

Color Vision: Encoding Mechanisms, Evolution, and Deficiencies

Yuhao Zhu

Department of Computer Science
Department of Brain and Cognitive Sciences
University of Rochester

yzhu@rochester.edu
<https://yuhaozhu.com/>
<https://horizon-lab.org/>

Contents

1 Color Encoding at Photoreceptors	2
1.1 From Light Spectrum to Cone Responses	2
1.2 Cone Excitation Space, Spectral Locus, and HVS Gamut	4
2 Metamerism and Color Matching Experiments	5
2.1 Trichromatic Color Matching and Its Linear System Perspective	5
2.2 Color Matching Experiments and Color Matching Functions	6
2.3 Connecting Color Matching Functions and Cone Fundamentals	9
3 Post-Receptoral Color Encoding: Opponent Processes	13
3.1 Hue Cancellation Experiment	14
3.2 Light-Dark Mechanism and Luminous Efficiency Function	16
3.3 Neural and Physiological Basis	18
3.4 There are Many Inconvenient Truths	22
4 Evolution of Color Vision	26
4.1 The Rise of Trichromacy	26
4.2 The Rise of Scotopic vs. Photopic Vision	29
5 Color Vision Deficiencies	30
5.1 Models of Deficient Color Vision	31
5.2 CVD Assistive Technologies	35
5.3 Genetic Basis of CVD	36

1 Color Encoding at Photoreceptors

Newton presumably did the famous experiment where he showed that a beam of white light is really a mixture of photons at different wavelengths, and each wavelength gives a different color percept. Color is very much our subjective sensation. What is the physical reality is the spectral power distribution of light. In Newton's words: "*rays of Light in falling upon the bottom of the eye excite vibrations in the retina. Which vibrations, being propagated along the solid fibres of the optick Nerves into the Brain, cause the sense of seeing.*" [Newton, 1704].

1.1 From Light Spectrum to Cone Responses

As we have seen before in the class, there are three classes of cone, each with a different spectral sensitivity function or a cone fundamental. When discussing colors, it is customary to normalize these functions to peak at unity and define them at an "equal-energy" (rather than "equal-quantal") basis. Since each function is normalized within itself, the relative sensitivity between different wavelengths within a cone type is still maintained, but we cannot really compare the results between cones after normalization. Figure 1 compares the absolute, equal-quantal cone fundamentals with the normalized, equal-energy cone fundamentals.

A normalized, equal-energy sensitivity function tells us the relative amount of photon absorption given a unit power at each wavelength. For instance, the normalized L cone response is 1 at 570 nm and 0.4 at 630 nm. This means that given two lights that have the same power/energy, one with photons only at 570 nm and the other with photons only at 630 nm, the fraction of photons absorbed in the 630 nm light is about 40% of that in the 570 nm light.

Critically, this also means if we have a 570 nm light at 1 W and a 630 nm light at 2.5 W, the two lights would cause the same amount of pigment excitations in L cones. If we had only L cones, these two lights would be seen as the exact same light, because the HVS will receive the exact amount of electrical responses — according to the Principle of Univarience. This explains why we could not see colors at night, when only rods are functioning.

Of course in reality most humans have three classes of cones, so what *is* the signal we receive? Given the Spectral Power Distribution (SPD) of a light $\Phi(\lambda)$, we can calculate the total number of photon absorptions for each cone type, given by:

$$L = \int_{\lambda} L(\lambda)\Phi(\lambda)d\lambda \quad (1a)$$

$$M = \int_{\lambda} M(\lambda)\Phi(\lambda)d\lambda \quad (1b)$$

$$S = \int_{\lambda} S(\lambda)\Phi(\lambda)d\lambda \quad (1c)$$

where $L(\lambda)$, $M(\lambda)$ and $S(\lambda)$ represent the cone sensitivity functions. The fact that we can directly multiply $\Phi(\lambda)$ with, say, $L(\lambda)$ is a result of defining $L(\lambda)$ on an equal-energy/power basis. The L/M/S values we calculate represent the total number of photon absorptions given an incident light. You would know why we care about photon absorption: it is equivalent to

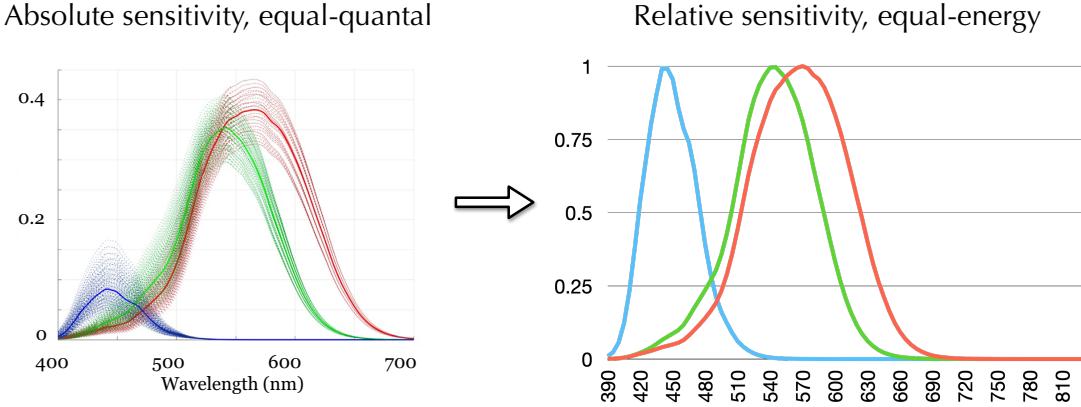


Figure 1: Physiological measurements give us absolute spectral sensitivities at an equal-quantal basis (left), but in color science each cone fundamental function is usually normalized to peak at unity and then converted to an equal-energy form (right). Left: from [Wandell et al. \[2022, Fig. 10a\]](#); Right: data from CIE 2006 “physiologically-relevant” LMS functions, which is based on [Stockman et al. \[1999\]](#) and [Stockman and Sharpe \[2000\]](#).

pigment excitation up to a constant scaling factor, and pigment excitations produce electrical signals that our brain actually receives. We sometimes simply call the L/M/S value the **cone responses** or **tristimulus values** of a light, but you should know that they do not represent the actual magnitude of the electrical responses of the cones, since the magnitude is not linearly proportional to absorption as we have discussed before.

In actual computation we discretize the spectra and perform summation rather than integration. We also limit the summation to within the [380 nm, 780 nm] range, since the cone fundamentals are practically 0 beyond that range. Assuming that we are quantizing the spectra at a 1-nm interval, the cone responses are linearly related to the light spectrum by:

$$\begin{bmatrix} L(380), L(381), \dots, L(780) \\ M(380), M(381), \dots, M(780) \\ S(380), S(381), \dots, S(780) \end{bmatrix} \times \begin{bmatrix} \Phi(380) \\ \Phi(381) \\ \vdots \\ \Phi(780) \end{bmatrix} = \begin{bmatrix} L \\ M \\ S \end{bmatrix} \quad (2)$$

We can see that this is a huge dimensionality reduction. That is, our brain receives only the three-dimensional cone responses not the actual spectrum of the light, which is of a much higher dimension. This is the basis of the **trichromatic theory of color vision**: color is a three-dimensional system. The theory was first proposed by [Young \[1802\]](#), who conjectured that there are three types of receptors, and later rediscovered, popularized, and extended by Hermann von Helmholtz in the later part of the nineteenth century.

The huge dimensionality reduction also means there are infinitely many lights (with different SPDs) that will be seen as having the same color, as long as they cause the same cone responses.

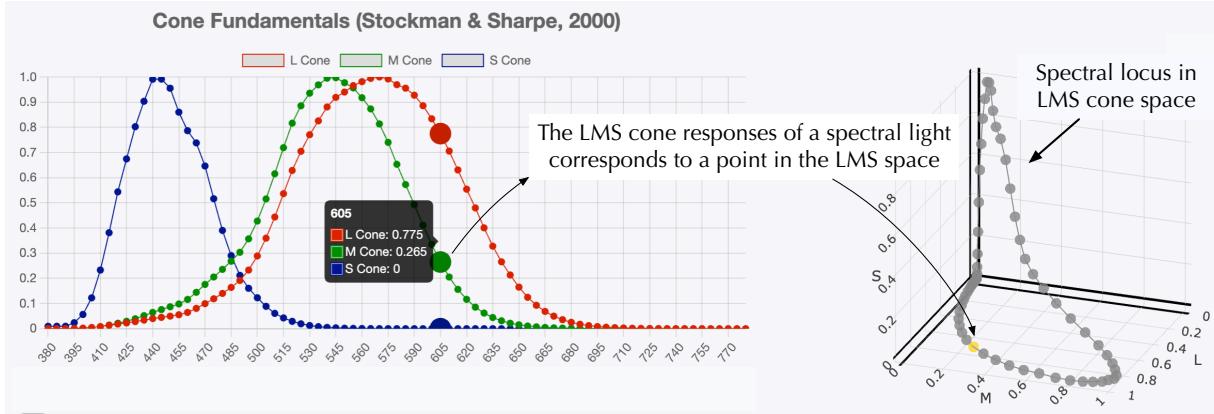


Figure 2: Spectral locus in LMS cone space. From the interactive tutorial in [Zhu \[2022a\]](#).

One way to understand this is if we try to solve the system of linear equations in Equation 2 given $[L, M, S]^T$, with the constraint that the Φ vector must be non-negative everywhere (since power cannot be negative), we will generally get infinitely many solutions, since it is an *under-determined* system. The fact that multiple physically different lights can end up having the same color is called **metamerism**, and these lights are called **metamers** of each other.

1.2 Cone Excitation Space, Spectral Locus, and HVS Gamut

The cone fundamentals essentially give us a color space, which we call the **LMS cone space** or **cone excitation space**. A color space allows us to geometrically interpret a color as a point in the coordinate system. In the cone space, the color of a light is interpreted as the amount of responses in each of the three cone classes produced by the light (as calculated by Equation 2).

The **spectral locus** is a curve on which each point represents the color of a spectral light at a wavelength. Figure 2 shows the spectral locus in the LMS cone space on the right, and the cone fundamentals on the left. The L, M, and S cone response of a spectral light at, for instance, 605 nm are 0.775, 0.265, and 0, which corresponds to the point [0.775, 0.265, 0] in the cone space. Connecting these points for all the spectral lights gets us the spectral locus in the LMS space.

We know a color corresponds to a point in the cone space, but does an arbitrary point in the cone space correspond to a real color? The answer is *no*. For instance, if a point has a negative coordinate it obviously could not be a color of a real light, since that a negative cone response would require negative power in the light. Also, [1, 0, 0] would also not a real color, since there is no real light that can produce only L cone response but no responses from M and S cones — if you examine the cone fundamentals carefully. We call these colors **imaginary colors**, since they cannot be produced by physically realizable lights, where the power must be non-negative at any wavelength.

In principle, a $[L, M, S]$ point corresponds to a real color if Equation 2 has a non-negative solution for Φ . The total set of $[L, M, S]$ points that have a non-negative Φ solution corresponds

to all the colors that humans can see, which is called the **gamut** of the human visual system. Geometrically, if a point in the cone space cannot be constructed through a *positive*, linear combination of the points on the spectral locus, it then is not a real color, since the SPD of a real light must be a positive, linear combination of the SPDs of the spectral lights.

For instance, the line segment connecting two points on the spectral locus contains real colors that can be produced by mixing some amount (i.e., positive linear combinations) of the two spectral lights. Of course we can apply this iteratively: once you get a real color through combining spectral colors, itself can then be used as a basic color to create other colors. [Zhu \[2022c\]](#) is an interactive tutorial that visualizes the HVS gamut in the cone space (and others), which you are invite to go through.

2 Metamerism and Color Matching Experiments

Perhaps the main implication of the trichromatic theory color is that one can, in theory, produce the color of any light by mixing three other colors. We will take a linear system perspective to give a mathematical explanation of this, and show a famous experiment that empirically confirmed this.

