Contents

List of Figures

Chapter 1

Wavelength Encoding

1.1   Introduction

Sir Isaac Newton's sketch in Figure 1.1 summarizes his investigations into the properties of light. In these experiments, Newton separated daylight into its fundamental components by passing it through a prism and creating a rainbow. Newton's demonstration that light can be decomposed into rays of different wavelength is at the foundation of our understanding of light and color.

To perform these experiments, Newton placed a shutter containing a small hole in the window in his room at Cambridge. The light emerging from the hole in the window shutter served as a point source to illuminate his apparatus. The key elements of the apparatus are featured prominently in the center of the figure: the lens and prism. Newton's drawing shows that when the daylight passed through the prism, it formed an image of a rainbow on his wall. With two experimental manipulations, he showed that the components of the rainbow were fundamental constituents of light. In the upper left of the sketch, we see a series of holes that Newton drilled in the wall permitting part of the rainbow to continue through to a second prism. This ray of light was cast upon a second surface, but the new image did not produce a second rainbow; rather, as Newton wrote:

> the color of the light was never changed in the least. If any part of the red light was refracted, it remained totally of the same red color as before. No orange, no yellow, no green or blue, nor other new color was produced by that refraction. (Newton, Opticks)

From this experiment, Newton concluded that the pass through the first prism had separated the daylight into its fundamental components. No further change was
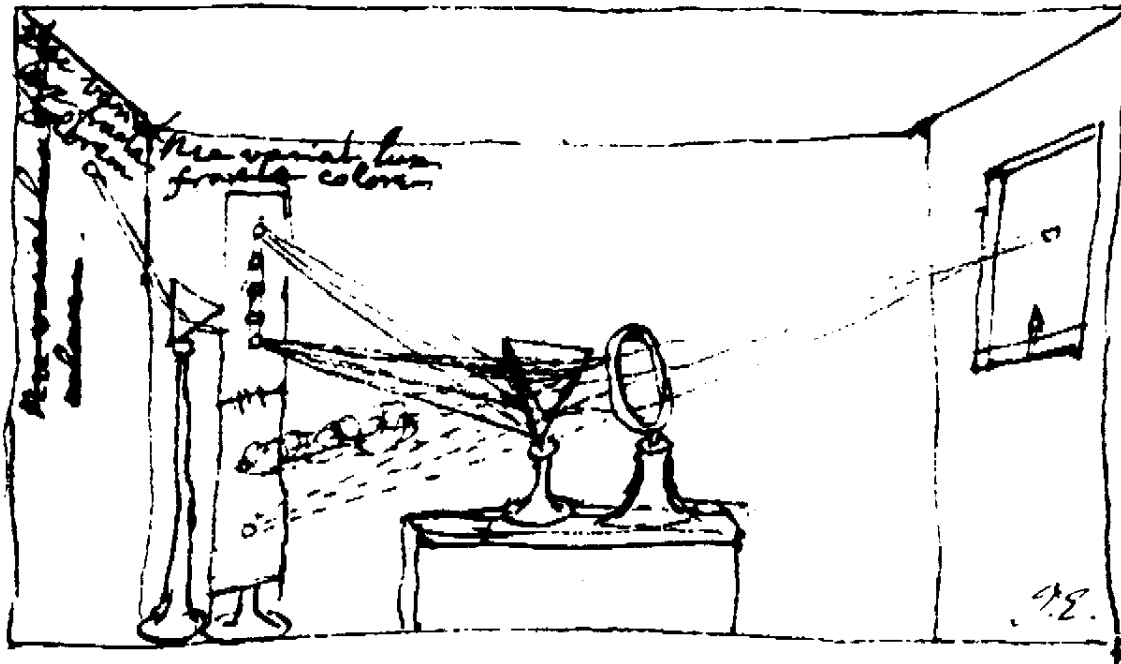
Figure 1.1: *Newton's summary drawing* of his experiments with light. Using a point source of light and a prism, Newton separated sunlight into its fundamental components. By reconverging the rays, he also showed that the decomposition is reversible.

observed when the ray passed through a second prism.

At the bottom of the sketch Newton illustrated that the decomposition is reversible: passing light through the prism does not destroy the character of the light. To show this Newton converged the rays following their passage through the prism to form a new image; he found that the color of the image same is the same as that of the source. Newton concluded that

> light being transmitted through the parallel surfaces of two prisms ... if it
> suffered any change by the refraction by one surface, it lost that
> impression by the contrary refraction of the other surface. (Newton,
> Opticks)

From the second experiment, he concluded that passage through the prism had not destroyed, but merely revealed, the character of the light.

We now know that Newton succeeded in decomposing the sunlight into its *spectral* components. each with its own characteristic wavelength. The prism separates the rays because the prism bends each wavelength of light by a different amount. (See the section on Snell's law in Chapter **??**). When we see the spectral components

(a)

(b)

Movable slit

Prism

Lens

Lens

Source

Sensor

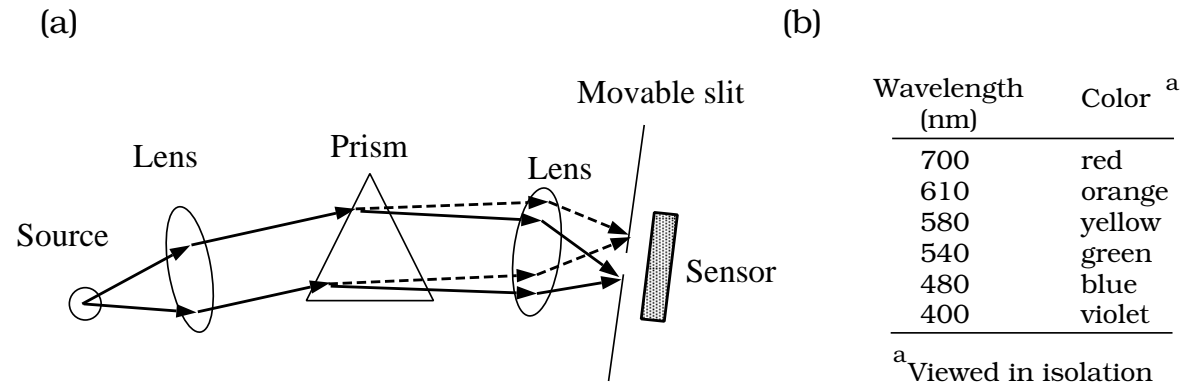| Wavelength (nm)	Color [a]
700	red
610	orange
580	yellow
540	green
480	blue
400	violet

[a]Viewed in isolation

Figure 1.2: *A spectroradiometer* is used to measure the spectral power distribution of light. (a) A schematic design of a spectroradiometer includes a means for separating the input light into its different wavelengths and a detector for measuring the energy at each of the separate wavelengths. (b) The color names associated with the appearance of lights at a variety of wavelengths are shown (After Wyszecki and Stiles, 1982).

separately, they each have a different color appearance. Light with relatively long wavelengths appears red when viewed against a dark background. Light with relatively short wavelengths appears blue when viewed against a dark background. Shorter wavelengths of light are refracted more strongly than longer wavelengths. A spectral light, with energy only at a single wavelength is also called a *monochromatic light*.

Newton's apparatus suggests a simple device we might build to measure the amount of power a light has in each of the different wavelength bands. As illustrated on the top of Figure 1.2, by proper use of lenses and prisms, we can form a focused image of the spectral components in an image plane with a movable slit placed in front of a photodetecting sensor. To measure the energy at different wavelengths, we move the slit passing only some of the spectral components at each position, and thus we measure the energy of the source at different wavelengths of light. In the visible region, the wavelength of light is on the order of a few hundred billionths of a meter, or *nanometers* (nm).

The *spectral power distribution* of a light is the function that defines the power in the light at each wavelength. In the modern theory of physics, the wavelength of light can be thought of in two different ways. We describe the light as if it were a continuous wave as it passes through a medium. When the light exchanges energy with some material, say by giving up its energy to be absorbed, we describe the light as if it were a discrete object called a *photon* or *quantum* of light. The amount of energy given up by the photon is predicted by the wavelength of the light.

The experimental aspect of light measurement that makes it useful and predictable
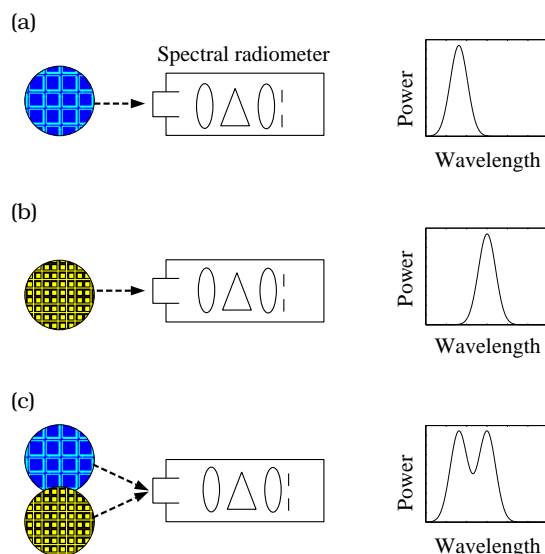
Figure 1.3: *The measurement of light spectral power distributions* satisfies the principle of superposition. The spectral power distributions of two lights measured separately are shown in (a) and (b) and together in (c). The spectral power distribution of the mixture is the sum of the individual measurements, thus demonstrating that superposition holds true.

is that the measurement satisfies the principle of superposition. We can demonstrate the superposition of light measurement as follows. First, measure the spectral power distributions of two lights separately. Then, mix the two lights together and measure again. The spectral power distribution of the mixture will be the sum of the first two spectral power distributions. This property of light mixture is illustrated in Figure 1.3. Superposition is a crucial property of light measurement because it implies that we can measure the energy of a light at each wavelength separately, and then combine the individual measurements to predict spectral power distribution when the spectral components are mixed together.

Suppose we wish to measure the spectral power distribution of a light source. How many wavelengths should we measure? Or, equivalently, how finely do we have to sample along the wavelength dimension? The answer to this question is important for both practical and theoretical reasons because the number of samples can be quite large. For example, to sample the visible spectrum from 400 nm to 700 nm in 1 nm steps, we need about 300 measurements. To sample in 10 nm steps, we need about 30 measurements.

The answer to this sampling question depends on the same set of issues as the sampling questions we addressed in Chapter **??** on the spatial sampling of the retinal image by the photoreceptor mosaics. If the energy in the light varies rapidly
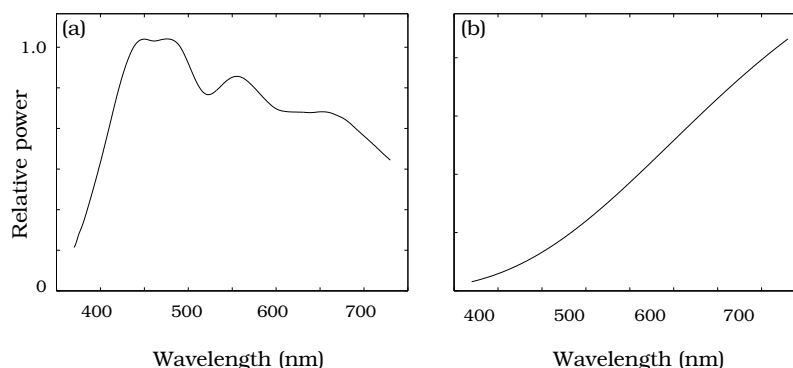
Figure 1.4: *The spectral power distribution* of two important light sources are shown: blue skylight (a) and the yellow disk of the sun (b).

as a function of wavelength, then we may have to sample quite finely to measure accurately; if the functions vary slowly, then only a few measurements are necessary. Also, the precision of the representation requires that we know how sensitive the photopigments in the are to rapid changes in the energy as a function of wavelength. It is difficult to make accurate generalizations about how spectral power distributions vary as a function of wavelength, but it is believed widely that for practical purposes we can approximate spectral power distributions using smooth, regular functions as shown in Figure 1.4. Also, it is known that the photopigments integrate broadly across the wavelength spectrum. Consequently, international standards organizations suggest making measurements every 5 nm to achieve an excellent representation of the signal. Practical measurements often rely on measurements spaced every 10 or 20 nm. We will consider this issue much more completely when we review color appearance, in Chapter **??**.

