend on the magnitudes of signals such as C_1 and $C_2 - C_3$, chroma could depend on the magnitudes of signals such as C_1/A_n and $(C_2 - C_3)/A_n$ where the subscript n again indicates that the signal is for the white.

Lightness and chroma are therefore defined as follows:

Lightness

The brightness of an area judged relative to the brightness of a similarly illuminated area that appears to be white or highly transmitting. (Adjectives: *light* and *dark*.)

Chroma

The colourfulness of an area judged in proportion to the brightness of a similarly illuminated area that appears to be white or highly transmitting. (Adjectives: *strong* and *weak*.)

Because these two attributes are defined with reference to a 'similarly illuminated area' they apply only to *related colours*, that is, colours perceived to belong to areas seen in relation to other colours. They do not apply to *unrelated colours*, that is, colours perceived to belong to areas seen in isolation from other colours. Self-luminous colours, such as light sources, are usually perceived as unrelated colours; colours produced by objects reflecting light in ordinary viewing conditions are usually perceived as related colours. A television display is of itself self-luminous, but, within the picture area, related colours can be seen if the portrayal is of illuminated objects. Equally the colours of objects within the picture area may be seen as related to the colours of objects within the room; this is especially true when the level of illumination in the room approaches, or even exceeds, that of the television display. A transmitting colour can be perceived as a related colour if seen in suitable relationship to other areas, as in a stained glass window in a church or in a photographic transparency; it is for this reason that the words 'highly transmitting' are included in the above definitions.

If we consider the case of the tomato on the plate again, because the tomato is a solid object, the level of illumination will vary considerably over its surface, being high where the light falls on it perpendicularly, low where it falls on it at a glancing angle, and even lower for those parts in shadow. The brightness of a similarly illuminated white can then be readily judged only for a few areas, and hence lightness and chroma can only be evaluated in these areas. It is also possible however, to judge colourfulness relative to the brightness of the same area, instead of relative to that of a similarly illuminated white; when this is done the attribute is called *saturation*. This attribute can be readily judged at all parts of the tomato; hence, saturation, together with hue, can then be used to judge the uniformity of colour over the surface of the tomato. Saturation could depend on signals such as C_1/A and $(C_2 - C_3)/A$, where A is the achromatic signal for the same area as that of the colour, rather than for the white. Saturation is then defined as follows:

Saturation

Colourfulness of an area judged in proportion to its brightness.

Because the judgement of saturation does not require the concept of a similarly illuminated white, it is applicable to both related and unrelated colours. Consider, as an example of an unrelated colour, a red traffic-light signal seen first directly, and then reflected in a piece of plane glass, such as a shop window. When seen directly, the red signal will usually look quite bright and colourful: but its reflection will look both less bright and less colourful. However, it will still look red, not pink, because its lower colourfulness will be judged in proportion to its lower brightness and it will be perceived to have the same saturation. Thus recognition of the colour of the traffic-light will depend on its hue and saturation, not on its hue and colourfulness.

In photopic conditions of fairly high levels of illumination, the rod response, S, is usually regarded as negligibly small; in this case, when ρ, γ, and β are in constant ratios to one another, $C_1/(C_2 - C_3)$ will be constant, and this implies constant hue; and C_1/A,

and $(C_2 - C_3)/A$ will be constant, and this implies constant saturation; but the brightness will not be constant (which is in accord with the discussion of the 500 nm and 600 nm lights in Section 1.5).

1.10 DEFECTIVE COLOUR VISION

It has long been known that some observers have colour vision that is markedly different from the average of most observers. Such people are popularly known as 'colour blind', but a more appropriate term is *colour defective*, because, in most cases, what is involved is a reduction in colour discrimination, not its complete loss.

There are various types of colour deficiency, and their exact causes are still a matter of some debate (Ruddock and Naghshineh, 1974; Nathans, Merbs, Sung, Weitz, and Wang, 1992; Birch, 2001; Carroll, Neitz, Hofer, Neitz, and Williams, 2004). However, the most likely causes for the various categories are given below in brackets:

Protanopia

No discrimination of the reddish and greenish contents of colours, with reddish colours appearing dimmer than normal. (ρ pigment missing.)