2.1 Trichromatic Color Matching and Its Linear System Perspective

We can produce, in theory, any color by mixing three other colors, which we call the **primary colors**. Here is the mathematical intuition. Let's say the SPDs of the three primary lights are $R(\lambda)$, $G(\lambda)$, $B(\lambda)$. What is the power of each of the primary lights we need to produce the color of a target light $\Phi(\lambda)$? For the color of the mixed light to match that of the target light, their corresponding cone responses must match:

$$\begin{bmatrix} \sum R(\lambda)L(\lambda), & \sum G(\lambda)L(\lambda), & \sum B(\lambda)L(\lambda) \\ \sum R(\lambda)M(\lambda), & \sum G(\lambda)M(\lambda), & \sum B(\lambda)M(\lambda) \\ \sum R(\lambda)S(\lambda), & \sum G(\lambda)S(\lambda), & \sum B(\lambda)S(\lambda) \end{bmatrix} \times \begin{bmatrix} r \\ g \\ b \end{bmatrix} = \begin{bmatrix} \sum \Phi(\lambda)L(\lambda) \\ \sum \Phi(\lambda)M(\lambda) \\ \sum \Phi(\lambda)S(\lambda) \end{bmatrix}, \quad (3)$$

where r, g, b represent the power of the three primary lights. This system in general has one unique solution because we have the same number of unknowns (r, g, b) as the number of equations. Each of the three equations constrains the cone-response matching of one class of cones. This means there is a single unique way to mix three primary lights to produce the color of an arbitrary target light.

What if we have more than three primary lights? We will end up with an *under-determined* systems (e.g., three equations but four unknowns if given four primary lights), which means there are infinitely many ways to mix the primaries to produce the target color. If we have only two primaries, we end up with an *over-determined* system, where there is in general no solution.

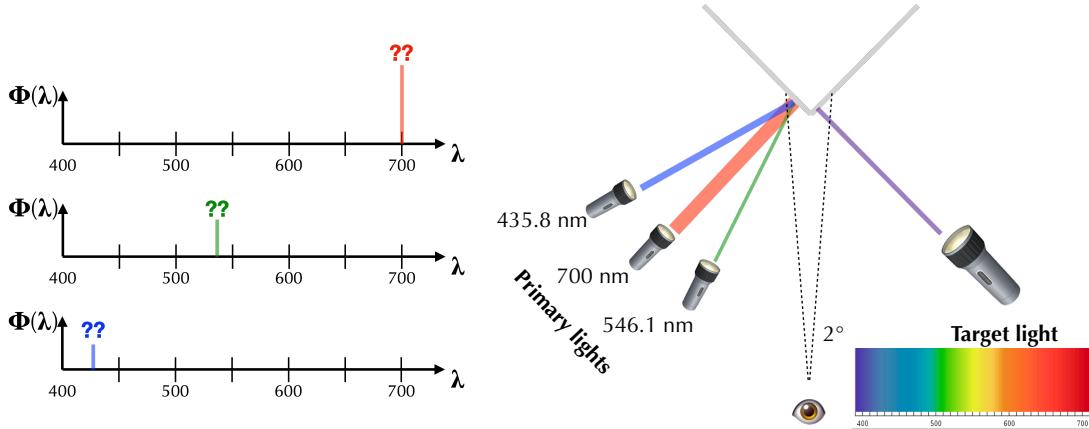


Figure 3: Color matching experiment setup. In CIE 1931 standardization of the experiment, the primary lights are spectral lights at 435.8 nm, 546.1 nm, and 700 nm, and they swept the visible spectrum [380 nm, 780 nm] at a 5-nm interval as the target light. Note that CIE 1931 did not do any actual experiments; they synthesized the data from Wright and Guild.

2.2 Color Matching Experiments and Color Matching Functions

Equation 3 gives a mathematical explanation for trichromatic color matching, but it requires knowing the cone fundamentals, which, as we have seen before in the class, were not experimentally measured until mid 20th century, first through microspectrophotometry [Marks et al., 1964; Brown and Wald, 1964; Dartnall et al., 1983] and then through suction electrode [Schnapf et al., 1987]. But even without the cone fundamentals, nothing prevents us to perform an actual experiment to find the amount of primaries for producing a color. Thomas Young apparently had no interest in such an experiment [Mollon, 2003]. Maxwell [1857] is believed to be the first to undertake an actual color matching experiment in the 19th century, but he did the experiments using rotating discs painted with different colors, relying on the temporal integration of the HVS.

Modern color matching experiments started with Wright and Guild [Wright, 1928, 1929, 1930; Guild, 1931]. International Commission on Illumination (CIE) in 1931 standardize the color matching experiment and synthesized Wright's and Guild's data (without any additional experiments) to obtain what is now known as the CIE 1931 RGB Color Matching Functions. This process is discussed in detail in [Broadbent, 2004, 2008; Service, 2016; Zhu, 2020]. We give an account for the key elements here; the experimental setup is illustrated in Figure 3.

Observers are presented with a 2° visual field. They are given three primary lights, which in the CIE 1931 standard are **spectral lights** (lights that have photons at only one single wavelength; also called monochromatic lights) at wavelengths 435.8 nm, 546.1 nm, and 700 nm. The three primary lights are pointed at the same point on one side of the visual field. On other side of the visual field is the target light. Their goal is to adjust the power of each of the three primary lights so that the colors from the two sides of the visual field match. CIE 1931 swept

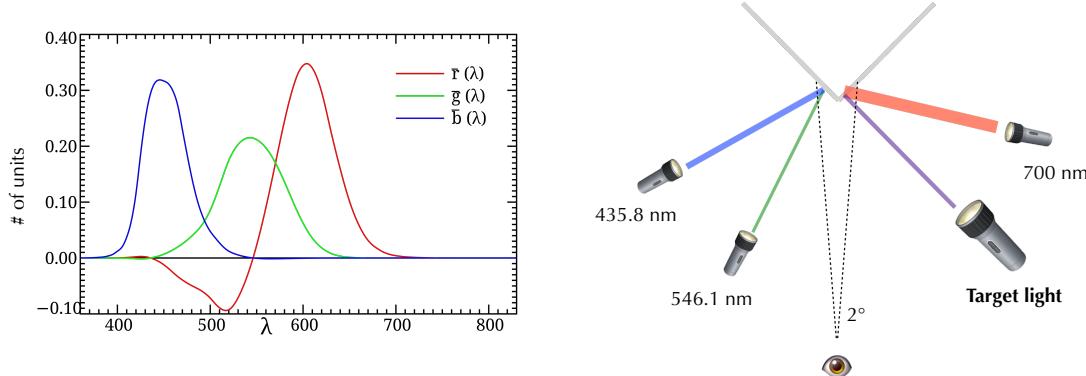


Figure 4: Left: CIE 1931 RGB Color Matching Functions (CMFs); from [Marco Polo \[2007\]](#). The y -axis shows the number of units needed of each primary so that the mixture matches the color at each wavelength (x -axis) on an equal-energy basis. The unit system is so defined that mixing equal amount (the number of units) of the three primaries produces the color of the equal-energy white, whose SPD is a constant over the entire spectrum. Right: the negative values in the CMFs indicate that the corresponding primary light is to be mixed with the target light in order to match the color of the mixture of the other primaries.

the entire visible spectrum for the target light at a 5-nm interval.

Color Matching Functions Require a Unit System and a White Point

The results obtained through the color matching experiments are shown in Figure 4 (left panel). The three curves are collectively called the **CIE 1931 RGB Color Matching Functions (CMFs)**. Intuitively, the CMFs tell us the amount of primaries needed to match the color at each wavelength. But the devil is in the details. Let's carefully walk through what this plot actually shows.

The y -axis represents the number of units required of each primary so that the mixture matches the color at a given wavelength at x -axis. What is a unit? The unit system is so defined that mixing the three primaries in equal units produces the color of the Equal-Energy White (EEW), whose SPD is a constant across the spectrum.

There are two judgement calls here. First, CIE 1931 decided that EEW was going to be the “white” color in their RGB color space. In general, however, there is no single color that we universally define as white, so if you were to design a color space you get to pick whatever color that you think is white in the color space. But an intuitive choice of a white color is one that is **achromatic** (colorless), a color that, subjectively, can only be described of having a certain level of gray but that has no apparent hue. Daylights at different times of a day are perceptually achromatic and could be used as the white point in a color space. The daylight colors are shown to be very similar to the colors of black-body radiation at different temperatures [[Judd et al., 1964](#)], shown in Figure 5.

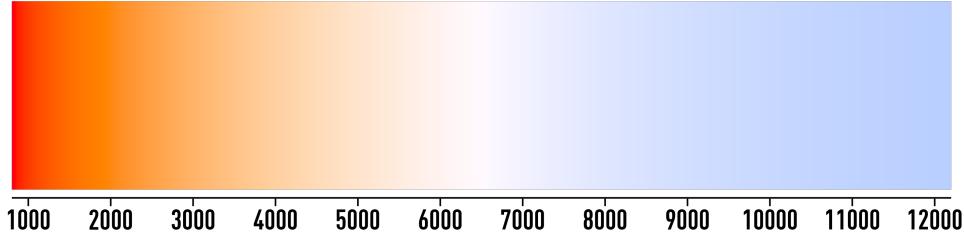


Figure 5: Color from black-body radiation at different temperatures (x -axis; unit: Kelvin). CIE Standard Illuminant D65 approximates the SPD of a noon daylight; its color is similar to that of a 6500K black-body radiation. From [Bhutajata \[2015\]](#).

You probably do not see most the colors in Figure 5 as achromatic on the display right now, but when you are in an environment illuminated by one of these colors, e.g., outdoor at noon, you do see the illuminant as achromatic; this is because of **chromatic adaptation**, a topic we will discuss later in the class. Briefly, the human visual system is evolved to adapt to different daylight colors so that when you spend enough time under such an illuminant you will see the illuminant as achromatic. The adaptation to other colors, however, is weak (or “incomplete” in chromatic adaptation parlance)¹, so it probably does not much sense to pick other colors as the white point if you want your user to see your white as achromatic. CIE has standardized a set of what they call Standard Illuminants (D series), each of which approximates a different daylight color. For instance, the D65 standard illuminant approximates noon daylight and is similar to the color of a black-body radiation at a temperature of 6500K. Many common color spaces, such as the sRGB color space, uses D65 as the white point.

Second, CIE 1931 RGB space, and virtually all color spaces that I know of, define units so that white, however defined, must be produced by an equal-unit mixture of the primaries. This, again, is a judgement call. One could totally design a color space where white is produced by mixing, say, 2 units of Red and 1 unit of Green and Blue — nothing wrong with that. It is just more intuitive for most people that white is produced by equal amount of the primaries.

The x -axis in Figure 4 is defined on an equal-energy/power basis. That is, the CMFs is interpreted as showing the amount (units) of the primaries needed to produce spectral lights of equal power. So if we actually mix the three primaries at each wavelength as indicated by the CMFs, we will get a set of spectral lights that have the same power.

What Does a Negative Unit Mean?

If you observe Figure 4 carefully you will see that some CMFs are negative over certain ranges. For instance, the red CMF is negative at 500 nm. This is perhaps a bit surprising, but mathematically it is entirely possible that some values in $[r, g, b]^T$ are negative when solving Equation 3. Physically, however, what does it mean to have a negative amount/power of primary light?

¹After all, artificial lights are a very recent thing in the scale of evolution, so our HVS has not had a chance to adapt to non-daylight colors yet, if ever.

The right panel in Figure 4 provides the intuition. It turns out that it is impossible to find a combination of the three primary lights to match the color of a spectral light at 500 nm. What does provide a match is to add a little red primary to the target light, and then we can find a combination of the primaries such as the blue and green mixture has the same color of the target light and red primary mixture.

In fact, if you examine the CMFs, you will see that there is a negative contribution from a primary at all but three wavelengths — the only three exceptions are the wavelengths of the three primaries (where two of the primary contributions are zero and the other is positive). This means that no spectral light color (except the three special cases) can be physically produced by mixing the three primaries.

Representing Colors Using CMFs

Given a set of CMFs, we can describe the color of a light with a SPD $\Phi(\lambda)$ using the following equation:

$$\begin{bmatrix} \bar{r}(380), \bar{r}(381), \dots, \bar{r}(780) \\ \bar{g}(380), \bar{g}(381), \dots, \bar{g}(780) \\ \bar{b}(380), \bar{b}(381), \dots, \bar{b}(780) \end{bmatrix} \times \begin{bmatrix} \Phi(380) \\ \Phi(381) \\ \vdots \\ \Phi(780) \end{bmatrix} = \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (4)$$

where $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ are the CMFs, and R , G , and B are the amount of the three primaries needed to match the color of $\Phi(\lambda)$.

