1
Human Color Vision

Color appearance models aim to extend basic colorimetry to specify the perceived color of stimuli in a wide variety of viewing conditions. To fully appreciate the formulation, implementation, and application of color appearance models, several fundamental topics in color science must first be understood. These are the topics of the first few chapters of this book. Since color appearance represents several of the dimensions of our visual experience, any system designed to predict correlates to these experiences must be based, to some degree, on the form and function of the human visual system. All of the color appearance models described in this book are derived with human visual function in mind, although most also include some empirical modeling of the visual system as a "black box." It becomes much simpler to understand the formulations of the various models if basic visual anatomy, physiology, and performance of the visual system are understood. Thus, this book begins with a treatment of the human visual system.

As necessitated by the limited scope available in a single chapter, this treatment of the visual system is an overview of the topics most important for an appreciation of color appearance modeling. The field of vision science is immense, complex, and fascinating. Readers are encouraged to explore the literature and the many useful texts with differing perspectives on human vision in order to gain further insight and details. Of particular note are the review paper on the mechanisms of color vision by Lennie and D'Zmura (1988), the text on human color vision by Kaiser and Boynton (1996), the more general text on the foundations of vision by Wandell (1995), the comprehensive treatment by Palmer (1999), and edited collections on color vision by Backhaus *et al.* (1998) and Gegenfurtner and Sharpe (1999). Other interesting and more recent texts on vision include the extensive and complete volume by Chalupa and Werner (2004), the revision of Dowling's (2012) classic on the retina, Livingstone's (2002) interesting treatment of the relationships between art and biology of seeing, Mausfeld and Heyer's (2003) book focused on perception,

Color Appearance Models, Third Edition. Mark D. Fairchild.

Schwab's (2012) discussion of the evolution of vision, and Valberg's (2005) revised edition covering all of vision, but with some more focus on color. General texts on sensation and perception, such as Wolfe *et al.* (2012), are also excellent sources for learning fundamental aspects of the human visual system. Johnsen (2012) provides a slightly different perspective on visual systems and other optical phenomena in nature. The material that is briefly summarized in this chapter is treated in more detail in those references.

1.1 OPTICS OF THE EYE

Our visual perceptions are initiated and strongly influenced by the anatomical structure of the eye. Figure 1.1 shows a schematic representation of the optical structure of the human eye with some key features labeled. The human eye can be thought of as acting like a camera. The cornea and lens act together like

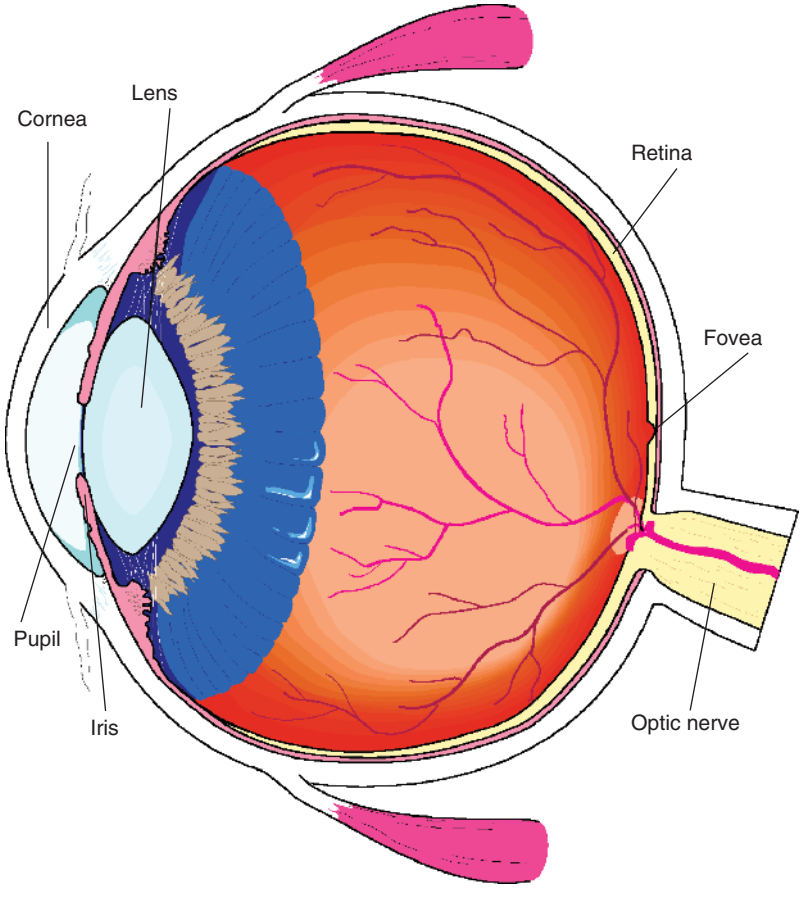


Figure 1.1 Schematic diagram of the human eye with some key structures labeled

a camera lens to focus an image of the visual world on the retina at the back of the eye, which acts like the image sensor (*e.g.*, CCD) of a camera. These and other structures have a significant impact on our perception of color.

The Cornea

The *cornea* is the transparent outer surface of the front of the eye through which light passes. It serves as the most significant image-forming element of the eye since its curved surface at the interface with air represents the largest change in index of refraction within the eye's optical system. The cornea is avascular, receiving its nutrients from marginal blood vessels and the fluids surrounding it. Refractive errors, such as nearsightedness (myopia), farsightedness (hyperopia), or astigmatism, can be attributed to variations in the shape of the cornea with respect to the location and the shape of the retina. These refractive errors are sometimes corrected with laser surgery to reshape the cornea.

The Lens

The *lens* serves the function of accommodation. It is a layered, flexible structure that varies in index of refraction. It is a naturally occurring gradient-index optical element with the index of refraction higher in the center of the lens than at the edges. This feature serves to reduce some of the aberrations that might normally be present in a simple optical system.

The shape of the lens is controlled by the ciliary muscles. When we gaze at a nearby object, the lens becomes "fatter" and thus has increased optical power to allow us to focus on the near object. When we gaze at a distant object, the lens becomes "flatter" resulting in the decreased optical power required to bring more distant objects into sharp focus. As we age, the internal structure of the lens changes, resulting in a loss of flexibility. Generally, when the age of about 45–50 years is reached, the lens has completely lost its flexibility and observers can no longer focus on near objects (this is called presbyopia, or "old eye"). It is at this point that most people must resort to reading glasses or bifocals.

Concurrent with the hardening of the lens is an increase in its optical density. The lens absorbs and scatters short-wavelength (blue and violet) energy. As it hardens, the level of this absorption and scattering increases. In other words, the lens becomes more and more yellow with age. Various mechanisms of chromatic adaptation generally make us unaware of these gradual changes. However, we are all looking at the world through a yellow filter that not only changes with age, but is significantly different from observer to observer. The effects are most noticeable when performing critical color matching or comparing metameric color matches with other observers. The effect is particularly apparent with purple objects and nearly

monochromatic stimuli such as the primaries of wide-gamut displays. Since an older lens absorbs most of the blue energy reflected from a purple object but does not affect the reflected red energy, older observers will tend to report that the object is significantly more red than reported by younger observers. Important issues regarding the characteristics of lens aging and its influence on visual performance are discussed by Pokorny *et al.* (1987), Werner and Schefrin (1993), and Schefrin and Werner (1993) and in the Commission Internationale de l'Éclairage (CIE) (2006) report on physiological color matching functions.

The Humors

The volume between the cornea and the lens is filled with *aqueous humor*, which is essentially water. The region between the lens and the retina is filled with *vitreous humor*, which is also a fluid, but with a higher viscosity similar to that of gelatin. Both humors exist in a state of slightly elevated pressure (relative to air pressure) to assure that the flexible eyeball retains its shape and dimensions in order to avoid the deleterious effects of wavering retinal images. The flexibility of the entire eyeball serves to increase its resistance to injury. It is much more difficult to break a structure that gives way under impact than one of equal "strength" that attempts to remain rigid. Since the indices of refraction of the humors are roughly equal to that of water, and those of the cornea and lens are only slightly higher, the rear surface of the cornea and the entire lens have relatively little optical power (in comparison with the front surface of the cornea).

The Iris

The *iris* is the sphincter muscle that controls pupil size. The iris is pigmented, giving each of us our specific eye color. Eye color is determined by the concentration and distribution of melanin within the iris. The pupil, which is the hole in the middle of the iris through which light passes, defines the level of illumination on the retina. Pupil size is largely determined by the overall level of illumination, but it is important to note that it can also vary with nonvisual phenomena such as arousal. (This effect can be observed by enticingly shaking a toy in front of a cat and paying attention to its pupils.) Thus it is difficult to accurately predict pupil size from the prevailing illumination. In practical situations, pupil diameter varies from about 3 to 7 mm. This change in pupil diameter results in approximately a five-fold change in pupil area and therefore retinal illuminance. The visual sensitivity change with pupil area is further limited by the fact that marginal rays are less effective at stimulating visual response in the cones than central rays (the Stiles–Crawford effect). The change in pupil diameter alone is not sufficient

to explain excellent human visual function over prevailing illuminance levels that can vary over 10 orders of magnitude or more.

The Retina

The optical image formed by the eye is projected onto the retina. The *retina* is a thin layer of cells, approximately the thickness of tissue paper, located at the back of the eye and incorporating the visual system's photosensitive cells and initial signal processing and transmission "circuitry." These cells are neurons, part of the central nervous system, and can appropriately be considered a part of the brain. The photoreceptors, rods and cones, serve to transduce the information present in the optical image into chemical and electrical signals that can be transmitted to the later stages of the visual system. These signals are then processed by a network of cells and transmitted to the brain through the optic nerve. More detail on the retina is presented in "The retina."

Behind the retina is a layer known as the *pigmented epithelium*. This dark pigment layer serves to absorb any light that happens to pass through the retina without being absorbed by the photoreceptors. The function of the pigmented epithelium is to prevent light from being scattered back through the retina, thus reducing the sharpness and contrast of the perceived image. Nocturnal animals give up this improved image quality in exchange for a highly reflective tapetum that reflects the light back in order to provide a second chance for the photoreceptors to absorb the energy. This is why the eyes of a deer, or other nocturnal animal, caught in the headlights of an oncoming automobile appear to glow. They are acting like very efficient retro-reflectors by focusing the light from the car they are looking at through the animal's eyes and right back to the car itself.

The Fovea

Perhaps the most important structural area on the retina is the fovea. The *fovea* is the area on the retina where we have the best spatial and color vision. When we look at, or fixate, an object in our visual field, we move our head and eyes such that the image of the object falls on the fovea. As you are reading this text, you are moving your eyes to make the various words fall on your fovea as you read them. To illustrate how drastically spatial acuity falls off as the stimulus moves away from the fovea, try to read the preceding text in this paragraph while fixating on the period at the end of this sentence. It is probably difficult, if not impossible, to read the text that is only a few lines away from the point of fixation. The fovea covers an area that subtends about 2° of visual angle in the central field of vision. To visualize 2° of visual angle, a general rule is that the width of your thumbnail, held at arm's length, is approximately 1° of visual angle. (Also, the moon and

sun each subtend almost exactly 0.5° of visual angle in the sky, an interesting coincidence that enhances the possibility of the Earth having both complete lunar and solar eclipses.)

The Macula

The fovea is also protected by a yellow filter known as the macula. The *macula* serves to protect this critical area of the retina from intense exposures to short-wavelength energy. It might also serve to reduce the effects of chromatic aberration that cause the short-wavelength image to be rather severely out of focus most of the time. Unlike the lens, the macula does not become more yellow with age. However, there are significant differences in the optical density of the macular pigment from observer to observer and in some cases between a single observer's left and right eyes. The yellow filters of the lens and macula, through which we all view the world, are the major source of variability in color vision between observers with normal color vision.