1.2   Scotopic Wavelength Encoding

What information do we encode about the spectral power distribution when rods initiate vision, under *scotopic* conditions? We can answer this question by an experiment designed to measure how well people can discriminate different spectral power distributions. In the *scotopic matching* experiment, we present an observer with two lights, side by side in a *bipartite* field. One side of the field contains the *test* light; it may have any spectral power distribution whatsoever. The second side of the field contains the *primary* light; it has a fixed relative spectral power distribution and can vary only by an overall intensity factor. The observer's task in the *scotopic matching experiment* is to adjust the primary light intensity so that the primary light appears indistinguishable from the test light. The observer can adjust only the

intensity of the primary light, so when the match is achieved the spectral power distributions of the test and primary lights that match are still different.

Under scotopic conditions, observers can adjust the primary intensity so that the primary matches any test light. Since subjects can always make this match, we have a simple answer to our question: The rods encode nothing about the relative spectral density of a light. An observer can adjust the intensity of a primary light to match the appearance of a test light with any spectral power distribution. The relative spectral power distribution is immaterial, all that matters is the relative intensities of the two lights.

Matching: Homogeneity and superposition

We can learn more about scotopic wavelength encoding by studying the quantitative properties of the matching experiment. To characterize the matching experiment completely, we must be able to predict how a subject will adjust the primary intensity to match any test light. We treat the experiment as a transformation by identifying the spectral power distribution of the test light as the input and the intensity of the primary light as the output. A quantitative description of the experiment tells us how to map the input to the output.

Naturally, we first ask whether we can characterize the matching experiment transformation using linear systems methods. Denote the spectral power distribution of the test and primary lights using the vectors $\mathbf{t}$ and $\mathbf{p}$ respectively. The n_λ entries of these vectors describe the power at each of the n_λ sample wavelengths. To test linearity, we evaluate whether the scotopic matching experiment satisfies the linear systems properties of homogeneity and superposition. We can evaluate these properties from the following experimental tests:

- (Homogeneity) If $\mathbf{t}$ matches $e\mathbf{p}$, will $a\mathbf{t}$ match $a(e\mathbf{p})$?

- (Superposition) If $\mathbf{t}$ matches $e\mathbf{p}$, and $\mathbf{t}'$ matches $e'\mathbf{p}$, will $\mathbf{t} + \mathbf{t}'$ match $(e\mathbf{p}) + (e'\mathbf{p})$?

An hypothetical test of homogeneity is shown in Figure 1.5. The separate panels show the intensity of the test light on the horizontal axis and the intensity of the matching primary light on the vertical axis. Each panel plots the results using spectral test lights at a series of wavelengths and a 510 nm primary light. In the scotopic matching experiment the data will fall on a straight line, consistent with the prediction from homogeneity. The slope of the line defines the relative scotopic sensitivity to the test and the primary lights. For example, in panel (c) the
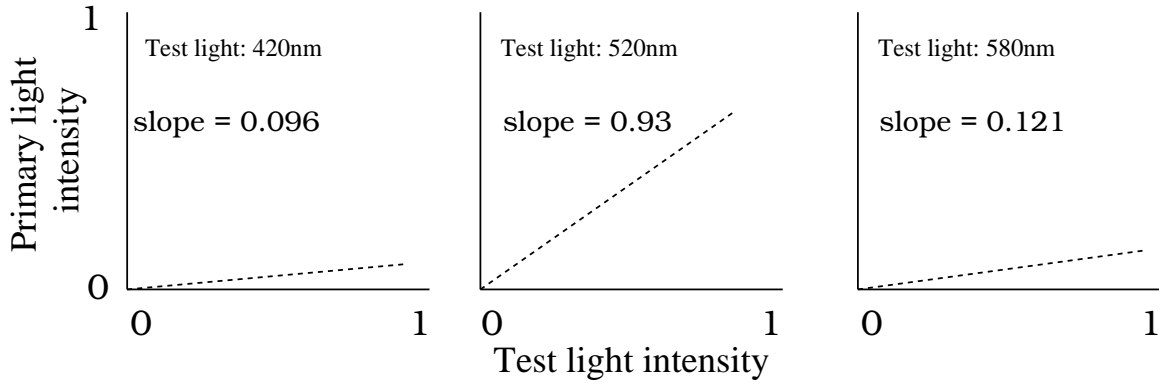
Figure 1.5: *Hypothetical scotopic matching experiment.* The horizontal scale measures the intensity of a monochromatic test light and the vertical scale measures the intensity a matching 510 nm primary light. Since the scotopic matching experiment satisfies homogeneity, the data will fall along a straight line. The slope of the line defines the relative scotopic sensitivity to each test wavelength.

hypothetical results from an experiment with a 580 nm test light are shown. The slope of the line shows that we need 8.3 units of energy at 580 nm to have the same effect as one unit of energy at 510 nm. Hence, the light at 510 nm is 8.3 times more effective, per unit energy, than the light at 580 nm.

Because the scotopic matching experiment is linear, there must be a system matrix, $\mathbf{R}$ that maps the input ($\mathbf{t}$, the test spectral power distribution), to the output (e, the primary light intensity). The system matrix, call it $\mathbf{R}$, must have one row and n_λ columns. The test light, system matrix, and primary intensity are related by the product, $e = \mathbf{Rt}$.

We can relate the measurements in the scotopic matching experiment to the entries of the system matrix as follows. Write the matrix product $\mathbf{Rt}$ as a summation over the sample wavelengths,

$$e = \sum_{i=1}^{i=n_\lambda} R_i t_i \tag{1.1}$$

Suppose we use a monochromatic test light of unit intensity, that is, an input $\mathbf{t}$ that has only a single non-zero wavelength, $(0, 0, ..., 0, 1, 0, ...0)^T$. Then Equation 1.1 becomes simply $e = R_i t_i$. This shows that the slope of the line relating the monochromatic test intensity, t_i, to the primary intensity, e, is the system matrix entry, R_i. Hence, we can estimate the system matrix from the slopes of the experimental lines we measure in the test of homogeneity shown in Figure 1.5.

Figure 1.7 is a graphical method of representing the system matrix of the scotopic matching experiment. The curve shows the entries of $\mathbf{R}$ as a function of wavelength,

Figure 1.6: *Matrix tableau of the scotopic matching experiment.* The primary light intensity, e, equals the product of the $1 \times n_\lambda$ scotopic matching system matrix, $\mathbf{R}$, and the $n_\lambda \times 1$ test light spectral power distribution, $\mathbf{t}$.

interpolated from experimental measurements at many sample wavelengths. The curve is called the *scotopic sensitivity function*.

Once we measure the system matrix, $\mathbf{R}$, we can predict whether any pair of lights will match under scotopic conditions. Figure 1.6 shows how we use the system matrix to predict the intensity of a primary light needed to match a test light. Suppose we have two test lights, $\mathbf{t}$ and $\mathbf{t}'$. Two lights will match when they are matched by the same intensity of the primary light. So, these two lights will match when $\mathbf{R}\mathbf{t} = \mathbf{R}\mathbf{t}'$.

Uniqueness

The hypothetical experiment illustrated in Figure 1.5 assumed a 510 nm primary light. Suppose that we perform the scotopic matching experiment using a different primary light. How will this effect the system matrix, $\mathbf{R}$?

We can answer this question by a thought experiment. Call the second primary light $\mathbf{p}'$. We can set a match between the new primary light, $\mathbf{p}'$ and the first primary light $\mathbf{p}$. We will find that there is some scalar, k, such that $k\mathbf{p}'$ matches $\mathbf{p}$, and we expect that whenever $a\mathbf{p}$ matches a test light, $\mathbf{t}$, then $a[k\mathbf{p}']$ will match $\mathbf{t}$. In particular, since $R_i\mathbf{p}$ matches the ith monochromatic test light, we expect that $R_i k\mathbf{p}'$ will match the ith monochromatic test light as well. It follows that the entries of the new system matrix will be kR_i, equal to the original except for a constant scale factor, k. Hence, the new system matrix will be $k\mathbf{R}$, and we say that the estimate of $\mathbf{R}$ is unique up to an unknown scale factor.
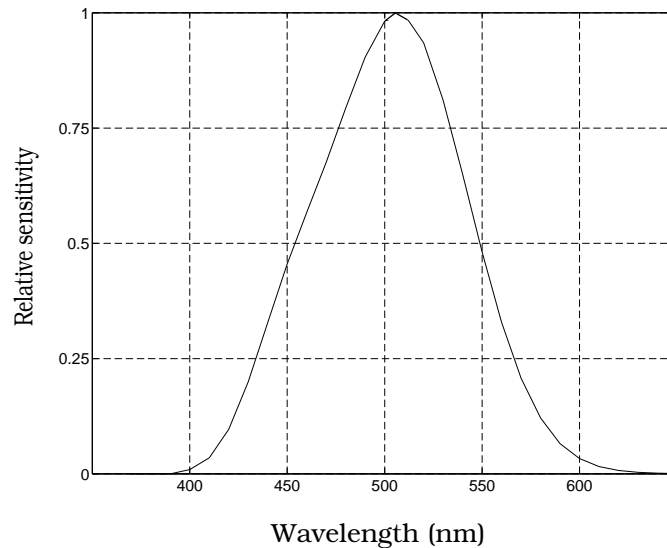
Figure 1.7: *The scotopic spectral sensitivity function* defines the human wavelength sensitivity under scotopic viewing conditions. The curve is a plot of the entries of the scotopic system matrix.

The Biological Basis of Scotopic Matching

The scotopic matching experiment is remarkable in its simplicity. We can understand the biological basis of the experimental matches by studying the properties of the rod photopigment, *rhodopsin*. Figure 1.8 is a photograph of the photopigment contained in the rod outer segments. In part (a) of the figure the photopigment is photographed in the eye of an alligator. Because the back of the alligator's eye contains a white reflective surface, called the *tapetum*, it is possible to see the color of the rod photopigment. Cats too have a white tapetum, which is why cats eyes appear to glow so brightly when they catch the beam of a car's headlights. The alligator shown in the picture had been kept in a dark closet for 24 hours so that the photopigment would be fully regenerated and easy to photograph. The closet was opened briefly, a flash picture taken, and then I suppose the door was shut again. Whew.

Rod photopigment is present in much higher density than any of the cone photopigments. Thus, researchers have been able to isolate and extract the rod photopigment for fifty years, whereas the cone photopigments have only become available recently through the methods of genetic engineering (Merbs and Nathans, 1992). Characteristically, when the rod photopigment is exposed to light, it undergoes a series of rapid changes in chemical state (Hubbard and Wald, 1951; Wald and Brown, 1956; Wald, 1968). Whenever a quantum of light is absorbed by
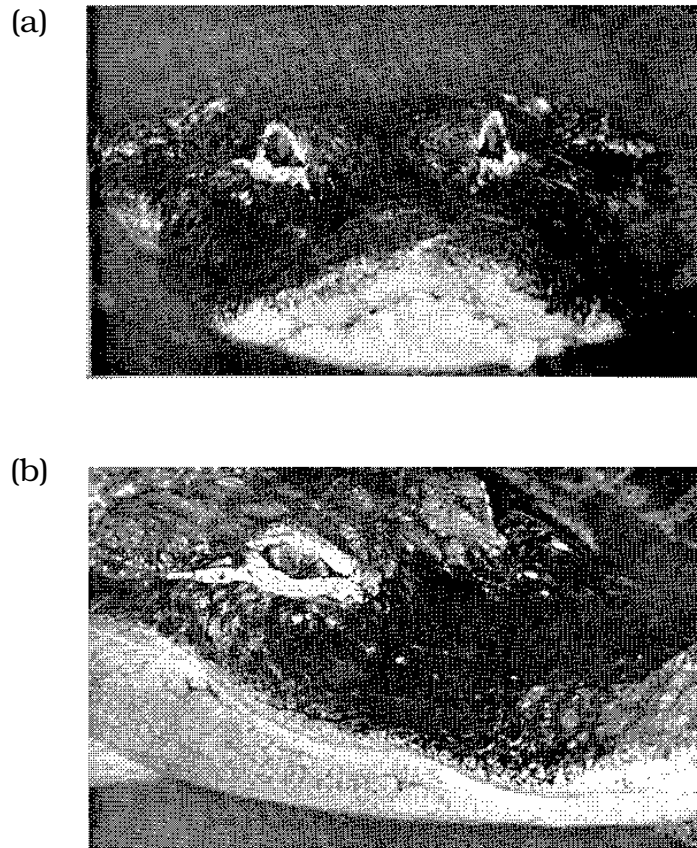
(a)

(b)

Figure 1.8: *Rhodopsin in the rod photoreceptors* of an alligator eye in the dark-adapted and light-adapted states. (a) After being in the dark, the unbleached rhodopsin is visible as a purple substance. (b) After exposure to light, the rhodopsin is bleached and we see only the white, reflective tapetum. (From Rushton, 1962).

the rhodopsin photopigment, it undergoes a specific sequence of events resulting in the decomposition of the rhodopsin molecule into opsin and vitamin A. Figure 1.8b shows the same alligator after its eye has been exposed to light. The rhodopsin is been broken into two parts and the resulting products are clear, rather than purple. In this state, the white tapetum of the eye is evident. It is the wavelength selectivity of the rhodopsin photopigment that provides the biological basis of scotopic matching. The relationship between the behavioral experiment and the properties of the rod photopigment is based on an important property called *univariance*.