The CMFs give us another color space, where the color of a light is interpreted as the amount of primary lights needed to match the color of the light. Of course, if we choose a different set of primary lights, we might end up with a new set of CMFs and a new RGB color space.

2.3 Connecting Color Matching Functions and Cone Fundamentals

CMFs and cone fundamentals both yield trichromatic color vision, so they must be inherently related, as they are just different ways to describing the same thing. We show the two are linearly related in theory and the measurement data of the two match well, too.

Deriving Color Matching Functions From Cone Fundamentals

Given the cone fundamentals, we can derive the CMFs based on the linear system shown in Equation 3. The interactive tutorial by Zhu [2022a] walks through the process, which you are invited to go over, and we will describe the main steps here.

In order to construct the CMFs, we have to match the colors of all the spectral lights, which means we have to specify cone-response matching at each wavelength. Using the basic idea of Equation 3, we have:

$$\begin{bmatrix} \sum R(\lambda)L(\lambda), & \sum G(\lambda)L(\lambda), & \sum B(\lambda)L(\lambda) \\ \sum R(\lambda)M(\lambda), & \sum G(\lambda)M(\lambda), & \sum B(\lambda)M(\lambda) \\ \sum R(\lambda)S(\lambda), & \sum G(\lambda)S(\lambda), & \sum B(\lambda)S(\lambda) \end{bmatrix} \times \begin{bmatrix} r(380), \dots, r(780) \\ g(380), \dots, g(780) \\ b(380), \dots, b(780) \end{bmatrix} = \begin{bmatrix} L(380), \dots, L(780) \\ M(380), \dots, M(780) \\ S(380), \dots, S(780) \end{bmatrix}, \quad (5)$$

where $L(\lambda)$, $M(\lambda)$, and $S(\lambda)$ are the cone fundamentals; $L(\lambda_0)$ is the L cone response of the spectral light at a particular wavelength λ_0 ; $[r(\lambda_0), g(\lambda_0), b(\lambda_0)]^T$ represents the (to-be-solved-for) power of each primary needed to match the color of the spectral light at λ_0 ; $R(\lambda)$, $G(\lambda)$, and $B(\lambda)$ are the SPDs of the primary lights used in the CIE 1931 Color Matching Experiment. The first matrix is a constant matrix given a particular set of CMFs, and we will denote it as the \mathbf{M} matrix. We can solve the system of equations by inverting the first matrix:

$$\begin{bmatrix} r(380), \dots, r(780) \\ g(380), \dots, g(780) \\ b(380), \dots, b(780) \end{bmatrix} = \mathbf{M}^{-1} \times \begin{bmatrix} L(380), \dots, L(780) \\ M(380), \dots, M(780) \\ S(380), \dots, S(780) \end{bmatrix}. \quad (6)$$

To get the CMFs, however, we need to turn the power measure to a unit measure. Recall the requirement that white must be produced by equal units of the primaries. We calculate the power of each primary needed to produce the EEW; let's denote the solution $[r_w, g_w, b_w]^T$:

$$\begin{bmatrix} r_w \\ g_w \\ b_w \end{bmatrix} = \mathbf{M}^{-1} \times \begin{bmatrix} L_w \\ M_w \\ S_w \end{bmatrix}, \quad (7)$$

where $[L_w, M_w, S_w]^T$ denotes the total L, M, and S cone responses of EEW. For the so-calculated $[r_w, g_w, b_w]$ to represent equal units, the last step is to scale $[\bar{r}(\lambda), \bar{g}(\lambda), \bar{b}(\lambda)]^T$ at each λ by $[r_w, g_w, b_w]$:

$$\begin{bmatrix} \bar{r}(380), \dots, \bar{r}(780) \\ \bar{g}(380), \dots, \bar{g}(780) \\ \bar{b}(380), \dots, \bar{b}(780) \end{bmatrix} = \begin{bmatrix} r_w, 0, 0 \\ 0, g_w, 0 \\ 0, 0, b_w \end{bmatrix} \times \begin{bmatrix} r(380), \dots, r(780) \\ g(380), \dots, g(780) \\ b(380), \dots, b(780) \end{bmatrix} \quad (8a)$$

$$= \begin{bmatrix} r_w, 0, 0 \\ 0, g_w, 0 \\ 0, 0, b_w \end{bmatrix} \times \mathbf{M}^{-1} \times \begin{bmatrix} L(380), \dots, L(780) \\ M(380), \dots, M(780) \\ S(380), \dots, S(780) \end{bmatrix} \quad (8b)$$

$$= \mathbf{T}_{lms2rgb} \times \begin{bmatrix} L(380), \dots, L(780) \\ M(380), \dots, M(780) \\ S(380), \dots, S(780) \end{bmatrix}, \quad (8c)$$

where $[\bar{r}(\lambda), \bar{g}(\lambda), \bar{b}(\lambda)]^T$ gives us the unit measure, i.e., the values of the CMFs, at each λ .

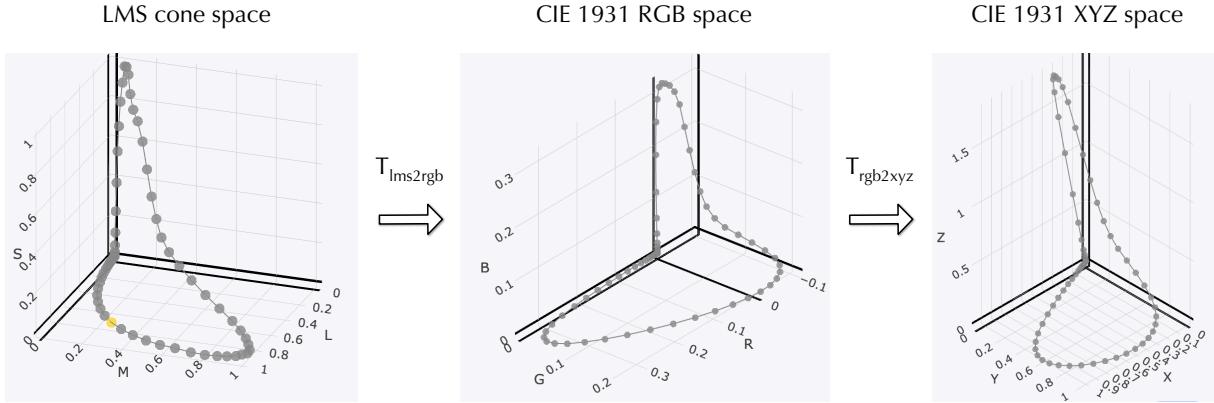


Figure 6: The spectral locus in the LMS cone space, CIE 1931 RGB space, and CIE 1931 XYZ space. The color spaces are a linear transformation away from each other. From the interactive tutorials in Zhu [2022a] and Zhu [2022d].

Cone Space and RGB Space are Related by a Linear Transformation

The rightmost matrix in Equation 8c is the cone fundamentals written out in the matrix form, and the leftmost matrix in Equation 8c is the CMFs written out at discrete wavelengths. So Equation 8c essentially describes a linear transformation from the cone fundamentals to the RGB CMFs, where the transformation is dictated by $\mathbf{T}_{lms2rgb}$. We can look at this in two ways. One, we can think of the cone fundamentals as the CMFs in the cone space: they tell us how much of each cone responses we need to match the color of a spectral light. Two, just like how we can construct the spectral locus from cone fundamentals, the RGB CMFs also give us a way to construct the spectral locus — in the RGB space. $\mathbf{T}_{lms2rgb}$ essentially transforms these two representations of the spectral locus, and this is visualized in Figure 6.

There is something deeper: $\mathbf{T}_{lms2rgb}$ not only transforms the spectral locus, it transforms the entire coordinate system from the cone space to the RGB space. In other words, it transforms every single color in the LMS space to its corresponding coordinates in the CIE 1931 RGB space. The way to think about this is to ask: given that the cone space and the CIE 1931 RGB space provide two ways to represent the color of a light Φ , how is the cone-space representation $[L_c, M_c, S_c]$ and the RGB-space representation $[R_c, G_c, B_c]$ related? Using Equation 2, Equation 4, and Equation 8c, it is easy to see that they are related by a linear transformation through $\mathbf{T}_{lms2rgb}$:

$$\begin{bmatrix} R_c \\ G_c \\ B_c \end{bmatrix} = \mathbf{T}_{lms2rgb} \times \begin{bmatrix} L_c \\ M_c \\ S_c \end{bmatrix}. \quad (9)$$

Of course nothing prevent us from applying an arbitrary linear transformation to the LMS cone space (or the RGB space) to get to another color space. As we will see in the Colorimetry

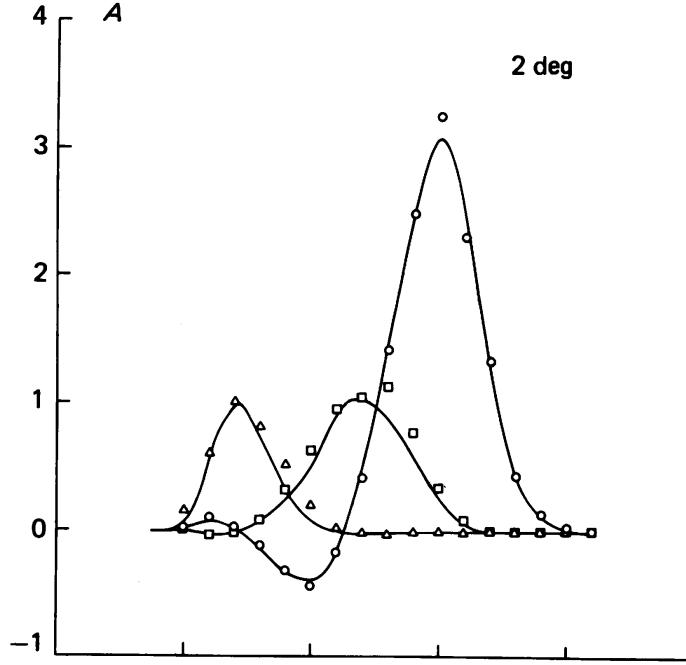


Figure 7: Smooth curves are the CMFs from [Stiles and Burch \[1955\]](#), which uses a different set of primaries and white point than those used in the CIE 1931 RGB CMFs. The markers are the predicted CMFs based on the cone fundamentals measured from macaques. From [Baylor et al. \[1987\]](#), Fig. 4A].

chapter soon, if you choose a different set of primaries and/or white light, you will end up with a color space that is one linear transformation away from the cone space. The right panel in Figure 6 shows one such space, called the CIE 1931 XYZ space. There are good reasons why we need the XYZ space and we will discuss the significance of it later. For now, just observe how the spectral locus in the XYZ space is transformed from that in the cone space.

Cone Responses Fully Explain Psychophysical Color Matching

The CMFs can be both experimentally measured and calculated if we know the cone fundamentals (through a linear transformation), but do the mathematical estimation and the measurement data match? If so, we can say that the physiological process of encoding light power as cone responses can fully account for the color matching experiments in psychophysics.

[Baylor et al. \[1987\]](#) performed one such comparison and showed the two sets of data matched very well. The results are shown in Figure 7, where the smooth curves are from [Stiles and Burch \[1955\]](#), which uses a different set of primaries and white point than those used in the CIE 1931 RGB CMFs. The markers are the predicted CMFs through a linear regression from the cone

fundamentals measured from macaques, after accounting for ocular and macular absorptions².

In fact, the modern versions of the cone fundamentals are constructed so that they are precisely a linear transformation away from some RGB CMFs. For instance, the CIE 2006 “physiologically-relevant” LMS functions (based on Stockman et al. [1999] and Stockman and Sharpe [2000]) are constructed by 1) first experimentally measuring the cone fundamentals in psychophysics (from color-vision deficient observers), 2) calibrating the results with a set of RGB CMFs in Stiles and Burch [1959] (which uses a different set of primary lights from the CIE 1931 RGB CMFs) to derive a best-fit linear transformation, and 3) applying the linear transformation to the CMFs to derive a “clean” set of cone fundamentals.

3 Post-Receptoral Color Encoding: Opponent Processes

Cone-response encoding can perfectly explain the trichromatic theory of color vision, where any color can be mixed from three other colors. The trichromatic theory of color has a perfect neural basis: human visual system has three classes of cones, so color is a three-dimensional system. But the trichromatic theory is not concerned with our subjective experience of color that we encounter on a daily basis. Here are two examples that highlight the difference between perceptual color experience and physical color mixing.

