OVERVIEW

Day vision and night vision are two separate modes of visual perception and the visual system shifts from one mode to the other based on ambient light levels. Each mode is primarily mediated by one of two photoreceptor classes in the retina, i.e., cones and rods. In day vision, visual perception is primarily cone-mediated and perceptions are chromatic. In other words, color vision is present in the light levels of daytime. In night vision, visual perception is rod-mediated and perceptions are principally achromatic. Under dim illuminations, there is no obvious color vision and visual perceptions are graded variations of light and dark. Historically, color vision has been studied as the salient feature of day vision and there has been emphasis on analysis of cone activities in color vision. Night vision has historically been studied in terms of rod activity and considerations of the shift from day vision to night vision.

This chapter will review basic aspects of rods and cones and neural pathways that process rod and cone information. Measurement of sensitivity during dark adaptation is discussed as the established measure of the shift between day vision (cone vision) and night vision (rod vision). Clinical assessment of rod and cone sensitivities using dark adaptation function as a means of assessing retinal disease is also discussed. Color vision is discussed in terms of experimental paradigms and theoretical considerations and variations in human color vision are described. Evaluation of color vision can be helpful in understanding the underlying mechanisms of some retinal diseases and suggestions for clinical evaluation of color vision are offered. In the last section of this chapter, new developments in color vision research are discussed.

Table 10.1 The dynamic range of the human visual system

<table>
<thead>
<tr>
<th>Visual environments</th>
<th>Starlight</th>
<th>Moonlight</th>
<th>Indoor lighting</th>
<th>Sunlight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photopic luminance (log cd/m²)</td>
<td>-6</td>
<td>-4</td>
<td>-2</td>
<td>0</td>
</tr>
<tr>
<td>Light category</td>
<td>Scotopic</td>
<td>Mesopic</td>
<td>Photopic</td>
<td></td>
</tr>
<tr>
<td>Photoreceptors</td>
<td>Rods only</td>
<td>Rods and cones</td>
<td>Cones only</td>
<td></td>
</tr>
<tr>
<td>Visual function</td>
<td>Rod absolute threshold</td>
<td>Cone absolute threshold</td>
<td>Rod saturation begins</td>
<td>Damage possible</td>
</tr>
</tbody>
</table>

Rod and Cone Functions

Differences in the anatomy and physiology (see Chapters 4, Autofluorescence imaging, and 9, Diagnostic ophthalmic ultrasound) of the rod and cone systems underlie different visual functions and modes of visual perception. The rod photoreceptors are responsible for our exquisite sensitivity to light, operating over a 10⁸ (100 millionfold) range of illumination from near total darkness to daylight. Cones operate over a 10¹¹ range of illumination, from moonlit night light levels to light levels that are so high they bleach virtually all photopigments in the cones. Together the rods and cones function over a 10¹⁴ range of illumination. Depending on the relative activity of rods and cones, a light level can be characterized as photopic (cones alone mediate vision), mesopic (both rods and cones are active), or scotopic (rods alone mediate vision). In the literature, the terms photopic vision and scotopic vision are used to reflect cone and rod vision, respectively. Table 10.1 shows this overlapping range of activity.

The distribution of rods and cones in the retina (see Chapter 4, Autofluorescence imaging) is also reflected in visual function. The greatest sensitivity to light occurs in the midperiphery of the visual field, which has a predominance of rods, while high-acuity and good color vision are mediated by the fovea, which has a predominance of cones. Nonetheless, the entire retina, with the exception of a very small area within the fovea, is capable of mediating night vision, and color vision is present throughout the visual field with daylight stimulation of the entire retina. The following will introduce rod and cone differences in light adaptation, spectral sensitivity, and spatial/temporal sensitivity.
**Light adaptation**

Photoreceptors, whether they are rods or cones, respond well to only a small range of variations in illumination within a steady adapting background. However, adaptation mechanisms adjust photoreceptor sensitivity so that this small range of responses is always centered near the current adaptation level, even though adaptation levels can vary over a wide range. This behavior forms the basis for the large operating range of the visual system.

It is possible to measure a threshold for the perception of an increment in light on a large, steady background field. As the background light level is increased, the increment threshold starts to increase. Rods and cones behave somewhat differently in this regard. For the rod system, as shown in Fig. 10.1A, the increment threshold increases steadily over almost a thousand-fold range. With further increases in background adaptation levels, an increment is not detected, no matter how much additional test light is presented as an increment, due to rod saturation. In comparison, the cone system, as shown in Fig. 10.1B, shows a continuous steady increase in the increment threshold with increases in background illumination, even at levels that bleach almost the entire amount of available photopigment. The portion of the curve that rises linearly with illumination levels is called the Weber region (Fig. 10.1). In the Weber region, an incremental light can be detected when it is a constant proportion (i.e., the Weber fraction) of the background light level. Different photoreceptor systems have a characteristic Weber fraction. Cones have lower Weber fractions than rods and the M and L cones have lower Weber fractions than S cones. Under optimal conditions, the cone system can detect a light level difference of 1%, while rods need a light change of 20%.

In addition to photoreceptor adaptational properties, other factors, including pupil size, the temporal and spatial summation characteristics, and photopigment depletion, can also contribute to extend the operating range of the visual system over a large luminance range. While some adaptation operates within the photoreceptors themselves, other properties of adaptation may reflect the effects of the complex neural circuitry of the retina.

**Spectral sensitivity**

Day vision is primarily mediated by three types of cone photoreceptors with different but overlapping spectral sensitivities. Each is identified by the relative position of the peak in spectral sensitivity. The three cone types are called the long-, middle-, and short-wavelength-sensitive (L, M, and S) cones. When overall sensitivity to light is measured at the light-adapted fovea, a broad sensitivity spectrum peaking near 555 nm is found. This sensitivity spectrum represents the combined activity of the L and M cones and is called the V(λ) function. When sensitivity to light is measured in the dark-adapted peripheral retina, where rods dominate, a broad-sensitivity spectrum is found with a peak sensitivity at 507 nm. This rod spectral sensitivity function is called V′(λ) (Fig. 10.2A). Both V(λ) and V′(λ) functions have practical significance and have been accepted by the International Commission of Illumination as representative of human vision relative luminous efficiency at photopic and scotopic levels. They are also used to relate luminous (perceived energy of light) to radiant (emitted light) energy.