Deuteranopia

No discrimination of the reddish and greenish contents of colours, without any colours appearing appreciably dimmer than normal. (γ pigment missing; the similarity of the shapes of the ρ and γ spectral sensitivity curves on the short wavelength side of their peaks preserves approximately normal brightness of colours.)

Tritanopia

No discrimination of the bluish and yellowish contents of colours, without any colours appearing appreciably dimmer than normal. (β pigment missing; the very small contribution of the β cones to the achromatic signal preserves approximately normal brightness of colours.)

Cone monochromatism

No colour discrimination, but approximately normal brightnesses of colours. (γ and β pigments missing; that is, a combination of deuteranopia and tritanopia.) Or no colour discrimination, and brightness confined mainly to the blue part of the spectrum (ρ and γ pigments missing, that is, a combination of protanopia and deuteranopia).

Rod monochromatism

No colour discrimination, and brightnesses typical of scotopic vision. (No cones present.)

Protanomaly

Some reduction in the discrimination of the reddish and greenish contents of colours, with reddish colours appearing dimmer than normal. (ρ cones having a spectral sensitivity curve shifted along the wavelength axis towards that of the normal γ cones, because of molecular changes in the ρ pigment.)

Deuteranomaly

Some reduction in the discrimination of the reddish and greenish contents of colours, without any colours appearing abnormally dim. (γ cones having a spectral sensitivity curve shifted along the wavelength axis towards that of the normal ρ cones, because of molecular changes in the γ pigment.)

Tritanomaly

Some reduction in the discrimination of the bluish and yellowish contents of colours, without any colours appearing appreciably dimmer than normal. (β cones having a spectral sensitivity curve shifted along the wavelength axis towards that of the normal γ cones, because of molecular changes in the β pigment.)

The degree of reduction in colour discrimination in the cases of protanomaly, deuteranomaly, and tritanomaly (referred to as *anomalous trichromatism*) varies from only slight differences from normal observers, to nearly as great a loss as occurs in protanopia, deuteranopia, and tritanopia (referred to as *dichromatism*). Colour matches made by normal observers are accepted by dichromats, but not always by anomalous trichromats.

The occurrence of these different types of defective colour vision varies enormously, and is different for men and for women. The following figures are estimates for Western races, based on various surveys.

Type	Men %	Women %
Protanopia	1.0	0.02
Deuteranopia	1.1	0.01
Tritanopia	0.002	0.001
Cone monochromatism	Very rare	Very rare
Rod monochromatism	0.003	0.002
Protanomaly	1.0	0.02
Deuteranomaly	4.9	0.38
Tritanomaly	Rare	Rare
Total	8	0.4

The nature of the colour confusions likely to be made by colour defective observers will be described in terms of colorimetry in Section 3.6. Detection of colour deficiency is usually carried out by means of confusion charts, such as the widely used *Ishihara* charts (Dain, 2004), the *City University Colour Vision Test* (Barbur, Harlow, and Plant, 1994), and the *Cambridge (HRR) Colour Vision Test* (Bailey, Neitz, Tait, and Neitz, 2004), or the more diagnostic *Farnsworth-Munsell 100 Hue Test* (Farnsworth, 1943); but accurate classification of the type of deficiency usually requires the use of other test methods (see, for instance, Fletcher and Voke, 1985; Birch, 2001). Most cases of colour deficiency are hereditary, but progressive tritanomaly is also acquired with certain diseases.