W. Rushton emphasized that when a photopigment molecule absorbs light, the effect upon the photopigment is the same no matter what the wavelength of the absorbed light might be. Thus, even though quanta at 400 nm possess more energy than quanta at 700 nm, the sequence of rhodopsin reactions to absorption of a 400 nm quantum is the same as the sequence of reactions to a 700 nm quantum. Rushton used the word *univariance* for this principle to remind us that a single photopigment makes only a single-variable response to the incoming light. The photopigment maps all spectral lights into a single-variable output, the *rate of absorptions*. The response of a single photopigment does not encode any information about the relative spectral composition of the light. This explains why we cannot discriminate between lights with different spectral power distributions under scotopic viewing conditions (Rushton, 1965; Naka and Rushton, 1966).

Univariance does not mean, however, that the photopigment responds equally well to all spectral lights. The photopigment is much more likely to absorb some wavelengths of light than others. Univariance asserts that once absorbed, however, all quanta have same visual effect.

We can measure the probability of absorption using the experimental apparatus shown in Figure 1.9. We place a thin layer of photopigment on a clear plate of glass. We create a monochromatic light by passing the light from an ordinary source through a *monochromator*. The monochromator can be constructed using prisms or diffraction gratings to separate the incident light into its separate wavelengths, much as in Newton's original experiments. We measure the amount of monochromatic light passed through the photopigment and the glass plate by means of a photodetector at the rear of the apparatus. We then move the glass plate upwards, to remove the photopigment from the light path, and measure again. The difference in the photodetector signal measured in these two conditions is proportional to the amount of light absorbed by the photopigment.

If only a thin layer of photopigment is present, the experimental measurements of the absorptions will satisfy homogeneity and superposition. To test homogeneity, we increase the intensity of the test light. We will find that the number of absorptions will increase proportionately over a significant range. To test
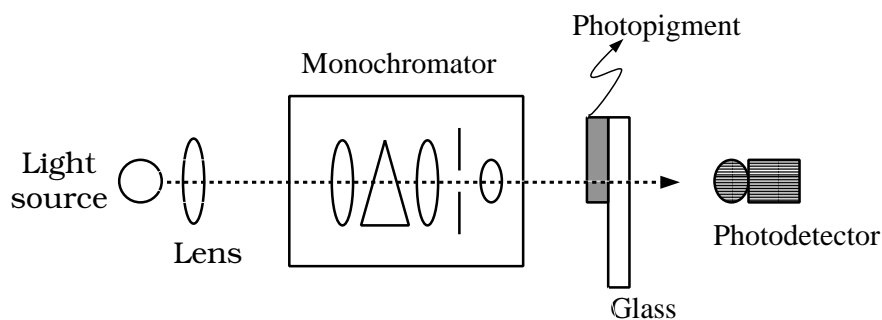
Figure 1.9: *An apparatus to measure the spectral absorption of a photopigment.* Using the monochromator, one can select light at one wavelength from the light source. To estimate the fraction of photons absorbed by the photopigment at that wavelength, we divide the number of photons detected through the glass and photopigment by the number detected after passing through the glass alone.
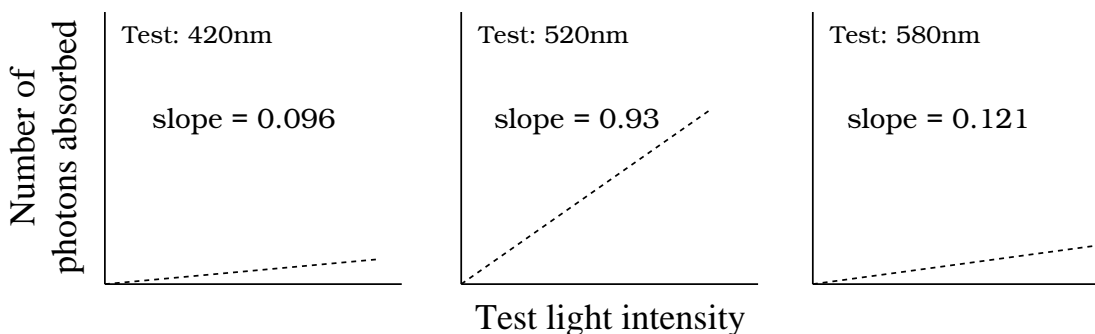
Figure 1.10: *Rhodopsin absorptions at different wavelengths.* The number of absorptions in a thin layer of photopigment are proportional to the intensity of the input light and thus satisfy the principle of homogeneity. The slope of the linear relationship between the light intensity and the number of absorptions describes the fraction of photon absorptions. The slope varies with the wavelength of the test light, thus defining the photopigment wavelength sensitivity.

Figure 1.11: *Comparisons of scotopic matching and rhodopsin wavelength sensitivity.* The filled circles show human rhodopsin absorption measured as in Figure 1.9. The open circles show human scotopic sensitivity, corrected for light loss at the lens and optical media. (Source: Wald and Brown, 1958)

superposition, we measure the photopigment absorptions to a test light $\mathbf{t}$ to be a, and the number of absorptions to a second light $\mathbf{t}'$ to be a'. When we superimpose the two input lights, we will measure $a + a'$ absorpt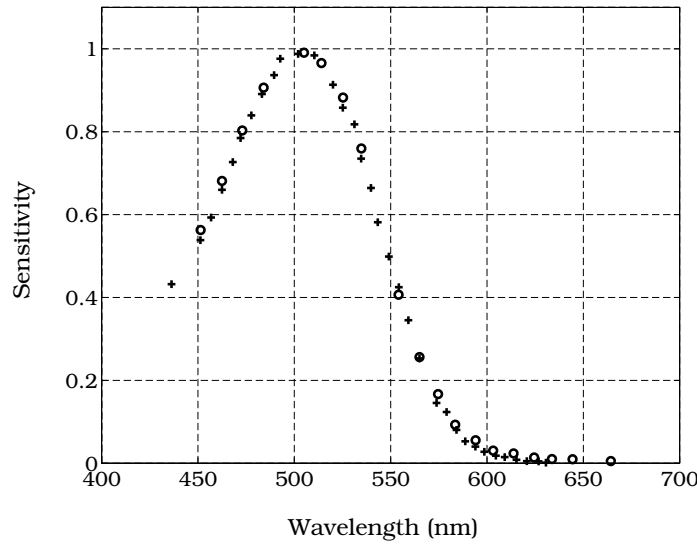ions. Since the measurement process is linear, we can estimate the system matrix of this absorption process, $\mathbf{A}$, just as we measured the system matrix, of the scotopic matching experiment, $\mathbf{R}$.

We can predict the matches in the scotopic matching experiment from the absorptions of the rhodopsin photopigment. A test and primary light match in the scotopic matching experiment when the two lights create the same number of absorptions in the rhodopsin photopigment. We can demonstrate this by comparing the system matrices of the scotopic matching experiment and the rhodopsin absorption experiment. After we correct for the effects of the wavelength sensitive elements of the eye, mainly the lens, we can plot the system matrices of the scotopic matching experiment $\mathbf{R}$, and the rhodopsin absorption experiment, $\mathbf{A}$, on the same graph. Wald and Brown (1958) made this comparison in the graph in shown Figure 1.11. The filled circles in the graph plot the the measurements of the system matrix from the rhodopsin absorption experiment, $\mathbf{A}$. The completely open circles plot estimates of the entries of $\mathbf{R}$ after correcting for the fact that the lens absorbs a significant amount of light in the short-wavelength part of the spectrum.

The agreement between the measurements of the rhodopsin photopigment and the scotopic matching experiment confirm a simple model of the observer's behavior.

Under scotopic viewing conditions the observer's perception of the two halves of the bipartite field depends on a signal initiated by the rod photopigment absorptions. The two sides of the field appear identical when the rhodopsin absorption rates on the two sides of the bipartite field are equal. During the experiment, then, the observer adjusts the intensity of the matching light to equalize the rod absorption rates on the two sides of the bipartite field. Since the absorption of the light is a linear process, the observer's behavior is linear, too.

The precise quantitative match between the scotopic matches and the rod photopigment make a very strong connection between performance and biological encoding of the light. This type of precise quantitative relationship between behavior and the biological encoding of light serves as a good standard to use when we consider the relationship between behavior and biology in other conditions.

1.3   Photopic Wavelength Encoding

When the cones initiate vision, under photopic conditions, we do encode some information about the relative spectral power distribution of the incident light. Changes in the relative spectral power distributions result in changes of the color appearance of the light. Several of the important properties of color appearance can be traced to the way cone photoreceptors encode the relative spectral power distribution of light[1].

We will relate the human ability to discriminate lights to the properties of the cones just as we did with the rods. First, we will review the matching experiments that characterize how well people can discriminate between lights with spectral power distributions. When we measure under photopic conditions, the experiment is called the *color-matching* experiment. The color-matching experiment is the foundation of color science and of direct significance to many color applications (see the Appendix). Second, we will relate the properties of the color-matching experiment to the properties of the cone photopigments. The analysis of photopic wavelength encoding parallels the analysis of scotopic wavelength encoding. The main differences are that (a) we must keep track of the photopigment absorptions in three cone photopigments rather than the single rod photopigment, and (b) until quite recently the cone photopigments were not present in sufficient quantity to define their properties with any certainty (Merbs and Nathans, 1992). Hence, the problem of relating color-matching and the cone photopigments was solved using other indirect biological measurements. We can learn a great deal from studying the logic

---

[1]N.B. Color appearance is not a simple consequence of the spectral power distribution of the incident light. We will discuss color appearance broadly in Chapter **??**.
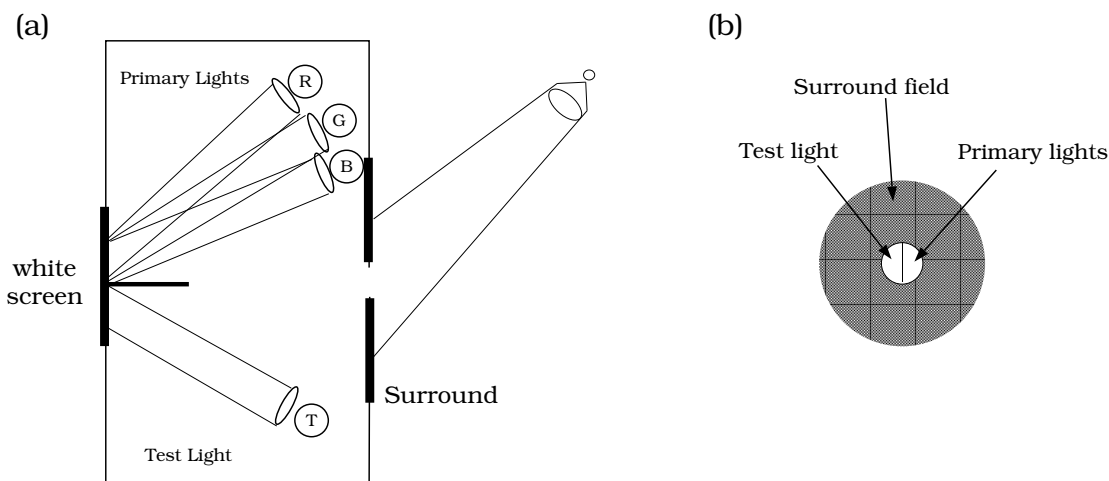
(a) (b)

Figure 1.12: *The color-matching experiment.* The observer views a bipartite field and adjusts the intensities of the three primary lights to match the appearance of the test light. (a) A side-view of the experimental apparatus. (b) The appearance of the stimuli to the observer. (After Judd and Wyszecki, 1975.)

of these methods.