The Optic Nerve

A last key structure of the eye is the optic nerve. The optic nerve is made up of the axons (outputs) of the ganglion cells, the last level of neural processing in the retina. It is interesting to note that the optic nerve is made up of approximately one million fibers carrying information generated by approximately 130 million photoreceptors. Thus there is a clear compression of the visual signal prior to transmission to higher levels of the visual system. A one-to-one "pixel map" of the visual stimulus is never available for processing by the brain's higher visual mechanisms. This processing is explored in greater detail below. Since the optic nerve takes up all of the space that would normally be populated by photoreceptors, there is a small area in each eye in which no visual stimulation can occur. This area is known as the *blind spot*.

The structures described above have a clear impact in shaping and defining the information available to the visual system that ultimately results in the perception of color appearance. The action of the pupil serves to define retinal illuminance levels that, in turn, have a dramatic impact on color appearance. The yellow-filtering effects of the lens and macula modulate the spectral responsivity of our visual system and introduce significant inter-observer variability. The spatial structure of the retina serves to help define the extent and nature of various visual fields that are critical for defining color appearance. The neural networks in the retina reiterate that visual perception in general, and specifically color appearance, cannot be treated as simple point-wise image processing problems. Several of these important features are discussed in more detail in the following sections on the retina, visual physiology, and visual performance.

1.2 THE RETINA

Figure 1.2 illustrates a cross-sectional representation of the retina. The retina includes several layers of neural cells beginning with the photoreceptors, the *rods* and *cones*. A vertical signal processing chain through the retina can be constructed by examining the connections of photoreceptors to bipolar cells, which are in turn connected to ganglion cells, which form the optic nerve. Even this simple pathway results in the signals from multiple photoreceptors being compared and combined. This is because multiple photoreceptors provide input to many of the bipolar cells, and

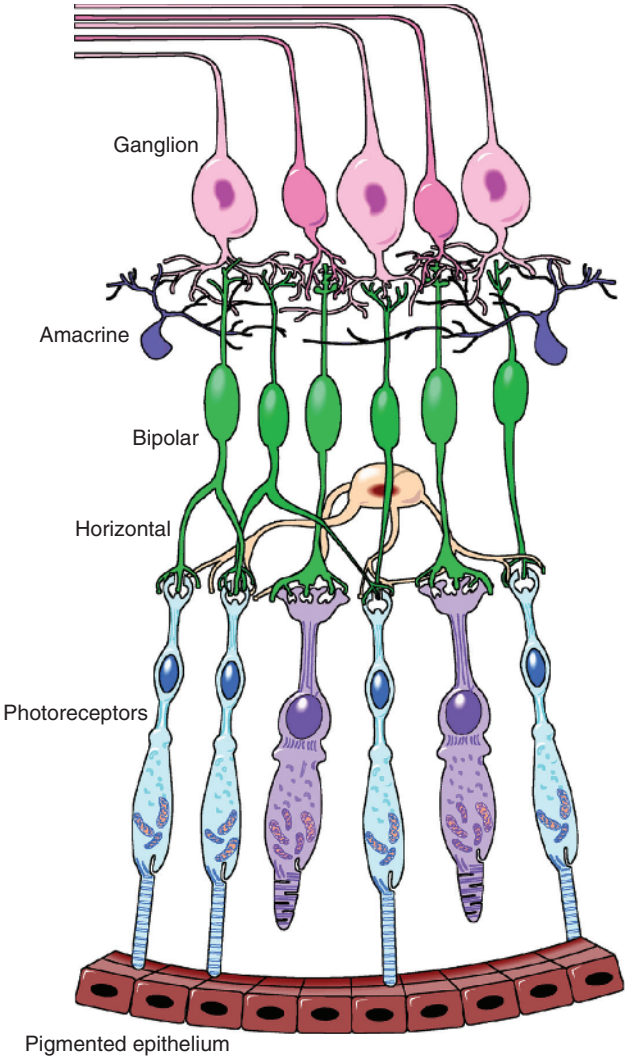


Figure 1.2 Schematic diagram of the "wiring" of cells in the human retina

multiple bipolar cells provide input to many of the ganglion cells. More importantly, this simple concept of retinal signal processing ignores two other significant types of cells. These are the *horizontal cells*, which connect photoreceptors and bipolar cells laterally to one another, and the *amacrine cells*, which connect bipolar cells and ganglion cells laterally to one another. Figure 1.2 provides only a slight indication of the extent of these various interconnections.

The specific processing that occurs in each type of cell is not completely understood and is beyond the scope of this chapter. However, it is important to realize that the signals transmitted from the retina to the higher levels of the brain via the ganglion cells are not simple point-wise representations of the receptor signals, but rather consist of sophisticated combinations of the receptor signals. To envision the complexity of the retinal processing, keep in mind that each synapse between neural cells can effectively perform a mathematical operation (add, subtract, multiply, and divide) in addition to the amplification, gain control, and nonlinearities that can occur within the neural cells. Thus the network of cells within the retina can serve as a sophisticated image computer. This is how the information from 130 million photoreceptors can be reduced to signals in approximately one million ganglion cells without loss of visually meaningful data.

It is interesting to note that light passes through all of the neural machinery of the retina prior to reaching the photoreceptors. This has little impact on visual performance since these cells are largely transparent and in fixed position, thus not perceived. It also allows the significant amounts of nutrients required, and waste produced, by the photoreceptors to be processed through the back of the eye.

Rods and Cones

Figure 1.3 provides a representation of the two classes of retinal photoreceptors: rods and cones. Rods and cones derive their respective names from their prototypical shape. Rods tend to be long and slender while peripheral cones are conical. This distinction is misleading since foveal cones, which are tightly packed due to their high density in the fovea, are long and slender, resembling peripheral rods.

The more important distinction between rods and cones is in visual function. Rods serve vision at low luminance levels (*e.g.*, less than $1\,cd/m^2$) while cones serve vision at higher luminance levels. Thus the transition from rod to cone vision is one mechanism that allows our visual system to function over a large range of luminance levels. At high luminance levels (*e.g.*, greater than $100\,cd/m^2$), the rods are effectively saturated and only the cones function. In the intermediate luminance levels, both rods and cones function and contribute to vision. Vision when only rods are active is referred to as *scotopic vision*. Vision served only by cones is referred to as *photopic vision*, and the term *mesopic vision* is used to refer to vision in which both rods and cones are active at intermediate luminance levels.

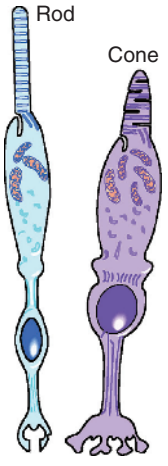


Figure 1.3 Illustrations of prototypical rod and cone photoreceptors

Rods and cones also differ substantially in their spectral sensitivities as illustrated in Figure 1.4(a). There is only one type of rod receptor with a peak spectral responsivity at approximately 510 nm. There are three types of cone receptors with peak spectral responsivities spaced through the visual spectrum.

The three types of cones are most properly referred to as L, M, and S cones. These names refer to the long-wavelength, middle-wavelength, and short-wavelength sensitive cones, respectively. Sometimes the cones are denoted with other symbols such as RGB or ργβ suggestive of red, green, and blue sensitivities. As can be seen in Figure 1.4(a) this concept can be misleading, and the LMS names are more appropriately descriptive. Note that the spectral responsivities of the three cone types are broadly overlapping, a design that is significantly different from the "color separation" responsivities that are often built into physical imaging systems. Such non-overlapping sensitivities, often incorporated in imaging systems for practical reasons, are the fundamental reason that accurate color reproduction is often difficult, if not impossible, to achieve.

The three types of cones clearly serve color vision. Since there is only one type of rod, the rod system is incapable of color vision. This can easily be observed by viewing a normally colorful scene at very low luminance levels. Figure 1.4(b) illustrates the two CIE spectral luminous efficiency functions, the $V'(\lambda)$ function for scotopic (rod) vision and the $V(\lambda)$ function for photopic (cone) vision. These functions represent the overall sensitivity of the two systems with respect to the perceived brightness of the various wavelengths. Since there is only one type of rod, the $V'(\lambda)$ function is identical to the spectral responsivity of the rods and depends on the spectral absorption of *rhodopsin*, the photosensitive pigment in rods. The $V(\lambda)$ function, however, represents a combination of the three types of cone signals rather than the responsivity of any single cone type.

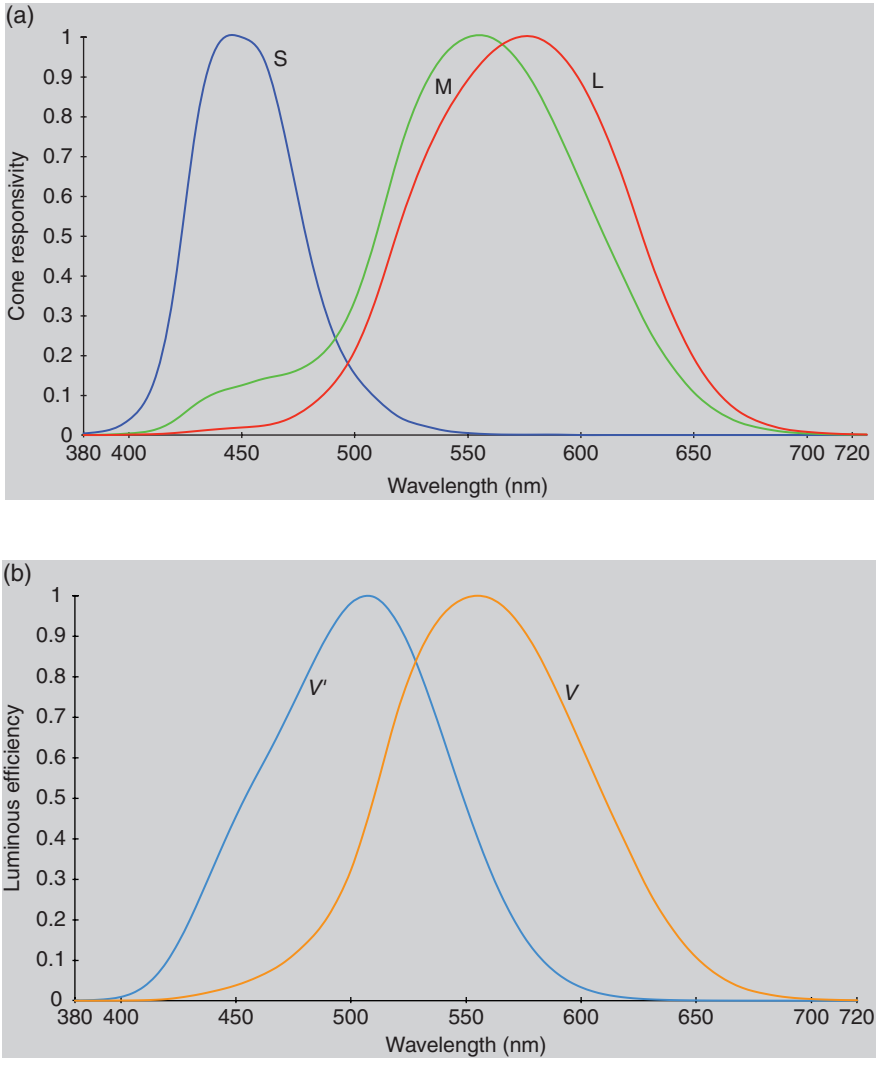


Figure 1.4 (a) Spectral responsivities of the L, M, and S cones and (b) the CIE spectral luminous efficiency functions for scotopic, $V'(\lambda)$, and photopic, $V(\lambda)$, vision

Note the difference in peak spectral sensitivity between scotopic and photopic vision. With scotopic vision we are more sensitive to shorter wavelengths. This effect, known as the Purkinje shift, can be observed by finding two objects, one blue and the other red, that appear the same lightness when viewed in daylight. When the same two objects are viewed under very low luminance levels, the blue object will appear quite light while the red object will appear nearly black because of the scotopic spectral sensitivity function's sensitivity to blue energy and almost complete lack of sensitivity to red energy.