First, when we see an orange color, we feel that it has a little bit of yellow in it and a little bit of red in it. Even though there are many ways to produce orange, some of which do not require mixing yellow and red lights, we cannot help but perceptually feel that orange combines yellow and red. Second, when we mix a red light with a green light, we get yellow, but perceptually, if we stare at yellow, most people would not say that yellow has contributions from red or green.

Hering [1878, 1964] hypothesized that, perceptually, there are four primary hues, which form two opposing pairs. Opposing hues cannot co-exist, perceptually, in a color. Any hue can be produced by combining two non-opposing hues. The four hues are: the Yellow and Blue opposing hues and the Red and Green opposing hues. Hering also considered light-dark as another opposing pair: no color can be simultaneously light and dark. In his theory, color vision is still a three-dimensional system, where the three axes are: Yellow-Blue axis, Red-Green axis, and light-dark axis. Any color, a point in this 3D space, is produced by mixing some amount of Red *or* Green, some amount of Yellow *or* Blue, and some level of lightness.

The opponent theory seems to contradict the trichromatic theory, which was dominant for the most part of the history — because it has both a solid psychophysical and neural basis. First, the color matching experiment quantitatively shows that behaviorally humans could match a color by mixing three other colors. In contrast, Hering had only a qualitatively description of perceptual mixing. His description was something like “*after this blue comes blue of increasing redness... (blue violet, red violet, purple red), until the last trace of blueness vanishes in a true red.*” [Hering, 1964, p. 41]. To Hering’s theory’s rescue, Jameson and Hurvish per-

²One subtlety is that Baylor et al. [1987] used suction electrode to measure electrical responses, so they obtained only the relative absorbance not the absolute absorption of the pigments. So what they actually ended up doing is to use the psychophysical CMFs to fit the peak axial absorption and calculate the cone fundamentals, and show that the regressed CMFs from the so-obtained cone fundamentals match that from psychophysics.

formed an experiment, called **hue cancellation experiment**, providing the first quantitative, psychophysical evidence of the opponent processes [Jameson and Hurvich, 1955; Hurvich and Jameson, 1957].

Second, the trichromatic theory has a clear neural and physiological basis (i.e., wavelength encoding by cone responses), and the physiological data match the behavioral data very well, as shown before. So a natural question is: are there neural mechanisms that can account for the opponent processes and, if so, how does that mechanism relate to the encoding mechanisms by the cone photoreceptors?

It turns out that we do need a set of new neural mechanisms to start accounting for the opponent processes. Not only do these new mechanisms *not* contradict the cone encoding mechanisms, they build on top of the cone encodings and operate post-receptorially. Schrödinger [1925, 1994] synthesized the earlier *zone theory* by von Kries [1905] and argued that the trichromatic theory and the opponent processes were nothing more than different stages of color encoding in the visual system. That said, while these new neural mechanisms seem to have what it takes to form the basis for the behavioral opponent observations, they do not fully explain those observations yet; the link between the two is still very much an open research question.

The rest of this section will discuss the hue cancellation experiment and the quest for a neural and physiological basis in more detail.

3.1 Hue Cancellation Experiment

In a landmark study, Jameson and Hurvich [1955] (while working for Eastman Kodak in Rochester) quantitatively measured the perceptual color opponency using a behavioral experiment. The participant is given a test light and is asked to first judge whether the light appeared blue-ish or yellow-ish. If the test light is judged to be blue-ish, the participant is then given a yellow-ish *cancellation light* (e.g., a spectral light at 588 nm), and is asked to adjust the intensity of the cancellation light so that the mixture of the test and cancellation light perceptually appears neither blue nor yellow. If the test light is judged to be yellow-ish, the participant is then asked to adjust the power of a blue-ish cancellation light (e.g., a spectral light at 467 nm) so that the test-cancellation mixture is again neither blue nor yellow. We sweep the spectrum from about 400 nm to 700 nm for the test light of equal energy, and record the energy of yellow or blue cancellation light needed at each step.

The result for one subject is shown in Figure 8 A, where the *y*-axis is showing the intensity of the yellow and blue cancellation light, i.e., the strength of blue-ness and yellow-ness of the test light. For the reference, we attached a colorbar showing roughly the color of the test light between 400 nm and 700 nm, but take this color visualization as a huge grain of salt, since it is almost certain that your display will not be able to actually render the colors of the spectral lights.

Unsurprisingly, we get two peaks, one in the blue range and the other at the yellow range, indicating, respectively, that the participant need a lot of the yellow and blue cancellation lights in those two regions. The test light at about 500 nm requires no cancellation light, indicating light there, which roughly has a green-ish color is yellow-blue neutral: it naturally looks neither blue nor yellow.

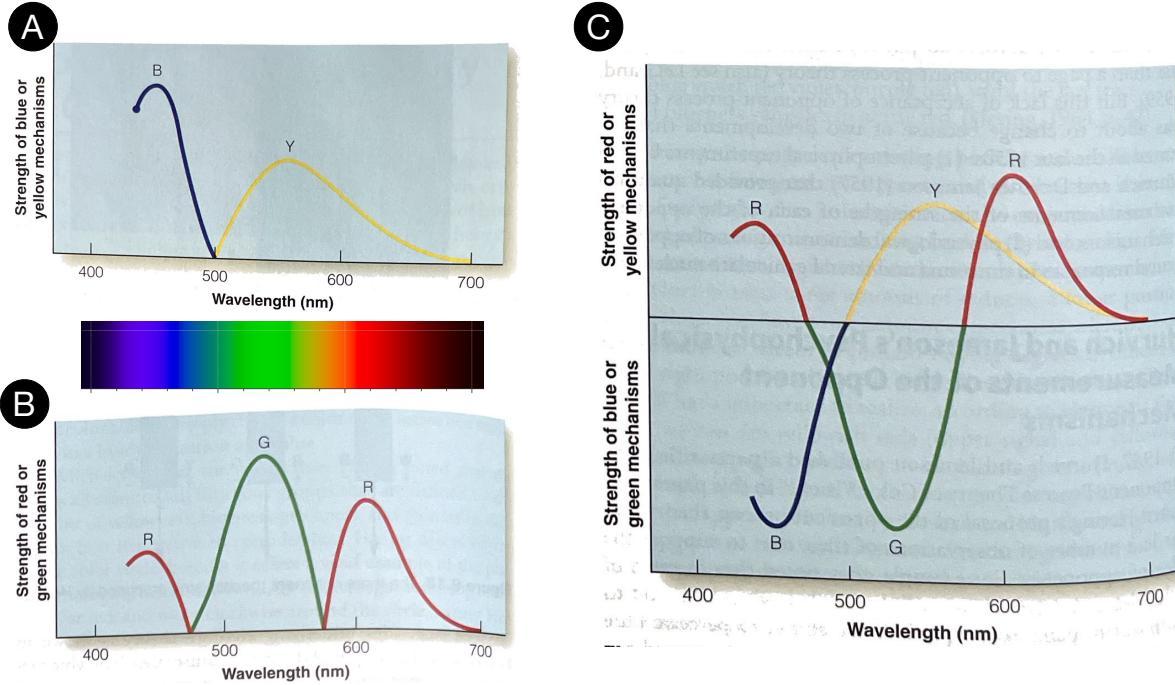


Figure 8: Measurements from the hue cancellation experiment in [Jameson and Hurvich \[1955\]](#). (A) the Blue-Yellow measurement; the y -axis shows the intensity of the Yellow/Blue cancellation light, i.e., the relative strength of the “Blue-ness” and “Yellow-ness” in the test light. (B) the Red-Green measurement; notice the two zero-crossings for Green. (C) The same data as A and B except we invert the Blue and Green curves so the y -axis is interpreted as the strength of Red-ness and Yellow-ness. Adapted from [Goldstein and Brockmole \[2017\]](#), Fig. 9.20 – 9.22]. The colorbar is generated using the Colour Python package [\[NumFOCUS\]](#).

Jameson and Hurvich then repeated the same experiment, but this time measuring the red-green opponent process, where the two cancellation lights are a 700 nm red-ish light and a 490 nm green-ish light. The results are in Figure 8 B, where the y -axis indicates the amount of red-ness and green-ness in the test light. Two observations are worth noting. First, while it is unsurprising that long-wavelength lights have a strong red component, it is perhaps surprising that short-wavelength lights appear red-ish too. But that become less surprising when we realize that short-wavelength lights (shorter than pure blue) appear violet, which perceptually is a red-ish blue. Second, because of the two red-ish regions over the spectrum, the entire red-green curve has two zero-crossings, one at about 470 nm and the other near 570 nm: pure blue and pure yellow look neither green nor red.

Figure 8 C summarizes the two sets of data by inverting the blur section of the curve in A and the green section of the curve in B. That way, the y -axis can be simply interpreted as the relative strength of red-ness and yellow-ness over the spectrum.

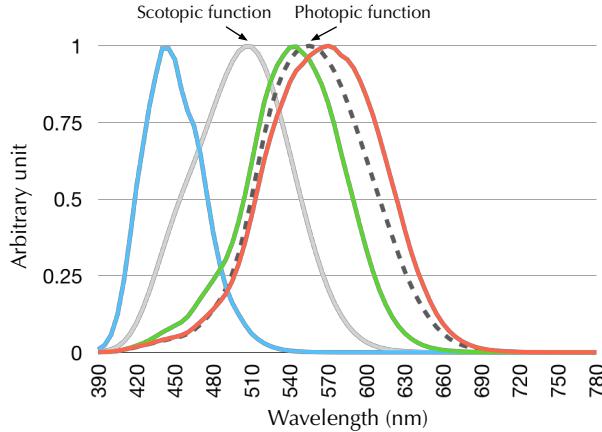


Figure 9: The grey solid curve is the scotopic luminous efficiency function (CIE 1951 standard; based on [Wald \[1945\]](#) and [Crawford \[1949\]](#)). The grey dashed curve is the photopic luminous efficiency function (CIE 2008 “physiologically-relevant” 2-deg function; based on [Sharpe et al. \[2005, 2011\]](#)) The other three curves are the cone fundamentals, shown for the reference.

3.2 Light-Dark Mechanism and Luminous Efficiency Function

[Hurvich and Jameson \[1957\]](#) also performed a measurement of the white-black (light-dark) opponent process, asking participant to assess the “whiteness” of spectral lights between 400 nm and 700 nm of equal power. A more modern method to measure the luminance mechanism is heterochromatic flicker photometry, where we alternate between a test light and a fixed reference light at a frequency of, say, 25 Hz. We adjust the intensity of the test light so that the alternation produces no visual flickering, at which point we say the two lights produce the same level of luminance [[Sharpe et al., 2005, 2011](#)]. We again sweep the entire visible spectrum for the test light, and record the relative intensity at each step. The so-obtained function is called the **luminance efficiency function** (LEF). The dashed gray curve in Figure 9 shows a modern version of the photopic LEF (the so-called CIE 2008 “physiologically-relevant” 2-deg function)³.

The way to interpret the LEF is that the y -axis is inversely proportional to the light power at each wavelength needed to produce the same level of perceptual brightness. The photopic LEF at 509 nm is about 0.5, half of that at 555 nm. It means we need twice as much power at 509 nm to produce the same level of brightness as that at 555 nm. It also explains the word “efficiency” in the name: if a wavelength needs less power to produce a criterion level of brightness, the wavelength is more efficient in its use of power. The way LEF is obtained, however, does *not* permit us to interpret the result as the relative brightness at different wavelengths. That is,

³In later research by Jameson and Hurvich, their white-black function was made equal to the CIE 1924 luminous efficiency function [[Hurvich and Jameson, 1955](#), p. 604], which is known to have severe flaws at low wavelengths and which is later corrected by [Judd \[1951\]](#) and [Vos \[1978\]](#). Compared to the Judd and Vos corrections, the function shown here has the advantage of being “physiologically relevant” in that the LEF is a linear combination of the cone fundamentals, whereas both the CIE 1924 LEF and its later corrections are not intentionally designed to be a linear combinations of anything.