Figure 1.12 shows a simple apparatus that can be used to perform the color-matching experiment. The observer views a bipartite visual field with the test light on one side. The test light may have any spectral power distribution. The second half of the bipartite field contains a mixture of *three* primary lights. Throughout the experiment, the relative spectral power distribution of each primary light is constant; only the absolute level of the primary lights can be adjusted. The observer's task is to adjust the intensities of the three primary lights so that the two sides of the bipartite field appear identical.

When the observer has completed setting an appearance match, the lights on the two sides of the bipartite field are not physically the same. The test light can have any spectral power distribution, while the mixture of primaries can only have a limited number of spectral power distributions determined by the possible weighted sums of the three primary light spectral power distributions. Lights that are photopic appearance matches, but that are physically different, are called *metamers*. Figure 1.13 contains a pair of spectral power distributions that match visually but differ physically, i.e. a pair of metamers.

The metamers in Figure 1.13 illustrate that even under photopic viewing conditions we fail to discriminate between very different spectral power distributions. To understand the behavioral aspects of photopic wavelength encoding, we must try to predict which spectral power distributions we can discriminate. The first question
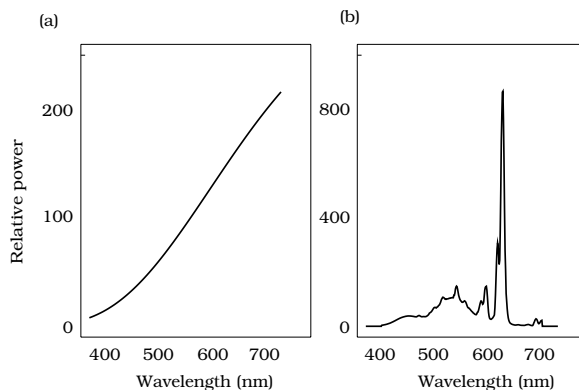
(a)   (b)

Figure 1.13: *Metameric lights.* Two lights with these spectral power distributions appear identical to most observers and are called *metamers.* The curve in part (a) is an approximation to the spectral power distribution of the sun. The curve in part (b) is the spectral power distribution of a light emitted from a conventional television monitor whose three phosphor intensities were set to match the light in (a) in appearance.

we ask is whether we can predict performance in the photopic color-matching experiment using linear systems methods.

We can define the experimental measurements in the color-matching experiment in direct analogy with the definitions we used in the scotopic matching experiment. The input variable in the color-matching experiment is the light t, just as in scotopic matching. In the color-matching experiment, however, the subject's responses consist of three numbers, not just one. So, we record the responses using a three-dimensional vector, e. The entries of e are the intensities of the three primary lights (e_1, e_2, e_3). To test superposition in the color-matching experiment we follow the logic illustrated in Figure 1.14. We obtain a match to a t by adjusting the primary intensities to the levels in e. We then obtain a match to t' by adjusting the three primary intensities to e'. We test additivity by verifying that the match to $t + t'$; is $e + e'$. Photopic color-matching satisfies homogeneity and superposition. We honor the person who first understood the importance of superposition in color-matching by calling this empirical property *Grassmann's additivity law.*

Because the color-matching experiment linearly maps the physical stimulus t to the primary intensities, e, there must be a system matrix that maps the input vector t to the output vector e. Figure 1.15 shows the input-output relationship for the photopic color-matching experiment in matrix tableau. We will call the $3 \times n_\lambda$ system matrix C.

We can estimate the system matrix C from the color-matches in the same way as we estimated the scotopic system matrix: by setting matches to a collection of monochromatic test lights with unit intensity. Since the vector representing a

Test   Primaries

(a)

matches   +   +   **t** matches **e**

(b)

matches   +   +   **t'** matches **e'**

(c)

                        t + t' matches **e + e'**
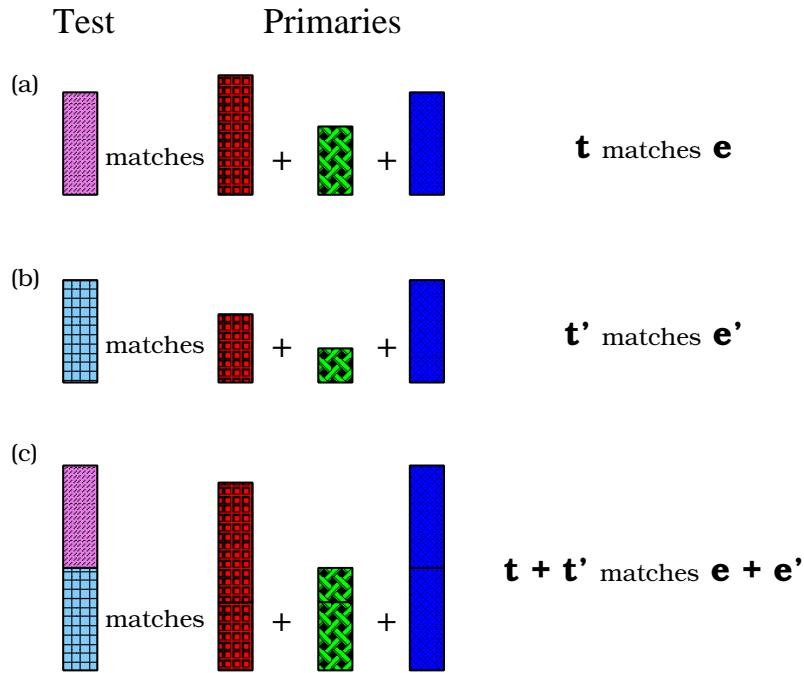
matches   +   +

Figure 1.14: *The color-matching experiment satisfies the principle of superposition.* In parts (a) and (b) test lights are matched by a mixture of three primary lights. In part (c) the sum of the test lights is matched by the additive mixture of the primaries, demonstrating superposition.

$$\text{Three primary intensities} \quad \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} = \begin{pmatrix} \text{Color-matching functions} \end{pmatrix} \begin{pmatrix} \text{Test spectral power distribution} \end{pmatrix}$$

$$\mathbf{e} = \mathbf{Ct}$$

Figure 1.15: *Matrix tableau of color-matching.* The photopic color-matching experiment defines a linear mapping from the test light spectral power distribution to the intensity of the three primary lights. The rows of the $3 \times n_\lambda$ system matrix are called the *color-matching functions*. These functions can be estimated by setting matches to many different test lights and solving a set of linear equations.

monochromatic test light is zero at each entry but one, the product of the system matrix and the monochromatic test light vector equals a single column of the system matrix. Thus, by matching a series of unit intensity monochromatic lights, we can define each of the columns of the system matrix, $\mathbf{C}$.

It is also useful to think of the system matrix in terms of its rows, which are called the *color-matching functions*. Each row of the matrix defines the intensity of a single primary light that was set to match the monochromatic test lights. We will relate the rows of the photopic system matrix to the properties of the cone photopigments just as we related the single row of the scotopic system matrix to the rhodopsin photopigment. However, to make the connection between the cone photopigments and the color-matching functions will require a little more work.

Measurements of the Color-Matching Functions

Two important caveats arise when we measure the color-matching functions. These are only a minor theoretical nuisance, but they have important implications for the laboratory experiment and for practical applications.

The first issue concerns the selection primary lights. We should chose lights that are visually *independent*: that is, no additive mixture of two of the primary lights should be a visual match to the third primary. This is an obvious but important constraint: it would be unreasonable to choose the second primary light that looked the same as the first except for an intensity scale factor. This choice would be foolish since we could always replace the second light by an intensity-scaled version of the first primary light, adding nothing to the range of visual matches we can obtain. Similarly, a primary that can be matched by a mixture of the first two adds nothing. We must choose our primary lights so that they are independent of one another.

Even among collections of primary lights that are independent, some are more convenient than others. Empirically, it turns out that no matter which primary lights we choose, there will always be some test lights that cannot be matched by an additive mixture of the three primaries. To match these test lights, we must move one or even two of the primary lights from the matching side of the bipartite field to the test side of the bipartite field. Thus, ordinarily we obtain a visual matches of the form

$$\mathbf{t} = e_1\mathbf{p}_1 + e_2\mathbf{p}_2 + e_3\mathbf{p}_3. \tag{1.2}$$

Shifting one of the primaries to the other side of the bipartite field means that our match has the form
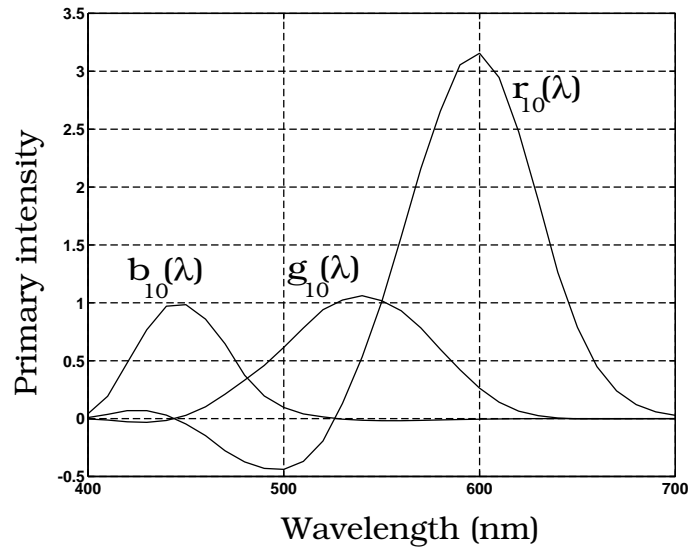
Figure 1.16: *The color-matching functions are the rows of the color-matching system matrix. The functions measured by Stiles and Burch (1959) using a $10°$ bipartite field and primary lights at the wavelengths 645.2 nm, 526.3 nm, and 444.4 nm with unit radiant power are shown. The three functions in this figure are called $\bar{r}_{10}(\lambda)$, $\bar{g}_{10}(\lambda)$, and $\bar{b}_{10}(\lambda)$.*

$$\mathbf{t} + e_1 \mathbf{p}_1 = e_2 \mathbf{p}_2 + e_3 \mathbf{p}_3. \tag{1.3}$$

To a mathematician Equation 1.3, is the same as

$$\mathbf{t} = -e_1 \mathbf{p}_1 + e_2 \mathbf{p}_2 + e_3 \mathbf{p}_3. \tag{1.4}$$

Hence, when we encode the intensity of the primary light that has been shifted to the other side of the test field we denote the match using a negative intensity value[2].

Figure 1.3 plots color-matching functions measured by Stiles and Burch (1959) using three monochromatic primary lights at 645.2 nm, 525.3 nm and 444.4 nm. Each function describes the intensity of one of the primary lights used to match various monochromatic test lights. Notice that the intensity of the red primary, at 700nm, is negative over a region of test light wavelengths, indicating that over this range of test lights the 700 nm primary light was added to the test field.

---

[2]Changing the sign of the primary intensity is a trivial matter for the theorist. It is a nuisance in the laboratory, however, and usually impossible in technological applications such as color displays. Thus, the issue of selecting primaries to minimize the number of times we must make this adjustment is of great practical interest.

The color-matching functions are extremely important in color technology, such as creating images on color monitors and color printers. I review the application of these methods to color monitors in the Appendix.

Uniqueness of the Color-Matching Functions

Suppose two research groups measure the color-matching functions using different sets of primary lights. One group measures the color-matching functions using three primary lights $\mathbf{p}_i$, while the second group uses a different set of primary lights, $\mathbf{p}_i'$. Because the groups use different primary lights, they will measure different sets of color-matching functions, $\mathbf{C}$ and $\mathbf{C}'$. How will the two sets of color matching functions be related?