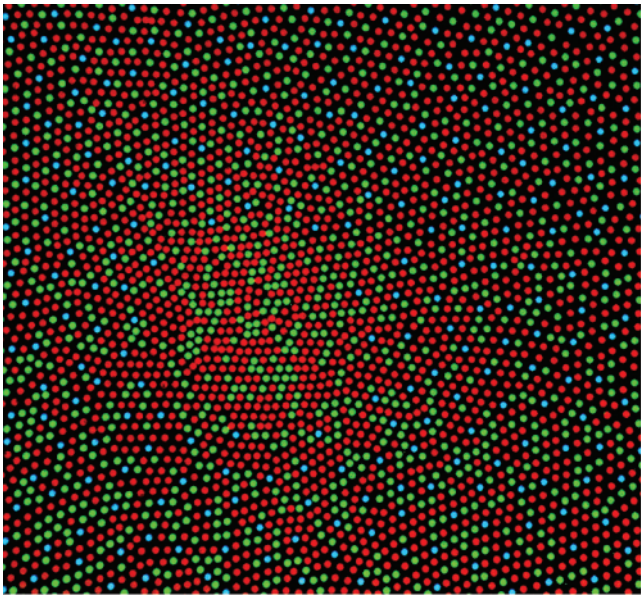
Another important feature about the three cone types is their relative distribution in the retina. It turns out that the S cones are relatively sparsely

populated throughout the retina and completely absent in the most central area of the fovea. There are far more L and M cones than S cones, and there are approximately twice as many L cones as M cones. The relative populations of the L : M : S cones are approximately 40 : 20 : 1. These relative populations must be considered when combining the cone responses (plotted with individual normalizations in Figure 1.4(a)) to predict higher-level visual responses. Figure 1.5 provides a schematic representation of the foveal photoreceptor mosaic with completely inaccurate and false coloring to represent a hypothetical distribution with the L cones in red, M cones in green, and S cones in blue. Figure 1.5 is presented simply as a convenient visual representation of the cone populations and should not be taken literally.

As illustrated in Figure 1.5, there are no rods present in the fovea. This feature of the visual system can also be observed when trying to look directly at a small dimly illuminated object, such as a faint star at night. It disappears since its image falls on the foveal area where there are no rods to detect the dim stimulus. Figure 1.6 shows the distribution of rods and cones across the retina. Several important features of the retina can be observed in Figure 1.6. First, note the extremely large numbers of photoreceptors. In some retinal regions, there are about 150000 photoreceptors per square millimeter of retina! Also note that there are far more rods (around 120 million per retina) than cones (around 7 million per retina). This might seem somewhat counterintuitive since cones function at high luminance levels and produce high visual acuity while rods function at low luminance levels and produce significantly reduced visual acuity (analogous to low-speed fine-grain photographic film vs a high-speed coarse-grain film). The solution to this apparent mystery lies in the fact that single cones feed into ganglion cell signals while rods pool their responses over hundreds of receptors (feeding into a single ganglion cell) in order to produce increased sensitivity at the expense of acuity. This also partially explains how the information from so many receptors can be transmitted through one million ganglion cells. Figure 1.6 also illustrates that cone receptors are highly concentrated in the fovea and more sparsely populated throughout the peripheral retina while there are no rods in the central fovea. The lack of rods in the central fovea allows that valuable space to be used to produce the highest possible spatial acuity with the cone system. A final feature to be noted in Figure 1.6 is the blind spot. This is the area, 12–15° from the fovea, where the optic nerve is formed and there is no room for photoreceptors.

Figure 1.7 provides some stimuli that can be used to demonstrate the existence of the blind spot. One reason the blind spot generally goes unnoticed is that it is located on opposite sides of the visual field in each of the two eyes. However, even when one eye is closed, the blind spot is not generally visible. To observe your blind spot, close your left eye and fixate the cross in Figure 1.7(a) with your right eye. Then adjust the viewing distance of the book until the spot to the right of the cross disappears when it falls on the blind spot. Note that what you see when the spot disappears is not a black region. Rather it appears to be an area of blank paper. This is an example of a phenomenon known as filling-in. Since your brain no longer

(a)



(b)

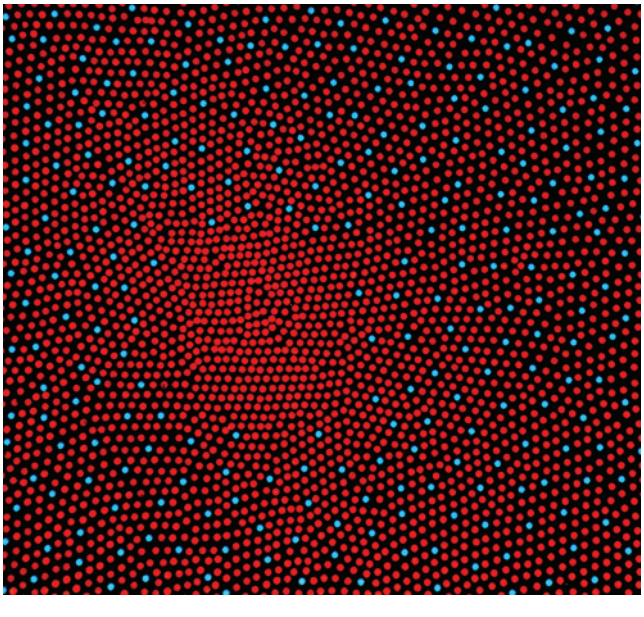


Figure 1.5 (a) A representation of the retinal photoreceptor mosaic artificially colored to represent the relative proportions of L (colored red), M (green), and S (blue) cones in the human retina. Modeled after Williams *et al.* (1991). (b) The same representation for a hypothetical deuteranope whose M cones contain L-cone photopigments (or whose M cones have been replaced with L cones)

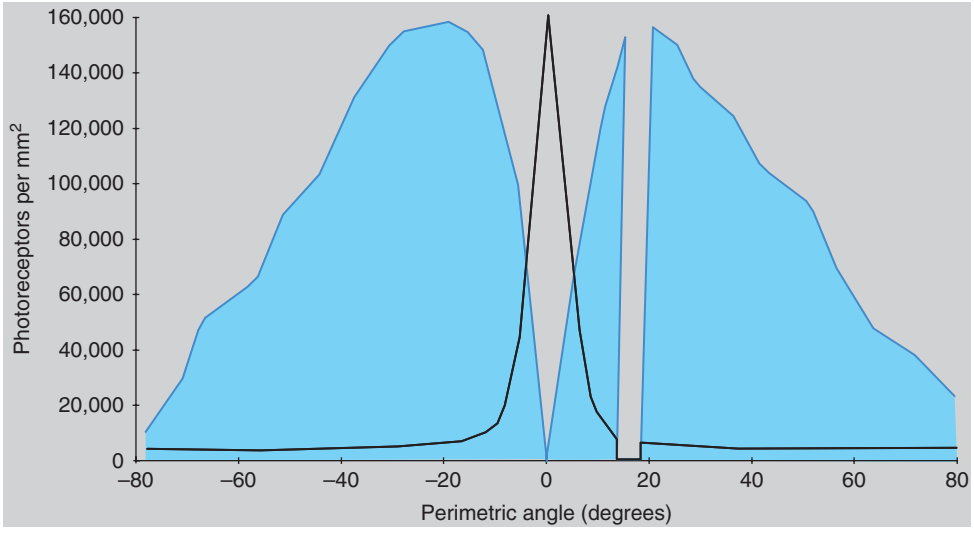


Figure 1.6 Density (receptors per square millimeter) of rod and cone photoreceptors as a function of location on the human retina

Figure 1.7 Stimuli used to illustrate the presence of the blind spot and "filling-in" phenomena. Close your left eye. Fixate the cross with your right eye and adjust the viewing distance until (a) the spot falls on your blind spot or (b) the gap in the line falls on your blind spot. Note the perception in that area in each case

has any signal indicating a change in the visual stimulus at that location, it simply fills in the most probable stimulus, in this case a uniform white of the paper. The strength of this filling-in can be illustrated by using Figure 1.7(b) to probe your blind spot. In this case, with your left eye closed, fixate the cross with your right eye and adjust the viewing distance until the gap in the line disappears when it falls on your blind spot. Amazingly the perception is that of a continuous line since that is now the most probable visual stimulus. If you prefer to perform these exercises using your left eye, simply turn the book upside down to find the blind spot on the other side of your visual field.

The filling-in phenomenon goes a long way to explain the function of the visual system. The signals present in the ganglion cells represent only local

changes in the visual stimulus. Effectively, only information about spatial or temporal transitions (*i.e.*, edges) is transmitted to the brain. Perceptually this code is sorted out by examining the nature of the changes and filling-in the appropriate uniform perception until a new transition is signaled. This coding provides tremendous savings in bandwidth to transmit the signal and can be thought of as somewhat similar to run-length encoding that is sometimes used in digital imaging.

Intrinsically Photosensitive Retinal Ganglion Cells

Within the past decade, the properties and roles of intrinsically photosensitive retinal ganglion cells (ipGRC) are beginning to be understood. These are ganglion cells that are directly photosensitive due to a unique photopigment within their cellular structure, known as melanopsin. Thus, rather than simply transmitting the signals from rod and cone photoreceptors, these ganglion cells transmit photo-signals produced intrinsically. The spectral responsivity of ipRGCs is quite broad and peaks at roughly 480 nm, between the S-cones (approximately 440 nm) and rods (approximately 505 nm), with a width similar to the $V(\lambda)$ function. Thus, ipRGCs represent a third class of photoreceptors in the retina and one that could have an interesting impact on how color appearance and chromatic adaptation are studied and modeled.

These cells have been implicated in a number of visual functions including modulation of circadian rhythms (Rea 2011), control of pupilary response, visual responses, and adaptation. The impacts of modification of these responses can include ailments such as seasonal affective disorder, obesity, cancer, and respiratory illness. Clearly ipRGCs are important to our welfare and visual performance and it remains to be seen how their understanding might improve our ability to predict color appearance.

1.3 VISUAL SIGNAL PROCESSING

The neural processing of visual information is quite complex within the retina and becomes significantly, if not infinitely, more complex at later stages. This section provides a very simplistic overview of the paths that some of this information takes. It is helpful to begin with a general map of the steps along the way. The optical image on the retina is first transduced into chemical and electrical signals in the photoreceptors. These signals are then processed through the network of retinal neurons (horizontal, bipolar, amacrine, and ganglion cells) described above. The ganglion cell axons gather to form the optic nerve, which projects to the lateral geniculate nucleus (LGN) in the thalamus. The LGN cells, after gathering input from the ganglion cells, project to visual area one (V1) in the occipital lobe of the cortex. Also note that there are more cells projecting from the cortex down to the LGN and vice versa. Therefore, the LGN clearly plays some role in modulating visual signals based on feedback from higher levels. At this

point, the information processing begins to become amazingly complex. At least 30 visual areas have been defined in the cortex with names such as V2, V3, V4, MT, *etc.* Signals from these areas project to several other areas and vice versa. The cortical processing includes many instances of feed-forward, feedback, and lateral processing. Somewhere in this network of information, our ultimate perceptions are formed. A few more details of these processes are described in the following paragraphs.

