Chapter 7

The Trichromacy of Color Vision

Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1</td>
<td>System properties: The trichromacy of color vision</td>
<td>162</td>
</tr>
<tr>
<td>7.2</td>
<td>A mathematical model of trichromacy</td>
<td>168</td>
</tr>
<tr>
<td>7.3</td>
<td>The search for the Fundamentals of color vision</td>
<td>173</td>
</tr>
<tr>
<td>7.4</td>
<td>Trichromacy as a causal story: As good as it gets!</td>
<td>175</td>
</tr>
<tr>
<td>7.5</td>
<td>A Design question: Why three and only three cone types?</td>
<td>176</td>
</tr>
<tr>
<td>7.6</td>
<td>Color vision deficiencies: Dichromacies and anomalies</td>
<td>176</td>
</tr>
<tr>
<td>7.7</td>
<td>Genetics of color vision</td>
<td>179</td>
</tr>
<tr>
<td>7.8</td>
<td>Spatial mosaics for the S, M and L cones</td>
<td>179</td>
</tr>
<tr>
<td>7.9</td>
<td>Photopic spectral sensitivity</td>
<td>184</td>
</tr>
<tr>
<td>7.10</td>
<td>Summary: Photoreceptors and the second transformation</td>
<td>187</td>
</tr>
</tbody>
</table>

In Chapters 2 and 3 we introduced some psychophysical facts concerning how our visual perceptions change (or don’t change) with changes in the wavelength and intensity of light. At low light levels (scotopic vision), all wavelengths of light look the same whitish color. Some patches of light will look brighter than others, but given only variations in physical intensity, lights of all wavelengths can be made to look identical. That is, in scotopic vision metamer sets include all wavelengths of light, and all wavelength information is lost.

In Chapter 5 we developed a model to explain this fact. The model assumes that only a single photoreceptor type, the rods, is functional at scotopic light levels. A single photoreceptor type cannot preserve wavelength information because wavelength information is discarded in the transduction process. Each rod sums quantal catches linearly, and any two lights that lead to the same total quantal catch in the rods will be metameric – they will look identical to a human subject.

When stimuli are at photopic light levels, however, subjects experience color variations, as discussed in Chapter 3. Thus, they readily discriminate among lights of different wavelengths because lights of different wavelengths differ in perceived color, and the colors are sustained (at least approximately) across variations of intensity. Using a bumblebees can fly argument, we therefore know immediately that photopic vision cannot be based on a single univariant photoreceptor class like the rods. The physiological model we adopted for scotopic vision must be rejected for photopic vision because it fails to account for the system property – the preservation of wavelength information.
You may be surprised to learn, however, that even in photopic vision there are metamer sets – sets of lights of different wavelength compositions and intensities that are indistinguishable from each other. The nature of these psychophysiologically defined sets of lights is described by a psychophysical law, the law of trichromacy, which will be presented in detail below. And the question is: why do these metamer sets occur?

To address this question we depart from our practice of avoiding mathematical formulations, and present a mathematical model of trichromacy. We do this because the model is a particularly simple use of algebra (three simultaneous linear equations in three unknowns), yet it is elegant and sufficient to the modeling task. Moreover, the historical interplay between the psychophysical law and the mathematical model, leading on to the discovery of the physiological and genetic entities that instantiate the model, provides one of the loveliest examples of progressive explanation in vision science (see Mollon, 2003, for a historical account).

In addition, trichromatic matches provide a new and interesting example of the use of identity matching, and thereby of the Identity family of linking propositions. The Identity family was first introduced in Chapter 2 in our account of thresholds and scotopic matching. Watch for the use of Identity propositions as we go along.

7.1 System properties: The trichromacy of color vision

7.1.1 Three facts about wavelength discrimination

Let us begin by introducing three psychophysical facts about wavelength discrimination. First, as already discussed, we can discriminate among lights of different wavelengths because different wavelengths look different colors. Second, some mixtures of physical wavelengths can be discriminated from other mixtures and from any single wavelength selected from the spectrum. Whites, purples, and desaturated colors (pink, baby blue, light green, etc.) are examples of colors that arise only from wavelength mixtures. They are called non-spectral (or extra-spectral) colors, meaning that we cannot match them to any individual wavelength.

7.1.2 Metamer sets in photopic vision

But third – and this you may find surprising – metamers occur in photopic as well as in scotopic vision. That is, even at photopic light levels there are sets of lights of very different wavelength compositions and intensities that look identical. The membership in these metamer sets is initially counterintuitive and very odd.

For example, Figure 7.1 shows a plot of the complementary wavelengths of light. In vision science, the term complementary wavelengths is used to describe pairs of wavelengths that look white when mixed together in proper proportions. This diagram tells us that many different mixtures of wavelengths all look white1. Moreover, by varying the intensities of the different mixtures you

---

1 Unfortunately the term “white” is used in both physics and psychophysics, and this causes confusion as usual. In physics the term white light is often taken to mean an equal energy mixture of all wavelengths, like the last mixture in Figure 7.1A. This definition is troublesome in vision science, because many different physical stimuli (wavelength mixtures) actually appear white, as the rest of Figure 7.1A shows. In fact, a whitish appearance tells us remarkably little about the wavelength composition of a light. To DT’s knowledge there is no word in either physics or psychophysics for “physical stimuli that are members of the perceptual white metamer set”. The closest phrase is “metameric to an equal energy light”.

7.1. SYSTEM PROPERTIES: THE TRICHROMACY OF COLOR VISION

For each wavelength from about 570 nm to 680 nm, it is possible to achieve a white-appearing spot by combining that wavelength in the proper proportion with the properly chosen complementary wavelength, which will fall somewhere between about 430 and 500 nm. For properly chosen intensities, the row of lights shown here would all look white, and would be perceptually indiscriminable. The last spot on the right is a mixture of equal energies of all wavelengths. This is usually the wavelength composition we assume a white-appearing spot to have, but obviously this assumption can be wrong. In addition to those shown, many other mixtures would also look white – for example, any of many mixtures of three wavelengths, four wavelengths, and so on. (The plus sign in the circle is the symbol for superposition: we are superimposing one light on another.) [Modified from Cornsweet (1970, Fig. 10-5, p. 232); after Sinden (1923)].
could match them all in perceived brightness, with the result that they would all look identical, even though they are very different physically. And there is an infinite number of other combinations of three or more wavelengths that all look white. Similarly, we can make a set of stimuli of many different wavelength compositions that all look identical and a particular shade of yellow; another set that all look identical and a particular shade of light blue; and so forth. Each of these sets of lights is a metamer set.