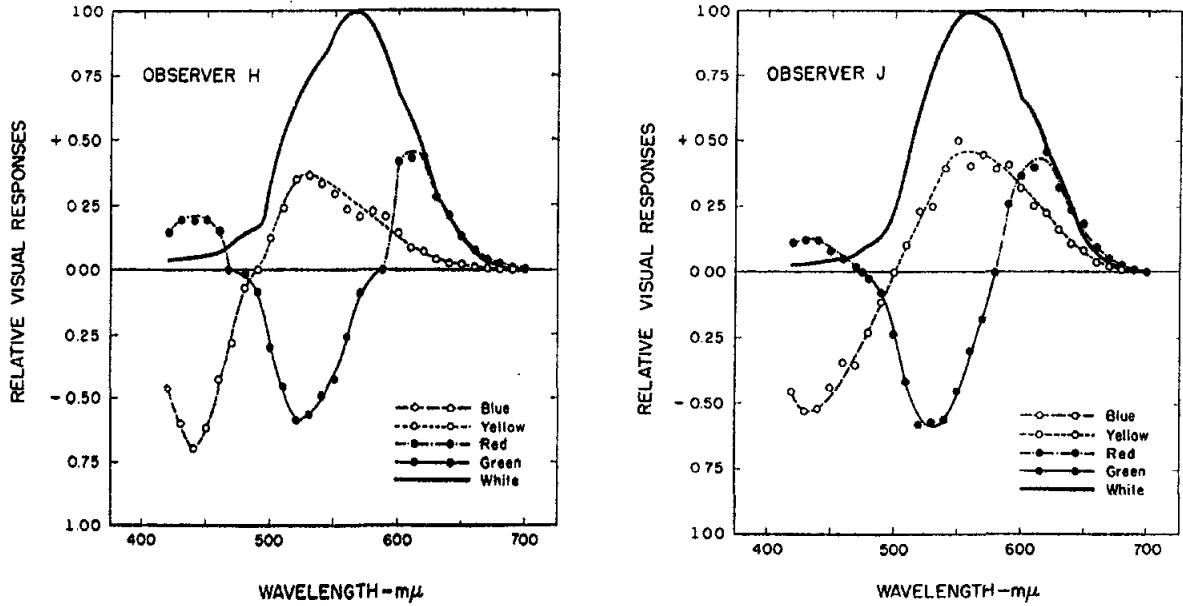


Figure 10: The solid curve is the white-black measurement, indicating the amount of whiteness in a light across the spectrum. The white-black curve in theory matches the luminous efficiency function. The two plots are for two participants. From [Hurvich and Jameson \[1957, Fig. 4\]](#).

555 nm is not twice as bright as 509 nm. This is similar to our interpretation of the cone fundamentals.

For comparison, the gray curve in Figure 9 is the scotopic LEF. The CIE 1951 scotopic LEF synthesizes the psychophysical measurements from [Wald \[1945\]](#) and [Crawford \[1949\]](#). Both used a threshold method where they measured the light intensity at each wavelength needed to produce a just detectable flash. Note that the photopic LEF peaks at about 555 nm and the scotopic LEF peaks at about 507 nm.

As a result, the relative brightness of longer-wavelength colors and shorter-wavelength colors are inverted when our vision transitions from a cone-mediated photopic vision to a rod-mediated scotopic vision. This phenomenon is called the **Purkinje shift**. In the words of [Glassner \[1995, p. 21\]](#), “When the sun is still above the horizon, your cones are active, and the yellow flower will appear lighter than the leaves because yellow is closer to peak of the photopic sensitivity curve than dark green. When the sun has set and light levels are lower, your rods are the principal sensors. The scotopic sensitivity curve is more responsive in the shorter wavelengths, so the green leaves will now appear relatively lighter than the yellow flower, though both will of course be much darker due to the lower amount of incident light.”

Figure 10 puts the three opponent measurements in one plot (the two plots are for two separate participants). Compare this plot with the cone fundamentals in Figure 1. Once again, a light with its SPD can be reduced to three-dimensional point, using Equation 1, except 1) instead of the three cone fundamentals we use the three opponent functions and 2) instead of getting the

three cone responses we get the strength of the three opponent mechanisms. Effectively, the hue cancellation curves and the light-dark curve construct a new three-dimensional color space. We call this the **hue-opponent** space, and we will return to this space in Chapter 3.4 and discuss how this space relates to the colorimetric spaces we have discussed so far.

3.3 Neural and Physiological Basis

The hue cancellation experiment solidifies Hering's opponent theory, and the next nature step in the scientific quest is to understand what underlying neural and physiological mechanisms can account for the opponent processes.

There are Both Spectrally-Opponent and Non-Opponent Neurons

There are RGC and LGN neurons that show opponent properties. [Svaetichin \[1953, 1956\]](#) and [Svaetichin and MacNichol Jr \[1958\]](#) are the first to identify opponent neurons in a fish retina; they recorded from horizontal cells. [De Valois et al. \[1958, 1966\]](#) measured the responses of LGN neurons in macaques using monochromatic lights, and found spectral opponent neurons, which get excited or inhibited depending on the wavelengths. [A–D](#) in Figure 11 show the recordings of four classes of opponent cells. [A](#) shows a class of LGN cells whose firing rate exceeds the spontaneous rate under long-wavelength, red-ish lights and whose firing rate drops below the spontaneous rate under short-wavelength, blue-ish lights. These cells are denoted +R-G (red-ON/green-OFF) cells. [B](#), [C](#), and [D](#) show that there exists +G-R, +B-Y, and +Y-B cells, respectively.

[De Valois et al. \[1966\]](#) also identified non-opponent cells, whose responses are universally inhibited or excited across the spectrum, as shown in [E](#) and [F](#) in Figure 11, respectively. These neurons are still wavelength-sensitive, but their responses are either universally excited or universally inhibited across the spectrum, unlike the spectrally-opponent neurons whose responses change polarity across the spectrum.

There are Potential Neural Circuitries for Opponent and Non-opponent Cells

What are some of the underlying visual pathways that could potentially give rise to these spectral tuning curves? Recall that LGN cells/RGCs have antagonistic Receptive Fields (RFs), and the antagonism seems to be a perfect mechanism to implement the opponent process. This suggests that in order to understand the opponent cells we must study their RF structures.

Much of the early work is done by [Wiesel and Hubel \[1966\]](#). While De Valois and his collaborators used diffuse lights to illuminate a large visual field, [Wiesel and Hubel \[1966\]](#) used both small spot lights that stimulated the center of the RF and larger lights that covered the entire RF. By comparing the responses under these two stimuli across different wavelengths (and white), they suggested potential RF structures of both opponent and non-opponent cells in macaque LGN. [Derrington et al. \[1984\]](#) designed a clever experiment that explicitly tied cone responses to LGN cell responses and thus more directly revealed the RF structure.

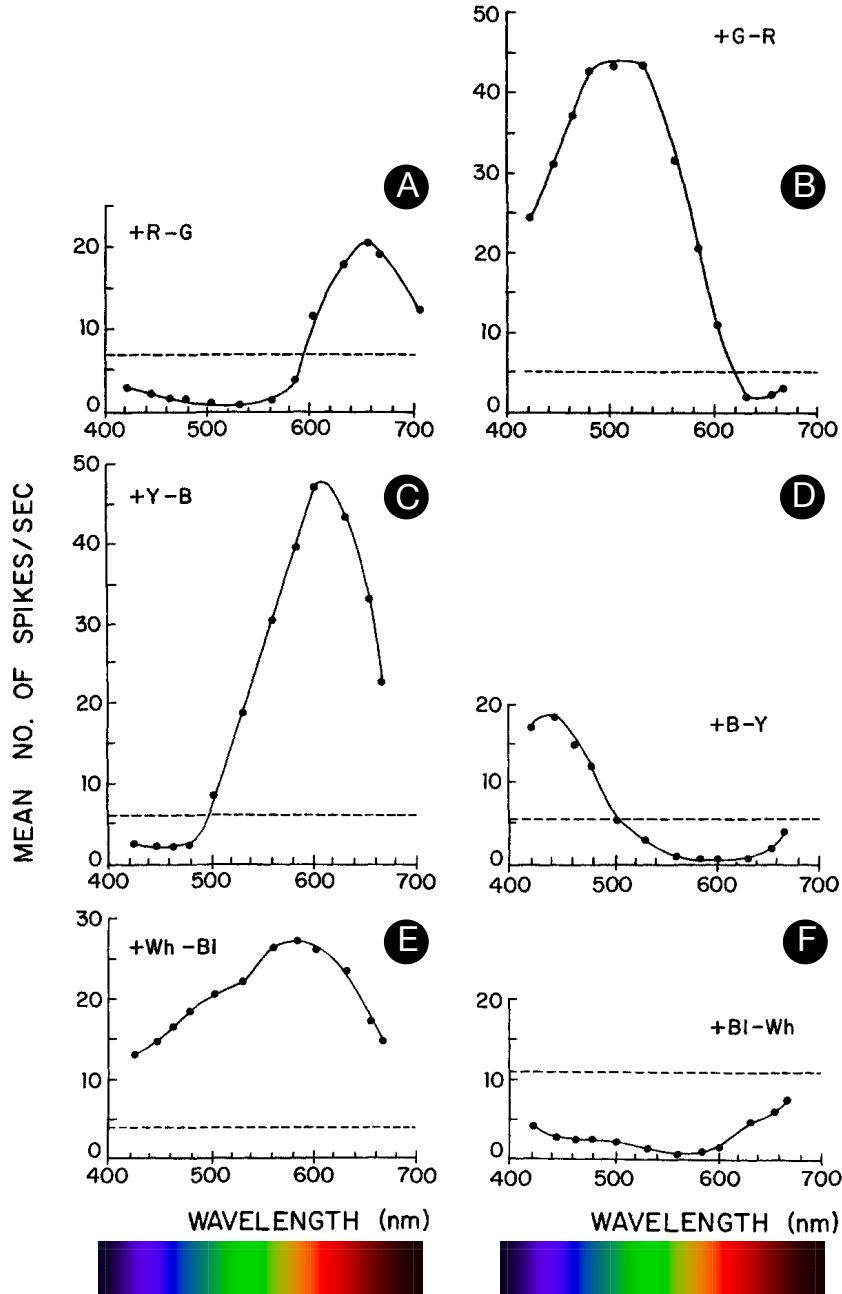


Figure 11: Responses of six typical classes of LGN neurons to incremental flashes of varying wavelengths. y -axis shows the spikes/second under spectral lights of equal energy. Each curve represents a particular energy level. (A): these cells are excited (activity exceeds the spontaneous firing rate) by red hues and inhibited by green hues, denoted +R-G cells. (B): +G-R cells. (C): +B-Y cells. (D): +Y-B cells. (E): non-opponent excitatory cell. (F): non-opponent inhibitory cell. From [DeValois and DeValois \[1990, Fig. 7.5\]](#), which is adapted from [De Valois et al. \[1966, Fig. 9–12, 15–16\]](#).

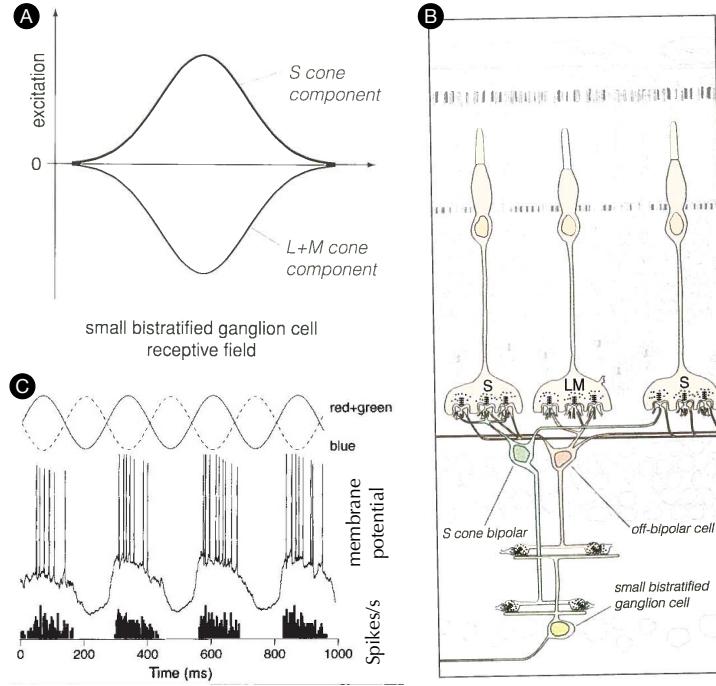


Figure 12: The small bistratified RGCs might be the substrate for the Y-B pathway. A: illustration of the receptive field structure of a small bistratified RGC, which is S-on and L/M-off (there are also S-off and L/M-on ones); from [Rodieck \[1998\]](#), p. 348. B: a small bistratified RGC receives excitatory inputs from S cones through the S-cone bipolar cells and inhibitory inputs from L and M cones through another class of bipolar cells; from [Rodieck \[1998\]](#), p. 346. C: membrane potential and spike rate of small bistratified cells under periodic, out-of-phase blue-yellow lights; adapted from [Dacey and Lee \[1994\]](#), Fig. 3C].