Light incident on the retina is absorbed by photopigments in the various photoreceptors (and ipRGCs). In rods, the photopigment is rhodopsin. Upon absorbing a photon, rhodopsin changes in structure, setting off a chemical chain reaction that ultimately results in the closing of ion channels in its cell walls which produce an electrical signal based on the relative concentrations of various ions (*e.g.*, sodium and potassium) inside and outside the cell wall. A similar process takes place in cones. Rhodopsin is made up of opsin and retinal. Cones have similar photopigment structures. However, in cones the "cone-opsins" have slightly different molecular structures resulting in the various spectral responsivities observed in the cones. Each type of cone (L, M, or S) contains a different form of "cone-opsin." Figure 1.8 illustrates the relative responses of the photoreceptors as a function of retinal exposure.

It is interesting to note that these functions show characteristics similar to those found in all imaging systems. At the low end of the receptor responses there is a threshold, below which the receptors do not respond. There is then a fairly linear portion of the curves followed by response saturation at the high end. Such curves are representations of the photocurrent at the receptors and represent the very first stage of visual processing after optical absorption. These signals are then processed through

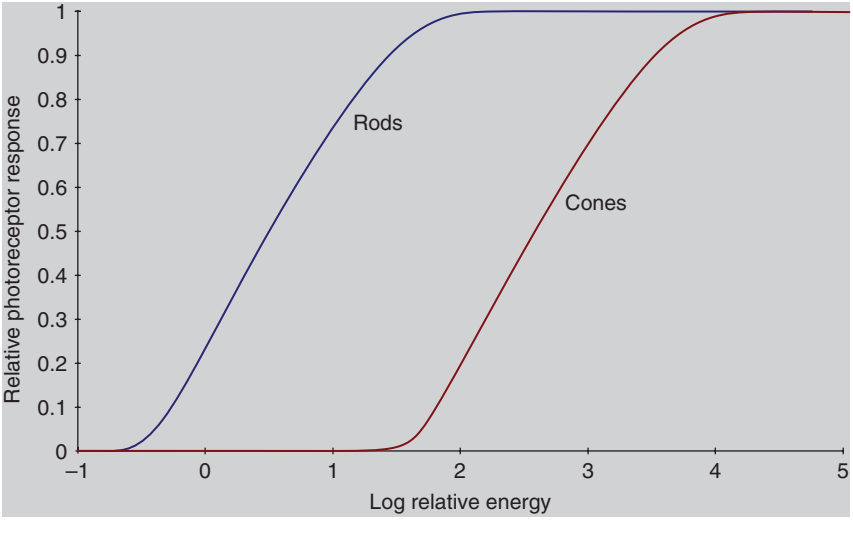


Figure 1.8 Relative energy responses for the rod and cone photoreceptors

the retinal neurons and synapses until a transformed representation is generated in the ganglion cells for transmission through the optic nerve.

Receptive Fields

For various reasons, including noise suppression and transmission speed, the amplitude-modulated signals in the photoreceptors are converted into frequency-modulated representations at the ganglion cell and higher levels. In these, and indeed most, neural cells the magnitude of the signal is represented in terms of the number of spikes of voltage per second fired by the cell rather than by the voltage difference across the cell wall. To represent the physiological properties of these cells, the concept of receptive fields becomes useful.

A *receptive field* is a graphical representation of the area in the visual field to which a given cell responds. In addition, the nature of the response (*e.g.*, positive, negative, spectral bias) is typically indicated for various regions in the receptive field. As a simple example, the receptive field of a photoreceptor is a small circular area representing the size and location of that particular receptor's sensitivity in the visual field. Figure 1.9 represents some prototypical receptive fields for ganglion cells. They illustrate center-surround antagonism, which is characteristic at this level of visual processing. The receptive field in Figure 1.9(a) illustrates a positive central response, typically generated by a positive input from a single cone, surrounded by a negative surround response, typically driven by negative inputs from several neighboring cones. Thus the response of this ganglion cell is made up of inputs from a number of cones with both positive and negative signs. The result is that the ganglion cell does not simply respond to points of light but serves as an edge detector (actually a "spot" detector). Those familiar with digital image processing can think of the ganglion cell responses as similar to the output of a convolution kernel designed for edge detection.

Figure 1.9(b) illustrates that a ganglion cell response of opposite polarity is equally possible. The response in Figure 1.9(a) is considered an on-center ganglion cell while that in Figure 1.9(b) is called an off-center ganglion cell. Often on-center and off-center cells will occur at the same spatial location, fed by the same photoreceptors, resulting in an enhancement of the system's dynamic range.

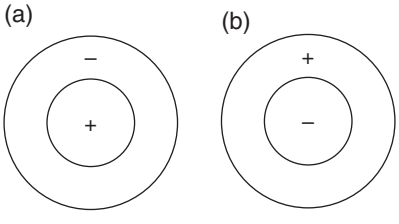


Figure 1.9 Typical center-surround antagonistic receptive fields: (a) on-center and (b) off-center

Note that the ganglion cells represented in Figure 1.9 will have no response to uniform fields (given that the positive and negative areas are balanced). This illustrates one aspect of the image compression carried out in the retina. The brain is not bothered with redundant visual information; only information about changes in the visual world is transmitted. This spatial information processing in the visual system is the fundamental basis of the important impact of the background on color appearance. Figure 1.9 illustrates spatial opponency in ganglion cell responses. Figure 1.10 shows that in addition to spatial opponency, there is often spectral opponency in ganglion cell responses. Figure 1.10(a) shows a red–green opponent response with the center fed by positive input from an L cone and the surround fed by negative input from M cones (or combinations of cone types if wired randomly). Even random wiring of ganglion cells with a single cone in the center will result in some form of spectral opponency. Figure 1.10(b) illustrates the off-center version of this cell. Thus, before the visual information has even left the retina, processing has occurred with a profound effect on color appearance.

Figure 1.9 and Figure 1.10 illustrate typical ganglion cell receptive fields. There are other types and varieties of ganglion cell responses, but they all share these basic concepts. On their way to the primary visual cortex, visual signals pass through the LGN. While the ganglion cells do terminate at the LGN, making synapses with LGN cells, there appears to be a one-to-one correspondence between ganglion cells and LGN cells. Thus, the receptive fields of LGN cells are identical to those of the ganglion cells. The LGN appears to act as a relay station for the signals. However, it most certainly serves some visual function since there are numerous neural projections from the cortex back to the LGN that could serve as some type of switching or adaptation feedback mechanism. The axons of LGN cells project to visual area one (V1) in the visual cortex.

Processing in Area V1

In area V1 of the cortex, the encoding of visual information becomes significantly more complex. Much as the outputs of various photoreceptors are combined and compared to produce ganglion cell responses, the outputs

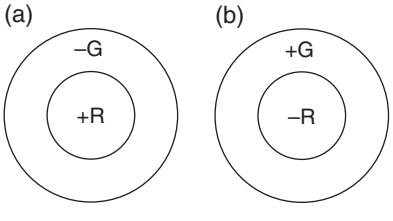


Figure 1.10 Examples of (a) red–green and (b) green–red spectrally and spatially antagonistic receptive fields

of various LGN cells are compared and combined to produce cortical responses. As the signals move further up in the cortical processing chain, this process repeats itself with the level of complexity increasing very rapidly to the point that receptive fields begin to lose meaning. In V1, cells can be found that selectively respond to various types of stimuli, including

- Oriented edges or bars
- Input from one eye, the other, or both
- Various spatial frequencies
- Various temporal frequencies
- Particular spatial locations
- Various combinations of these features.

In addition, cells can be found that seem to linearly combine inputs from LGN cells and others with nonlinear summation. All of these various responses are necessary to support visual capabilities such as the perceptions of size, shape, location, motion, depth, and color. Given the complexity of cortical responses in V1 cells, it is not difficult to imagine how complex visual responses can become in an interwoven network of many visual areas.

Figure 1.11 schematically illustrates a small portion of the connectivity of the various cortical areas that had been identified around the turn of the century. Bear in mind that Figure 1.11 is showing connections of areas, not cells. There are on the order of 10^9 cortical neurons serving visual functions. At these stages it becomes exceedingly difficult to explain the function of single cortical cells in simple terms. In fact, the function of a single cell might not have meaning since the representation of various perceptions must be distributed across collections of cells throughout the cortex. Rather than attempting to explore the physiology further, the following sections will describe some of the overall perceptual and psychophysical properties of the visual system that help to specify its performance.

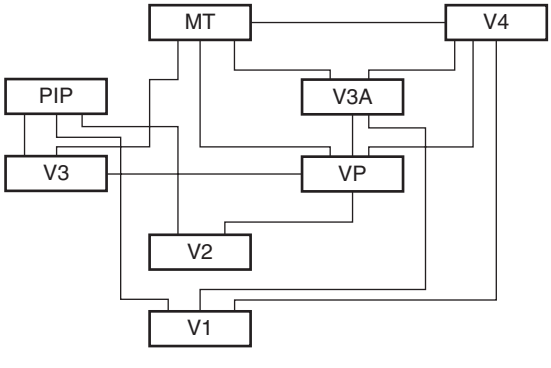


Figure 1.11 Partial flow diagram to illustrate the many streams of visual information processing in the visual cortex. Information can flow in both directions along each connection

1.4 MECHANISMS OF COLOR VISION

Historically, there have been many theories that attempt to explain the function of color vision. A brief look at some of the more modern concepts provides useful insight into current concepts.

Trichromatic Theory

In the later half of the nineteenth century, the trichromatic theory of color vision was developed based on the work of Maxwell, Young, and Helmholtz. They recognized that there must be three types of receptors, approximately sensitive to the red, green, and blue regions of the spectrum, respectively. The trichromatic theory simply assumed that three images of the world were formed by these three sets of receptors and then transmitted to the brain where the ratios of the signals in each of the images were compared in order to sort out color appearances. The trichromatic (three-receptor) nature of color vision was not in doubt, but the idea of three images being transmitted to the brain is both inefficient and fails to explain several visually observed phenomena.

Hering's Opponent Colors Theory

At around the same time, Hering (1920) proposed and discussed an opponent colors theory of color vision based on many subjective observations of color appearance. These observations included appearance of hues, simultaneous contrast, afterimages, and color vision deficiencies. Hering noted that certain hues were never perceived to occur together. For example, a color perception is never described as reddish-green or yellowish blue, while combinations of red and yellow, red and blue, green and yellow, and green and blue are readily perceived. This suggested to Hering that there was something fundamental about the red–green and yellow–blue pairs causing them to oppose one another. Similar observations were made of simultaneous contrast in which objects placed on a red background appear greener, on a green background appear redder, on a yellow background appear bluer, and on a blue background appear yellower. Figure 1.12 demonstrates the opponent nature of visual afterimages. The afterimage of red is green, green is red, yellow is blue, and blue is yellow. (It is worth noting that afterimages can also be easily explained in terms of complementary colors due to adaptation in a trichromatic system. Hering only referred to light–dark afterimages in support of opponent theory, not chromatic afterimages.) Lastly, Hering observed that those with color vision deficiencies lose the ability to distinguish hues in red–green or yellow–blue pairs.

All of these observations provide clues regarding the processing of color information in the visual system. Hering proposed that there were three types of receptors, but Hering's receptors had bipolar responses to

Figure 1.12 Stimulus for the demonstration of opponent afterimages. Fixate upon the black spot in the center of the four-colored squares for about 30 seconds, and then move your gaze to fixate the black spot in the uniform white area. Note the colors of the afterimages relative to the colors of the original stimuli

light–dark, red–green, and yellow–blue. At the time, this was thought to be physiologically implausible, and Hering's opponent theory did not receive appropriate acceptance.