We can answer this question by the following thought experiment. First, create a matrix whose columns contain the spectral power distributions of the first group's primary lights, and call this matrix $\mathbf{P}$. The spectral power distribution of a mixture of the primaries, with primary intensities $\mathbf{e}$ is the weighted sum of the columns. We can express this mixture using the matrix product $\mathbf{Pe}$. Now, we can use the color-matching functions to predict when a test light light will match the mixture of three primaries. The test and primaries will match when

$$\mathbf{Ct} = \mathbf{CPe}. \tag{1.5}$$

Suppose the second group of researchers can also established matches to this test light using their primaries. To describe their measurements, we create a second matrix whose columns contain the spectral power distributions of the second group's primary lights, $\mathbf{P}'$. Call the primary intensities used to match the test with the second primaries is $\mathbf{e}'$. Since the light $\mathbf{P}'\mathbf{e}'$ is a visual matched to the test light, we know that

$$\mathbf{Ct} = \mathbf{CP}'\mathbf{e}' \tag{1.6}$$

By combining Equations 1.5 and 1.6, we find that the two vectors of primary intensities, $\mathbf{e}$ and $\mathbf{e}'$, are related by a linear transformation,

$$\mathbf{e} = (\mathbf{CP})^{-1}\mathbf{CP}'\mathbf{e}'.$$

Finally, recall that the vectors in the columns of the color-matching functions are the primary intensity settings necessary to match the monochromatic lights. We have just shown that the primary intensities used to make matches with the two different sets of primaries are related by a linear transformation. Hence, the system matrices containing the color-matching functions $\mathbf{C}$ and $\mathbf{C}'$ must be related a linear transformation.

With a little more algebra, one can show that the color-matching functions are related by the following linear transformation:

$$\mathbf{C} = (\mathbf{CP'})\mathbf{C'}. \tag{1.7}$$

The 3×3 matrix relating the two sets of color-matching functions, $\mathbf{CP'}$, has a simple empirical interpretation; its columns contain the intensities of the new primaries needed to match the original primaries. To see this, remember that each column of $\mathbf{P'}$ is the spectral power distribution of one of the primary lights, $\mathbf{p}_i'$. Thus, the first column of $\mathbf{CP'}$ is the vector of intensities of the first group of primaries needed to match $\mathbf{p}_1'$. Similarly the second and third columns of $\mathbf{CP'}$ contain the intensities of the first group of primaries needed to match the corresponding primaries in $\mathbf{p'}$. The matrix $\mathbf{CP'}$ contains three columns equal to the primary intensities of $\mathbf{p}_i$ needed to match the new primary lights, $\mathbf{p}_i'$.

The photopic color-matching functions are not unique; when we measure using different sets of primaries we will obtain different color-matching functions. But, the color-matching functions are not completely free to vary either, since different pairs of color-matching functions will always be related by a linear transformation. We say that the color-matching functions are unique up to a free linear transformation.

A Standard Set of Color-Matching Functions

When the members of the Committe Internationale d'Eclairage (CIE; an international standards organization) met in 1931, they were fully aware that the color-matching functions were not unique. To facilitate communication about color, the CIE defined a standard system of color representation based on one particular set of color-matching functions, that everyone should uses. This set of color-matching functions defines the *XYZ tristimulus coordinate system*. The color-matching functions in this system are called $\bar{x}(\lambda), \bar{y}(\lambda)$ and $\bar{z}(\lambda)$ respectively. They define one of the many possible system matrices of the color-matching experiment. Figure 1.17 shows the three standard color-matching functions, $\bar{x}(\lambda), \bar{y}(\lambda)$ and $\bar{z}(\lambda)$.

The standard color-matching functions were chosen for several reasons. First, $\bar{y}(\lambda)$ is a rough approximation to the brightness of monochromatic lights of equal size and duration. A second important reason is that the functions are non-negative which simplified some aspects of the design of instruments to measure the tristimulus coordinates. But, as with any standards decision, there are some irritating aspects of the XYZ color-matching functions as well. One serious drawback is that there is no set of physically realizable primary lights that by direct measurement will yield the color-matching functions. Primary lights that would yield these functions would

Figure 1.17: *The XYZ standard color-matching functions.* In 1931 the CIE standardized
a set of color-matching functions for image interchange. These color-matching func-
tions are called $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$. Industrial applications commonly describe the
color properties of a light source using the three primary intensities needed to match
the light source that can be computed from the XYZ color-matching functions.

have to have negative energy at some wavelengths and cannot be instrumented.
Another problem is that these early estimates have been improved upon.
Specifically, D. B. Judd (1951) noted that the functions are inaccurate in the
short-wavelength region and he proposed a modified set of functions that are often
used by scientists, although they have not displaced the industrial standard. Also,
and perhaps most significantly, there is very little that is intuitive about the XYZ
color-matching functions. Although they have served us quite well as a technical
standard, and are understood by the mandarins of our discipline, they have served
us quite poorly in explaining the discipline to new students and colleagues.

The Biological Basis of Photopic Color-matching

Just as we can explain the scotopic color-matching experiment in terms of the light
absorption properties of the rhodopsin photopigment, we also would like to explain
the photopic color-matching experiment in terms of the light absorption properties
of the cone photopigments. We related the photopigments and the behavior by
studying the system matrices of the two experiments. We found that two lights were
scotopic matches when $\mathbf{Rt} = \mathbf{Rt}'$, and we then showed that the entries in the $1 \times n_\lambda$
scotopic matching matrix, $\mathbf{R}$, was the same as the rhodopsin absorption function $\mathbf{A}$.

$$\text{Cone} \atop \text{absorptions} \begin{pmatrix} L \\ M \\ S \end{pmatrix} = \begin{pmatrix} \text{L cone wavelength sensitivity} \\ \text{M cone wavelength sensitivity} \\ \text{S cone wavelength sensitivity} \end{pmatrix} \begin{pmatrix} \text{Test spectral} \\ \text{power distribution} \end{pmatrix}$$

$$\mathbf{r = Bt}$$

Figure 1.18: *Cone photopigments and the color-matching functions.* If we measure the wavelength sensitivity of each of the cone photopigments, we can create $3 \times n_\lambda$ system matrix to describe the cone absorptions. Then, we can evaluate whether the cone absorption system matrix can serve to explain the results of the color-matching experiment.

For photopic vision, we use the same general approach. But, there are two complications: there are three cone photopigments, not just one; the photopic matching matrix is not unique.

Extending our analysis to account for three cone photopigments instead of one rod photopigment is straightforward. We measure the absorption properties of each of the three cone photopigments, and we create a system matrix, **B**, with three rows to define the three cone photopigment absorption functions. This matrix generalizes the rhodopsin system matrix **A**. Then, we compare the cone absorption system matrix, **B**, with the color-matching experiment system matrix, **C**. We must evaluate whether the cone photopigment matrix can explain the color-matching data.

From our analysis of color-matching, we know that the color-matching system matrix is not unique; there is a collection of equivalent system matrices, all related by a linear transformation. Hence, to evaluate whether the cone absorption matrix can explain the color-matching experiment, we must evaluate whether the color-matching system matrix, **C**, is related to **B** by a linear transformation. Our next task, then, is to measure the cone absorption system matrix, **B**.

Measuring Cone Photocurrents

Currently, the best estimate of the cone photopigment absorptions is derived from measurements of the cone photocurrent, that is the change in the current flow through the membrane of individual cones as they are stimulated by light. Relating
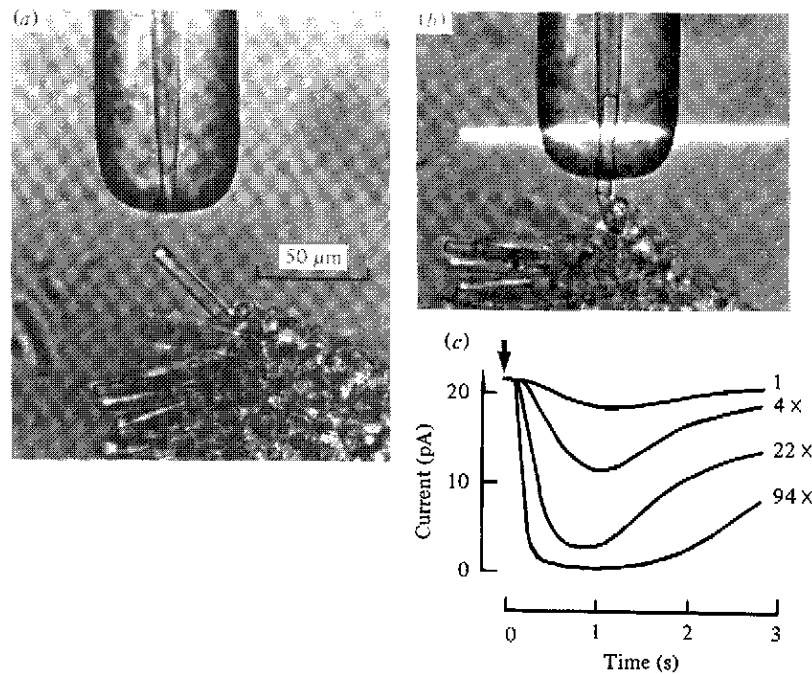
Figure 1.19: *Measurement instruments for detecting the photocurrent generated by a toad photoreceptor.* The image shows a portion of the retina suspended in solution. A single photoreceptor from this retinal section has been drawn into the microelectrode and is being stimulated by a beam of light passing transversely through the photoreceptor and microelectrode (Image courtesy of D. Baylor).
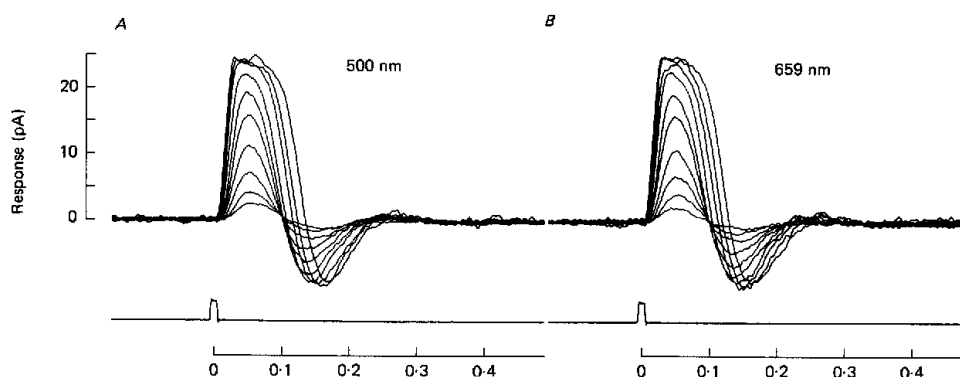
Figure 1.20: *The timecourse of cone photocurrent* in response to a brief test flash is biphasic. The amplitude of the photocurrent response increases with the stimulus intensity. The response functions are the same for different wavelengths of light, such as at 500 nm and 659 nm in parts (a) and (b), respectively. The stimulus timecourse is shown below the photocurrent plots. (Source: Baylor Nunn and Schnapf, 1987)

the photocurrent to the photopigment absorptions requires some careful analysis because the photocurrent depends nonlinearly on the photopigment absorptions in the cone. In this section we will develop new theoretical methods to interpret the nonlinear cone photocurrent measurements and infer the linear properties of the cone photopigments.

Baylor, Nunn and Schnapf (1987; Baylor, 1987) were the first to measure the cone photocurrents in the macaque retina. The macaque has three types of cones and its behavior on most color tasks is quite similar to human behavior Thus, the comparison between the properties of the macaque photoreceptors and human behavior is a good place to begin (DeValois and Jacobs, 1971).

To measure the cone photocurrent, Baylor, Nunn and Schnapf removed the retina from the eye and chopped into fine pieces about 100 μm across. The pieces are placed in a chamber containing special solutions that support the metabolism of the cells. Even though the retina has been dissected from the eye and chopped into pieces, the electrical response of the photoreceptors remain vigorous for several hours. Baylor and his colleagues recorded the photocurrent of individual cells using the experimental technique pictured in Figure 1.19. The figure shows glass micro-pipette approaching a single photoreceptor. The inner diameter of the micropipette is between 2 and 6 μm, only ten times as wide as the wavelength of visible light. A single photoreceptor outer segment is held inside the micropipette, and there is a thin ray of light passing transversely through the photoreceptor and stimulating it.