---

**Fig. 10.1** The increment threshold functions for scotopic (rod) and photopic (cone) vision as a function of background illumination. (A) Rod increment thresholds measured at 9° in the parafovea. The dashed line has a slope of 1. The portion where the curve has unit slope (in parallel to the dashed line) is the Weber region, followed at higher levels by the region of rod saturation. To the right is shown the fraction of rod photopigment bleached. The rods are saturated before there is substantial photopigment bleaching. (B) Cone increment thresholds measured at the fovea. To the right is shown the fraction of photopigment bleached. The Weber region extends to luminances (6 million tds) that bleach virtually all the cone photopigment. (Reproduced with permission from Enoch JM. The two-color threshold technique of Stiles and derived component color mechanisms. In: Jameson D, Hurvich L, editors. Visual psychophysics: handbook of sensory physiology, vol. VII/4. Berlin: Springer-Verlag; 1972.)
Chapter 10
Color Vision and Night Vision

VISUAL PATHWAYS FOR ROD AND CONE FUNCTIONS

Retinal pathways
Rod signals are conveyed by two primary neural pathways that are dependent on the illumination level. One pathway is via ON rod bipolars, AII amacrine cells, and ON and OFF cone bipolars, which are all cells in the retina. This is a temporally sluggish pathway that mediates rod vision at low scotopic light levels. The second pathway transmits rod information via rod–cone gap junctions and ON and OFF cone bipolar cells in the retina. This is a fast pathway that mediates vision at higher scotopic and mesopic light levels. A third insensitive rod pathway between rods and OFF cone bipolars has been identified in rodents but, thus far, not in primates. The significant point here is that rods and cones share neural pathways and have joint inputs to retinal ganglion cells.

Retinogeniculate pathways
There are three major neural retinogeniculate pathways in primates that convey retinal information to the visual cortex. The pathways are named after the layers of the lateral geniculate nucleus (LGN) that receive input from distinct types of ganglion cells and project to different areas of the primary visual cortex. The MC layer of the LGN receives inputs from parasol ganglion cells. The MC pathway processes the summed output of the L and M cones to signal luminance information. The parvocellular (PC) layer of the LGN receives input from midget ganglion cells. The PC pathway mediates spectral opponency of L and M cones (discussed later) to signal chromatic information. The koniocellular (KC) layer of the LGN, which receives input from small bistratified as well as other ganglion cells, detects changes in

Estimates of the spectral sensitivities of the three cone types have been obtained from a variety of psychophysical procedures. One approach was to derive the spectral sensitivities of cones from analysis of color-matching data. Another approach used spectral bleaching lights to depress the responses of two cone types relative to the third so that measurements of the spectral sensitivity of the third cone type could be isolated. Figure 10.2B shows the relative spectral sensitivities of the three cone types. The S cones are most sensitive to light near 445 nm, with sensitivity declining rapidly at longer wavelengths. At 555 nm and longer wavelengths, the S cones are virtually unresponsive to light. The M and L cones have overlapping spectral sensitivities that span the entire visible spectrum. The M cones peak in sensitivity near 543 nm, while the L cones peak near 566 nm. The differential spectral sensitivity functions of the L, M, and S cones provide the foundation of early spectral processing.

Spatial and temporal resolution
Compared with the cone system, the rod system has poorer spatial resolution (acuity). For an observer with 20/20 photopic acuity, scotopic acuity would be about 20/200 (10 times worse than photopic acuity). The rod system also has poorer temporal resolution, which refers to the ability to perceive a physically alternating light as steady or flickering in time. The transitional temporal frequency at which the light appears from flickering to steady is called the critical fusion frequency (CFF). CFFs increase with light adaptation level, reaching a maximum of 20 Hz for rods and 55–60 Hz for cones. This means that flickering lights can be perceived at higher frequencies in brighter light conditions. Interestingly, dark-adapted rods can suppress cone-mediated flicker detection and this suppression mainly occurs in the magnocellular (MC) pathway (explained below).
S-cone signals compared to the sum of the L- and M-cone signals. These three pathways mediate different aspects of vision, with the MC pathway mainly carrying out luminance and motion processing, the PC pathway mainly processing red-green color, acuity, and shape information, and the KC pathway mainly handling blue-yellow color processing.

The sharing of neural pathways between rods and cones implies that rods should have input to the MC, PC, and KC pathways. Indeed, physiological studies have shown that there is strong rod input to the MC pathway, but weak input to the PC pathway. A demonstration of rod input to the KC pathways is less clear. An earlier study did not find rod input to the bi-stratified ganglion cells in the parafovea, while two recent studies demonstrated strong rod input in the peripheral retina. Figure 10.3 shows a schematic diagram of the visual pathways conveying rod and cone inputs to the MC, PC, and KC ganglion cells, which would then produce signals that are projected into the cortex to mediate different aspects of visual perception.

**DARK ADAPTATION FUNCTIONS: ASSESSMENT OF THE SHIFT FROM DAY VISION TO NIGHT VISION**

Measurements of sensitivity thresholds during adaptation to darkness have produced a characteristic biphasic function with an initial segment that is attributed to cone responses and a subsequent segment attributed to rod responses. Figure 10.4 shows a characteristic dark adaptation function measured in the peripheral retina. Thresholds decrease quickly initially and this rapid recovery is attributed to cones. Thresholds then reach a plateau in about 5 minutes and remain invariant for another 5 minutes (cone plateau, reflecting cone absolute thresholds). Then, there is a second rapid decrease in thresholds due to sensitivity recovery of the rods, referred to as the rod–cone break, to a new plateau that is reached in 40-50 minutes (rod plateau, reflecting rod absolute thresholds).