Modern Opponent Colors Theory

In the middle of the twentieth century, Hering's opponent theory enjoyed a revival of sorts when quantitative data supporting it began to appear. For example, Svaetichin (1956) found opponent signals in electrophysiological measurements of responses in the retinas of goldfish (which happen to be trichromatic!). DeValois *et al.* (1958) found similar opponent physiological responses in the LGN cells of the macaque monkey. Jameson and Hurvich (1955) also added quantitative psychophysical data through their hue-cancellation experiments with human observers, which allowed measurement of the relative spectral sensitivities of opponent pathways. These data, combined with the overwhelming support of much additional research since that time, have led to the development of the modern opponent theory of color vision (sometimes called a stage theory) as illustrated in Figure 1.13.

Figure 1.13 illustrates that the first stage of color vision, the receptors, is indeed trichromatic as hypothesized by Maxwell, Young, and Helmholtz. However, contrary to simple trichromatic theory, the three "color-separation" images are not transmitted directly to the brain. Instead the neurons of the retina (and perhaps higher levels) encode the color into opponent signals. The outputs of all three cone types are summed (L+M+S) to produce an

achromatic response that matches the CIE $V(\lambda)$ curve as long as the summation is taken in proportion to the relative populations of the three cone types. Differencing of the cone signals allows construction of red–green (L – M + S) and yellow–blue (L + M – S) opponent signals. The transformation from LMS signals to the opponent signals serves to decorrelate the color information carried in the three channels, thus allowing more efficient signal transmission and reducing difficulties with noise. The three opponent pathways also have distinct spatial and temporal characteristics that are important for predicting color appearance. They are discussed further in Section 1.5.

The importance of the transformation from trichromatic to opponent signals for color appearance is reflected in the prominent place that it finds within the formulation of all color appearance models. Figure 1.13 includes not only a schematic diagram of the neural "wiring" that produces opponent responses, but also the relative spectral responsivities of these mechanisms both before and after opponent encoding.

Adaptation Mechanisms

However, it is not enough to consider the processing of color signals in the human visual system as a static "wiring diagram." The dynamic mechanisms of adaptation that serve to optimize the visual response to the particular viewing environment at hand must also be considered. Thus an overview of the various types of adaptation is in order. Of particular relevance to the study of color appearance are the mechanisms of dark, light, and chromatic adaptation.

Dark Adaptation

Dark adaptation refers to the change in visual sensitivity that occurs when the prevailing level of illumination is decreased, such as when walking into a darkened theater on a sunny afternoon. At first the entire theater appears completely dark, but after a few minutes one is able to clearly see objects in the theater such as the aisles, seats, and other people. This happens because the visual system is responding to the lack of illumination by becoming more sensitive and therefore capable of producing a meaningful visual response at the lower illumination level.

Figure 1.14 shows the recovery of visual sensitivity (decrease in threshold) after transition from an extremely high illumination level to complete darkness. At first, the cones gradually become more sensitive until the curve levels off after a couple of minutes. Then, until about 10 minutes have passed, visual sensitivity is roughly constant. At that point, the rod system, with a longer recovery time, has recovered enough sensitivity to outperform the cones and thus the rods begin controlling overall sensitivity. The rod sensitivity continues to improve until it becomes asymptotic after about 30 minutes.

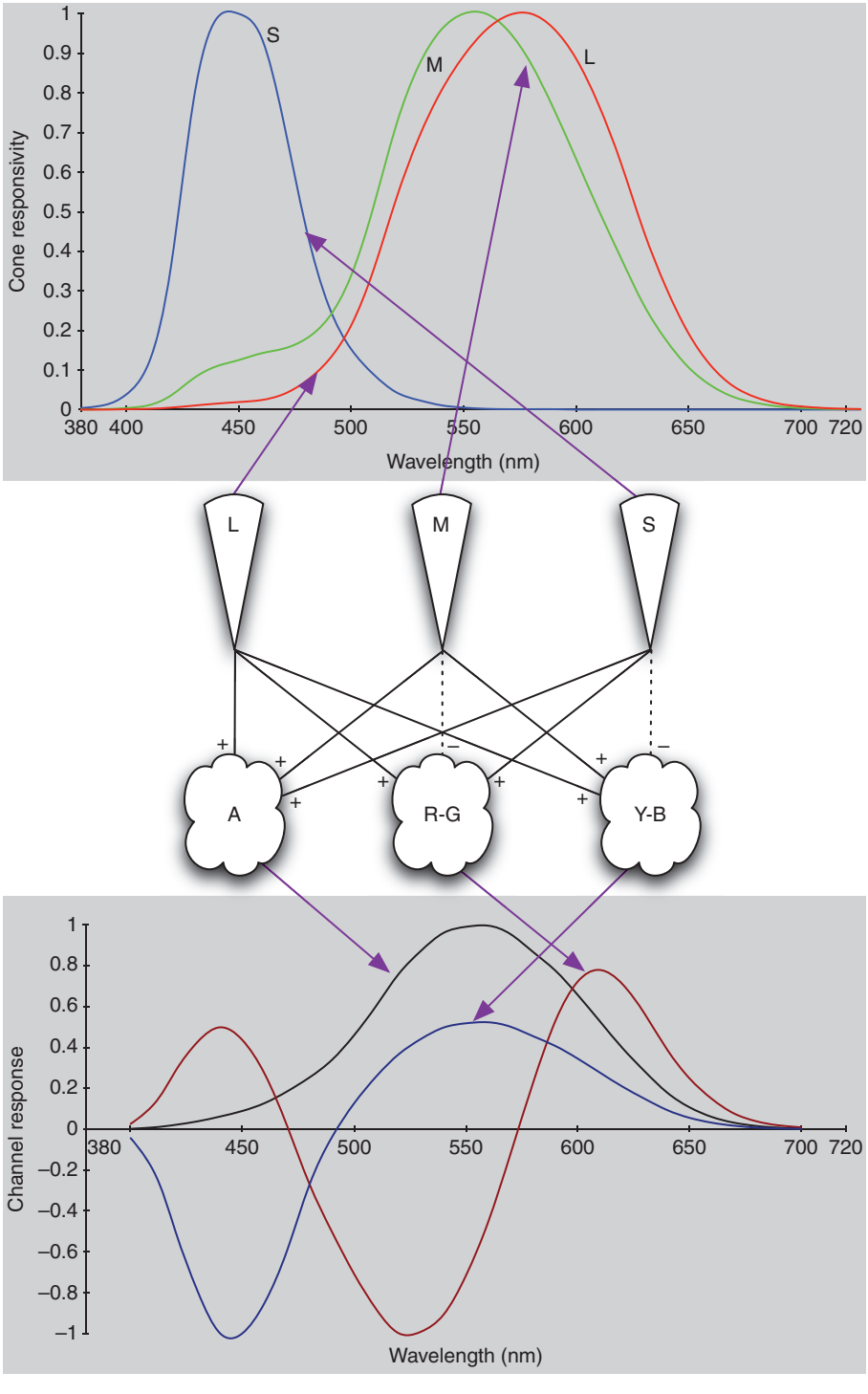


Figure 1.13 Schematic illustration of the encoding of cone signals into opponent colors signals in the human visual system

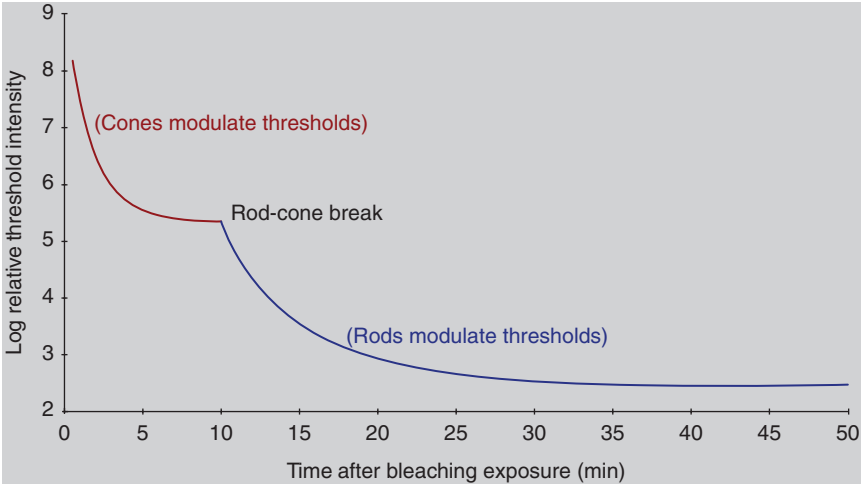


Figure 1.14 Dark-adaptation curve showing the recovery of threshold after a bleaching exposure. The break in the curve illustrates the point at which the rods become more sensitive than the cones

Recall that the five-fold change in pupil diameter is not sufficient to serve vision over the large range of illumination levels typically encountered. Therefore, neural mechanisms must produce some adaptation. Mechanisms thought to be responsible for various types of adaptation include the following:

- Depletion and regeneration of photopigment
- The rod–cone transition
- Gain control in the receptors and other retinal cells
- Variation of pooling regions across photoreceptors
- Spatial and temporal opponency
- Gain control in opponent and other higher level mechanisms
- Neural feedback
- Response compression
- Cognitive interpretation.

Light Adaptation

Light adaptation is essentially the inverse process of dark adaptation. However, it is important to consider it separately since its visual properties differ. Light adaptation occurs when leaving the darkened theater and returning outdoors on a sunny afternoon. In this case, the visual system must become less sensitive in order to produce useful perceptions since there is significantly more visible energy available.

The same physiological mechanisms serve light adaptation, but there is an asymmetry in the forward and reverse kinetics resulting in the time course of light adaptation being on the order of 5 minutes rather than

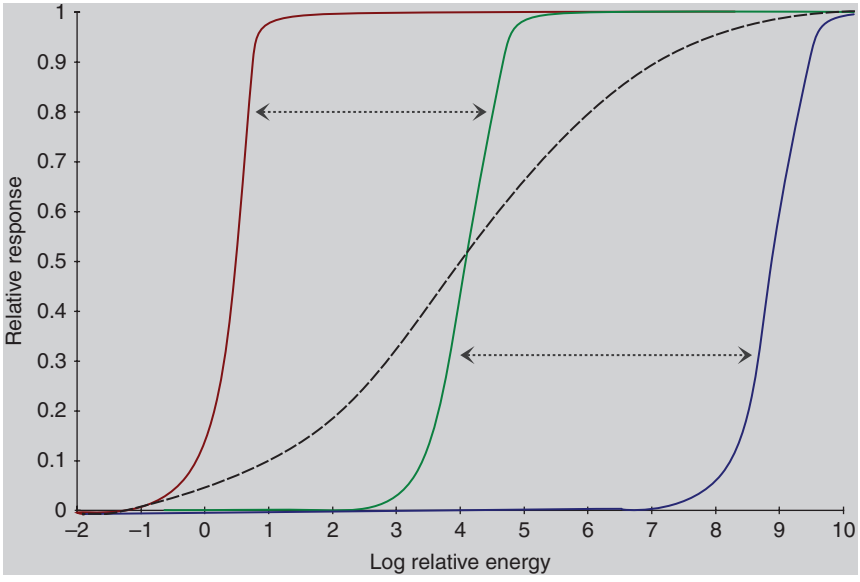


Figure 1.15 Illustration of the process of light adaptation whereby a very large range of stimulus intensity levels can be mapped into a relatively limited response dynamic range. Solid curves show a family of adapted responses. Dashed curve shows a hypothetical response with no adaptation

30 minutes. Figure 1.15 illustrates the utility of light adaptation. The visual system has a limited output dynamic range, say 100 : 1, available for the signals that produce our perceptions. The world in which we function, however, includes illumination levels covering at least 10 orders of magnitude from a starlit night to a sunny afternoon. Fortunately, it is almost never important to view the entire range of illumination levels at the same time. If a single response function were used to map the large range of stimulus intensities into the visual system's output, then only a small range of the available output would be used for any given scene. Such a response is shown by the dashed line in Figure 1.15. Clearly, with such a response function, the perceived contrast of any given scene would be limited and visual sensitivity to changes would be severely degraded due to signal-to-noise issues.