Figure 1.20 shows the result of stimulating the photoreceptor with a brief impulse of

light. The curves illustrate the membrane photocurrent following a brief light flash. The curves in part (a) of the figure plot the response to 500 nm light at a range of intensities. The curves in part (b) plot the response to 659 nm light at a range of intensities. Before the stimulus is presented, there is a steady inward flow of current consisting of a stream of positively charged sodium ions entering the photoreceptor through ion channels in the cell membrane. This steady level in the absence of light is called the *dark current*. It represents a baseline level and is denoted as zero in the graph. The plotted values are *biphasic*, varying both above and below the baseline.

When the photopigment absorbs light from the flash, the inward flow of sodium ions is slowed. The sodium current in darkness reduces the negative electrical polarization of the cell interior. When light blocks the inward flow, the negative voltage difference between the inside and outside of the cell increases. Thus, the initial photoreceptor response to light is a *hyperpolarization*. After the initial blockage of inward flowing sodium current, the current flow is actively restored. The mechanism that restores balance overcompensates; during the second phase of the response the total photocurrent flow reverses direction. Thus the photocurrent response first flows in one direction and then the opposite direction, leading to the biphasic impulse response.

In this experiment the test light is the input, t, and the cone photocurrent response is the output. We can evaluate whether the input-output relationship satisfies one of the requirements of a linear system, homogeneity, from the graphs in Figure 1.20. Suppose the input signal is t and the photocurrent response is i, aa vector representing the photocurrent as a function of time following the stimulus. To test homogeneity we should measure the response to the scaled the input, kt. If the system is linear, then we expect that the photocurrent response will be ki. From a visual inspection of the curves in Figure 1.20 we can see that homogeneity fails. There are two features of the curves that should make this evident to you. First, notice that as the test intensity increases, the peak deviation reaches a maximum of about 25 pA and then saturates. Saturation is inconsistent with a strictly linear relationship between input intensity and output photocurrent. A second way to see that linearity fails is to consider the point of the biphasic response at which the output crosses the zero level at baseline. If the output photocurrent is proportional to the input intensity, points with a zero response level should always have a zero response level: multiplying zero by any intensity still yields zero. Hence, we expect that the zero-crossing should not change its position as we increase the test intensity. This prediction is true for lower test intensities, but as the input intensity increases to fairly high levels, the zero-crossing shifts its position in time.

How surprising: Human performance in the color-matching experiment satisfies the principles of a linear system, homogeneity and superposition, yet the cone
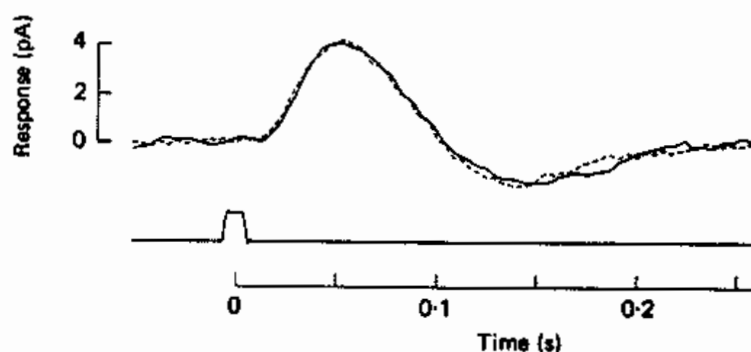
Figure 1.21: *The principle of univariance* states that a photon absorption leads to the same neural response, no matter what the wavelength of the photon. The principle predicts that two stimuli at different wavelengths can be adjusted to equate the photocurrent response throughout its timecourse. This is shown here as the match between photocurrents in response to 550 nm (dashed) and 659 nm (solid) test lights set to a nine to one intensity ratio (Source: Baylor et al., 1987).

photocurrent responses a part of the chain of biological events that mediate the behavior, fail the simplest tests of linearity. How can the behavior be linear when the components mediating the behavior are nonlinear? We will answer this question in the following section. The answer is given specifically for color-matching, but the principles we will review are quite general. They will be helpful again when we consider the relationship between behavior and other neural responses throughout this book.

Static Nonlinearities: Photocurrents and Photopigments

By comparing the sets of photocurrent responses on the top and bottom of Figure 1.20, it appears that as we vary the level of the test signal we sweep out the same set of curves. The similarity of the measured photocurrent responses to the two test lights suggests that we can perform a color-matching experiment at the level of the photocurrent response. We can perform such an experiment by choosing a test light and a primary light and adjusting the intensity of the primary light light until the photocurrent responses of the test and primary are the same. The curves in Figure 1.21 show one example of such a match using a primary light at 500 nm and a test light at 659 nm.

The physiological preparation is very delicate and it is difficult to keep the photoreceptors alive and functioning. This makes it impossible to set full photocurrent matches for arbitrary test and primary combinations. But, it is possible
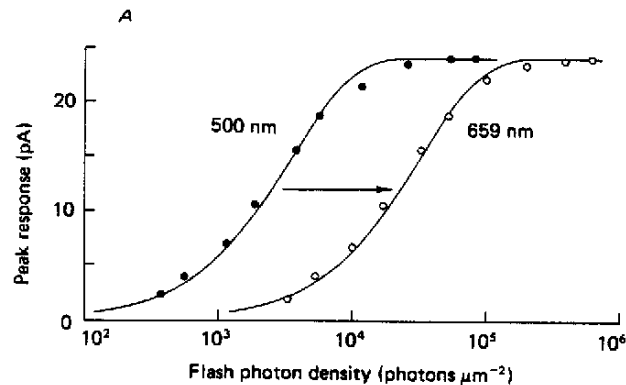
Figure 1.22: *A matching experiment using the cone photocurrent as response.* Lights at different wavelengths have equivalent effects on the cone photocurrent when the light intensity ratio is set properly. For this cone, the 659 nm light must be nine times more intense than the 500 nm light to have an equivalent effect (Source: Baylor et al., 1987).

to carry out an efficient approximation of the full experiment. The two curves in Figure 1.22 summarize the photocurrent responses to a 659 nm test light and the 500nm primary light at a series of different intensity levels. The data points shows the peak value of the photocurrent response as a function of intensity; the peak value summarizes the full photocurrent timecourse. The smooth curves drawn through the points interpolate the peak response at any intensity level. From these interpolated measurements, we can estimate the intensity levels needed to obtain complete matches between the test and primary lights.

If the matching experiment performed at the level of the photocurrent satisfies homogeneity, the intensity of the test and primary lights that match should be proportional to one another. We can estimate the intensity of the test and primary lights that match at different response levels by drawing a horizontal line across the graph and noting the intensity levels that produce the same peak photocurrent. The curves through the two sets of data points in Figure 1.22 are parallel on a logarithmic intensity axis, so that the intensities of pairs of points that match are separated by a constant amount. Since the axis is logarithmic, equal separation implies that when the photocurrents match the test and primary light intensities are in a particular ratio, precisely as required by homogeneity. Hence, the photocurrent matching experiment satisfies homogeneity even though the photocurrent response itself is nonlinear.

From the separation between the two curves, we see that more 659 nm photons are needed than 500 nm photons to evoke the same response. For this pair of wavelengths the curves are separated by 0.97 log units that corresponds to a ratio of 9.3. It takes 9.3 times as many 659 nm quanta to equal the photocurrent produced by
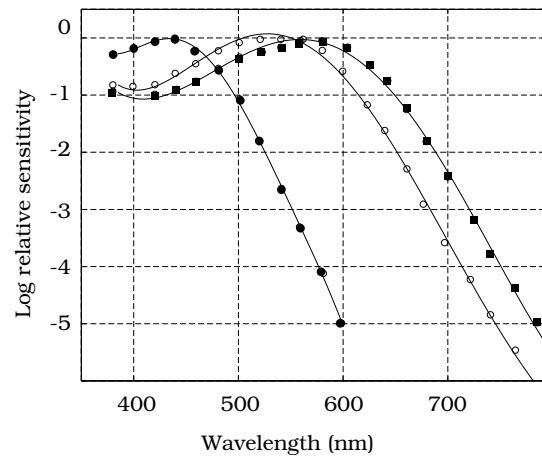
Figure 1.23: *Cone photocurrent spectral responsivities.* The measurement range spans a factor of one million. The S cone sensitivity at short wavelengths is high compared to behavioral measurements because in behavioral conditions the lens absorbs short wavelength light strongly. (After Baylor, 1987).

a number of 500 nm quanta. By repeating this experiment using test lights at many different wavelengths, we can estimate the complete spectral responsivity curves for the cone photoreceptors. From a collection of such measurements we can estimate the wavelength sensitivity of the cone receptor. The wavelength sensitivity is due to the properties of the cone photopigment, so in this way we can derive the cone photopigment absorption function from the photocurrent measurements.

The reader will not be surprised to learn that Baylor, Nunn and Schnapf found cones with three distinct spectral response functions: these measurements are plotted in Figure 1.23. Notice that the vertical axis spans six logarithmic units, so that they measured sensitivities varying over a factor of one million, an extraordinary technical measurement achievement.

Static Nonlinearities: The principle

We can analyze the photopigment sensitivity from the photocurrent response because the nonlinear relationship between the test light and the photocurrent signal is very simple: The photons are absorbed by a linear process; the linear encoding is followed by a nonlinear process that converts the photopigment absorption rate into membrane photocurrent. The properties of the nonlinear process are independent of the linear encoding step, and thus we call this process a *static nonlinearity*. When a system is a linear process followed by a static nonlinearity, we can characterize the

system performance completely.

It is worth spending a little time thinking about why we can characterize this type of nonlinear system. First, consider the linear process of photopigment absorption. There is a photopigment system matrix, say, $\mathbf{A}$, that maps the test light into a photon absorption rate, $\mathbf{At}$. Second, the static nonlinearity converts the photopigment absorption rate into a peak photocurrent response. Together, these two processes define the nonlinear system response, $F(\mathbf{At})$.

When we set a match between the peak photocurrent from the test light and the primary light, we establish an equation of the form

$$F(\mathbf{At}) = F(a\mathbf{Ap}), \tag{1.8}$$

where a is the intensity of the primary light needed to match the test light. Since the nonlinear function F is monotonic, we can remove it from both sides of the Equation and write

$$\mathbf{At} = a\mathbf{Ap}. \tag{1.9}$$

From this equation we see that there is a linear relationship between the primary and test light intensities, since if $\mathbf{t}$ matches $a\mathbf{p}$, then $k\mathbf{t}$ will match $ka\mathbf{p}$. Thus, even if a system has a static nonlinearity, the system's performance in a matching experiment will satisfy the test of homogeneity. We can also show that in a matching experiment a system with a static nonlinearity will satisfy superposition.

Cone Photopigments and Color-matching

How well do the spectral sensitivity of the cone photopigments predict performance in the photopic color-matching experiment? We predict that two lights are metamers when they have the same effect on the three types of cone photopigments. To answer how well the cone photopigments predict the color-matching results, we can perform the following calculation.

Create a matrix, $\mathbf{B}$, whose three rows are the cone photopigment spectral sensitivities. We use this matrix to calculate the cone absorptions to a test light, so that $\mathbf{Bt}$, is a 3×1 vector containing the cone sensitivities to the test light. We predict that two lights $\mathbf{t}$ and $\mathbf{t}'$ will be a visual match when they have equivalent effects on the cone photopigments. Thus, two lights will be metamers when

$$\mathbf{Bt} = \mathbf{Bt}'.$$

It follows that the cone absorption matrix $\mathbf{B}$ is a system matrix for the color-matching experiment. This is precisely what we mean when we say that the
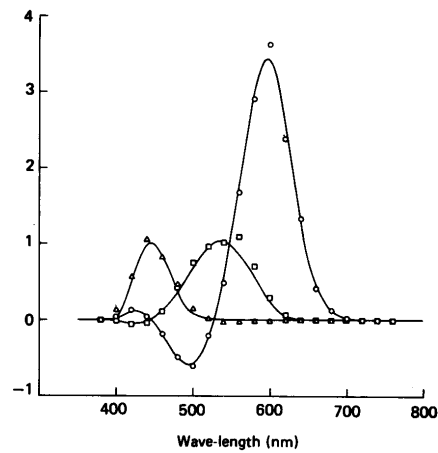
Figure 1.24: *Comparison of cone photocurrent responses and the color-matching functions.*
The cone photocurrent spectral responsivities are within a linear transformation of
the color-matching functions, after a correction has been made for the optics and inert
pigments in the eye. The smooth curves show the Stiles and Burch (1959) 10° color-
matching functions. The symbols show the matches predicted from the photocurrents
of the three monkey cones. The predictions included a correction for absorption by
the lens and other inert pigments in the eye (Source: Baylor, 1987).

cone photopigments can explain the color-matching experiment. Earlier in this
chapter, we showed that the color-matching functions are all related by a 3×3 linear
transformation. Thus, there should be a linear transformation, say $\mathbf{Q}$, that maps the
cone absorption curves to the system matrix of the color-matching experiment,
namely $\mathbf{C} = \mathbf{QB}$.