The shape of the dark adaptation function depends on testing parameters, including retinal location, wavelength, and temporal and spatial characteristics. The effects of these parameters on dark adaptation curves can be understood by the differences between rods and cones in terms of their distributions, spectral sensitivity, and spatial/temporal resolution characteristics. For instance, because there are only cones in the fovea, dark adaptation measured at the fovea using a small test light reveals a rapid monophasic branch attributable to the cones. On the other hand, because the rod and cone systems have similar absolute sensitivities at long wavelengths, dark adaptation measured with long-wavelength lights is monophasic, resembling the cone function. As the test wavelength is changed to shorter wavelengths, a biphasic curve emerges because the rods show greater absolute sensitivity than cones at shorter wavelengths.

**Clinical evaluation using dark adaptation functions**

It is known that certain retinal disorders may selectively affect rods (e.g., retinitis pigmentosa) or cones (e.g., cone dystrophies). Clinically, rod and cone functions can be evaluated electrophysiologically by measuring rod and cone electroretinograms (ERG: see Chapter 7, Electrogenesis of the electroretinogram) or psychophysically, by measuring dark adaptation functions. Dark adaptation functions quantify the ability of the rod and cone systems to recover sensitivity (i.e., regenerate photopigment) after exposure to light. The recovery is faster for cones, but the absolute level of sensitivity is greatest for rods. Variations in
sensitivity and sensitivity recovery times can be used to characterize retinal disorders.

Clinical evaluation using dark adaptation functions involves a measure of the cone and rod absolute thresholds and the time of the rod–cone break. Specifically, rod absolute thresholds have been used as a psychophysical supplement to ERG measurement for night-blindness evaluation. The instrument for dark adaptation rod absolute threshold measurement is called a dark adaptometer and the most widely used is a Goldmann–Weekers dark adaptometer (Haag–Streit). This instrument is old, however, and finding replacement lamps is difficult. A new light-emitting diode-based dark adaptometer has recently become commercially available (LKC Technologies scotopic sensitivity tester-1: SST-1). The SST-1 dark adaptometer can determine a full dark adaptation curve as well as full-field scotopic sensitivities.

**COLOR VISION**

Color vision refers to our ability to perceive colors based on spectral variations in light absorbed by the photoreceptors. Color vision includes both chromatic discrimination and color appearance appreciation. Color matching and color discrimination experimental tasks are two fundamental psychophysical procedures that have provided theoretical insights into the nature of color vision and have also been developed for clinical diagnosis of color vision. Color matching and color discrimination results, however, do not address questions of color appearance, for example, why an object appears red. Color appearance is far more complex because it depends on not only the chromatic properties of an object but also the spatial, temporal, and spectral characteristics of the neighboring objects. ¹³³ Neural processing beyond the retina is required for color appearance perceptions.

**Color matching**

**Color matching as the foundation for the theory of trichromacy**

The psychophysical procedure in which an observer sets a mixture of three primary lights to match the color of a test stimulus is called color matching. It has been known since the 19th century that different colors perceived by humans can be specified by an economical three-variable (trichromatic) system. In the 1800s, Thomas Young and Hermann von Helmholtz proposed that there must be three kinds of physiological entities in the eye accounting for this trichromacy and their theory is called the Young–Helmholtz trichromatic theory of color vision. We now know the basis for trichromacy is the existence of three cone types in the retina. Color matching was the means to uncover the spectral sensitivity functions of the three different cone types (Fig. 10.1B). ¹³³

**Color-matching experimental techniques and data**

Theoretically, a color match occurs when the photoreceptor quantal catch for the test stimulus and the quantal catch for the stimulus that is a combination of the three primary lights are the same. Color matches are unperturbed by changes in luminance levels, as long as no significant photopigment bleaching occurs, which means that the basic nature of trichromacy exists under wide variations in light levels. In a classical color-matching experiment, three spectral lights are chosen as the three primaries and the precise setting of the relative percentages of the three primaries to match a test light is the data for that match. When all three primaries are added together, they can be set to match a neutral white.

Procedurally, due to the mechanics of testing optical equipment, the spectral test color and one primary are made to appear in one field and the other two primaries appear in a second field with the two fields usually appearing as two halves of a circle. The observer’s task in a color-matching experiment is to make the two fields appear identical in color. Color-matching data are usually represented in one of two ways. In the first, the amounts of energy of each primary can be plotted as a function of the test wavelength. These are called color-matching or color mixture functions. In such plots, the primary that is added to the test color is given a negative value, and the other two primaries are given positive values. The second plotting method is to normalize the value of each matching primary relative to the sum of the three primaries, leading to the sum of the normalized primaries being equal to 1. Therefore, a plot of one primary value against a second primary value is sufficient to display all the information in the color-matching data. This kind of plot is called a chromaticity diagram, which shows the relative contributions of each spectral primary needed to match any spectral light.

**The CIE colorimetric system**

This spectral primary-based chromaticity diagram has proven to be very useful as a generalized color specification system. However, linear transformation is needed to compare data collected with different choices of primaries. In 1931, the Commission Internationale d’Eclairage (CIE) standardized the spectral-primary-based chromaticity system by adopting three imaginary primaries (X, Y, Z primaries) that are out of the spectral locus and therefore do not exist physically. The choices of the imaginary primaries were based on two important considerations: first, to ensure that the chromaticities of all physical lights have positive values and, second, to relate colorimetric functions to the previously adopted luminosity function, V(λ). In the system, the color-matching function for the Y primary is identical to V(λ) of the photometric system as designed.

Figure 10.5A shows the 1931 CIE chromaticity chart, in which the coordinates of the Y primary [y = Y/(X + Y + Z)] are plotted against those of the X primary [x = X/(X + Y + Z)]. The spectrum loci (their wavelengths are indicated on the graph) form the horseshoe-shaped curve. Equal-energy-spectrum (EES) light is plotted in the center, with the coordinates of x = 0.3333 and y = 0.3333. A straight line connects 400 nm to 700 nm for purples, which result from mixtures of short- and long-wavelength light. All possible lights occur within the boundaries of the spectrum locus and the purple line. Highly saturated colors occur near the locus and desaturated (pale) colors occur near the white point.