1.11 COLOUR PSEUDO-STEREOPSIS

When saturated red and blue lettering is viewed on a dark or black background, some observers perceive the red letters as standing out in front of the plane of the paper, and the blue letters lying behind it, even though the letters are all actually in the plane of the paper. However, although a majority of observers see this phenomenon, there is a minority for whom the reverse effect occurs, with the red letters receding and the blue letters advancing; and there is a third, and still smaller group, who see all the letters in the same plane as the paper. This effect is caused by a combination of the chromatic aberration of the eye and the fact that the pupils of the eyes are not always central with respect to their optical axes. This is illustrated in Figure 1.5. In the left hand diagram, the pupils are displaced outwards relative to the optical axes, A, so that the rays from an object at O are dispersed by refraction with images of blue (B) light in the two eyes being closer to one another than in the case of red (R) light. The red light therefore appears to emanate from an object that is closer than an object from which blue light appears to emanate. In the right hand diagram, the pupils are displaced inwards relative to the optical axes, A, with the result that a blue object now appears nearer than a red object. In the centre diagram, the pupils are concentric with the optical axis, A, and no effect occurs.

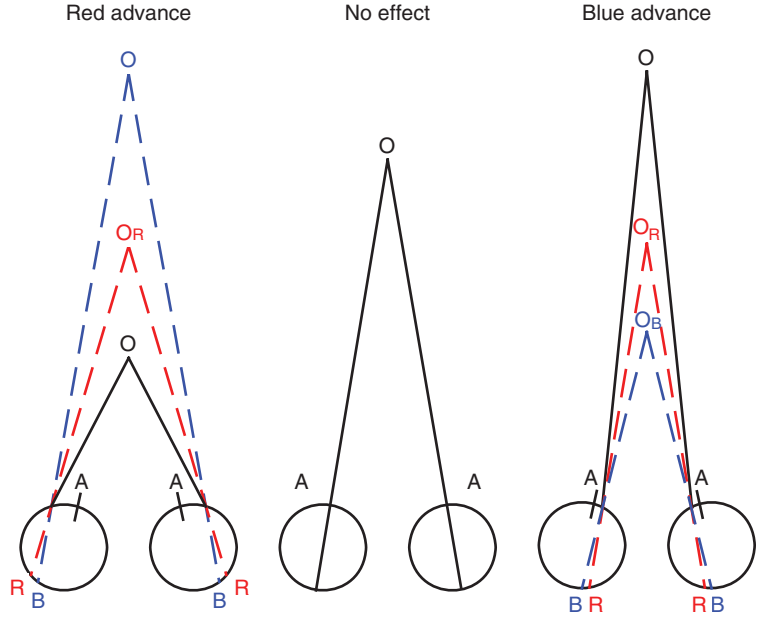

Figure 1.5 Illustration (not to scale) of the basis for colour pseudo-stereopsis

REFERENCES

Bailey, J.R., Neitz, M., Tait, D.M., and Neitz, J., Evaluation of an updated HRR color vision test, *Visual Neuroscience*, **21**, 431–436 (2004).

Barbur, J.L., Harlow, A.J., and Plant, G.T., Insights into the different exploits of colour in the visual cortex, *Proc Roy. Soc. London B*, **258**, 327–334 (1994).

Birch, J., *Diagnosis of Defective Colour Vision*, 2^{nd} *Ed*., Butterworth-Heinemann, Edinburgh, Scotland (2001).

Crawford, B.H., The scotopic visibility function, *Proc. Phys. Soc. B*, **62**, 321–334 (1949).

Carroll, J., Neitz, M., Hofer, H., Neitz, J., and Williams, D.R., Functional photoreceptor loss revealed with adaptive optics: an alternate cause of color blindness, *Proceedings of the National Academy of Sciences*, *U.S.A.*, **101**, 22, 8461–8466 (2004).

Dain, S.J., Clinical colour vision tests, *Clinical Experimental Optometry*, **87**, 276–293 (2004).

Dartnall, H.J.A., Bowmaker, J.K., and Mollon, J.D., Human visual pigments: microspectrophotometric results from the eyes of seven persons, *Proc. Roy. Soc. Lond. B*, **220**, 115–130 (1983).

Estévez, O., *On the Fundamental Data-Base of Normal and Dichromatic Colour Vision*, Ph.D. Thesis, University of Amsterdam, Holland (1979).