In ordinary experience we usually don’t notice the existence of metamer sets, because metamers are such perfect perceptual facsimiles of each other that the fact that they are physically different passes unnoticed. But here’s an example. Outside DT’s old office, there was a light fixture that consisted of two light bulbs inside a translucent globe. Around Christmas time one year, someone took out the two ordinary light bulbs and replaced them with a “red” bulb (i.e. a bulb that emitted a band of long wavelengths, say, above 620 nm) and a “greenish-yellow” bulb (i.e. a bulb that emitted a band of middle wavelengths, say, between 530 and 560 nm). When the globe was replaced, one half of the globe looked red and the other half greenish-yellow. But between the two, there appeared a band of very distinct and saturated yellow.

Why was the yellow band there? Not because the band was illuminated by a light of an isolated wavelength that looks yellow (say, 575 nm), but because it contained just the right mixture of light from the “red” and “green” bulbs. The mixture of wavelengths coming from the band belonged to the yellow-appearing metamer set. But most people who walked by would not have even wondered why the yellow band was there. Of those who wondered, most would probably have assumed that the globe must have contained a source of 575 nm light. Only a few would have guessed that nothing but broadband “red” and “green” bulbs were hidden inside the globe.

To emphasize again the oddity of color mixture, notice that the appearances of combinations of wavelengths of light is very different from the sounds we hear when we combine sound waves. When we create a series of vibrations of different temporal frequencies in the air, we hear a series of tones of different pitches. When we play these tones together, we hear chords that still perceptually contain the original tones; we don’t hear an intermediate tone, much less a completely novel tone or no tone at all. Why is the mixing of lights so different?

### 7.1.3 The psychophysical law of trichromacy

As it turns out, metamer sets are not as arbitrary as they originally seem. They follow a particular rule, called the law of trichromacy (tri = three).

Figure 7.2 shows a laboratory set-up for demonstrating trichromacy. We assemble a set of three slide projectors, fitted out with devices for allowing continuous variation of their intensities. We put a narrow-band color filter in front of each, and overlap the three beams on a projection screen to make a patch of light, A. In other words, patch A is a mixture of three lights of different wavelengths, $\lambda_1$, $\lambda_2$, and $\lambda_3$. A useful set of choices (cf. Figure 7.4) is to let $\lambda_1 = 460$ nm (which looks predominantly blue), $\lambda_2 = 530$ nm (which looks predominantly green), and $\lambda_3 = 650$ nm (which looks predominantly red)$^2$. We also set up a fourth projector to make a second patch of light, B. We then use a variety of color and neutral filters in turn, to make patch B appear any color and brightness that we choose.

---

$^2$The three lights we use as the mixture set are sometimes called primaries. However, this term is used in different ways in other contexts. Suffice it to say that the “primary” colors you learned about in kindergarten make use of a different meaning of the term.
7.1. SYSTEM PROPERTIES: THE TRICROMACY OF COLOR VISION

The law of trichromacy, informally stated, is that by varying only the intensities of the three wavelengths in patch A, we can make a perfect perceptual match to (almost) any other light in patch B; that is, we can make $A \equiv B$ (A is metameric to B). Different colors and brightness in patch B will require different intensities of the three wavelengths in patch A, but an exact perceptual match will almost always be possible.

In fact, we can exactly match any light in patch B if we are allowed to move one of the three wavelengths from patch A to patch B. This additional provision leads to the formal statement of the law of trichromacy: Given any four lights, we can arrange them with three in patch A and one in patch B, or two in patch A and two in patch B; vary the intensity of any three of them, and end up with a perfect perceptual match between the two patches. Figure 7.3 shows a simulation of the basic trichromacy demonstration.

In 1928, W.D. Wright carried out a classic study of trichromatic color matching. Wright tested 10 subjects with mixtures of 460, 530, and 650 nm in patch A. He set up patch B with each of the different individual wavelengths of light in turn. For each wavelength in patch B, the subjects adjusted the intensities of the three lights in patch A to make metameric matches between the two patches.

Wright’s data are shown in Figure 7.4. In the short wavelength range, below about 460 nm, color changes very little with wavelength. These lights all look predominantly violet, and each can be nearly matched with the 460 nm (“blue”) primary alone. However, a small amount of the 650 nm (“red”) primary must be added, and a small amount of the 530 nm (“green”) primary must be subtracted – moved to the other side and mixed with the light in patch B – in order to make the matches. At 460 nm, of course, the 460 nm primary provides an exact match. Between 460 and 650 nm, perceived hue changes more rapidly with wavelength, as do the proportions of the different primaries required for matches. Notably, the 650 nm (“red”) primary must be subtracted for all wavelengths between 460 and 530 nm, as must the 460 (“blue”) primary for all wavelengths between 530 and 650 nm.
CHAPTER 7. THE TRICHROMACY OF COLOR VISION

Figure 7.3: COLOR PLATE. A simulation of color mixture. In this figure, the three superimposed beams of Figure 7.2 have been partially separated in space. The outer crescents simulate the colors of each of the three original wavelengths. The outer triangles simulate the colors of combinations of two wavelengths, and the central triangle simulates the color resulting from the combination of all three wavelengths. (Figure 7.3 is only a simulation and not a true demonstration, because the colors of the various segments will be simulated with the broadband pigments used in printing, rather than being made from narrow wavelength bands.)

Trichromacy is a remarkable and puzzling system property of photopic vision. Why do wavelength mixtures behave the way they do? Why are the metamer sets as they are? Why are three lights enough? The answer lies in our visual systems.

7.1.4 Reprise on the Converse Identity proposition

Let us do the exercise of ferreting out a linking proposition. The Identity family of linking propositions was introduced in Chapter 2 (Figure 2.7B), in the context of measurements of matching and thresholds. Trichromatic metamers are sets of stimuli that are very different physically but appear identical perceptually. That is, they are another case in which subjects are carrying out a matching task. The data are perceptual, so to explain them physiologically we will be trying to reason from perception to physiology. Thus the two available Identity propositions are the Contrapositive and the Converse.

Moreover, the basic perceptual observation is that patches A and B match, so the relevant linking proposition is the Converse: perceptual identity implies physiological identity. Thus, if the theorist assumes the truth of the Converse Identity proposition, metamer matches imply that the signals that arise from a set of metameric stimuli are rendered physiologically identical somewhere within the visual system.