Before getting into the details, it is worth reminding ourselves that studying LGN cells and RGCs is equivalent, since different classes of RGCs project to distinct LGN layers with virtually the same RFs: midget RGCs project to the Parvocellular layers (P cells) in the LGN (forming the P pathway/stream), parasol RGCs project to the Magnocellular layers (M cells) in the LGN (forming the M pathway/stream), and bistratified RGCs project to the Koniocellular layers (K cells) in the LGN (forming the K pathway/stream).

The visual pathway for the Y-B opponent cells seems to be clear. [Derrington et al. \[1984\]](#) showed that some LGN cells receive antagonistic inputs from S cone vs. L and M cones. [Dacey and Lee \[1994\]](#) later identified that the small bistratified RGCs (which project to the K cells in the LGN) are responsible for carrying such signals. The small bistratified RGCs are excited by S cone responses and inhibited by L and M cone responses (or vice versa). Since blue-ish lights produce strong S cone responses and red/green lights produce strong L/M cone responses (recall red + green is yellow), it stands to reason that if a cell is excited by S cones and inhibited by L and M cones, it would give a vigorous on-response under blue lights and a vigorous off-response

under yellow lights, producing the kind of yellow-ON/blue-OFF spectral tuning curve that we see in Figure 11 C.

Figure 12 A illustrates the potential Receptive Field (RF) of a yellow-ON/blue-OFF small bistratified cell, and B shows the neutral circuitry that gives rise to such an RF (but also see [Field et al. \[2007\]](#)). The small bistratified RGC have a center-only RF, which receives excitatory responses from a S-cone bipolar cell and inhibitory responses from another class of bipolar cells that are connected to L and M cones. [Dacey and Lee \[1994\]](#) records both the membrane potential and the spiking rate of a small bistratified RGC, shown in C, under periodic, out-of-phase blue and yellow (red+green) lights. The cell's responses are the strongest under maximum yellow light (maximum excitatory S cone responses) and minimum blue lights (minimum inhibitory L and M cone responses).

[Derrington et al. \[1984\]](#) showed that most of the midget RGCs (and thus P cells in LGN) are either excited by L cone responses and inhibited by M cone responses (L-ON/M-OFF) or the other way around. Given that, loosely, L cones are excited by red-ish lights but not so much by green-ish lights and M cones behave oppositely, it stands to reason that L-ON/M-OFF cells produce vigorous on-responses (above spontaneous rate) under red lights and vigorous off-responses (below spontaneous rate) under green lights, giving a spectral tuning curves shown in Figure 11 A.

The actual RF structure of these cells takes two forms [[Wiesel and Hubel, 1966](#)]. Some of these cells have a center-surround RF, so there are four combinations: L+/M- (L center-ON/M surround-OFF), L-/M+, M-/L+, and M+/L-. Other midget RGCs have no center-surround arrangement. The excitatory and inhibitory regions have the same spatial extent. Either way, signals from the L cones and M cones are antagonistic in these cells.

Finally, the parasol RGCs (and thus M cells in LGN) seem to be the most probable source for the luminance opponent mechanism [[Lee et al., 1988](#)]. These cells do have a center-surround RF but the L cones and M cones contribute to both the center and the surround [[Wiesel and Hubel, 1966](#)]; S cones seem to be contribute little, if any, to these cells [[Lennie et al., 1993](#)]. When the total excitation by the L and M cones to the center out-weighs the total inhibition to the surround, the entire cell appears to be excited by L and M cone responses, giving a broadband, non-opponent spectral tuning curve in Figure 11 E; otherwise we see a tuning curve like Figure 11 F.

A Cone-Opponent Model for Color-Opponent Mechanisms

It is clear that there are cells that receive opponent cone signals; the spectral tuning curves of these cells seem to largely account for the perceptual opponent mechanisms. Based on these observations, [Derrington et al. \[1984\]](#) proposed a *cone-opponent* color space, which is now commonly used (in color science and, to a large extent, visual neuroscience) to give a first-order approximation of the perceptual *color-opponent* processes. The color space is now famously known as the DKL color space (named after the three authors; the L is Peter Lennie, who was once on the faculty at University of Rochester and served as the Provost).

- The Y-B channel is given by $aS - (bL + cM)$, where a , b , and c are all positive values rep-

resenting the contributions of the S, L, and M cones to the Y-B opponent process. It is generally said that this signal is delivered by the Koniocellular pathway.

- The R-G channel is given by $dL-eM$, where d and e are all positive values representing the contributions of the L and M cones to the R-G opponent process. This opponent signal is generally said to be delivered by the Parvocellular pathway.
- The Light-Dark or luminance channel is given by $fL+gM$, where f and g are all positive values representing the contributions of the L and M cones to the luminance channel. This luminance channel is meant to represent the LEF (Chapter 3.2), which generally is believed to be delivered by the Magnocellular pathway.

The DKL space operates not on raw cone responses but on response *contrasts* with respect to a perceptually neutral/achromatic color. The inherent assumption is that the achromatic color should have no strength in any of the three cone-opponent channels and be the origin in the cone-opponent space. The achromatic color depends on an observer's state of chromatic adaptation, a topic we will discuss later in the class. People usually fit data to regress the values of the free parameters, and the exact values depend on which cone fundamentals are used and the normalization convention. [Brainard \[1996\]](#) describes one such procedure.

Since the cone-opponent model operates on (contrast of) cone responses, a common theory of color vision is that it is a two-stage process: the wavelength encoding by cone photoreceptors followed by opponent encoding of cone responses post-receptorially. While the cone response encoding can perfectly explain the color matching experiments as we have seen earlier, the cone opponent encoding is only an approximation of the hue cancellation experiments, as we will see next.

3.4 There are Many Inconvenient Truths

The cone-opponent model is a good approximation for behavioral color-opponent mechanisms, but there are many inconsistencies between these two. Reconciling the two and thus elucidating how humans perceptually code opponent hues is still an open research question.

P and K Pathways Do Not Fully Account For R-G and Y-B Opponent Processes

The opponent neurons clearly have what it takes to start accounting for the perceptual opponent processes, but the spectral tuning curves of those neurons have only a weak correlation with the hue cancellation curves. Thus, it is unlikely that excitation and inhibition in opponent neurons cause our perception of red-green and blue-yellow opponency.

The most jarring difference appears in the R-G process. The R-G hue cancellation curve (Figure 8 C) shows two perceptually neutral colors, as there are two zero-crossings. However, the spectral tuning curve of the R-G neurons (Figure 11 A B) shows only one zero-crossing. These neurons do not predict the R-G neutral color in the short-wavelength range and, by extension, cannot explain the fact that short-wavelength violet-ish lights appear to have a red hue. [Derrington et al. \[1984\]](#) (also see [Wandell \[1995, Fig. 9.18\]](#)) shows a great deal of variation of

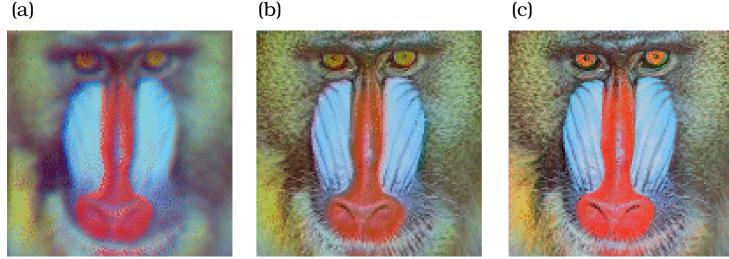


Figure 13: We take an image, decouple it into three channels: luminance, red-green, and blue-yellow. We then spatially blur one of the channels while keeping the other two channels unchanged and then reconstruct the image. Our vision is much more sensitive to spatial blurring in the luminance channel (a) than is to blurring in the red-green channel (b) and in the blue-yellow channel (c). This is the basis of chroma subsampling used in modern image and video compression algorithms. From [Wandell \[1995, Fig. 9.23\]](#).

the spectral tuning property within P cells, making them even less certain as the sole candidate for R-G opponent mechanism.

In fact, people have shown that the perceptual R-G hue cancellation data can be fit by $a'L - b'M + c'S$, where a' , b' , and c' are cone contributions [[Poirson and Wandell, 1993](#); [Bäuml and Wandell, 1996](#)]. Intuitively, the contribution by S cones in the short-wavelength range could give rise to a positive response there. However, there is no physiological evidence that L cone and S cone responses combine at some point in the visual pathway, suggesting the phenomenological nature of these models.

Even though the K pathway clearly shows the capability of carrying S vs. L+M signals, the latter do not accurately predict Y-B neutral signals and, thus, do not fully account for the Y-B hue opponency. That is, a color that leads to a null response (no significant increase or decrease compared to the spontaneous response rate) in the L-M channel is not perceptually pure yellow or pure blue [[Shevell and Martin, 2017, Fig. 4f](#)]. Similarly, a color that causes a null response in the S-(L+M) channel is not perceptually pure red or pure green. That is, null-response colors in the DKL cone-opponent space are not perceptually neutral in the hue-opponent space, implying fundamental discrepancies between cone-opponent and hue-opponent spaces.

M Pathway Does Not Fully Account For Luminance

The Magnocellular pathway (starting from the parasol RGCs) is said to be responsible for the dark-light opponent cells, but that poses a dilemma. We know that parasol RGCs have large RFs. A large RF is equivalently to applying an aggressive low-pass filter to the optical image; as a result, the M pathway has a low spatial acuity. So if the M pathway is fully responsible for mediating our luminance perception, we should be insensitive to spatial blurring (low-pass filtering) in the luminance signal. But the result is the opposite: our vision is very sensitive to spatial blurring in the luminance channel (but relatively insensitive to blurring in the two color opponent channels).

This is illustrated in Figure 13, where we take an image, decouple it into three channels: luminance, red-green, and blue-yellow. We then spatially blur one of the channels while keeping the other two channels unchanged and then reconstruct the image. Our vision is much more sensitive to spatially blurring in the luminance channel (a) than is to blurring in the red-green channel (b) and in the blue-yellow channel (c); in fact, this is the basis of **chroma subsampling**, a key step in modern image and video compression algorithms. This suggests that the M pathway alone cannot be exclusively responsible for our luminance perception.

[Gouras and Zrenner \[1979\]](#) also shows that P cells, which are ordinarily thought of as L-M spectrally-opponent, could also give a LEF-like spectral tuning curve as if it acts as the luminance channel. The reason is that the surround signals reach a cell later than do the center signals, so at a high frequency the out-of-phase center-surround signals can actually come in the same phase.

Hue-Opponent Space is Not a Linear Transformation from Cone Space

It is perhaps not surprising, by now, that if there is a color space that can fully account for the perceptual coding of opponent hues, it is never going to be a linear transformation from the LMS space (or any other space that is a linear transformation away from the LMS space, e.g., the CIE 1931 XYZ space or the DKL cone-opponent space).

As we have seen above, for instance, the DKL space [[Derrington et al., 1984](#)], which is a linear transformation from the LMS cone space, does not fully account for the perceptual opponent processes, e.g., does not predict any unique hue. People have shown that one can construct a linear transformation from the LMS space that can accurately predict three of the four unique perceptual hues by fitting data from psychophysical measurements that do not presuppose the existence of opponent mechanisms [[Poirson and Wandell, 1993](#); [Bäuml and Wandell, 1996](#)], but they cannot predict the fourth unique hue. [Schrödinger \[1925\]](#) also estimated a linear transformation between the cone response space and the hue-opponent space based on the four unique hues, but the transformation could not accurately predict the achromatic color (also see the commentary by Zaidi in [Schrödinger \[1994\]](#)).