The 1931 CIE system was based on 2° field color-matching functions that were derived from color matches of many observers. The averaged color-matching function from these observers was treated as the standard. Therefore, an observer with the standard color-matching function is referred to as the standard observer. Since the color-matching data are affected by the size of the stimulus field presented to the observer, in 1964, the CIE also adopted a large-field XYZ system that was based on the 10° field standard observer color-matching functions. For large stimuli, such as those generated by Ganzfeld for ERG measurements, it is recommended to use the 1964 CIE chromaticities to reflect more accurately the large field color-matching functions.
**Cone chromaticity space**

The CIE colorimetric system is valuable for light specifications; however, psychophysical experiments using the CIE system cannot yield results that allow easy interpretation of the underlying physiological mechanisms. When the physiological mechanisms of color vision are of interest, a cone chromaticity space that can represent cone stimulations, as well as the postreceptoral pathways, is preferred.

The concept of cone chromaticity space appeared in the early 20th century. It was not until 1979 that MacLeod and Boynton published a cone chromaticity space based on modern estimates of the cone spectral sensitivities. In the MacLeod and Boynton cone chromaticity space (Fig. 10.5B), the horizontal axis \[L/(L+M)\] represents the variation of relative L- versus M-cone stimulation at equiluminance, while the vertical axis \[S/(L+M)\] represents the variation of S-cone stimulation. The space normalizes in that \(S/(L+M) = 1\) for spectral light of 400 nm. Later, a relative cone troland space that normalizes \(S/(L+M)\) for EES light to be 1 was proposed to link cone excitations with retinal illuminance, which is measured in trolands. Another spectral opponency space normalizes the cone chromaticity based at the EES-white to reflect both cone contrasts and postreceptoral opponency signals. These cone chromaticity spaces are a major breakthrough for vision research because neurons in the PC and KC pathways show preferred responses to stimuli along the two axes of the cone chromaticity spaces. Therefore, psychophysical experiments can be designed to infer the functions of the postreceptoral pathways by generating stimuli along the two theoretical axes.

**Chromatic discrimination**

Chromatic discrimination refers to the ability of an observer to discriminate two colors. Chromatic discrimination has been investigated using three approaches: wavelength discrimination, purity discrimination, and chromaticity discrimination (reviewed by Pokorny and Smith). Chromatic discrimination is usually measured at a constant luminance level to avoid potential interactions between the luminance pathway (MC pathway) and the chromatic pathway (PC or KC pathway).

**Wavelength discrimination**

In a typical wavelength discrimination experiment, the stimulus consists of two equiluminant semicircular fields, one filled with a narrow band of spectral light to serve as the standard field and the other as the comparison field. The observer is instructed to change the wavelength of the comparison field to achieve a just noticeable difference (JND) from the standard wavelength. Typical wavelength discrimination thresholds, as a function of standard wavelengths, form a skewed “W” shape with two minima, one at 490 nm and the other at 580 nm.

**Purity discrimination**

There are two ways to measure purity discrimination. The first method measures the minimum amount of spectral light the observer adds into a white field to achieve a JND from the same white in another juxtaposed field. Discrimination thresholds measured in this way are the largest at 570 nm and the smallest at 400 nm. The second method measures the minimum amount of white light added into a spectral light to achieve a JND from...
the spectral light. Purity discrimination thresholds measured in this way do not vary much with variations in the wavelength of the spectral light.

**Chromaticity discrimination**

Early attempts to measure chromaticity discrimination included measurement of the minimum variation needed in chromaticity to achieve a JND from any point in the CIE diagram. Another measure of chromatic discrimination involved derived discrimination ellipses using the standard deviations of repeated color matches at a set of chromaticities in the CIE diagram.

In modern chromaticity discrimination experiments, discrimination thresholds are obtained for test field chromaticities varied along the two cardinal axes of a cone chromaticity space in a steady adaptation field. A cone chromaticity space allows that discrimination can be mediated by the L/M cones only (L/M cone discrimination) or by the S cones only (S-cone discrimination). Chromatic discrimination was found to be the best at the adaptation chromaticity and then deteriorated with increasing chromatic contrast between the test and adapting fields. Chromatic discrimination data can be explained adequately by a model based on primate ganglion cell responses in the PC and KC pathways.

**Color appearance**

In color-matching and color discrimination experiments, observers determine whether two colors appear the same or different but they do not determine which color is perceived. Color appearance has three perceptual dimensions: hue, saturation, and brightness. Hue is the perceptual dimension that differs from white, such as red, orange, green, blue, yellow, purple, and pink. Saturation indicates how different the hue is from white. For instance, colors on the spectral locus are highly saturated compared with desaturated colors near white (Fig. 10.5A). Brightness is the perceptual dimension related to luminance variance.

An important feature of color appearance is that colors do not simultaneously appear red and green, nor do they simultaneously appear blue and yellow. However, colors do appear as mixtures of red and yellow or red and blue and they also appear as mixtures of green and yellow or green and blue. Further, human observers can separately abstract the qualities of redness–greenness or blueness–yellowness, respectively. (Reproduced with permission from Hurvich LM, Jameson D. Some quantitative aspects of an opponent-colors theory. II. Brightness, saturation and hue in normal and dichromatic vision. J Opt Soc Am 1955;45:602–16.)

**VARIATIONS IN HUMAN COLOR VISION**

Abnormal color vision that is either inherited or acquired is present in about 4.5% of the population. Congenital color vision defects are stationary over the lifespan and do not result from other visual problems. These color vision defects have been studied extensively and their classification is well established based on psychophysical and genetic works. The most common are the congenital X-chromosome-linked red–green color vision defects, which have been associated with alterations in the gene sequences encoding the opsins on the X chromosome.