The causal story for trichromacy then becomes a locus and coding question. Where within the visual system do the signals originating from the physically different stimuli become identical, and by means of what physiological processes and computations?
Figure 7.4: Data from a trichromatic color mixture experiment. The abscissa shows the wavelength of light in patch B. The ordinate shows the proportions of 460, 530, and 650 nm lights required to make the two patches metameric. The lines show fits to the data from 10 subjects. Note the “subtraction” of the different primaries in different wavelength regions. [Modified from Boynton (1979, Fig. 5.18, p. 149), data from Wright (1928).]
7.2 A mathematical model of trichromacy

7.2.1 Assume three Fundamentals

The mathematical model of trichromacy starts with the set of mathematical or physiological assumptions shown schematically in Figure 7.5. The model assumes that photopic vision is served by three Fundamentals – three mathematical/physiological entities with different spectral sensitivity curves. The peak sensitivities of the three Fundamentals are assumed to differ, but the ranges are assumed to overlap substantially (for simplicity, they overlap entirely in Figure 7.5A). The model also assumes that each Fundamental forms a linear summation of the signals resulting from different wavelengths of light.

For concreteness, in the following pages we identify the Fundamentals with three types of cones. But it’s interesting to notice that the mathematical model of trichromacy preceded any direct evidence of the numbers of cone types or their spectral sensitivity curves. We will return to this point below.

As was the case for rods in Chapter 2, we can represent each cone type with a funnel that counts the quanta it catches, without keeping track of their wavelengths (Figure 2.9, in which the marbles can now be identified as quanta). In the case of photopic vision there are three funnels, each with its own counter. The three-funnel analogy is shown in Figure 7.5B.

Now we need to develop some symbols. To identify each Fundamental (cone type) with the wavelength range of its maximum sensitivity, the three Fundamentals will be called L, M, and S. The letters L, M, and S in italics will denote the quantum catch rates generated in the L, M, and S cones respectively. Because we are dealing with two patches of light, A and B, there will be two sets of cones (two sets of funnels in the analogy), one for patch A and one for patch B. Let the quantum catches resulting from patch A be $L_A$, $M_A$, and $S_A$, and from patch B be $L_B$, $M_B$, and $S_B$ respectively.

7.2.2 The condition for metamerism

By hypothesis, metamsers occur when lights of different wavelength composition yield identical quantum catches in each of the three hypothetical photoreceptor types. That is,

$$\text{if } L_A = L_B \text{ and } M_A = M_B \text{ and } S_A = S_B, \text{ then } A \equiv B.$$

This statement can be called the condition for metamerism. But under what circumstances is the condition for metamerism satisfied? Is it a fool’s dream, or a realistic basis for a model?

7.2.3 The color equations

What does light of a given wavelength, $\lambda$, do when it encounters a three pigment system like that shown in Figure 7.5? It makes a triplet (a set of three) quantum catch rates, one in each of the three cone types. For any given wavelength such as $\lambda_1$, the heights of the three curves at that wavelength tell us the probabilities of absorption of a quantum by each cone type. Let,

---

3The different cone types have historically been called “red cones”, “green cones” and “blue cones”. Vision scientists avoid this terminology, in order to maintain the clear separation of perceptual and physiological terms. If the terminology is sloppy, it is easy to think that color vision is simple – we see red because we have “red cones”! We use color names to refer to perceived colors, and a different set of names – S, or short-wavelength-sensitive, M, or mid-wavelength-sensitive, and L, or long-wavelength sensitive – to refer to cones.
7.2. A MATHEMATICAL MODEL OF TRICROMACY

Figure 7.5: A mathematical model of trichromacy. A. Completely overlapping spectral sensitivity curves assumed for the three Fundamentals for purposes of illustration. B. Extension of the funnel analogy to the case of trichromacy. Each funnel captures a broad range of wavelengths with probabilities determined by its width at each wavelength, and counts its total quantum catch without regard to wavelength. The result is a set of three variables, \( S \), \( M \), and \( L \), whose values are determined by summing the quantal catches within each cone type.

\[
S = \sum Q_{\lambda} s_{\lambda} \quad M = \sum Q_{\lambda} m_{\lambda} \quad L = \sum Q_{\lambda} l_{\lambda}
\]
• $l_1$ = the height of the $L$ curve at $\lambda_1$,
• $m_1$ = the height of the $M$ curve at $\lambda_1$, and
• $s_1$ = the height of the $S$ curve at $\lambda_1$.

In addition, let $l_2$, $m_2$, and $s_2$, and $l_3$, $m_3$, and $s_3$ be similarly defined. Since we assumed the shapes of the three pigment curves in Figure 7.5, all of these values of curve heights are constants in the equations we will develop below.

Now we need symbols for the intensity of the lights from each of the three projectors; that is, for the rate of arrival of quanta of each wavelength $\lambda_1$, $\lambda_2$, and $\lambda_3$. These intensity values are just the intensities of the three projectors in Figure 7.2. The intensities of projectors 1, 2, and 3 will be called $Q_1$, $Q_2$, and $Q_3$ respectively. Since the subject varies these intensities to make the metameric matches, the $Q$'s will turn out to be the variables in the equations we will develop below.

We are now in a position to write expressions for the rate of quantal absorptions from each of the three wavelengths in each of the three cone types. The rate of quantal absorptions of the wavelength $\lambda_1$:

• by the $L$ cones is: $Q_1l_1$,
• by the $M$ cones is: $Q_1m_1$, and
• by the $S$ cones is: $Q_1s_1$.

and similarly for $\lambda_2$ and $\lambda_3$. Moreover, to calculate the total quantum catches in each cone type in response to any wavelength mixture, we just add up the quantum catches from all of the available wavelengths for each photoreceptor:

• $L$ cone quantum catch is $L = \sum Q_\lambda l_\lambda$,
• $M$ cone quantum catch is $M = \sum Q_\lambda m_\lambda$, and
• $S$ cone quantum catch is $S = \sum Q_\lambda s_\lambda$.

Now let’s return to our two patches, $A$ and $B$. Patch $A$ is composed of three wavelengths, $\lambda_1$, $\lambda_2$, and $\lambda_3$, with variable intensities $Q_1$, $Q_2$ and $Q_3$. We can now write three equations to describe the three cone quantum catch rates produced by patch $A$:

$$L_A = Q_1l_1 + Q_2l_2 + Q_3l_3,$$
$$M_A = Q_1m_1 + Q_2m_2 + Q_3m_3,$$
$$S_A = Q_1s_1 + Q_2s_2 + Q_3s_3.$$

What about patch $B$? Patch $B$ is a light of any chosen wavelength composition and intensity. For any specific choice of wavelength composition and intensity, patch $B$ generates a specific triplet of cone signals, $L_B$, $M_B$, and $S_B$. Once the wavelength composition is chosen, these entities are constants.