Baylor, Nunn and Schnapf made this comparison[3]. They found a linear
transformation to convert their cone photopigment measurements into the
color-matching functions. Figure 1.24 shows the color-matching functions along
with the linearly transformed cone photopigment sensitivity curves. From the
agreement between the two data sets, we can conclude that the photopigment
spectral responsivities are a satisfactory biological basis to explain the photopic
color-matching experiment.

Why this is a big deal

The methods we have used to connect cone photopigments and color-matching are a
wonderful example of how to relate physiological and behavioral data precisely. To

---

[3]After correcting for the absorptions by the lens and inert pigments in the eye.

make the connection between the behavioral and physiological data we have had to reason through some challenging issues. Still, we have obtained a close quantitative agreement between the behavioral and physiological measurements.

Notice that as we began our analysis, the properties of the neural measurements and the behavioral measurements appeared different. The linearity of the color-matching experiment contrasts with the nonlinearity of the photocurrent response. But by comparing stimuli that cause equal-performance responses, rather than comparing behavioral matches with raw photocurrent levels, we can see past the dissimilarities and understand their profound relationship. In this case, we know how to connect these two different measurements and the simplicity and clarity of the relationship is easy to see. It makes sense, then, to ask what we can learn from this successful analysis that might help us when we move on to try to relate other behavioral and biological measurements.

We should remember that the relationship between behavior and biology may not always be found at the level of the measurements that are natural within each discipline. Direct comparisons between the intensity of the primary lights and the photocurrent signals do not help us to explain the relationship, even though each measure is natural within its own experiment. To make a deep connection we needed to look at the structural properties of the experiment. When we perform the color-matching experiment, we learn about the equivalence of different stimuli. This equivalence is preserved under many transformations. Thus, we succeed at comparing the behavior and the biology when we compare the results at this level, although they seem different when we use other quantitative measures.

How do we know when we have the right set of biological and behavioral measures? There are many related physiological measures we might use to characterize the photoreceptors, and there are many variants of the behavioral color-matching experiment. For example, we could have asked the subject whether the brightness of the test light doubles when we double the intensity (the answer is no). Or we could have asked the subject to assess the change in redness or greenness. Just as the input-output relationship of the photocurrent may violate linearity for intense stimuli, so too many behavioral measures violate linearity. Finding the right measures to reveal the common properties of the two data sets is in part science and in part art. We learn about connections between these disciplines by trying to recast our experiments using different methods until the relationships become evident.

As we study the neural response in more central parts of the nervous system, you may be tempted to interpret a physiological measurement as a direct predictor of some percept. The rate at which a neuron responds or the stimulus that drives a neuron powerfully are natural biological measures. Remember, however, that there is no simple relationship between the photocurrent response and the intensity level

of a primary light. We achieved a good link between the physiological and behavioral measures by structuring a theory of the information that is preserved in each set of experimental measurements. Understanding our measurements in terms of this level of abstraction – what information is present in the signal – is a harder but better way to forge links between different disciplines. Color science has been fortunate to have workers in both disciplines who seek to forge these links. We should take advantage of their experience when we relate behavior and biology in other domains.

1.4 Color Deficiencies

I have emphasized the fact that for most observers color-matching under the standard viewing conditions requires three primary lights to form a match and we call color vision *trichromatic*. There are some viewing conditions in which only two different primary lights are necessary. Under these viewing conditions, color vision is *dichromatic*. Finally, when only a single primary is required, as under rod viewing conditions, performance is *monochromatic*.

Small field dichromacy

Perhaps the most important of case of dichromacy occurs when we reduce the size of the bipartite field used in the color-matching experiment. If the field is greatly reduced in size, from 2 degrees to only 20 minutes of visual angle, then observers no longer need three independent primary lights: two primary lights suffice. Under these circumstances, observers act as if they have only two classes of photoreceptors rather than three.

Why should observers behave as if they had only two classes of receptors when the field is very small? If this observation surprises you, go back to Chapter **??** and re-read the section on the S cone mosaic. You will find that there are very few short-wavelength cones, and there are none in the central fovea. Oddly, we encode less about the spectral properties of the incident light in the central fovea than we record just slightly peripheral to the fovea. In the central fovea, people are dichromatic.

Dichromatic observers

Some observers find that they can perform the color-matching experiment using only two primary lights throughout their entire visual field. Such observers are called *dichromats*. The vast majority of dichromats are male. By studying the family relationships of dichromats, it has been found dichromacy is a sex-linked genetic trait (Pokorny, Smith, Verriest, and Pinckers 1979). Dichromatic observers can be missing the long-wavelength photopigment (*protanopes*), the middle-wavelength photopigment (*deuteranopes*), or short-wavelength photopigment (*tritanopes*). Tritanopes are much more rare than either protanopes or deuteranopes. The difference in the probabilities arises because the gene responsible for the creation of the short-wavelength photopigment is on a different chromosome (Nathans, et al., 1992).

It is possible to use the color-matching functions measured from dichromatic observers to estimate the photoreceptor spectral responsivities. Suppose we have two dichromatic observers: the first observer has only the L and M photoreceptors, and the second observer has only the L and S photoreceptors. Since the photoreceptor sensitivities are linearly related to the color matching functions, a weighted sum of the first observer's color-matching functions will equal the L cone absorption function, and a different weighted sum of the second observer's color-matching functions will equal the L cone absorption function, too. This establishes a linear equation we can use to estimate the L cone absorption function. Similarly, from a pair of dichromats who share only the M cones, we can estimate the M cone sensitivity, and so forth.

Pseudoisochromatic Plates   For some purposes, we do not need the complete results of a color matching experiment to learn about the observer's color vision. A much simpler test for dichromacy is to have a subject examine a set of colored images called the *Ishihara Plates*. These plates were designed based on the results of the color matching experiment and can be used to identify different types of dichromats based on a few simple judgments.

Each plate consists of a colored test pattern drawn against a colored background. The test and background are both made up of circles of random sizes; the test and background are distinguished only by their colors. The color difference on each plate is invisible to one of the three classes of dichromats. Hence, when a subject fails to see the test pattern, we conclude that the subject is missing that cone class.

Farnsworth-Munsell 100 Hue test   The *Farnsworth-Munsell 100 Hue test* is also commonly used to test for dichromacy. In this test, which is much more challenging
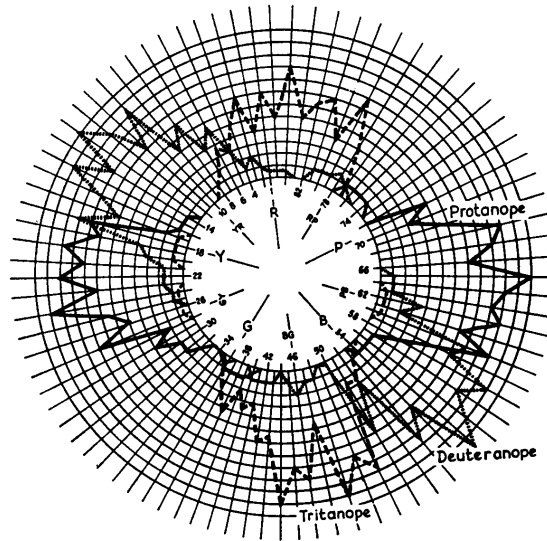
Figure 1.25: *Representing errors in the Farnsworth-Munsell 100 Hue test.* Each of the test objects, called *caps*, is assigned a position around the circle. The error score is indicated by the radial distance of the line from the center of the circle. Observers with normal color vision rarely have an error score greater than two. Errors characteristic of an observer missing the L cone photopigment (protanope), the M cone photopigment (deuteranope) and the S cone photopigment (tritanope) are shown. (Source: Kalmus, 1965).

than the Ishihara plates, the observer is presented with a collection of cylindrical objects, roughly the size of bottle caps and often called *caps*. The colors of the caps can be organized into a hue circle, from red, to orange, yellow, green, blue-green, blue, purple and back to red. Despite the name of the test, there are a total of 85 caps, each numbered according to its position around the hue circle. The color of the caps differ by roughly equal perceptual steps.

The observer's task is to take a random arrangement of the caps and to place them into order around the color circle. At the beginning of the task, four of the caps (1,23,43, and 64) are used to establish anchor points for the color circle. The subject is asked to arrange the remaining color caps "to form a continuous series of colors."

The hue steps separating the colors of the caps are fairly small; subjects with normal color vision often make mistakes. After the subject finishes sorting the caps, the experimenter computes an error for each of the 85 positions along the hue circle. The error is equal to the sum of the differences between the number on the cap and its neighbors. For example, in a correct series, the caps are ordered, say 1-2-3. In that case the difference between the cap in the middle and the one on the left is -1, and the one on the right is +1. The error score is 0 in this case. If the caps are ordered 1-3-2, the two differences are +2 and +1 and the error is 3. Normal observers do not produce an error greater than 2 or 3 at any location.

The subject's error scores are plotted at 85 positions on a circular chart as in Figure 1.25. An error score of zero plots at the innermost circle and increasing error scores plot further away from the center. Subjects missing the L cones (protanopes), M cones (deuteranopes), and S cones (tritanopes) show characteristically different error patterns that cluster along different portions of the hue circle.

Anomalous Observers

Dichromatic observers have only two types of cones. The slightly larger population of observers, who are called *anomalous*, have three types of cones and require three primaries in the color-matching experiment. The matches that they set are stable, but they are well outside of the range set by most of the population. These observers have cone photopigments that are slightly different in structure from most of the population, which is why they are called anomalous. The color-matching functions for anomalous observers are not within a linear transformation of the normal color-matching functions. This is equivalent to the experimental observation that lights that visually match for these observers do not match for normal observers, and vice versa.

Neitz, Neitz and Jacobs (1993) have argued on genetic grounds that many people

contain small amounts of the anomalous photopigments so that there are more than three cone photopigments present in the normal eye. Because the anomalous photopigments are not very different from the normal, it is hard to discern their presence in all but the most sensitive experimental tasks. They attribute the trichromatic behavior in the color-matching experiment to a neural bottleneck rather than a limit on the number of photopigment types. Since the differences between the normal and anomalous photopigments are very small, however, this hypothesis will be difficult to prove or disprove and it will have very little impact on color technologies.

The relationship between anomalous observers and normal observers parallels the relationship between color cameras and normal observers. The spectral responsivities of the color sensors in most color cameras differ from the spectral responsivity of the human cone photoreceptors. Worse yet, the camera sensors are not within a linear transformation of the cone photopigments. As a result, lights that cause the same effect on the camera, that is lights that are visual matches when measured at the camera sensors, may be discriminable to the human observer. Conversely, there will be pairs of lights that are visual matches but that cause different responses in the camera sensors. I will return to discuss this topic when we return to discuss color appearance in Chapter **??**.

1.5 Color Appearance

Color-matching provides a standard of precision to strive for when we analyze the relationship between behavior and physiology. The work in color-matching is also important because it has had an impact well beyond basic science, into engineering and technology that touch our lives.

The success of color-matching and its explanation is so impressive, that there is a tendency to believe that color-matching explains more than it does. The theory and data of photopic color-matching provide a remarkably complete explanation of when two lights will match. But, the theory is silent about what the lights look like.