Acquired color vision defects refer to abnormalities that
accompany eye diseases or drug toxicity. Acquired color vision defects are more variable and their classification is more difficult and less satisfactory. Color vision is often tested clinically with screening tests that allow identification of abnormalities and most screening tests are based on color discrimination and color-matching abilities.29

Color vision classifications

Color vision classifications are based on both the number of functioning cone types and the presence of abnormal cones. An observer with three functioning cone types is called a trichromat, an observer with two functioning cone types is a dichromat, and an observer with one functioning cone type is a monochromat (monochromacy sometimes is also termed achromatopsia in the literature since it is believed that vision based on a single cone type cannot produce color perception, assuming rods are not involved). An observer with normal color vision has three normal cone types, in terms of spectral sensitivity, and is called a normal trichromat. Observers are said to have defective color vision have at least one abnormal cone type or are missing at least a cone type with conventional color-matching techniques. Observers with rods only (lack of any cones) are called rod monochromat or complete achromatopsia.30

X-linked color vision defects have been recognized since the 18th century and were subdivided into two qualitatively different types: protan (“red-blind”) and deutan (“green-blind”). The term “protanope” is used for a dichromat who is thought to be missing L cones and the term “deuteranope” is used for a dichromat who is thought to be missing M cones, based on color-matching characteristics using a 2° visual field in the fovea. Anomalous trichromacy is a variation in color vision that is attributed to the presence of a cone type that is shifted in spectral sensitivity. A protanomalous observer has trichromatic color vision but the L cones have spectral sensitivity that is shifted to shorter wavelengths compared to normal L cones. A deuteranomalous observer also has trichromatic color vision but the M cones have spectral sensitivity that is shifted to longer wavelengths compared to normal M cones. A third qualitatively different type of color vision is tritan (“blue-blind”). Tritan color vision defects are thought to arise from variations in S cones.

The genes encoding the human photopigments

It has been known for many years that the spectral sensitivities of rods and cones reflect the absorption spectra of the visual photopigments. In a major advance, the genes encoding the opsin genes of the human visual photopigments were cloned and mapped in the human genome in 1983.27,31 The gene for rhodopsin was found on chromosome 3 and the human rhodopsin gene showed high homology (93.4%) to that of bovine rhodopsin.32 A visual photopigment gene for the opsin of S cones (OPN1SW gene) was found on chromosome 7. A tandem array of visual photopigment genes for the opsins of the M and L cones (OPN1MW and OPN1LW genes) was found on the X chromosome. The human opsin genes show about 45% homology between rhodopsin and any of the three cone photopigment genes and between the chromosome 7 and the X-chromosome pigment genes. This similarity among the photopigment genes suggests a common ancestor. The genes on the X chromosome have a high homology to each other (about 96%), suggesting a more recent evolutionary appearance. An unexpected finding was that of multiple genes in a tandem array on the X chromosome, which has been proposed to range from 2 to 6. Nathans et al.28 postulated that the multiple genes in the tandem array on the X chromosome arose as a result of unequal homologous recombination. Subsequent study has suggested there are polymorphisms among these genes in color-normal and color-defective individuals.27

The initial study of the opsin genes on the X chromosome was based on the long-standing conventional ideas about X-linked congenital color vision defects29; that is, protanopes are missing L cones and deuteranopes are missing M cones, and that the genetics of these types of defective color vision may be based on missing genes for either L or M cones. Specifically, protanopes were thought to be dichromats who lacked an L-cone gene and deuteranopes were dichromats who lacked an M-cone gene.

The initial work with the X-chromosome opsin genes, in attempting to link color vision defects with these genes, was carried out using restriction fragment length polymorphisms (RFLPs), an early molecular genetic technique that simplified the study of DNA sequences. The initial RFLP analysis of the X-chromosome genes was done on DNA from one observer with normal color vision and a few protanopes and deuteranopes. Comparisons were carried out to determine which of the fragments found in the normal observer’s RFLPs were missing in the protanopes and deuteranopes. A fragment found in the normal observer that was missing in the protanope was said to be part of the assumed missing L-cone gene. Similarly, a fragment found in the normal observer that was missing in the deuteranope was said to be part of the assumed missing M-cone gene. In the initial study, observers with dichromatic as well as anomalous trichromatic color vision defects showed the presence of hybrid genes (genes comprising the head of one type of gene and the tail of the other type of gene). The hybrid genes were said to be the basis for the altered spectral sensitivity of anomalous L or M cones associated with X-linked anomalous trichromacy (discussed previously).

Subsequent studies that have been carried out on both normal and X-linked defective color vision have their roots in the initial RFLP analysis based on the conventional idea of missing cone types and missing genes using one normal observer and a few color-defectives. The variations that have been found in subsequent studies on the visual pigment genes have correlated largely, but not absolutely, with phenotypes established by color vision testing. The current view is that the X-chromosome tandem array consists of one or more opsin genes encoding L-cone photopigment, followed by one or more opsin genes encoding the M-cone photopigment. The expression of the genes in the tandem array is believed to be governed by a stochastic process.27

There have been advances in molecular genetic technology and in the understanding of the human genome since the initial RFLP studies of the opsin genes. Knowledge derived from the Human Genome Project, which has spurred an understanding of the scarcity of genes in the human genome, and the developing knowledge of epigenetics, such as the editing processes of microRNAs, may contribute to future studies and understanding of the genetics of color vision variations.
CLINICAL EVALUATION OF COLOR VISION

Screening tests

Screening tests are rapid tests (requiring 2–3 minutes) and color-defective observers are identified due to their inability to see the difference between certain colors that are easily discriminated by normal observers. Screening tests can be administered to both children and adults.

Pseudoisochromatic plate tests

The most commonly used screening tests are the pseudoisochromatic plate tests. First introduced by Stillings, a pseudoisochromatic plate presents a figure composed of colored dots in a background of differently colored dots. Usually, the colors are chosen so that an X-linked color-defective observer does not see the figure that is easily seen by normal observers. The clearest designs use four sets of colors, chosen so that the normal observer sees one figure and the defective observer sees a different figure.

The majority of pseudoisochromatic plate tests (such as the Ishihara) were designed to identify observers with X-linked congenital color defects (i.e., protan or deutan color anomalies). The choice of colors was optimized to take advantage of the particular discrimination losses found in X-linked color vision defects and the tests are successful in detecting 90–95% of color-defective observers. Pseudoisochromatic plate tests cannot be used to identify acquired color vision defects, which are more likely to affect blue/yellow color vision. More recently developed pseudoisochromatic plates (such as the HRR and SPP2 plates) have been designed specifically for acquired color vision deficiencies, including tritan (blue/yellow) anomalies.