Now, the fundamental question is, by varying only the intensities $Q_1$, $Q_2$, and $Q_3$, can the triplet of values $L_A$, $S_A$, and $S_A$ be made identical to the triplet $L_B$, $M_B$, and $S_B$? That’s the condition for metamerism.
Assume for the moment that the two lights are metamers. Then by the condition for metamerism we can substitute \( L_B, M_B, \) and \( S_B \) for \( L_A, M_A, \) and \( S_A \) to produce:

\[
L_B = Q_1l_1 + Q_2l_2 + Q_3l_3, \\
M_B = Q_1m_1 + Q_2m_2 + Q_3m_3, \\
S_B = Q_1s_1 + Q_2s_2 + Q_3s_3.
\]

Notice that this set of three simultaneous equations contain three unknowns – the three intensities \( Q_1, Q_2, \) and \( Q_3. \) (The little \( l, m, \) and \( s \) are known because they are specified by the heights of the spectral sensitivity curves \( L, M \) and \( S, \) and \( L_B, M_B, \) and \( S_B \) are known because they are specified by the choice of the light \( B. \) Each of the three cone types contributes an equation, and each of the three wavelengths contributes a variable.

But remember that it is an elementary property of linear algebra that three simultaneous linear equations in three unknowns are guaranteed to have a solution, so we know that for any specified values of \( L_B, M_B, \) and \( S_B \) we can solve for the values of \( Q_1, Q_2, \) and \( Q_3. \) It follows that for any specified set of values of \( L_B, M_B, \) and \( S_B \) – that is, for any light in patch \( B – \) these equations can be solved. Thus logically we know that by varying only the radiances of the three wavelengths in patch \( A, \) patch \( A \) can be made metameric to light of any wavelength composition in patch \( B. \) But that’s the informal statement of the law of trichromacy! So in sum, the equations provide a sufficient mathematical model of the law of trichromacy.

But there’s one possible flaw in the argument. Remember that, although we are guaranteed a solution to three simultaneous equations in three unknowns, there is no guarantee that all of the values for \( Q_1, Q_2, \) and \( Q_3 \) will be positive. One or more of them might be negative. But real lights cannot have negative intensities, so how do we interpret the negative values? The answer is, in algebra, we move the negative term to the other side of the equation and make its value positive. In the matching experiment, we move the corresponding light, \( \lambda_1, \lambda_2, \) or \( \lambda_3, \) from patch \( A \) to patch \( B. \) From this convention results the formal statement of the law of trichromacy: Given any four lights, we can arrange them in two patches to make \( A \equiv B. \)

We now return to a more historically accurate picture. For the sake of concreteness we initially identified the three Fundamentals with three cone types, and the linear summation property with the loss of wavelength information in an individual photoreceptor caused by the properties of the transduction process. But historically, both the psychophysical fact of trichromacy and the mathematical model of trichromacy were established by about 1860, before we had any other evidence of the number of cone types, or the univariance of photoreceptors, or even any modern notion of quantum theory. It was a deep mathematical insight to see that a set of three simultaneous linear equations in three unknowns would provide a sufficient model for the perceptual fact of trichromacy, and to posit three Fundamentals with overlapping spectral sensitivity curves to provide the three variable system of simultaneous linear equations.

Finally, let’s return to the properties of wavelength discrimination with which we started this chapter, and review explicitly why they occur. First, we can discriminate among wavelengths of light different wavelengths keep their distinctive colors well across variations in intensity. As shown in Figure 7.5, each different wavelength sets up a different set of relative quantum catches, \( l \) vs. \( m \) vs. \( s, \) in the \( L \) vs. \( M \) vs. \( S \) cones. Putting it another way: wavelength information is lost in each individual cone type, but it is preserved in the ensemble of three cone types by the relative quantal catches among them. This ensemble code is the form in which wavelength information
passes through the photoreceptor level of processing. Similarly, information about the intensity of the light is preserved in the absolute quantum catches in the three kinds of cones.

Second, some mixtures of wavelengths can be discriminated from other mixtures and from any single spectral wavelength. Why? Because not all of the possible ratios of cone signals are created by individual wavelengths of light, and some mixtures of wavelengths create these novel ratios. For example, by inspection of Figure 7.5, there are no individual wavelengths that create L/M/S ratios of 1:1:1, or 2:1:2, and so on; but you can find mixtures of wavelengths that will do so. When mixtures of wavelengths create these ratios, non-spectral colors appear.

The third property, metamerism, blends into the law of trichromacy, and has already been explained in detail.

7.2.4 Reprise on the Initial Identity proposition

Here’s a second exercise on linking propositions. Whereas the inference from behavioral trichromacy to three Fundamentals rests on a Converse Identity proposition, the mathematical model from fundamentals to behavioral trichromacy rests on the Initial Identity proposition. The mathematical model begins by assuming (seemingly arbitrarily) the existence of three physiological entities – the three Fundamentals. We argued mathematically that three such entities, acting together, would process physical stimuli in just such a way as to create the metamer sets observed psychophysically in human subjects, and summarized by the law of trichromacy.

The Initial Identity proposition – that identical physiological states imply identical perceptual states – enters the argument because we are trying to reason from (assumed) physiological states to perceptual states. Assuming Initial Identity allows us to use physiological identity to infer perceptual identity, and thus use the model to provide an account of the psychophysical data.

7.2.5 Exact physiological implications of trichromacy

Now let’s step back a little. Clearly the system property of trichromacy, together with its mathematical model, places major constraints on physiological models of the visual system. Up until this point, for the sake of specificity and simplicity, we have identified the three mathematical Fundamentals with three cone types. But the true constraints are actually somewhat more general.

What the mathematical model of trichromacy actually suggests is that information available for discriminating among wavelengths and intensities of light passes through a serious bottleneck – or rather, three bottlenecks – somewhere on its way through the visual system. That is, there is a stage at which this information is limited to three variables. (A statistician would say the visual signal has only three degrees of freedom, and an engineer would say the system has only three information channels).

In the specific model we introduced (and in reality in foveal vision) this three-channel stage is instantiated by three kinds of cones with three different photopigments. But logically, behavioral trichromacy could as well come about from having several kinds of photoreceptors, but with the information reduced to three variables by passing through a three-channel stage somewhere later in the visual system. And as it happens, in peripheral vision we do have rods as well as the three cone types, and the reduction to three channels does come later.

A final thing to note is that even if the photoreceptors are the bottlenecks, the argument we gave above does not depend on assuming any particular set of shapes or wavelengths of maximum
sensitivity for the spectral sensitivity curves of the three photopigments. We assumed three Fundamentals, L, M, and S, and the heights of these three curves at particular wavelengths were constants that entered into the color equations. But logically we could have assumed any of many shapes for the spectral sensitivity curves, as long as they are consistent with color mixture data like that shown in Figure 7.4.