Farnsworth, D., The Farnsworth-Munsell 100-hue and dichotomous tests for color vision, *J. Opt. Soc. Am*., **33**, 568–574 (1943).

Fletcher, R., and Voke, J., *Defective Colour Vision*, Hilger, Bristol, England (1985).

Hofer, H., Carroll, J., Neitz, J., Neitz, M., and Williams, D. R., Organization of the human trichromatic cone mosaic, *J. Neurosci*., **25**(42), 9669–9679 (2005).

Hurvich, L.M., *Color Vision*, p.21, Sinauer Associates, Sunderland, Mass., U.S.A. (1981).

Judd, D.B. and Wyszecki, G., *Color in Business Science and Industry*, *3rd Ed*., p.388, Wiley, New York, NY, U.S.A. (1975).

Land, E.H., and McCann, J.J., Lightness and Retinex theory, *J. Opt. Soc. Amer*., **61**, 1–11 (1971).

McLaren, K., Newton's indigo, *Color Res. Appl*., **10**, 225–229 (1985).

Mollon, J.D., Color vision, *Ann. Rev. Psychol*., **33**, 41–85 (1982).

Mollon, J.D., and Bowmaker, J.K., The spatial arrangement of cones in the primate fovea, *Nature*, **360**, 677–679 (1992).

Nathans, J., Merbs, S.L., Sung, C.-H., Weitz, C.J., and Wang, Y., Molecular genetics of human visual pigments, In Campbell, A., ed., *Annual Review Genetics*, **26**, 403–424, Annual Reviews Inc., Palo Alto, CA., U.S.A. (1992).

Ruddock, K.H., and Naghshineh, S., Mechanisms of red-green anomalous trichromacy: hypothesis and analysis, *Mod. Prob. Ophthal*., **13**, 210–214 (1974).

Smith, V.C., and Pokorny, J., Spectral sensitivity of colorblind observers and the cone pigments, *Vision Res*., **12**, 2059–2071 (1972).

Stockman, A., Physiologically-based color matching functions, *IS&T and SID's* 16^{th} *Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, pp. 1–5, IS&T Springfield, VA, U.S.A. (2008).

Stockman, A., Sharpe, L. T., and Fach, C. C., The spectral sensitivity of the human short-wavelength cones, *Vision Res*., **39**, 2901–2927 (1999).

Stockman, A., and Sharpe, L. T., Spectral sensitivities of the middle- and long-wavelength sensitive cones derived from measurements in observers of known genotype, *Vision Res*., **40**, 1711–1737 (2000).

Wald, G., Human vision and the spectrum, *Science*, **101**, 653–658 (1945).

Walraven, P.L. and Bouman, M.A., Fluctuation theory of colour discrimination of normal trichromats, *Vision Res*., **6**, 567–586 (1966).

Williams, D., Color and the cone mosaic, *IS&T and SID's 14th Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, pp. 1–2, IS&T Springfield, VA., U.S.A. (2006).

GENERAL REFERENCES

Bass, M., ed., *Handbook of Optics, Vol. III, Vision and Vision Optics*, McGraw Hill, New York, NY, U.S.A. (2010).

Hunt, R.W.G., Colour terminology, *Color Res. Appl*., **3**, 79–87 (1978).

Gregory, Richard L., *Eye and Brain: The Psychology of Seeing*, 5th *revised Ed*., Oxford University Press, Oxford, England (1998).

Kaiser, P., and Boynton, R.M., *Human Color Vision*, 2nd *Ed*., Optical Society of America, Washington DC, U.S.A. (1996).

Mollon, J.D., Pokorny, J., and Knoblauch, K., (editors) *Normal and Defective Colour Vision*, *revised Ed*., Oxford University Press, Oxford, England (2003).

Valberg, A., *Light Vision and Color*, Wiley, Chichester, England (2005).

Wandell, Brian A., *Foundations of Vision: Behaviour, Neuroscience and Computation*, Sinauer Associates, Sunderland, MA., U.S.A. (1995).