On the other hand, light adaptation serves to produce a family of visual response curves as illustrated by the solid lines in Figure 1.15. These curves map the useful illumination range in any given scene into the full dynamic range of the visual output, thus resulting in the best possible visual perception for each situation. Light adaptation can be thought of as the process of sliding the visual response curve along the illumination level axis in Figure 1.15 until the optimum level for the given viewing conditions is reached. Light and dark adaptation can be thought of as analogous to an automatic exposure control in a photographic system.

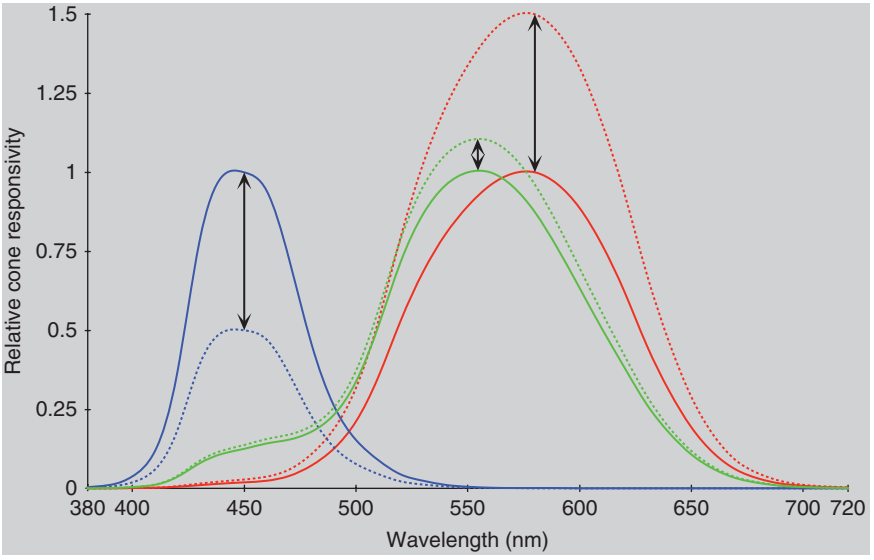


Figure 1.16 Conceptual illustration of the process of chromatic adaptation as the independent sensitivity regulation of the three-cone responsivities

Chromatic Adaptation

The third type of adaptation, closely related to light and dark adaptation, is chromatic adaptation. Again, similar physiological mechanisms are thought to produce chromatic adaptation. *Chromatic adaptation* is the largely independent sensitivity control of the three mechanisms of color vision. This is illustrated schematically in Figure 1.16, which shows that the overall height of the three-cone spectral responsivity curves can vary independently. While chromatic adaptation is often discussed and modeled as independent sensitivity control in the cones, there is no reason to believe that it does not occur in opponent and other color mechanisms as well.

Chromatic adaptation can be observed by examining a white object, such as a piece of paper, under various types of illumination (*e.g.*, daylight, fluorescent, and incandescent). Daylight contains relatively far more short-wavelength energy than fluorescent light, and incandescent illumination contains relatively far more long-wavelength energy than fluorescent light. However, the paper approximately retains its white appearance under all three light sources. This is because the S-cone system becomes relatively less sensitive under daylight to compensate for the additional short-wavelength energy and the L-cone system becomes relatively less sensitive under incandescent illumination to compensate for the additional long-wavelength energy.

Chromatic adaptation can be thought of as analogous to an automatic white-balance in video cameras. Figure 1.17 provides a visual demonstration of chromatic adaptation in which the two halves of the visual field are

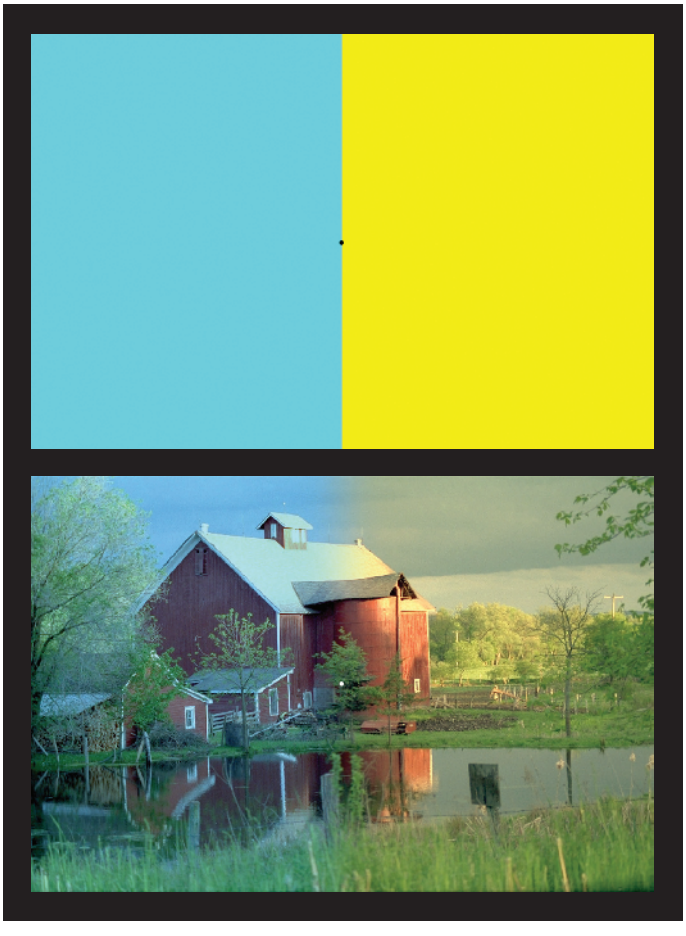


Figure 1.17 A demonstration of retinally localized chromatic adaptation. Fixate the black spot in between the uniform blue and yellow areas for about 30 seconds, and then shift your gaze to the white spot in the center of the barn image. Note that the barn image appears approximately uniform after this adaptation. Original barn image from Kodak Photo Sampler PhotoCD

conditioned to produce disparate levels of chromatic adaptation. Given its fundamental importance in color appearance modeling, chromatic adaptation is covered in more detail in Chapter 8.

Visual Mechanisms Impacting Color Appearance

There are many important cognitive visual mechanisms that impact color appearance. These are described in further detail in Chapters 6–8. They include memory color, color constancy, discounting-the-illuminant, and object recognition.

- *Memory color* refers to the phenomenon that recognizable objects often have a prototypical color that is associated with them. For example, most people have a memory for the typical color of green grass and can produce a stimulus of this color if requested to do so in an experiment. Interestingly, the memory color often is not found in the actual objects. For example, green grass and blue sky are typically remembered as being more saturated than the actual stimuli.
- *Color constancy* refers to the everyday perception that the colors of objects remain unchanged across significant changes in illumination color and luminance level. Color constancy is served by the mechanisms of chromatic adaptation and memory color and can easily be shown to be very poor when careful observations are made.
- *Discounting-the-illuminant* refers to an observer's ability to automatically interpret the illumination conditions and perceive the colors of objects after discounting the influences of illumination color.
- *Object recognition* is generally driven by the spatial, temporal, and light–dark properties of the objects rather than by chromatic properties (Davidoff 1991).

Thus once the objects are recognized, the mechanisms of memory color and discounting-the-illuminant can fill in the appropriate color. Such mechanisms have fascinating impacts on color appearance and become of critical importance when performing color comparisons across different media.

Clearly visual information processing is extremely complex and not yet fully understood (perhaps it never will be). It is of interest to consider the increasing complexity of cortical visual responses as the signal moves through the visual system. Single-cell electrophysiological studies have found cortical cells with extremely complex driving stimuli. For example, cells in monkeys that respond only to images of monkey paws or faces have been occasionally found in physiological experiments, and adaptation effects have been measured on stimuli as complex as the features of human faces. The existence of such cells begs the question of how complex a single-cell response can become. Clearly it is not possible for every perception to have its own cortical cell. Thus, at some point in the visual system, the representation of perceptions must be distributed with combinations of various signals producing various perceptions. Such distributed representations open up the possibilities for numerous permutations on a given perception, such as color appearance. It is clear from the large number of stimulus variables that impact color appearance that our visual system is often experimenting with these permutations.

1.5 SPATIAL AND TEMPORAL PROPERTIES OF COLOR VISION

No dimension of visual experience can be considered in isolation. The color appearance of a stimulus is not independent of its spatial and temporal characteristics. For example, a black and white stimulus flickering at

an appropriate temporal frequency can be perceived as quite colorful (Fechner–Benham colors). The spatial and temporal characteristics of the human visual system are typically explored through measurement of contrast sensitivity functions (CSFs). CSFs in vision science are analogous to modulation transfer functions (MTFs) in imaging science. However, CSFs cannot legitimately be considered MTFs since the human visual system is highly nonlinear and CSFs represent threshold sensitivity and not suprathreshold modulation. A CSF is defined by the threshold response to contrast (sensitivity is the inverse of threshold) as a function of spatial or temporal frequency. Contrast is typically defined as the difference between maximum and minimum luminance in a stimulus divided by the sum of the maximum and minimum luminances (called Michelson contrast), and CSFs are typically measured with stimuli that vary sinusoidally across space and/or time. Thus a uniform pattern has a contrast of 0 and sinusoidal patterns with troughs that reach a luminance of 0 have a contrast of 1.0 no matter what their mean luminance is.

Figure 1.18 conceptually illustrates typical spatial CSFs for luminance (black–white) and chromatic (red–green and yellow–blue at constant luminance) contrast. The luminance CSF is band-pass in nature, with peak sensitivity around five cycles per degree. This function approaches 0 at zero cycles per degree, illustrating the tendency for the visual system to be insensitive to uniform fields. It also approaches 0 at about 60 cycles per degree, the point at which detail can no longer be resolved by the optics of the eye or the photoreceptor mosaic. The band-pass CSF correlates with the concept of center-surround antagonistic receptive fields that would be most sensitive to an intermediate range of spatial frequency. The chromatic

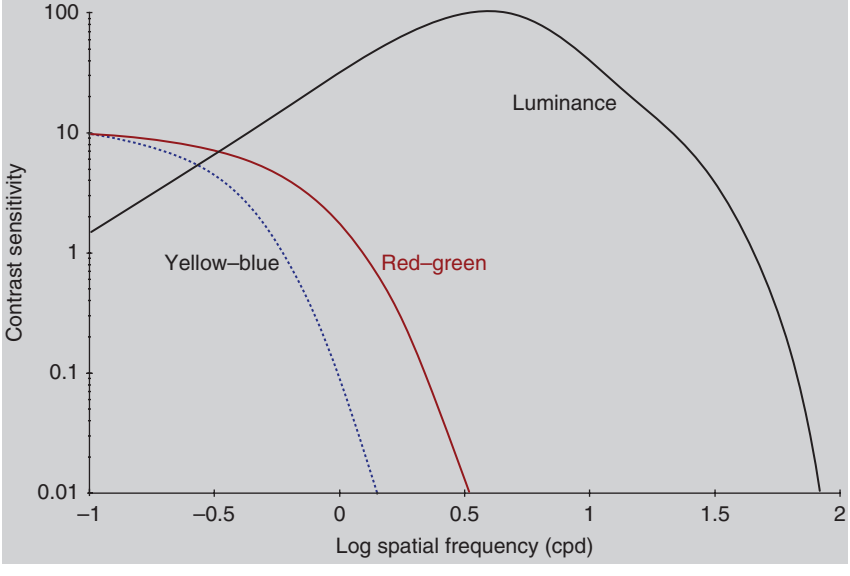


Figure 1.18 Spatial CSFs for luminance and chromatic contrast

mechanisms are of a low-pass nature and have significantly lower cutoff frequencies (Mullen 1985). This indicates the reduced availability of chromatic information for fine details (high spatial frequencies) that is often taken advantage of in image coding and compression schemes (*e.g.*, MPEG, NTSC, or JPEG).