7.3 The search for the Fundamentals of color vision

The psychophysical fact and the three-Fundamental model of trichromacy were both well established by about 1860 (Mollon, 2003). But the situation left vision scientists in a state of acute frustration. We knew that there are (probably) three cone types, but neither the psychophysics nor the mathematical model reveals their spectral sensitivity curves. Determining the spectral sensitivity curves of the three Fundamentals has thus been a fundamental challenge (pun intended) for vision scientists for 150 years, and scientists from several disciplines have set out to determine the shapes of these curves. We will review three historical approaches.

First, some of the earliest relatively accurate estimates of the Fundamentals were derived from psychophysical measurements. Of these the most successful approach was based on the assumption that certain “color-blind” individuals (see below) have lost one of the Fundamentals, but that their two remaining Fundamentals are identical to the Fundamentals of normal subjects. Without the interference of the third pigment, the available pigments in “color-blind” human subjects could be estimated psychophysically, and by hypothesis used to estimate the spectra of the normal pigments. Excellent psychophysically-based estimates of the sensitivity maxima and the curve shapes of the cone Fundamentals emerged in the 1970s. They are shown with data from other, later techniques in Figure 7.6B, and it can be seen that the data correspond closely. However, the challenge persisted of verifying these estimates with more direct measurements that were free of the assumption that normal pigments occur in color-deficient subjects.

In the next historical iteration, the technique of microspectrophotometry was developed. In microspectrophotometry, one dissociates the cells of an excised retina, until individual photoreceptors can be seen floating free under the microscope. One can then shine a tiny beam of light through an individual photoreceptor and onto a photocell, and the percent of light absorbed can be measured for each wavelength in turn.

Cone spectral sensitivity curves measured with microspectrophotometry are shown in Figure 7.6B. Microspectrophotometry confirms that the spectral sensitivity curves of individual cones are smooth and U-shaped, and these particular data suggest absorption maxima at about 420, 530, and 560 for the three cone types. Microspectrophotometry, however, is beset with signal/noise problems that limit the measurements to a relatively narrow range of wavelengths around the spectral maximum (notice the limited wavelength range of the squares in Figure 7.6B).

In the most recent assault, in the mid-1980’s, the problem of cone spectral sensitivities was attacked with suction electrodes. This technique has the advantage that the physiological response of the photoreceptor itself is used for the actual measurements. Since very small amounts of light are sufficient to produce measurable changes in photocurrents, the suction electrode produced an increase in sensitivity over earlier techniques, with a corresponding increase in the wavelength range over which meaningful measurements could be made. Data from an early suction electrode study are compared with those of earlier techniques in Figure 7.6B. The suction electrode data, plotted on
CHAPTER 7. THE TRICHROMACY OF COLOR VISION

Figure 7.6: The spectral sensitivities of the three cone types in color-normal subjects. A. L, M and S cone spectra from macaque retina, recorded with suction electrodes, shown on a linear ordinate. These empirical curves replace the hypothetical curves shown in Figure 7.5A. B. The three cone spectra estimated with three different techniques, shown on a log ordinate. The results of psychophysical (diamonds), microspectrophotometric (squares), and suction electrode (triangles) techniques are shown. The three sets of measurements agree remarkably well. [A replotted from Baylor, Nunn, and Schnapf (1984, courtesy of J. Schnapf); B modified from Lennie and D’Zmura (1988, Fig. 10, p. 344).]
a linear sensitivity axis, are shown in Figure 7.6A. These data replace the conceptual Fundamentals introduced in Figure 7.5A.

All of these techniques have evolved over DT’s scientific lifetime, and each decade has brought a surer answer. Nowadays there is excellent agreement from all of these very different kinds of measurements, and the cone spectral sensitivity curves are a solved problem. The most recent estimates suggest that the spectral maxima of the S, M, and L cones are very close to 430, 530, and 560 nm. Moreover, as needed to sustain the three channel mathematical model developed above, the spectral curves are broadly overlapping, at least between 400 and 550 nm, so that in this wavelength range, a single wavelength can set up a quantum catch rate in each of the three kinds of photoreceptors, and set up the ensemble code.

Notice, however, that the S cones have negligible sensitivity above about 550 nm. As a result, color-normal subjects actually have only two functional cone types in the mid to long wavelength spectral range. And in fact, color-normal subjects can match any wavelength above about 550 nm with a mixture of only two primaries, such as a 530 nm “green” and a 650 nm “red”.

7.4 Trichromacy as a causal story: As good as it gets!

Let us consider where we’ve been. First, psychophysical measurements (color matching) on human subjects quantified the puzzling system properties of metamerism and trichromacy. Second, a mathematical model was developed to explain these system properties. The model—a three Fundamentals with broadly overlapping spectra and linear summation of effects across wavelengths, expressed mathematically in three simultaneous linear equations with three unknowns—provides a sufficient account of the psychophysical data.

But the model also provided specific predictions about neural elements we should find within the visual system: three kinds of cones with different spectral maxima but overlapping spectral sensitivity curves. We then went looking for independent evidence for the existence and spectral properties of three kinds of cone photoreceptors, and found it. Moreover, the linear summation across wavelengths assumed to occur within each Fundamental finds its explanation in the loss of wavelength information in the transduction process. The causal story is complete, and tightly tied into the network of surrounding sciences.

The story of trichromacy also illustrates the difference between mathematical and physiological models. Mathematical modeling is challenging, and it’s a major achievement to invent a model that just exactly accounts for a set of psychophysical findings. But given a satisfying mathematical model, the question arises: will the model be instantiated in the real visual system? Vision scientists become keenly interested in finding physiological entities that embody the mathematical entities or parameters assumed by the model.

For DT, it’s thrilling to understand that three simultaneous equations in three unknowns provide an account of trichromacy. But it’s even more thrilling to learn that the hypothesized discarding of wavelength information within a Fundamental corresponds to the actual discarding of wavelength information by the transduction process, and that the mathematical property of adding up the terms in the linear equations across wavelength corresponds to the physiological property of combining the effects of individual cis-trans isomerizations across wavelength.

In sum, for DT, the causal story of trichromacy is vision science at its very best. Trichromacy and its explanation provide an important case example to which to aspire as one tries to invent causal stories to explain other system properties of vision.
7.5 A Design question: Why three and only three cone types?

A design question: Why did humans evolve to have exactly three cone types rather than two or four? It has been argued that ancestral primates had only two types – an S cone and a prototypical LM cone with a spectral maximum in the mid to long wavelength region. But since the S cones have negligible sensitivity above 550 nm, the ancestral primate would have had only a single pigment available above 550 nm, and would not have been able to discriminate among middle and long wavelengths, nor among surfaces that reflect different combinations of middle and long wavelengths. Roughly speaking, the ancestral primate would not have been able to discriminate among objects or surfaces that we perceive as yellow-greens, yellows, oranges and reds.