Other rapid tests of color vision

Other rapid tests of color vision involve sorting colored pieces. In principle, this approach can be used successfully with acquired color vision abnormalities since it does not involve the choice of a particular pair of colors that has a prediction based on common X-linked color vision defects.

Chromatic discrimination ability tests

Clinical assessment of color discrimination ability involves arrangement tests that require the observer to arrange a set of colored samples according to their similarity in color. If the samples are closely spaced in chromaticity (e.g., the Farnsworth–Munsell 100-hue test: Fig. 10.7), the task becomes one of fine chromatic discrimination. Tests involving fine chromatic discrimination are usually relatively time-consuming. If the samples are widely spaced in chromaticity (e.g., the Farnsworth panel D-15), the test evaluates color confusions that occur with defective color vision that would not be perceived by a normal observer. Tests with widely spaced colors are conducted rapidly and can even be used for screening. An arrangement test may use samples that differ only in chromaticity to test hue discrimination (e.g., the Farnsworth–Munsell 100-hue test, Farnsworth panel D-15, Farnsworth desaturated panel D-15, and Lanthony new color test), may vary in luminance only to test lightness discrimination (Verriest’s lightness discrimination test), or may vary only in grayness to test saturation discrimination (Sahlgren’s saturation test, Lanthony new color test). Arrangement tests are easy to administer but require the concept of abstract ordering, manual dexterity, and patience. As a result, they are rarely suitable for children under 10 years of age.

The best known of the arrangement tests is the Farnsworth–Munsell 100-hue test. The test includes 85 black plastic caps with inserted papers that vary in hue but have constant lightness and saturation. The caps are divided into four boxes with each box covering one-quarter of the color circle. Observers being tested arrange the caps in a natural color order according to their unique perception. An error occurs if the caps are misplaced from the ideal color order. A numeric score can be calculated and displayed on a polar graph. Age norms have been prepared giving the range of the expected total error scores for an unselected population as well as the expected intereye variability.

Observers with congenital color vision deficiencies make characteristic errors on arrangement tests because their chromatic discrimination ability is weakened or lost on particular axes in chromaticity space. Discrimination loss in acquired color vision defects is more variable. However, following the idea that discrimination of blueness content is, to a first order, independent of discrimination of redness–greenness, it is possible to partition the caps into those where correct ordering depends on normal function of the S- and M-cone opponent system (caps 1 through 12, 24 through 34, and 76 through 84) and those where correct ordering depends on normal function of the M- and L-cone opponent system (caps 13 through 33 and 55 through 75). The partitioned scores can be examined to determine whether an acquired color vision defect causes a particular type of discrimination loss; i.e., the S or L/M systems.

Importance of the test illuminant for plate and discrimination color vision tests

The plate and discrimination tests described above use reflective materials as colored test objects and the perceived color presented to an observer depends on the illuminating light as well as the reflective properties of the test materials. The original pseudoisochromatic plate tests were designed to be viewed under afternoon daylight in the northern hemisphere, and more recent tests have followed this design convention. Standardized illuminants (called illuminant C or illuminant D65) that closely simulate the spectra of afternoon daylight are more preferable to natural daylight, which may vary substantially in both spectrum and radiance with time of day and weather. Light sources that approximate these standard illuminants are commercially available and are suitable for use in illuminating clinical color vision tests. Most fluorescent light sources, however, do not accurately mimic the colors produced in natural light and, therefore, they are not appropriate illuminants for color vision tests.

Color-matching tests

Adjusting three primaries to match a test color is not intuitive for many observers and research experimental methods are therefore not appropriate for clinical purposes. To adapt color-matching procedures for clinical evaluation, simplified methods have been developed by using two-variable matches and rapid, less complicated tasks. The instruments that allow these matches to be carried out are called anomaloscopes and the color-matching paradigms are called equations. The equations are named after the researchers who first proposed or used them.
Anomaloscope color matching test using the Rayleigh equation

In a Rayleigh match, a spectral “yellow” test field (589 nm) is presented in one-half of a circular field with an adjustable brightness. In the half field, a mixture of two “green” and “red” spectral primaries (545 and 670 nm) is presented. The mixture field can appear green (545 nm), green–yellow, yellow, orange, or red (670 nm) as the ratio of the primaries is varied. The task of the observer is to adjust the primary red–green ratio and brightness of the yellow to make the two fields appear identical. Observers with normal color vision accept a narrow range of ratios near the middle of the red–green range, but color-deficient observers will pick ratios shifted away from this range, depending on the specific deficiency. Matching abnormalities accompanying ophthalmic disease or X-linked color vision variations may include a widened matching range or acceptance of an abnormal match.33

The Rayleigh equation assesses the normality of M- and L-cone functions, since S cones are not contributing to the match because they are unresponsive to the primary wavelengths.
A graphic display of the cone photopigment excitations allows the tester to evaluate which photopigments participated in an observer’s match. Figure 10.8A shows the expected yellow setting, based on the L- and M-photopigments, as a function of the red–green primary ratios. For each photopigment, the quantal catch in cones may be characterized by a straight line in a chromaticity diagram as the red–green primary ratio is changed. Two lines show the predicted responses of the L cones alone (dashed and labeled L) and the M cones alone (dotted and labeled M). The rod photopigment is responsive at these wavelengths as well and the predicted rod response (solid and labeled Rod) is also shown. The normal match occurs at or near the intersection of the L- and M-cone settings and the matching range is narrow.

Deuteranopes can make satisfactory matches throughout the L line. A deuteranope is thought to lack functional M cones in the retinal area responding to the stimuli. Studies on the X-chromosome opsin genes in deuteranopes show that many of these observers have only a single gene that is thought to encode the L photopigment. Protanopes can make matches throughout the extent of the M line. It has been speculated that the protanope lacks functional L cones in the retinal area responding to the
stimuli and genetic studies have suggested that these observers usually have only a single gene on the X chromosome that has been described as a hybrid gene composed of a piece of the normal L- and a piece of the normal M-opsin genes.