The reason is that perceptually unique red and green hues are not *collinear* with white, the achromatic color that is perceptually neutral in both the Y-B and R-G channel, i.e., does not appear yellow, blue, red, nor green⁴. That is, red, white, and green do not lie on a line. Why is this significant? Assuming there was a linear transformation T from the cone responses to the strengths of the hue-opponent mechanisms:

$$\begin{bmatrix} Y/B \\ R/G \\ \text{Lum} \end{bmatrix} = T \times \begin{bmatrix} L \\ M \\ S \end{bmatrix} \quad (10)$$

Both unique red hue ($[L_R, M_R, S_R]$) and unique green hue ($[L_G, M_G, S_G]$) have no yellow (or blue) hue, so their response in the Y-B channel response would be 0:

⁴Again, what is considered achromatic depends on the observer's adaptation state; there is no single

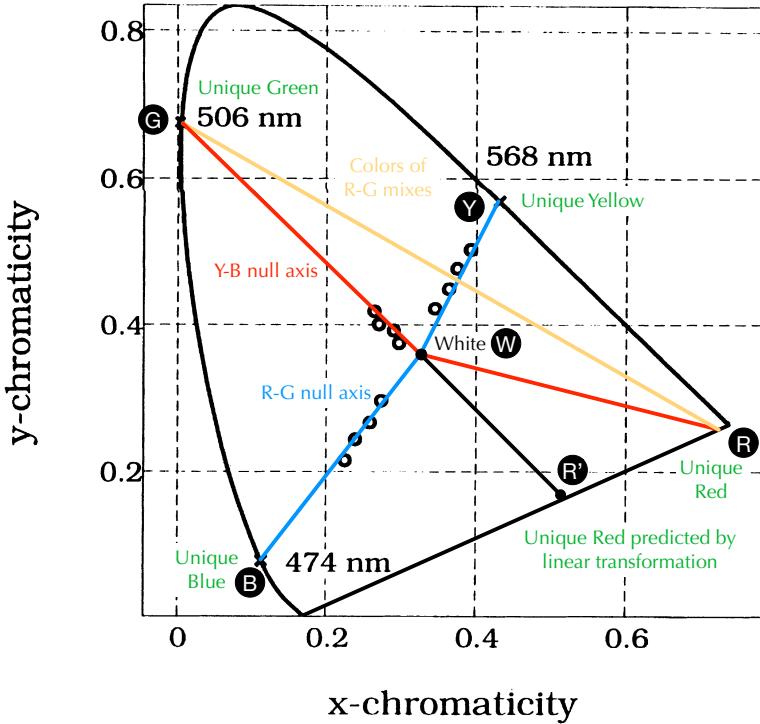


Figure 14: Circles are unique hues derived from psychophysics reported in [Bäuml \[1993\]](#). Fitting lines and extrapolating the lines give us estimations of unique hues that are spectral colors. Three of the four unique spectral hues (blue at 474 nm, green at 506 nm, and yellow at 568 nm) can be accurately predicted by a linear transformation constructed by [Bäuml and Wandell \[1996\]](#), but not the unique red hue. The fact that the red, white, and green are not collinear suggests that there is no linear transformation between the hue-opponent space and the cone space. Adapted from [Bäuml and Wandell \[1996, Fig. 12\]](#).

$$\begin{bmatrix} 0 \\ | \\ | \end{bmatrix} = T \times \begin{bmatrix} L_R \\ M_R \\ S_R \end{bmatrix}, \quad \begin{bmatrix} 0 \\ | \\ | \end{bmatrix} = T \times \begin{bmatrix} L_G \\ M_G \\ S_G \end{bmatrix} \quad (11)$$

Therefore, any mixture of the unique red hue and the unique green hue would not appear to have a yellow hue either:

$$\begin{bmatrix} 0 \\ | \\ | \end{bmatrix} = T \times \begin{bmatrix} aL_R + bL_G \\ aM_R + bM_G \\ aS_R + bS_G \end{bmatrix}, \quad (12)$$

where a and b are contributions of red and green to the mixed color. However, we know that when we mix red with green colors we get yellow. The fact that two colors without any yellow hue can generate a color that does have a yellow hue means the hue-opponent space cannot be a linear transformation from the LMS cone space.

Figure 14 illustrates this point with some real data. The empty markers are three sets of perceptually unique hues (which do not have to be spectral colors) measured psychophysically in Bäuml [1993]. When we fit a straight line across each set of unique hues and extrapolate the line we can estimate what spectral colors are unique hues (blue \textcircled{B} , green \textcircled{G} , and yellow \textcircled{Y}). No spectral color is seen as a unique red hue (all spectral red-ish colors appear to have a yellow hue), which requires a mixture of unique blue hue and a spectral red to cancel the yellow percept [Dimmick and Hubbard, 1939b; Larimer et al., 1975] (and also see the commentary by Zaidi in Schrödinger [1994]). Dimmick and Hubbard [1939a] measured that unique red hues \textcircled{R} are complementary to a spectral light at 494 nm; that is, spectral light at 494 nm, white \textcircled{W} , and unique red hues should fall on a straight line⁵.

Bäuml and Wandell [1996] constructed a linear transformation from the cone space to the hue-opponent space that can accurately predict the unique spectral hues of blue, green, and yellow. It is comforting, and corroborates others [Larimer et al., 1974], that blue \textcircled{B} , white \textcircled{W} , and yellow \textcircled{Y} are collinear, as would be required by a linear transformation from the cone space to the hue-opponent space: mixing colors that have no green or red hue will not give a color that does. But clearly the predicted red hue \textcircled{R}' deviates significantly away from red hue \textcircled{R} from actual measurements. If we connect the unique green hue \textcircled{G} and unique red hue \textcircled{R} , the line would not cross \textcircled{W} . This suggests that a simple linear transformation does not exist; at least the Y-B null-response axis is not linear with respect to the cone responses. Non-linear models have been proposed [Larimer et al., 1975; Shevell and Martin, 2017].

4 Evolution of Color Vision

See Lamb [2013, 2020, 2022], Lamb et al. [2007], Shichida and Matsuyama [2009], Jacobs [2009], and Bowmaker [2008, 1998] for comprehensive discussions. We provide a concise summary of what is relevant to our discussions in this chapter.

4.1 The Rise of Trichromacy

Proto-Vertebrates Had Four Cone Opsin Genes

When studying the evolution of vision and performing comparative studies of vision across species, we must understand the differences in all the molecules that participate in phototransduction and the subsequent neural circuitry. The photopigment/opsin itself is one of the most important components (e.g., pigments differ in their peak absorption wavelength) and its evolution is the most well-understood. It is main the focus of our discussion.

⁵Of course, it is conceivable the result might vary in population and depend on the adaptation state (i.e., what is considered white/achromatic).

A primordial opsin gene that existed in ancestral bilateria (before the separation of protostomes and deuterostomes), through duplications and mutations, gave rise to many opsin genes, some of which are even expressed in non-visual opsins (e.g., ipRGCs and RPEs). One of them is called the Long-Wavelength Sensitive (LWS) opsin gene. The LWS opsin gene was duplicated multiple times, each of which went through mutations, and eventually four opsin genes existed in proto-vertebrates. The sequence of duplications most likely went like the following. The LWS gene was duplicated, and the duplicated copy evolved to be Short-Wavelength Sensitive and is called the SWS opsin gene. The SWS gene was duplicated again. One of the copies is called the SWS1 gene. The other was duplicated once again. Of its two copies, one is called the SWS2 gene, and the other is called the RHL gene. RHL means “rod like”; it is given the name because its duplicate would later evolve to be expressed in rhodopsin (photopigment in rods).

As a result, proto-vertebrates (primitive animals from which the vertebrates evolved; think of them as ancestors to vertebrates but are not vertebrates themselves) possessed 4 opsin genes: LWS, SWS1, SWS2, and RHL. These genes were expressed in pigments that mediate photopic vision, just like modern cones, so we will simply call them cone opsins, but keep in mind that proto-vertebrate “cones” are most definitely different from modern vertebrate cones.

Generally, a modern form of a gene is different from its ancestral form, but for the simplicity we usually call them by the same name. This might be common source of confusion when studying evolutionary biology. For instance, modern vertebrates all have LWS opsin genes (that are slightly different); they are all evolved from, but most definitively different from, the ancestral LWS gene at the time of duplication that gave rise to the SWS branch. It is perhaps more rigorous to say “the ancestral LWS was duplicated; one copy evolved to become the modern LWS genes, and the other evolved to become the ancestral SWS gene, which was duplicated again; its two copies are the ancestral SWS1 gene and the ancestral SWS2 gene”, but you can see how cumbersome this would be. Even with this it is sometimes unclear at which point in the evolution does “ancestral” refer to.

Most Vertebrates are Tetrachromatic and Most Mammals are Dichromatic

Then something remarkable happened. During the Cambrian epoch and at around 530 to 500 million years ago (Mya), there were two rounds of Whole Genome Duplication (WGD), a.k.a., polyploidy. Each WGD made a complete copy of all the genes. So two rounds of WGD would quadruple the number of genes. Each WGD was followed by a relatively short period of genome instability with extensive loss of genes. As a result, not all four copies (a result of two rounds of WGD) of a gene were retained. WGD in general is a major source of speciation, and the two rounds of WGD were responsible for the explosion of new species at the end of the Cambrian epoch (Cambrian explosion). WGDs increased the number of genes and made more genes available for mutation. As a result, there was a sudden radiation of species and diversification of life. Perhaps most importantly to human evolution, vertebrates appeared after the first round (1R) WGD.

As far as the four cone opsin genes are concerned, all their copies were lost except two copies of RHL, which are now called RH1 and RH2. So after the two rounds of WGD, our ancestral vertebrates possessed five cone opsin genes: SWS1, SWS2, RH1, RH2, and LWS. RH1 evolved

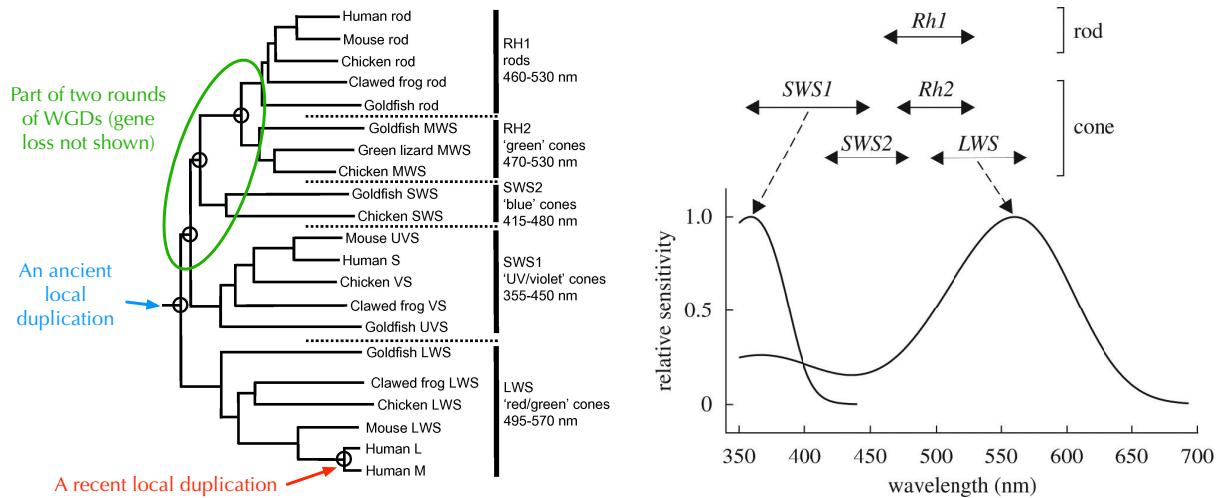


Figure 15: Left: Phylogenetic tree of vertebrate visual opsins; adapted from [Bowmaker \[2008, Fig. 2\]](#). Note how humans are trichromatic (three cone opsins), chicken (a non-mammalian vertebrate) is tetrachromatic (four cone opsins), and mouse (a mammal) is dichromatic (two cone opsins). Right: approximate spectral sensitivity of the five visual opsins; from [Jacobs \[2009, Fig. 1\]](#).

to be expressed in rhodopsins (pigments for rods) to mediate scotopic vision in vertebrates; the other four evolved too but retained their ability to mediate photopic vision. This is why most vertebrates are tetrachromatic. From the sequence of duplications, we can also deduce that the rod pigment evolved after all four classes of cone pigment were already present [[Okano et al., 1992](#)]. In fact, rod signals largely piggyback onto the pre-existing circuitry for cone signaling [[Lamb, 2016](#)].