The splitting of the LM prototype into two separate classes – L and M – probably occurred only 30-40 million years ago, and only in old-world primates. It has been proposed that such a trichromatic system allows us to discriminate red, orange, and yellow fruit from green trees, and to tell ripe from unripe fruit. Thus, the third cone type allowed ancestral primates to exploit important new food sources efficiently, and probably carried a selective advantage in evolutionary terms.

Why not keep on evolving, and have more than three cone types? After all, the larger the number of cone types, the smaller the metamer sets, and the more wavelength information is preserved. The speculation here is that, although it is easy to create trichromatic metamers in the lab, they rarely occur in nature. In nature most surfaces reflect broad bands of wavelengths and most light sources produce broad bands of wavelengths. Thus, most of the objects that produce lights in any one metamer set probably have relatively similar spectral characteristics. So additional photopigments might not allow us much more useful color discriminations than we can already make with three.

7.6 Color vision deficiencies: Dichromacies and anomalies

People whose retinas contain the three standard pigments are said to have normal color vision, or to be color-normal trichromats. But not everyone is color-normal. Some people are missing one of the three kinds of cones. Others still have three cone types, but one or more of the pigments is shifted in spectral sensitivity. Such changes make predictable changes in color discrimination capacities. Look back at Figure 7.5 as you read the next few paragraphs.

First, what would happen if you were missing one of the three cone photopigments? Suppose you were missing the L photopigment. In that case, you would have two rather than three funnels, and two rather than three cone output signals. You would have only two rather than three equations in your set of color equations, because the equation for quantum catches in the L cones would not be needed. Since there would be only two equations, you would need only two wavelengths in patch A to match any wavelength in patch B (two equations need only two unknowns to be guaranteed a solution). All of the metamer sets of a trichromatic individual would also be metamer sets for you; but your metamer sets would be larger than those of the trichromat – you would confuse patches of light that would be readily discriminable for your trichromatic friend, and you would probably be worse than they are at finding yellow, orange and red fruit in green trees.

This form of color vision deficiency is well known, and occurs quite frequently in human beings. Since the color vision system is reduced from three to two variables, such individuals are called dichromats (di = two). The two most common types are protanopia, in which the person is missing functional L cones (proto = first; protanopia = the first kind of color deficiency); and deuteranopia,
in which the person is missing functional M cones (deutero = second; deuteranopia = the second kind). Each of these types of color vision deficiency is sex-linked, and occurs in about 1% of the Caucasian male population. The third kind of dichromacy, tritanopia (tri = three; tritanope = the third kind), in which the person is missing functional S cones, is much rarer, and occurs with equal frequency in both males and females.

Second, what would happen if the spectral sensitivity of one of your photopigments were shifted along the wavelength axis? Suppose your L cone pigment were shifted toward your M cone pigment. How would your color equations change? Since the height of the L curve would have changed a little at each wavelength, all of the little l’s in the equation for the quantum catch in L cones would change. You would still be trichromatic, because your color vision would still be described by three equations in three unknowns. But for each wavelength composition of patch B, the change in values of l’s would make a change in the intensities of the three lights in patch A needed to make the match to patch B. That is, your color matches would be different than those of your color-normal friend. Similar changes would occur if the M or the S cone spectral sensitivity curve were shifted.

This form of color vision deficiency is also well known, and people with trichromatic vision but non-normal metamer sets are said to be color-anomalous. The color vision of people with a shifted L cone pigment is called protanomalous, while that of people with a shifted M cone pigment is called deuteranomalous. Protanomaly and deuteranomaly occur in about 1% and 3% of the Caucasian male population respectively. Tritanomalous color vision, which results from a shifted S cone pigment, is much rarer and occurs equally often in males and females.

In sum, in the Caucasian population about 8% (one in 12) of the male population and less than 1% of the female population have a color deficiency caused by losses or spectral shifts of either the L or the M cone photopigment. As a group, these forms of color vision are often called the red/green color deficiencies. Another small percentage (less than 1%) have tritan deficiencies – losses or anomalies of the S cone photopigment.

Color-normal and color-deficient individuals live in different perceptual worlds. To illustrate this point, we here discuss a clinical color mixture test that diagnoses among color-normal, dichromatic and anomalous trichromatic subjects. The test is the Rayleigh match, carried out with a device called an anomaloscope. The test is illustrated in Figure 7.7. In the anomaloscope field (Figure 7.7A), the subject sees a mixture of 550 and 670 nm lights (which ordinarily look green and red respectively to a color-normal subject) in one half of a circular field, and a 589 nm light (which ordinarily looks a slightly orangish yellow to a color-normal subject) is presented in the other half of the field. The subject is asked to vary the proportion of the 550 vs. 670 nm lights in the one half field, and the intensity of the 589 nm light in the other half field, to try to make a metameric match between the two halves of the field. Figure 7.7, panels B-D show a simulation of the outcomes of Rayleigh matches for color-normal, dichromatic, and anomalous trichromatic subjects, and panel E simulates the appearance of one anomalous trichromat’s match to a color-normal observer.

Red/green color deficiencies are so common that in any class of 30 school children, one or more of the boys is likely to have a red/green color deficiency. These children are likely to find it difficult to learn color names, use the “correct” color in drawing with crayons, and so on. An adult color-deficient individual can have trouble choosing two socks that match, and may wear color combinations that seem bizarre to his color-normal friends. If you argue with your friends over what colors things are, you may be a dichromat or an anomalous trichromat and not realize it.
Figure 7.7: COLOR PLATE. Simulated Rayleigh matches. Panel A shows the actual spatial layout of the anomaloscope fields. In panels B-E the two wavelengths in the left half-field have been shifted upward and downward for purposes of illustration. For a color normal subject (panel B), there will be a particular ratio of intensities of the 550 and 670 lights (say, 50:50) that is metameric to the 589 nm light. Both halves of the anomaloscope field will look slightly orangish yellow. For a dichromat (panel C), who cannot discriminate the 550 from the 670 light in the first place, the 589 nm field can be matched by any ratio of the 550 and 670 nm fields, including 100% of either of these wavelengths. In other words, lights that look red, yellow and green to color-normal subjects all look the same – are in the same metamer set – for a dichromat. For an anomalous trichromat (panel D), the normal trichromat’s metamers look different colors, but the 550 and 670 lights can be mixed in some other proportion (say 70:30) to match the 589 nm light. But (panel E) the half fields that match for the anomalous trichromat look very different to the color normal subject. [J. Neitz, personal communication].
7.7 Genetics of color vision

Each of the red/green color deficiencies described above shows an X-linked pattern of inheritance (the particular deficiency is passed from grandfather to grandson with the mother being a carrier). To geneticists this pattern suggests strongly that the L and M pigment genes are located on the X chromosome. Tritanopia and tritanomaly show an autosomal pattern of inheritance, suggesting that the S pigment gene is located on some other chromosome.