Often, students who are introduced to color-matching for the first time are surprised that the words brightness, saturation and hue never enter the discussion. The logic of the color-matching experiment, and what the color-matching experiment tells us about human vision, does not speak to color appearance. What we learn from color-matching is fundamental, but it is not everything we want to know. For many purposes we want to know An understanding of color-matching is necessary for an understanding color appearance; but, it is not a solution to the problem.
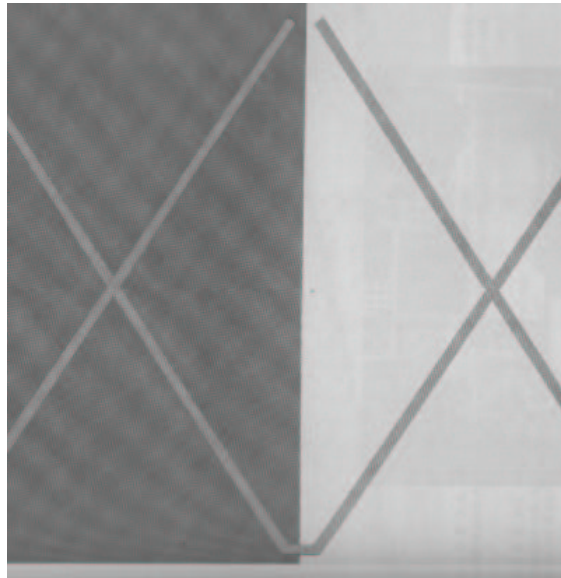
Figure 1.26: *Color-matching does not predict color appearance.* The X's are physically the same (notice where they join) and thus have the same effect on the photopigments; but, their appearance differs. The photopigment responses at a point do not determine the color appearance at that point. Appearance instead depends on the spatial structure in the image. (Source: Albers, 1975).

To emphasize the separation between color-matching and color appearance, consider the following experiment. Suppose that we form a color-match between two lights that are presented as a pair of crossing lines against one background. Such a pair is illustrated on the left hand side of Figure 1.26. On the left, the lines both appear gray. Now, move this pair of lines to a new background. Color-matching assures us that the two lights will continue to match one another as we move them about.

But we should not be assured that their appearance remains the same. For example, on the right of the figure we find that the pair of lights now have quite a different color appearance. By examining the point where the lines come together at the top of Figure 1.26, which is a painting by the artist Joseph Albers, you can see that the lines are physically identical on both sides of the painting.

Color-matching is different from color appearance. To build theories of color appearance we will need to incorporate experimental factors – such as the viewing context – that are not included in either the theory or experimental manipulations of the color-matching experiment. It is precisely because the important discoveries recounted in this chapter do not solve the problem of color appearance that the chapter is so oddly titled. We will review the topic of color in Chapter **??**.

Exercises

1. Our analysis of color encoding begins with the color-matching experiment. Make sure you can explain the highlights of this experiment.

   (a) Describe a procedure to measure the color-matching functions for an observer who is using three primary lights.

   (b) What constraints apply to your selection of primary lights?

   (c) What restrictions must you obey when you select the primary lights in a color-matching experiment?

   (d) Suppose you and your colleague measure a color-match to a test light, but you use different sets of primary lights. What will be the relationship between the color-matching intensities you find and those that your colleague finds?

2. When the eye is adapted to a steady light, the nervous system readjusts its visual sensitivity in a variety of ways. For example, when you walk into a dark theatre from the outdoors, at first you cannot see well. But, after some time, your visual system adjusts and it becomes easy to see the dim lights. Similarly, a light presented on a bright background is difficult to see but the same light presented on a weak background may be easy to see.

   In the following questions, think about the difference between a pair of lights that match one another versus what the lights look like – color-matching versus color appearance.

   (a) Suppose that we establish a pair of foveal lights as metamers by adjusting them to match on a zero (black) background. (Since the lights are viewed purely in the fovea, they are matched by the cones.) Now, suppose we view the metamers on an intense red background. Based on the theory that color matches are photopigment matches, will the two lights continue to be metamers?

   (b) Neitz, Neitz and Jacobs et al. (1991) claimed that some color normal males change their matches when the test lights are superimposed on red fields. In a related article, they argue that there are more than three types of cones in the human eye (Neitz et al., 1993). Read their articles and evaluate their claims.

   (c) Suppose that, in fact, lights continue to match when they are superimposed upon various backgrounds. Seen on the bright background, the two lights are only barely visible. Will the lights still match when they are presented in the dark, against no background? What if we present the lights in the periphery where there are many rods?

(d) When the two lights are seen on the bright background and on the dim background, will their *appearance* be unchanged?

3. Suppose we represent two lights by the three-dimensional vectors that represent each light's cone photopigment sensitivities, **a** and **b**. The vector difference between the two representation of the two lights is **d** = **a** − **b**. Finally, consider two lights **m** and **n** that also differ by this same vector **d** = **m** − **n**.

   (a) Suppose that we double the intensity of **a** and **b**. What happens to the vector representing each of the lights? What happens to the vector representing the difference between the scaled lights?

   (b) Suppose that we express the coordinates of these lights in another color space obtained by applying a linear transformation, T. What will be the vector difference between **a** and **b** in the new color space? What will be the vector difference between **m** and **n** in the new color space?

   (c) Do you think the two lights represented by **a** and **b** will be as discriminable as the lights **m** and **n**? Why or why not? Do you know of any experimental data to support your claim? Should we collect some?

4. For many practical applications, people wish to use only two dimensions to describe colored lights. Specifically, they wish to compare the direction of the three-dimensional vectors and ignore the length of the vectors. The reduction in dimension of the representation is usually done by introducing chromaticity coordinates via the following formula. Suppose the entries of the three-dimensional color vector are L, M, S. Then we define two chromaticity coordinates, l and m as

$$l = \frac{L}{L + M + S}$$

$$m = \frac{M}{L + M + S}.$$

   (a) Show that two vectors with color representations that differ by a scale factor have the same chromaticity coordinates.

   (b) Consider the following four lights $(1, 1, 0), (0.7, 1.0, 0.3), (0.3, 1.0, 0.7), (0, 1, 1)$. These lights are weighted mixtures of the two components, $(1, 1, 0)$ and $(0, 1, 1)$. Compute the chromaticity coordinates of lights that are formed as weighted sums of the lights. Plot them on a graph whose axes are l and m.

(c) Compute the general formula for the chromaticity of a pair of lights formed as the mixture

$$a(L, M, S) + b(L', M', S')$$

(d) Challenge. The chromaticity coordinates of this mixture of the two lights describe a set of points on the chromaticity diagram that depend on the weights, a and b. Call these points, $(l(a, b), m(a, b))$, and prove that they always fall on a straight line.

5. Table 1.1 lists estimates of the proportion of photons absorbed per second for unit intensity lights at different wavelengths in the human eye.

(a) How many photons will be absorbed during one second to a light at 500nm and 5 units of intensity? What about a light at 600 nm and 10 units of intensity? Answer for both receptor classes.

(b) How many photons will be absorbed in each receptor class when we present the superposition of the two lights? Again, answer for both receptor classes.

(c) How would you set the intensities of the 500nm and 600nm lights so that the absorptions to these lights equal the absorptions to a unit intensity 550nm light?

(d) Can you set the intensities of the 500nm and 600nm lights so that the absorption rate matches a 400nm light at 10 units of intensity?

6. Now, suppose you are studying the color-matching performance of a dichromat, a person with only the L and M cones. We can summarize the properties of the two receptor system using some simple drawings.

(a) Make a graph whose x-axis is the rate of absorptions by the first photoreceptor and whose y-axis is the rate of absorption by the second photoreceptor. On the graph plot the rate of absorptions to each of the unit-intensity monochromatic lights.

(b) On the same graph, plot the number of absorptions during one second to a 500nm light at 0.5 units of intensity and 2 units of intensity. Plot the number of absorptions during one second to a 600nm light at 0.5 units of intensity.

(c) On the same graph, plot the number of absorptions in one second to mixtures of 500nm and 600nm when their respective intensities are (0.2,0.8), (0.5,0.5), and (0.8,0.2).

Table 1.1: Photopigment Absorption Sensitivities (Source: Smith and Pokorny)

L Cones	M Cones	S Cones
0.004249	0.004602	0.174419
0.008655	0.009716	0.364341
0.015893	0.018921	0.662791
0.023446	0.031705	0.906977
0.030212	0.047814	1.000000
0.034461	0.063667	0.918605
0.041385	0.086167	0.802326
0.062785	0.130657	0.693798
0.102282	0.189210	0.468992
0.162392	0.267706	0.279070
0.263572	0.397597	0.166667
0.424233	0.596778	0.096899
0.618411	0.810534	0.046512
0.775138	0.944515	0.023256
0.885759	1.000000	0.011628
0.956412	0.989772	0.003876
0.995909	0.925850	0.003876
1.000000	0.809000	0.000000
0.967113	0.653030	0.000000
0.896459	0.478650	0.000000
0.796696	0.318844	0.000000
0.672069	0.194068	0.000000
0.531393	0.110458	0.000000
0.380960	0.058553	0.000000
0.257120	0.029660	0.000000
0.159559	0.014319	0.000000
0.091581	0.007159	0.000000
0.048308	0.003324	0.000000
0.025806	0.001534	0.000000
0.012431	0.000767	0.000000
0.006294	0.000256	0.000000

7. Answer the following questions about scotopic sensitivity.

    (a) Suppose you study the wavelength sensitivity of an observer under scotopic viewing conditions. At the end of the experiment you discover that the observer was wearing tinted contact lenses. The observer has to go on an extended holiday to Mars, but is willing to leave his contact lenses behind. What measurements do you need to make to correct your estimate of the observer's wavelength sensitivity?

    (b) There are some intensity ranges in which both rods and cones actively respond to lights. At those intensity levels, human observers are still trichromatic, even though there are four active receptor classes. How can this be?

    (c) Suppose we adjust a pair of lights so that they are metamers under scotopic vision. Will they be metamers under photopic vision?

    (d) Suppose we adjust a pair of lights so that they are metamers under photopic vision. Will they be metamers under scotopic vision?

    (e) A yellow daisy and a blue lilac may be perceived to be equally bright under scotopic conditions. Purkinje noticed that these lights are not equally bright under photopic conditions, where the yellow flower is perceived to be much brighter. This phenomenon is called the *Purkinje shift*. Explain the phenomenon.

8. Answer the following questions on the limits of color-matching.

    (a) Use a computer drawing program to make a pattern of fine yellow and blue lines. Make sure that the colors in the lines look blue and yellow when you are close to the monitor. Step away from the monitor three or four meters. What happens to the color appearance of the lines? Try the same with white and black lines. What happens to their appearance?

    (b) What optical effects could be playing a role in the experiment in part (a)?

    (c) Given what you know about the optics of the eye, do you think we will obtain the same color-matching functions if we repeat our experiments using a 10 cycles per degree sinusoidal pattern rather than a uniform 2 deg spot? What qualitative expectations do you have about the 10 cycles per degree sinusoidal pattern experiments?

    (d) Suppose you establish a metameric match. Then you put on a pair of sunglasses. Will the metameric match be preserved? Describe why or why not.

    (e) As we age, the wavelength transmissivity of our cornea and lens changes. What effect will this have on the color-matching functions?

9. In an abstract for a meeting, Knoblauch and McMahon (1993) described a test of a cure for dichromacy. The idea, which is also found in Tom Cornsweet's book, is simple. Dichromats should wear a tinted lens over one eye. This changes the spectral absorption of the photopigments in that eye, providing enough information in the photopigment absorptions to permit discrimination of lights that were previously identical at the photopigments.

   Now consider a dichromatic subject, Mr. X, as described by James Clerk Maxwell

   > By furnishing Mr X. with a red and a green glass, which he could distinguish only by their shape, I enabled him to make judgements in previously doubtful cases of a colour with perfect certainty. I have since had a pair of spectacles constructed with one eye-glass red and the other green. These Mr X. intends to use for a length of time, and he hopes to acquire the habit of discriminating red from green tints by their different effects on his two eyes. Though he can never acquire our sensation of red, he may then discern for himself what things are red, and the mental process may become so familiar to him as to act unconsciously like a new sense. (J.C. Maxwell, Trans. Roy. Soc. Edin. (1855), v. 21, p. 275-298, reprinted in Scientific Papers, W.D. Niven (ed.), Dover, New York).

   Do you agree with Maxwell that Mr. X's experience of color would be the same if he were to ear the tinted-lens glasses? Knoblauch and McMahon, who are protanopes, thought that the ability to perform discriminations did change when wearing the glasses. Even if you are not a dichromat, try this idea for yourself. Do you agree with their conclusions?