The low-pass characteristics of the chromatic mechanisms also illustrate that edge detection/enhancement does not occur along these dimensions. The blue–yellow chromatic CSF has a lower cutoff frequency than the red–green chromatic CSF due to the scarcity of S cones in the retina. It is also of note that the luminance CSF is significantly higher than the chromatic CSFs, indicating that the visual system is more sensitive to small changes in luminance contrast compared to chromatic contrast. However, this comparison is just conceptual as the dimensions of contrast in luminance are not easily related to chromatic contrast metrics. Changes in the measurement units, which can be made arbitrarily, can reverse the order of the CSF plots. The spatial CSFs for luminance and chromatic contrast are generally not directly incorporated in color appearance models although there is significant interest in doing so. Zhang and Wandell (1996) presented an interesting technique for incorporating these types of responses into the CIELAB color space calculations. Johnson and Fairchild (2003b) provide a more recent implementation of the model.

Figure 1.19 illustrates the spatial properties of color vision with a spatial analysis of a typical image. Figure 1.19(a) shows the original image. The luminance information is presented alone in Figure 1.19(b) and the residual chromatic information is presented alone in Figure 1.19(c). It is clear that far more spatial detail can be visually obtained from the luminance image than from the chromatic residual image. This is further illustrated in Figure 1.19(d), in which the image has been reconstructed using the full-resolution luminance image combined with the chromatic image after subsampling by a factor of four. This form of image compression produces no noticeable degradation in perceived resolution or color.

Figure 1.20 conceptually illustrates typical temporal CSFs for luminance and chromatic contrast. They share many characteristics with the spatial CSFs shown in Figure 1.18. Again, the luminance temporal CSF is higher in both sensitivity and cutoff frequency than the chromatic temporal CSFs, and it shows band-pass characteristics suggesting the enhancement of temporal transients in the human visual system. Again, temporal CSFs are not directly incorporated in color appearance models, but they might be of importance to consider when viewing time-varying images such as digital video clips that might be rendered at differing frame rates.

It is important to realize that the functions in Figure. 1.18 and Figure 1.20 are typical and not universal. As stated earlier, the dimensions of human visual perception cannot be examined independently. The spatial and temporal CSFs interact with one another. A spatial CSF measured at different temporal frequencies will vary tremendously and the same is true for a temporal CSF measured at various spatial frequencies. These functions also depend on other variables such as

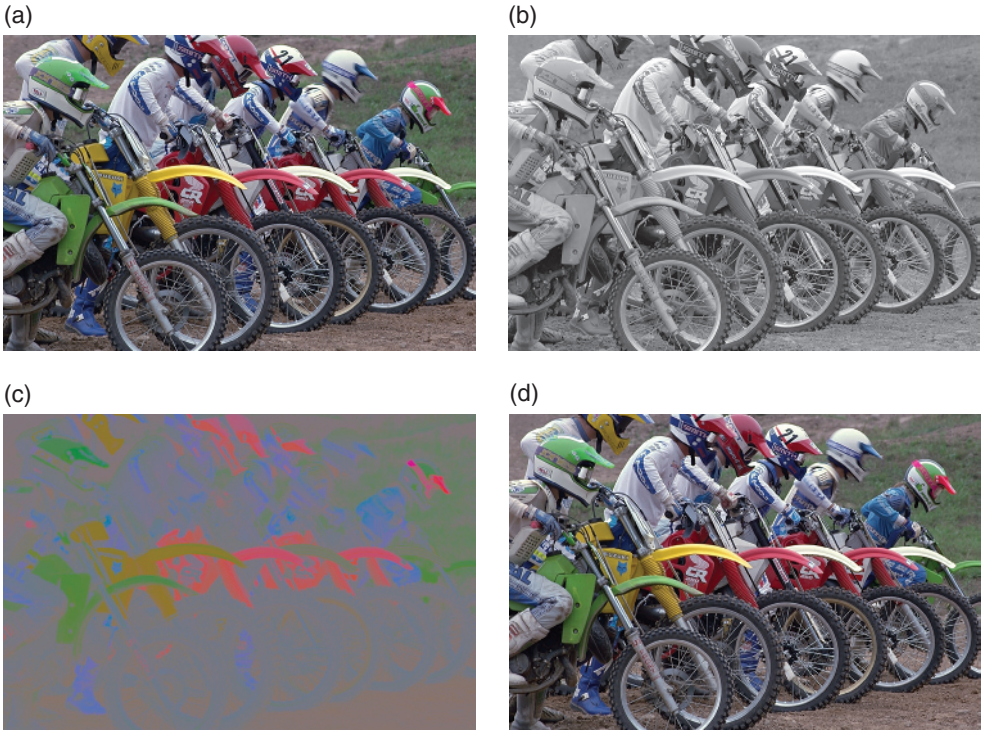


Figure 1.19 Illustration of the spatial properties of color vision: (a) original image, (b) luminance information only, (c) chromatic information only, and (d) reconstruction with full resolution luminance information combined with chromatic information subsampled by a factor of four. Original motorcycles image from Kodak Photo Sampler PhotoCD

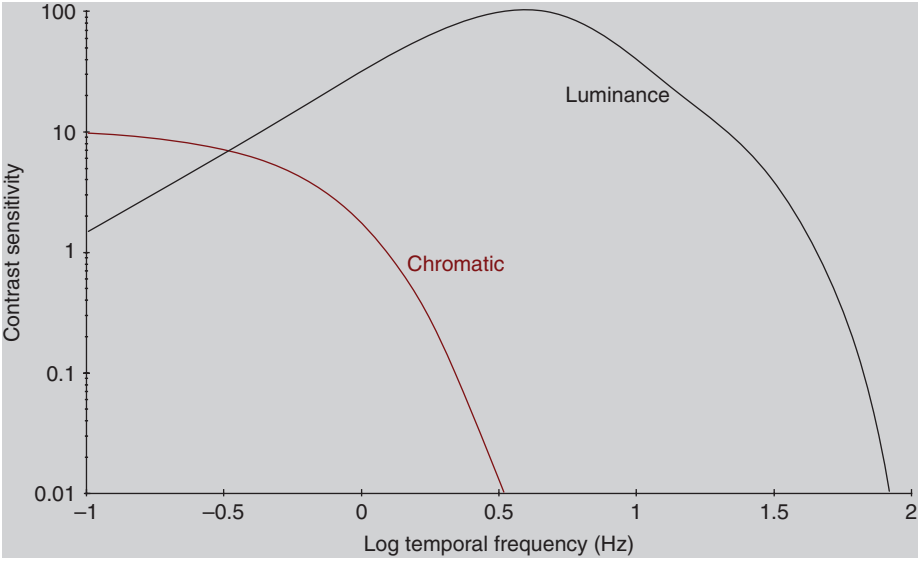


Figure 1.20 Temporal CSFs for luminance and chromatic contrast

luminance level, stimulus size, retinal locus, *etc*. See Kelly (1994) for a detailed treatment of some of these interactions.

The Oblique Effect

An interesting spatial vision phenomenon is the oblique effect. This refers to the fact that visual acuity is better for gratings oriented at 0° or 90° (relative to the line connecting the two eyes) than for gratings oriented at 45°. This phenomenon is considered in the design of rotated halftone screens that are set up such that the most visible pattern is oriented at 45°. The effect can be observed by taking a black-and-white halftone newspaper image and adjusting the viewing distance until the halftone dots are just barely imperceptible. If the image is kept at this viewing distance and then rotated 45°, the halftone dots will become clearly perceptible (since they will then be oriented at 0° or 90°). Figure 1.21 illustrates this with a halftoned image with an enlarged dot structure that has been rotated 45°. Find the distance at which you can just barely perceive the dot pattern in the upright image and note how the pattern is significantly more visible in the rotated version of the image. The dot pattern is at 45° (oblique) in the upright image and 0°/90° in the rotated image.

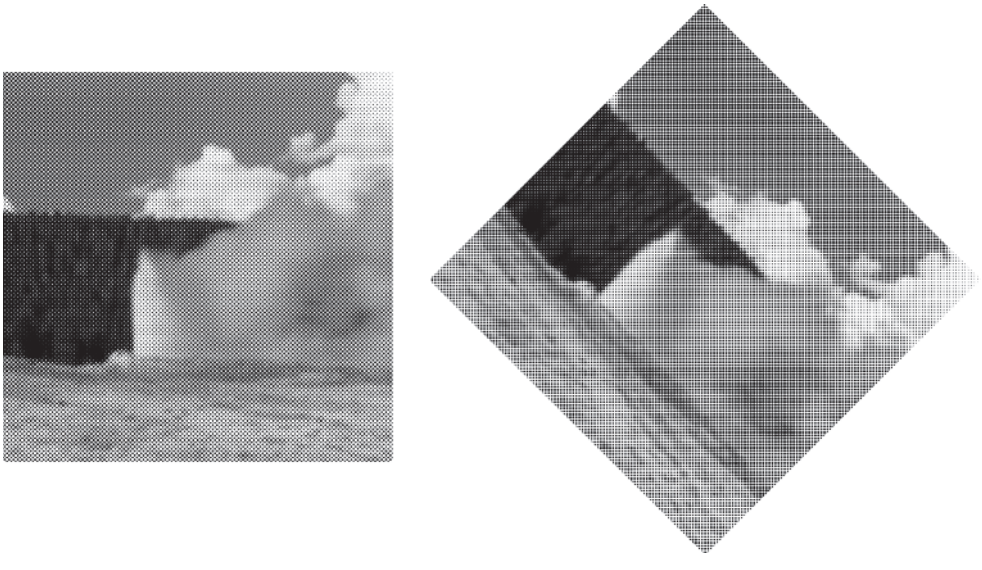


Figure 1.21 Demonstration of the oblique effect. Find the distance at which you can just barely perceive the dot pattern in the upright image and note the visibility of the pattern in the rotated image. The dot pattern is oriented at 45° (oblique) in the upright image to be less visible and 0°/90° in the rotated image, which makes it more visible

CSFs and Eye Movements

The spatial and temporal CSFs are also closely linked to the study of eye movements. A static spatial pattern becomes a temporally varying pattern when observers move their eyes across the stimulus. Noting that both the spatial and temporal luminance CSFs approach 0 as either form of frequency variation approaches 0, it follows that a completely static stimulus is invisible. This is indeed the case. If the retinal image can be fixed using a video feedback system attached to an eye tracker, the stimulus does disappear after a few seconds (Kelly 1994). (Sometimes this can be observed by carefully fixating an object and noting that the perceptions of objects in the periphery begin to fade away after a few seconds. The centrally fixated object does not fade away since the ability to hold the eye completely still has variance greater than the high spatial resolution in the fovea. Peripheral objects fade away since the spatial resolution of the visual system is poor in the periphery. This phenomenon is known as Troxler fading.)