In 1986, a team of geneticists and psychophysicists led by Jeremy Nathans isolated and sequenced the genes that control the production of each of the three human cone photopigments. The DNA encoding was characterized from both color-normal and red-green color-deficient individuals. This research was extremely exciting to vision scientists, since it took the search for the Fundamentals of color vision all the way to the molecular level. A recent review can be found in Neitz and Neitz.

Comparisons of the molecular structures of the four normal human photopigments are shown in Figure 7.8. As expected from the X-linked inheritance patterns of red/green color vision deficiencies, the genes for two photopigments were found next to each other on the X chromosome. Unexpectedly, many individuals had several rather than just one copy of the second gene in the sequence. Protanopes turned out to be missing the first gene of the sequence, which was therefore identified as the L cone pigment gene. Deuteranopes often had only the first gene of the sequence, and the second and later genes were therefore characterized as the M cone pigment genes. More complex genetic patterns — genes made up of pieces from both the L and the M pigment genes — were also found, especially in color-anomalous individuals. Moreover, polymorphisms were found in the normal L and M pigment genes, allowing an explanation of more subtle variations of color vision among color-normal individuals.

7.8 Spatial mosaics for the S, M and L cones

The numbers and distributions of rods and cones across the retina, shown in Figure 6.2, have been known for half a century. But what are the numbers and distributions of each of the three cone types?

These questions have been of major theoretical interest, but the answers proved elusive for many years. The S cone mosaic has been of interest because acuity is poor under conditions that isolate S cones, and it has therefore been speculated that there might be only a small number of S cones. The L and M cone mosaics have been of interest because there are individual differences in photopic spectral sensitivity curves, and it has been suspected that these might be due to individual differences in the proportions of L vs. M cones: the \( L/M \) cone ratio.

Within the last 20 years or so, the numbers and distribution of S cones has been determined with several different techniques. In the most definitive early study, Christine Curcio and her colleagues (Curcio, Allen, Sloan, Lerea, Hurley, Klock, and Milam, 1991) used a newly developed stain specialized to reveal the S cone opsin. Use of this stain exposes the whole S-cone matrix.

A sample of Curcio et al’s results are shown in Figure 7.9. In their data, S cones provide only about 10% of the cones in the human retina. In fact, although it does not show in the figure, S cones

---

4Geneticists (and the rest of us) sometimes slip into talking as though complex human behaviors can be attributed to single genes. A deuteranopic student in one of DT’s classes wrote a term paper on the genetics of color vision deficiencies. In parody of the single-gene assumption, his opening sentence was, “I have a gene for mismatched socks”.

Figure 7.8: Comparisons of the molecular sequences of rhodopsin and the three cone photopigments. Each black dot marks a difference between the two sequences being compared. Note the close similarity of the L vs. M pigments. This similarity is indicative of a recent evolutionary separation. [From Nathans, Thomas, and Hogness (1986, Fig. 11, p. 200).]
Figure 7.9: The matrix of S cones. The picture shows the distribution of S cones over a small region just off from the center of the fovea. The black dots are S cones; the open dots are L and M cones. [From Rodieck (1998, p. 210); after Curcio et al. (1991).]
are missing entirely from the central fovea; and they are spaced relatively far apart throughout the rest of the retina.

Why? The suspected design explanation for this sparse representation is that chromatic aberration defocuses the retinal image formed from short wavelength light, so that fine spatial sampling would be wasted in the S cone system. The S cones provide another example of the idea that “poor” optics – in this case chronically defocussed images, due to chromatic aberration – limit our acuity for short wavelength light, and do so through an evolutionary mechanism that matches the spacing of the photoreceptor matrix to the quality of the optical image.

The numbers and distributions of L and M cones are currently being attacked with the techniques of adaptive optics. In our discussion in Chapter 4 we suggested that one of the major uses of adaptive optics will be to look in through a person’s corrected optics and be able to see the structures in his living retina. Omitting many details, Figure 7.10 shows some of the very first pictures of the distributions of the three different cone types in a living human eye.

Moreover, the three cone mosaics shown in Figure 7.10 can be combined into a single image to reveal the subject’s overall retinal mosaic. Figure 7.11 shows pseudocolor images of the retinas of two different subjects. The proportions of S cones were about 5% in both retinas. But the L/M cone ratios differed considerably: 3.8 for JW vs. 1.2 for AN. Measurements of larger samples of subjects confirm the existence of this large range of ratios (Carroll et al., 2002). We will return to these individual differences immediately below.

More recently, adaptive optics have yielded answers to some of the classic questions concerning dichromacy. It has long been speculated that the loss of a photopigment gene could lead to the functional or actual loss of a type of cone: the L cones for protanopes, the M cones for deuteranopes, and the S cones for tritanopes. The images shown in Figure 7.12, taken with adaptive optics in the living eyes of two dichromats, show that this speculation is correct. The retina of the deuteranope shown in Figure 7.12a is missing its S cones, and the retina of the protanope shown in Figure 7.12b is missing its L cones.

These images also address a second question. There have been two rival theories for how the loss of a pigment plays itself out at the level of disabling the photoreceptors that would have contained that pigment. Loss theories suggest that the cones that would have contained the missing pigment are literally lost from the retina, leaving holes in the mosaic where the missing class of photoreceptors would have been. On the other hand, replacement theories suggest that the cones that would have contained the missing pigment are filled with a different pigment – for example, that a protanope would have its potential R and S cones both filled with the S cone pigment – so that the full complement of cones is retained in the retinal mosaic.

It turns out that both theories are probably right. In studies of the molecular genetics of color vision, two different kinds of genetic changes have been found in different dichromats. A dichromat can be missing the gene for the L or M cone pigment. Alternatively, he can have a mutation that makes the L or M cone pigment misshapen and therefore nonfunctional. Perhaps the absence of a gene leads to replacement, and a misshapen pigment leads to loss.