Other X-linked color-defective observers, called anomalous trichromats, have a wider range of acceptable matches that are displaced from the normal match. Such observers are trichromatic but have one cone photopigment with an abnormal spectral sensitivity. Deuteranomalous matches occur along the L line (indicating a normal L-cone photopigment) but are shifted to low red–green primary ratios. The psychophysical interpretation is that the deuteranomalous trichromat has an abnormal M-cone photopigment that is shifted so that its spectral sensitivity closely overlaps that of the normal L-cone photopigment. Protanomalous matches occur along the M line (indicating a normal M-cone photopigment) but are shifted to high red–green primary ratios. The interpretation of these psychophysical studies is that the protanomalous trichromat has an abnormal L-cone photopigment, shifted so that its spectral sensitivity closely overlaps that of the normal M-cone photopigment. Anomaloscope color matching using the Rayleigh equation is recognized as the only clinical method that allows definitive classification of the X-linked color vision defects.

Anomaloscope analysis is important in ophthalmic disease assessment since it can be used to recognize a number of clinical entities, such as rod-dominated vision characteristic of cone degenerations and incomplete and complete achromatopsias. These diseases are characterized by Rayleigh matches extending along the rod line (solid line in Fig. 10.7A). Matching abnormalities that are characteristic of choroidal disorders affecting the fovea lead to Rayleigh matches that extend along or just below the L line but are shifted to higher red primary ratios (pseudo-protanomaly). The discrimination loss characteristic of optic nerve disorders leads to matches that widen around the normal match on the L line.

**Anomaloscope color-matching test using the Moreland equation**

The Moreland equation is a match of a bicolor test field (480 and 580 nm) to a mixture of two primaries (440 and 500 nm). The test field appears blue–green to a normal observer, while appearance of the mixture field appearance ranges from violet, to blue, to blue–green, to green–blue, to green as the primary ratio is changed from mostly 440 nm to mostly 500 nm. The primary ratio is expressed in terms of the amount of the 440 nm (“violet”) primary required for the match. As with the Rayleigh equation, an observer with normal color vision can find a primary ratio and test field luminance at which both fields appear identical.

The Moreland equation assesses the normality of S-cone function. Figure 10.8B shows the predicted function for a single photopigment to the bicolor test field as a function of the violet–green primary ratio. The L and M cones share the same line (dotted and labeled as L/M). The S-cone response crosses this on the diagonal (dashed and labeled as S). The predicted rod (solid line labeled as Rod) line has the same direction as the L- and M-cone lines. The normal match occurs at the intersection of the S-cone line with the L/M line. Both congenital and acquired color vision defects can be recognized using the Moreland equation. The tritanope (an observer lacking S-cone function) has matches extending along the L/M line, as do observers...
with acquired defects that affect S-cone function. Point mutations in the S-cone opsin gene on chromosome 7 have been reported in affected individuals in tritan pedigrees.36

Patients with X-linked achromatopsia (i.e., S-cone monochromacy) have matches that extend along the S-cone line. The majority of X-linked achromatopedia shows major deletions of the region just preceding the cone opsin tandem gene array on the X chromosome in affected individuals.37 This region is thought to be a regulatory area controlling expression of the L- and M-cone opsin genes. Additionally, some X-linked achromatopedia shows only a single abnormal gene instead of the normal tandem array. Patients with complete achromatopsia have matches extending along the rod line.

Considerations in the use of anomaloscopes
Even though the color-matching task on an anomaloscope is simplified compared with usual research color-matching procedures, anomaloscope testing procedures are not easily explained and an observer may require some practice before being able to complete a match. Correct use of an anomaloscope requires extensive operator training and these instruments are therefore usually found only in research centers. However, if properly used, the anomaloscope is a diagnostic instrument of great power.

Computerized color vision tests
Computer-controlled color vision tests have been developed using a similar principle as that used for pseudoisochromatic plate tests or discrimination ability tests (such as the FM-100 hue test). Computerized tests are automated and therefore easy to use. Further, computerized tests can avoid the illuminating light issue that exists for reflective materials. However, the sensitivity and specificity of the computerized tests largely rely on the accurate presentation of color on computer monitors, including cathode ray tubes (CRTs) or liquid crystal displays (LCDs). A CRT is typically preferred over a LCD because CRTs have reliable temporal characteristics. To present color accurately, the displays have to be calibrated, including spectral distribution measurements and linearization. Caution must be taken when using a computerized color vision test that does not incorporate specific display calibration.

Color assessment and diagnosis (CAD) test
The CAD test was developed by Dr Barbur and coworkers at City University, London.38 The web-based CAD uses a color square that moves against a flickering luminance contrast noise. The color of the square changes along different chromaticity directions. Color-defective observers have difficulty seeing the square moving, whereas normal observers easily see the square move. The web-based CAD test requires a monitor to be balanced at around 9000 K and the ambient illumination must be kept at a minimum. It is reported that this web-based test has good sensitivity and specificity for red-green color deficiencies.39

Cambridge color test (CCT)
The CCT is a computer-controlled, easy-to-use color vision test. The CCT was developed by Drs Mollon, Reffin, and Regan at the University of Cambridge, England.40 The CCT system provides 14-bit color and luminance control on a calibrated CRT display. The stimulus is a Landolt C on an achromatic background. The chromaticity of the target C is varied along the protan, deutan, and tritan lines of a chromaticity diagram. The task is to indicate the position of the gap in the target C. A staircase procedure is used to measure discrimination thresholds along any of the three lines. The results are plotted as discrimination ellipses in a CIE space. Individuals with color vision deficiencies will have elongated discrimination ellipses along a protan, deutan, or tritan line and, therefore, this test can be used for all color vision deficiencies.