To avoid this rather unpleasant phenomenon in typical viewing, our eyes are constantly moving. Large eye movements take place to allow viewing of different areas of the visual field with the high-resolution fovea. Also, there are small constant eye movements that serve to keep the visual world nicely visible. This also explains why the shadows of retinal cells and blood vessels are generally not visible since they do not move *on* the retina, but rather move *with* the retina. The history of eye movements has significant impact on adaptation and appearance through integrated exposure of various retinal areas and the need for movements to preserve apparent contrast. Recent technological advances have allowed psychophysical investigation of these effects (Babcock *et al.* 2003).

1.6 COLOR VISION DEFICIENCIES

There are various types of inherited and acquired color vision deficiencies. Kaiser and Boynton (1996) provide a comprehensive overview of the topic while Rosenthal and Phillips (1997) provide some fascinating and useful insights. This section concentrates on the most common inherited deficiencies.

Protanopia, Deuteranopia, and Tritanopia

Some color vision deficiencies are caused by the lack of a particular type of cone photopigment. Since there are three types of cone photopigments, there are three general classes of these color vision deficiencies, namely protanopia, deuteranopia, and tritanopia. An observer with protanopia, known as a protanope, is missing the L-cone photopigment and therefore is unable to discriminate reddish and greenish hues since the red–green

opponent mechanism cannot be constructed. A deuteranope is missing the M-cone photopigment and therefore also cannot distinguish reddish and greenish hues due to the lack of a viable red–green opponent mechanism. Protanopes and deuteranopes can be distinguished by their relative luminous sensitivity since it is constructed from the summation of different cone types. The protanopic luminous sensitivity function is shifted toward shorter wavelengths. A tritanope is missing the S-cone photopigment and therefore cannot discriminate yellowish and bluish hues due to the lack of a yellow–blue opponent mechanism.

Anomalous Trichromacy

There are also anomalous trichromats who have trichromatic vision, but the ability to discriminate particular hues is reduced either due to shifts in the spectral sensitivities of the photopigments or contamination of photopigments (*e.g.*, some L-cone photopigment in the M-cones). Among the anomalous trichromats are those with any of the following:

- Protanomaly (weak in L-cone photopigment or L-cone absorption shifted toward shorter wavelengths)
- Deuteranomaly (weak in M-cone photopigment or M-cone absorption shifted toward longer wavelengths)
- Tritanomaly (weak in S-cone photopigment or S-cone absorption shifted toward longer wavelengths).

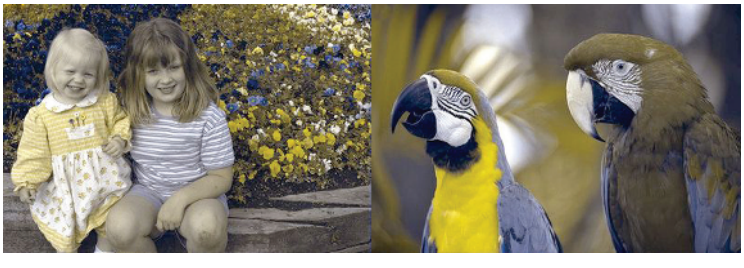
There are also some cases of cone monochromatism (effectively only one cone type) and rod monochromatism (no cone responses).

While it is impossible for a person with normal color vision to experience what the visual world looks like to a person with a color vision deficiency, it is possible to illustrate the hues that become indistinguishable. Figure 1.22 provides such a demonstration. To produce Figure 1.22, the two color-normal images (Figure 1.22(a)) processed according to a simulation algorithm published by Brettel *et al.* (1997) as implemented at www.vischeck.com to generate the images. More background on that model and the theory behind it can be found in Viénot *et al.* (1995). This allows an illustration of the various colors that would be confused by protanopes, deuteranopes, and tritanopes. The study of color vision deficiencies is of more than academic interest in the field of color appearance modeling and color reproduction. This is illustrated in Table 1.1 showing the approximate percentages of the population (European descent) with various types of color vision deficiencies.

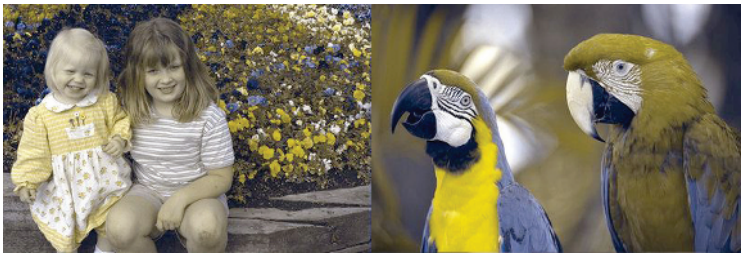
It is clear from Table 1.1 that color deficiencies are not extremely rare, particularly in the male population (about 8%), and that it might be important to account for the possibility of color-deficient observers in many applications.

(a)

(b)



(c)



(d)

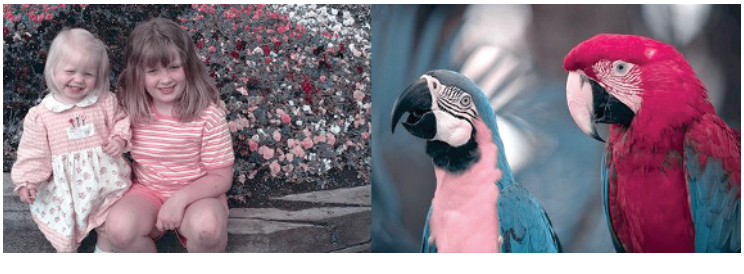


Figure 1.22 Images illustrating the color discrimination capabilities that are missing from observers with various color vision deficiencies: (a) original images, (b) protanope, (c) deuteranope, and (d) tritanope. Original birds image from Kodak Photo Sampler PhotoCD. Original girls image from the author. Images were processed at www.vischeck.com

Table 1.1 Approximate percent occurrences of various color vision deficiencies in populations of European descent

Type	Male %	Female %
Protanopia	1.0	0.02
Deuteranopia	1.1	0.01
Tritanopia	0.002	0.001
Cone monochromatism	~0	~0
Rod monochromatism	0.003	0.002
Protanomaly	1.0	0.02
Deuteranomaly	4.9	0.38
Tritanomaly	~0	~0
Total	8.0	0.4

Based on data in Hunt (1991a).

Color Vision Deficiencies and Gender

Why the disparity between the occurrence of color vision deficiencies in males and females? This can be traced back to the genetic basis of color vision deficiencies. It turns out that the most common forms of color vision deficiencies are sex-linked genetic traits.

The genes for photopigments are present on the X chromosome. Females inherit one X chromosome from their mother and one from their father. Only one of these need have the genes for the normal photopigments in order to produce normal color vision. On the other hand, males inherit an X chromosome from their mother and a Y chromosome from their father. If the single X chromosome does not include the genes for the photopigments, the son will have a color vision deficiency. If a female is color deficient, it means she has two deficient X chromosomes and all male children are destined to have a color vision deficiency. It is clear that the genetic "deck of cards" is stacked against males when it comes to inheriting deficient color vision. There are also multiple types of cones within classes (*e.g.*, L cones), and it is possible for women to have X chromosomes that express two different types of L cones. Such heterozygous females can actually be tetrachromatic in terms of number of distinct cone types. However, it appears that tetrachromatic females are still behaviorally trichromats, so the four cone signals must be combined into three channels at some point (Jordon *et al.* 2010). Neitz and Neitz (2000) provide a nice review of the genetics of color vision that can lead to tetrachromatic females, and Mancuso *et al.* (2009) also introduce the possibility of gene therapy for one-day treating color vision deficiencies.

Knowledge regarding the genetic basis of color vision has grown tremendously in recent years. John Dalton was an early investigator of deficient color vision. He studied his own vision, which was historically thought to be protanopic based on his observations, and came up with a theory as to the cause of his deficiencies. Dalton hypothesized that his color vision deficiency was caused by a coloration of his vitreous humor causing it to act

like a filter. Upon his death, he donated his eyes to have them dissected to experimentally confirm his theory. Unfortunately Dalton's theory was incorrect. However, Dalton's eyes have been preserved to this day in a museum in Manchester, UK. Hunt *et al.* (1995) performed DNA tests on Dalton's preserved eyes and were able to show that Dalton was a deuteranope rather than a protanope, but with an L-cone photopigment having a spectral responsivity shifted toward the shorter wavelengths. They were also able to complete a colorimetric analysis to show that their genetic results were consistent with the observations originally recorded by Dalton.

Screening Observers Who Make Color Judgments

Given the fairly high rate of occurrence of color vision deficiencies, it is necessary to screen observers prior to allowing them to make critical color appearance or color matching judgments. There are a variety of tests available, but two types, pseudoisochromatic plates and the Farnsworth–Munsell 100-Hue test, are of practical importance.

Pseudoisochromatic plates (*e.g.*, *Ishihara's Tests for Colour Blindness*) are color plates made up of dots of random lightness that include a pattern or number in the dots formed out of an appropriately contrasting hue. The random lightness of the dots is a design feature to avoid discrimination of the patterns based on lightness difference only. The plates are presented to observers under properly controlled illumination and they are asked to respond by either tracing the pattern or reporting the number observed. Various plates are designed with color combinations that would be difficult to discriminate for observers with the different types of color vision deficiencies. These tests are commonly administered as part of a normal ophthalmological examination and can be obtained from optical suppliers and general scientific suppliers. Screening with a set of peudoisochromatic plates should be considered as a minimum evaluation for anyone carrying out critical color judgments.

The Farnsworth–Munsell 100-Hue test, available through the Munsell Color company, consists of four sets of chips that must be arranged in an orderly progression of hue. Observers with various types of color vision deficiencies will make errors in the arrangement of the chips at various locations around the hue circle. The test can be used to distinguish between the different types of deficiencies and also to evaluate the severity of color discrimination problems. This test can also be used to identify observers with normal color vision, but poor color discrimination for all colors.

1.7 KEY FEATURES FOR COLOR APPEARANCE MODELING

This chapter provides a necessarily brief overview of the form and function of the human visual system concentrating on the features that are important in the study, modeling, and prediction of color appearance phenomena. What follows is a short review of the key features.

Important features in the optics of the eye include the lens, macula, and cone photoreceptors. The lens and macula impact color matching through their action as yellow filters. They impact inter-observer variability since their optical density varies significantly from person to person. The cones serve as the first stage of color vision, transforming the spectral power distribution on the retina into a three-dimensional signal that defines what is available for processing at higher levels in the visual system. This is the basis of metamerism, the fundamental governing principle of colorimetry.

The numerical distribution of the cones (L : M : S of about 40 : 20 : 1, but widely varying across observers) is important in constructing the opponent signals present in the visual system. Proper modeling of these steps requires the ratios to be accounted for appropriately. The spatial distribution of rods and cones and their lateral interactions are critical in the specification of stimulus size and retinal locus. A color appearance model for stimuli viewed in the fovea would be different from one for peripheral stimuli. The spatial interaction in the retina, represented by horizontal and amacrine cells, is critical for mechanisms that produce color appearance effects due to changes in background, surround, and level of adaptation.

The encoding of color information through the opponent channels along with the adaptation mechanisms before, during, and after this stage are perhaps the most important feature of the human visual system that must be incorporated into color appearance models. Each such model incorporates a chromatic adaptation stage, an opponent processing stage, and nonlinear response functions. Some models also incorporate light and dark adaptation effects and interactions between the rod and cone systems.

Lastly, the cognitive mechanisms of vision such as memory color and discounting-the-illuminant have a profound impact on color appearance. These and other color appearance phenomena are described in greater detail in Chapters 6–8.