These arguments are supported by the images in Figure 7.12. Two subjects are shown: NC, who is a deuteranope because of a mutant M pigment gene, and MM, who is a protanope because of a missing L pigment gene. The retina of the deuteranope NC is shown in Figure 7.12a. It shows L cones but no M cones, and a reduced overall number of cones, with “holes” between them, as predicted by loss theory. In contrast, the retina of the protanope MM is shown in Figure 7.12b. It shows M cones but no L cones, but a normal number of cones overall. Apparently the
7.8. SPATIAL MOSAICS FOR THE S, M AND L CONES

Figure 7.10: Distributions of the three cone types in a color-normal human retina. A. A 1° patch of the retina of the right eye of subject JW at 1° eccentricity. The small round dots are individual cone photoreceptors. In Panels B, C, and D the same region of retina was exposed to three different combinations of wavelengths of lights. The different combinations were selected to favor visualization of S cones (the dark spots in B), L cones (the dark spots in C), or, least successfully, M cones (barely visible as the dark spots in D). (In each case the new picture is subtracted from the picture in A. The grey stripe down the middle of A disappears in the subtraction process). [From Roorda and Williams (1999, Fig. 1, p. 520).]
L cones have been filled with M cone pigment, and retained in the retinal mosaic, as predicted by replacement theory. Thus ends a controversy that lasted a century – loss of a cone type occurs in some dichromats, and replacement of the missing photopigment by the available one occurs in others.

7.9 Photopic spectral sensitivity

As it turns out, knowing the actual spectral sensitivities of the three Fundamentals provides us with several theoretical bonuses. First, in Chapter 3 we introduced the photopic spectral sensitivity curve, $V(\lambda)$. Now that the spectral sensitivity curves of the actual L and M cones are available for use, it turns out that $V(\lambda)$ can be readily modeled by a weighted linear sum of L and M cone signals. This idea is illustrated in Figure 7.13, and the physiological model fits the psychophysical data well.

Second, we also mentioned in Chapter 3 that although vision scientists have adopted a standard curve for photopic spectral sensitivity, there are actually small but consistent individual differences in the empirical photopic spectral sensitivity curves for different individual subjects. It has long been speculated that these individual differences could come about from variations in L/M cone ratios among subjects. Inspection of Figure 7.13 reveals that independent sliding of the L and M spectral sensitivity curves up and down will allow interesting changes in the overall photopic curve, and this model provides reasonable fits to the known individual differences.

Moreover, we have had a chance to examine the cone mosaics of two color-normal subjects, JW and AN, in Figure 7.11. The L/M cone ratios differed markedly between these two retinas. The photopic spectral sensitivity curves of these two subjects were also measured, and the differences
Figure 7.12: [COLOR PLATE]. Cone mosaics in two dichromats. A. Deuteranope NC, missing M cones. B. Protanope MM, missing L cones. NC has a “patchy” retina, with some cones apparently missing, whereas MM appears to have the full complement of cones. [Carroll, Neitz, Hofer, Neitz, and Williams (2004, Fig. 4, page 8465).]
Figure 7.13: A model of $V(\lambda)$ based on a weighted sum of L and M cone inputs. A shows the calculation on a linear ordinate, and B shows it on a logarithmic ordinate. The curves labeled L (or log L) and M (or log M) show the spectral sensitivities of the L and M cones; the curves labeled $V(\lambda)$ (or log $V(\lambda)$) show the synthesis of $V(\lambda)$ from the sum of the L and M curves. For the average color-normal subject, the weighting needed to fit $V(\lambda)$ is about 2:1 for the L vs. M cone signals; that is, $V(\lambda) = 2L + M$. The differential weighting is incorporated into the diagram by increasing the height of the L curve with respect to the M curve. [Modified from Boynton (1979, Fig. 9.3, p. 307).]
are in the right direction to be modeled by the differences in L/M cone ratio.

And third, in Chapter 3 we also described the fact that \( V(\lambda) \) emerges from many different kinds of psychophysical experiments – flicker photometry, motion photometry and minimally distinct border judgments, among others. We argued that when a characteristic psychophysical “signature” emerges frequently from the data, it’s a good guess that that characteristic has a physiological instantiation. That is, vision scientists would be drawn to speculate that individual neurons that sum inputs from L and M cones, with spectral sensitivity curves corresponding to \( V(\lambda) \), will be formed within the visual system. However, neurons that embody this prediction are simply not present at the level of the photoreceptors. We are left to speculate that they will emerge at a later level of processing.

7.10  **Summary: Photoreceptors and the second transformation**

At the end of Chapter 4 we summarized the effects of the First Transformation – from the physical world to the retinal image. We argued that the first transformation rendered the retinal image **two-dimensional** and **low pass filtered**.

In addition to the optics, the incoming visual signal also encounters a stage of discrete sampling by the photoreceptors that rendered it (poetically) **pointillistic**. The discrete sampling stage could be considered either part of optical processing (since the signal is still carried by light), or part of processing by the photoreceptors (since they are the spatially discrete elements). As such, it could be considered part of either the first or the second transformation. We summarized it along with the first transformation at the end of Chapter 4.

We are now ready to summarize the (rest of the) second transformation: from the retinal image to the quantum catches in photoreceptors via the phototransduction process, and from the quantum catches to the photoreceptor outputs via a complex set of chemical and electrical information transmission processes. These two stages of processing are summarized in Figure 7.14. In combination these two stages create a spatial array of quantal catches in the rods and L, M and S cones, and transform it into a spatial array of synaptic outputs in the same four kinds of neurons.

The transduction process and the combination of four kinds of photoreceptors provide us with causal stories for some of the system properties of scotopic and photopic vision introduced in Chapters 2 and 3, as well as for the trichromacy of color vision discussed in the present chapter. The scotopic spectral sensitivity curve is maximal at about 500 nm, and wavelength information is lost in scotopic vision, because scotopic vision is served by rods and rods alone. Wavelength information is preserved in photopic vision, up to the limits described by trichromacy, because photopic vision is served by three and only three kinds of cones.

The information transmission process also leaves its mark. For example, at low light levels, each rod is so exquisitely sensitive that it produces a detectable signal in response to the absorption of a single quantum. This sensitivity enables human subjects to detect the absorption of only 5-10 quanta in an extended test field, and therefore of a single quantum in an individual rod. In addition, a saturating non-linearity, probably within the cone photoreceptors, provides a signal that allows the detection of interference fringes in the vicinity of 60 cy/deg. Of course, the properties of photoreceptors influence all aspects of vision, but additional examples are beyond our scope.

Information processing by the three cone types together also illustrates the concept of **ensemble codes** (or **pattern codes**). Because of the nature of transduction, each photoreceptor individually loses wavelength information. Yet, working together as an ensemble, the three types of photore-
Figure 7.14: Photoreceptors: the second transformation.

Photoreceptors preserve at least some information about the wavelength composition of each region of the retinal image. By comparing the signals from L, M and S cones, later levels of the system have access to a fair bit of wavelength information. The concept of pattern codes will recur frequently throughout the remainder of this book. Later we will see neurons that compare signals from the L, M and S cones, to create a new wavelength/color code.