The portal color sort test (PCST)
The PCST is based on the FM-100 hue test but uses only 36 colored “chips,” which significantly reduces the testing time. The chips are representative of the original 85 chips in the FM-100 hue test. The observer arranges the order of the chips according to color similarities and the computer provides automatic scores. The correlation between the PCST and FM-100 hue test is high for testing congenital color vision defects but is unknown for testing acquired defects.41

Smartphone/tablet applications for color vision screening
With the popularity of mobile communication devices (e.g., smartphones or tablets) increasing in medical settings, numerous applications have developed for these devices as tools for clinical testing/screening, patient education, or physician education and reference. Several applications use the pseudoisochromatic plate principle for color vision screening. These applications might be useful for a quick screening of color vision; however, these tools have not been validated and their sensitivity and specificity for color vision defect screening are not known. Further, there are many factors that can affect the test results. The most important is the display characteristics of the mobile device. These displays are not calibrated and they may not be homogeneous; therefore the tests for color vision screening are not necessarily “pseudoisochromatic.” Finally, the illumination in the office may impair the test reliability. Therefore, results from these applications need to be confirmed by comparison with other established tools for color vision screening.

Which test to use in a clinical setting?
Many clinicians want to have some means of testing color vision without necessarily acquiring expensive instruments and expertise needed for professional evaluation and diagnosis. Tests using colored papers (pigment tests) and the proper illuminant offer office clinicians the possibility of some color vision evaluation.

A screening plate test with the proper illuminant would be minimal equipment for testing color vision. Approximately 8–10% of American males have one of the X-linked color vision defects. Identification of color vision defects in children before they enter grade school allows the clinician to provide counseling to the parents. Many children with color vision defects have memories of being teased or ridiculed in the early grades and early testing allows appropriate counseling. Also, many males with color vision defects may inadvertently choose careers, for example, as pilots or firefighters, from which they will be barred due to their color vision status. It should be noted, however, that the screening plate tests that may be very useful for the common genetic color vision defects are less successful at identifying acquired color vision defects.

A more ambitious plan to test color vision in an office would be to combine a screening plate test with a test of color
discrimination. A discrimination test (e.g., the Farnsworth-Munsell 100-hue test) allows the clinician to follow changes in color vision over time, such as might occur with optic neuritis. Occasionally, a clinician will see a patient who has an extremely rare form of color defect and color vision testing may be informative. For example, cerebral achromatopsia is a fascinating case that likely arises from damage to higher-order visual processes. In these cases, a more complete color vision testing (such as spectral sensitivity and saturation discrimination evaluation) should be considered and case reports would be of wide interest in the medical community. Patients should be referred to a psychophysics laboratory for more visual function testing, such as contrast discrimination evaluation.

NEW DEVELOPMENTS IN COLOR VISION RESEARCH

Color vision has been studied considering genetics, evolution, physiology, and psychophysics. A few newer research areas related to color vision are provided here.

Gene therapy for color vision defects

Drs Jay and Maureen Neitz and coworkers have carried out pioneering studies on “curing” red–green color deficiency in dichromatic adult squirrel monkeys that were missing the L-cone opsin gene at birth. As gene therapy, the human L-cone opsin gene was delivered into the photoreceptor layers of the retinas of the monkeys. A few months after the introduction of the new opsin gene, these monkeys exhibited trichromatic color vision behavior with spectral sensitivity shifted, chromatic discrimination improved, and color perception enriched. An experiment involving gene therapy in humans would require approval from the National Institutes of Health Office of Recombinant DNA Activities (ORDA)/Recombinant DNA Advisory Committee (RAC) and the Food and Drug Administration and this would not be expected to be granted without a long and thorough approval process. However, the fundamental question of whether it is necessary to “cure” color deficiencies is currently the point of debate since the majority of dichromats live a normal life and most are not significantly affected by having the color vision deficiency.

Adaptive optics (AO) retinal imaging system

AO was initially used in astronomy to remove the effects of atmospheric distortion to improve the performance of telescopes and laser communication systems. An AO system has three components: (1) a wavefront sensor for ocular aberration measurement; (2) a deformable mirror for aberration correction; and (3) a control system that compares the sensor output and adjusts the deformable mirrors to achieve optimal resolution. This technology was first adopted for retinal imaging in the 1990s by Dr David Williams to reduce ocular aberrations in the eyes. After its initial introduction, the AO imaging system was considered to have great scientific and clinical application potential because it had the capability of imaging photoreceptors, the retinal pigment epithelium, retinal blood vessels and, potentially, ganglion cells at a high magnification. For instance, AO made it possible to measure color perception or cell responses in the postreceptoral pathways associated with tiny flashes of light stimulating a single cone in the eye. For color vision screening, in particular, the AO system can provide high resolution of the cone mosaic and it can show whether a particular type of cone (e.g., L cones) is missing in the retina. Combining this information with genetic studies is potentially informative because AO imaging can provide insights about cone distributions that can be related to the cone opsin genes. (See Chapter 5, Advanced imaging technologies, for more details about this topic.)

Rod and cone interactions in color vision

Duplex theory states that rods and cones independently contribute to different aspect of visual perception. However, rods and cones share common neural pathways from the retina to the brain and this provides a neural basis for rod–cone interactions in visual function, including color vision. Conventionally, rod vision has been considered to be achromatic. However, numerous psychophysical studies have indicated that rods contribute to color vision at either mesopic or even scotopic light levels. Psychophysical evidence for rod contributions to color vision comes from measurements of scotopic color contrast, photichromatic intervals during the course of dark adaptation following a light bleach, chromatic discrimination, and color-matching or color appearance methods using unique hue measurement or hue-scaling methods.

Recently, rod contributions to color vision were studied using a four-primary Maxwellian-view photostimulator that allowed independent control of rod and cone excitations at the same chromaticity, retinal locus, and light level. This new method has yielded new insights into rod contributions to color vision. Specifically, rods contribute to color percepts in a manner analogous to M-cone signals at all mesopic light levels and analogous to S-cone signals only at low mesopic light levels near cone thresholds. Also, the strength of rod contributions is linearly related to rod contrasts.

For online acknowledgments visit http://www.expertconsult.com

REFERENCES

ACKNOWLEDGMENTS

The preparation of this chapter is partially supported by a National Institutes of Health grant (R01 EY019651). I express my great appreciation to Drs Vivianne Smith and Joel Pokorny for allowing me access to the materials they prepared for an earlier version of the chapter in a previous edition of Retina. I thank Dr Margaret Lutze for her assistance in preparing this chapter.
52. Willmer EN. Low threshold rods and the perception of blue. J Physiol (Lond) 1949;111:17P.