Early mammals arose at the age of dinosaurs and were nocturnal, so they had no need for strong color vision and lost two of the four cone opsin genes. Only LWS and SWS1 genes were retained. This is why most mammals are dichromatic.

Figure 15 shows the phylogenetic tree of vertebrate visual opsins, of which there are five classes. I say “class” of opsin genes here because there are variations between, say, SWS1 gene in humans and that in mouse. Circles represent gene duplications. The first duplication on the ancestral LWS gene was local and gave rise to the SWS branch. The other duplications within the SWS branch are part of the two rounds of WGDs (which, recall, were followed by gene losses, which are omitted in the diagram). The local duplication within the LWS branch gave humans trichromacy, which we will discuss next.

LWS Gene Duplication in Catarrhini Gave Human Trichromacy

Then primates evolved from some mammals, first to prosimians, and then anthropoid (higher primates) split off from prosimians. Anthropoids include Platyrrhini (broad noses) and Catarrhini (narrow noses). Platyrrhini evolved to modern New World (South America) Monkeys.

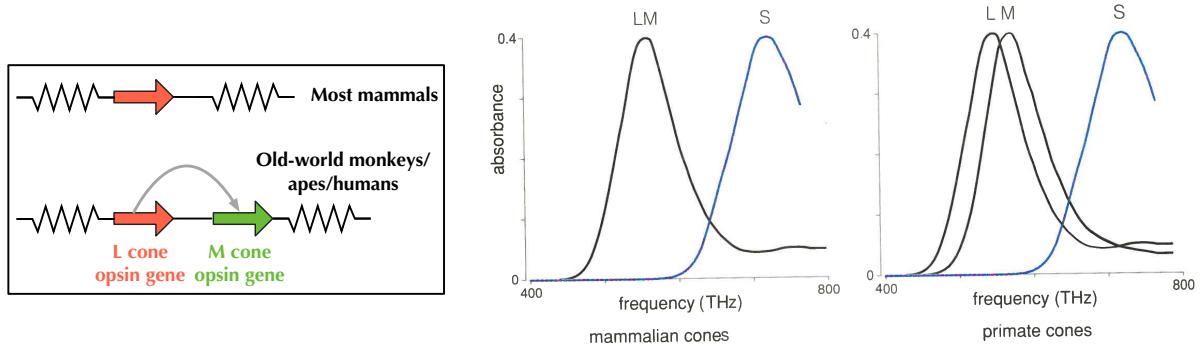


Figure 16: Left: LWS duplication gave two copies of the same gene in the X chromosome, and subsequent mutations to both copies give rise to modern L and M cone opsin genes. Middle and Right: spectral sensitivities of the two cone pigments in mammals (middle) and the three sensitivity spectra in primates (note the *x*-axis is frequency rather than wavelength); from Rodieck [1998, p. 218].

Catarrhini is the common ancestor of Old World (Africa and Asia) Monkeys, apes, and humans.

Trichromacy emerged in an early Catarrhini through gene duplication. The LWS gene is duplicated. One copy is evolved to be expressed in modern L cones, and the other evolved to be expressed in modern M cones. Since the duplication occurred “only” about 30 Mya, which is fairly recent from an evolutionary standpoint, the L and M cone pigments are very similar. 96% of the amino acids in the L cone opsin and M cone opsin are the same [Nathans et al., 1986]: they simply have not had much time to be mutated enough yet. With the SWS1 gene being expressed in S cones, all Catarrhini (including humans) are trichromatic. Figure 16 illustrates the duplication and the spectral sensitivities before and after the duplication.

What was the evolutionary pressure for the duplication that gave rise to the trichromacy? Since the duplication of LWS gene gave us medium-wavelength pigments that peak at green-ish lights, one interesting theory is that the duplication offered the ability to distinguish red-ish colors from green-ish colors in order to select ripe from unripe fruit (or ripe fruits against the background of green leaves), which had an obvious evolutionary advantage [Hunt et al., 1998; Bowmaker, 1998, p. 544].

4.2 The Rise of Scotopic vs. Photopic Vision

Only Jawed Vertebrates Have a “Modern” Scotopic Vision

Vertebrates arose after 1R. 1R is important for the rod-cone duplex vision in today’s vertebrates. RH1, which eventually is evolved to be expressed in rhodopsins, appeared after 1R. 1R also duplicated other non-opsin genes important for phototransduction (e.g., G proteins, PDEs, cGMPs, etc.) [Lamb, 2020]. These genes evolved to be expressed in distinct isoforms of many of the molecules that participate in the rod vs. cone phototransduction cascades.

Jawless and jawed vertebrates (from which humans evolved) split after 1R, so they both

possessed distinct scotopic and photopic vision. But there are differences in the scotopic vision between the two. Why? Because the second round (2R) WGD happened only to jawed vertebrates. 2R introduces a few new genes that allow jawed vertebrates to possess “true”, modern rod photoreceptors, where rod pigments are more thermally stable (lower dark noise) than cone pigments and regenerate faster than cone pigments [Lamb et al., 2007; Lamb, 2022]. In jawless vertebrates, rods also have a cone-like anatomical structure, but 2R changed the morphology of rods in jawed vertebrates so that their rods look differently than cones, e.g., having sealed-off discs. Interestingly, morphology apparently is not important for scotopic vision: northern hemisphere lampreys (a kind of jawless vertebrate) have rods that look like cones but physiologically behave like rods, e.g., very sensitive to lights, so they do mediate scotopic vision [Morshedian and Fain, 2015]. Whether you call these photoreceptors rods or cones is largely a matter of definition.

Scotopic Vision Predates Vertebrates

It is also interesting to note that proto-vertebrates, likely some chordate ancestors of ours, also possessed distinct photopic and scotopic vision way before vertebrates, except those ancestral scotopic photoreceptors utilized pinopsin [Okano et al., 1994], rather than rhodopsin, as their photopigments [Lamb, 2020; Sato et al., 2018]. Pinopsin evolved prior to 1R, and in fact was duplicated along with LWS cone opsin in proto-vertebrates from a Ciliary opsin (C-opsin).

After 1R, presumably RH1-expressed rhodopsins had better performance under scotopic condition so they gradually superseded pinopsins to be the pigments expressed in scotopic photoreceptors. Mammals do not possess pinopsins anymore, but some vertebrate species still use pinopsins for scotopic vision [Sato et al., 2018] and many vertebrates use pinopsins for non-vision functions such as regulating circadian rhythm (as they are expressed in the pineal organ) [Takanaka et al., 1998].

5 Color Vision Deficiencies

Normal color vision is trichromatic in that there are three classes of cone photoreceptors on the retina. If the retina is deprived of the functions of one of two cone classes, the color vision is no longer trichromatic. Instead of being represented in a three-dimensional space, a color would now be expressed in a two-dimensional or one-dimensional space. Individuals with two functioning classes of cones are **dichromatic**, and those with one functioning cone class are **monochromatic**. Dichromatic vision is further classified into three types: **Protanopia**, where L cones are missing, **Deutanopia**, where M cones are missing, and **Tritanopia**, where S cones are missing. An interesting form of dichromatic vision is called **small-field dichromacy**. At the central 20 arcmin of the fovea there are no S cones, so human vision is effectively dichromatic there. There are also individuals who have no functioning cones, only rods, and those people have **rod monochromacy**.

In addition to strict dichromatic vision, another important form of CVD is **anomalous trichromacy** vision, where individuals have all three cone types, but the spectral sensitivity

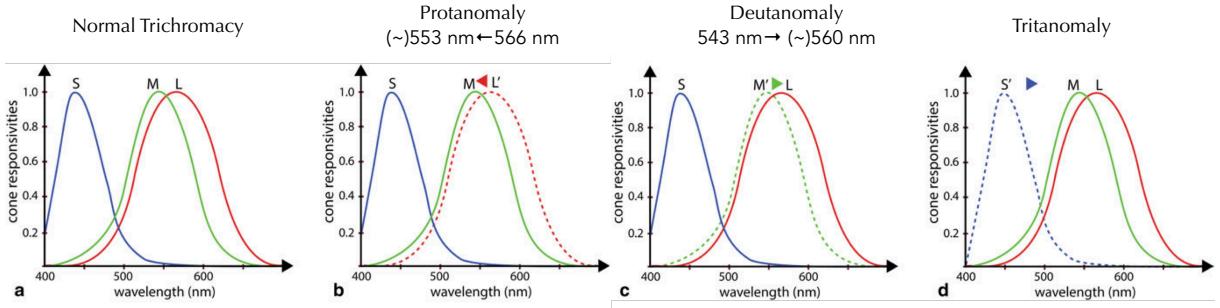


Figure 17: Illustration of how cone fundamentals shift under each anomalous trichromacy vision; adapted from [Milić et al., 2015].

of a cone type deviates from normal. **Protanomaly**, **Deuteranomaly**, and **Tritanomaly** are the names given to the three types of anomalous trichromatic vision. Figure 17 illustrates how the cone fundamentals change under different anomalous trichromacy compared to normal trichromacy. For Protanomalous and Deuteranomalous individuals (by far the most common anomalous trichromats), their L and M cone fundamentals are closer than normal, so the L and M cone excitations are very similar. In theory their color vision is still three-dimensional, but the L and M dimensions are very correlated, which weakens their color discrimination ability.

Finally, some females with anomalous trichromacy may have four cone classes on the retinal mosaic owing to random X chromosome inactivation. However, it is unclear whether their color vision is four-dimensional or they have a stronger color discrimination ability [Jordan et al., 2010; Simunovic, 2010].

We will primarily focus on dichromatic vision, building a phenomenological model for simulating dichromacy, and then discuss the genetic basis of CVD now that we have a good understanding of the evolution of color vision.

5.1 Models of Deficient Color Vision

Perhaps one of the most important questions in CVD is: what exactly do individuals with CVD actually see? In other words, how do we simulate a particular CVD? It is incredibly difficult to answer, since color is a subjective experience, and we can never be so certain of one another's subjective experience. With the help of some luck (yes, some luck, from evolution) and some psychophysics, there is at least a consensus how to model dichromatic vision. Taking Deutanopia as an example, the interactive tutorial Zhu [2022b] walks through a commonly used model. We will give the main intuitions here.

Confusion Lines

Dichromatic individuals see colors only in a 2D space, because they lack (the functionality of) one cone type. For instance, Deutanopes lack the M cones and, thus, any color is encoded only in the L and S cone excitations, resulting in a 2D color space. Therefore, any colors that

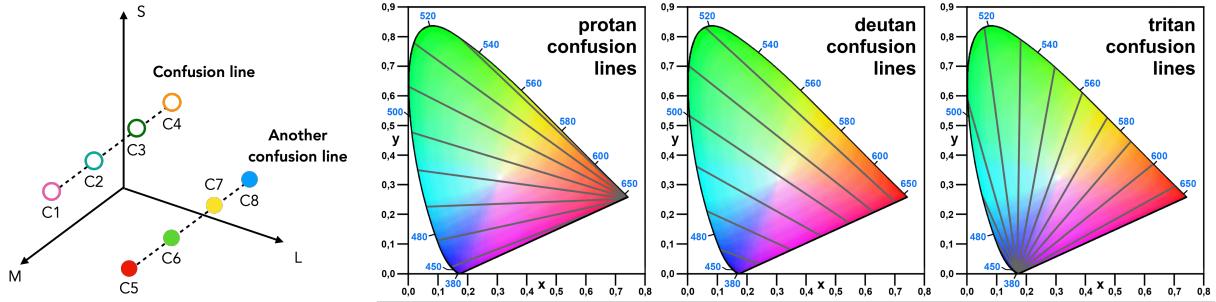


Figure 18: Left: two confusion lines for Deutanopia (missing functioning M cones). Right: confusion lines in the CIE 1931 xy-chromaticity space; from [Curran919 \[2022\]](#).

differ only in the M dimension are seen as the same color by a Deuteranope. A line parallel to the M dimension in the LMS space is called a Deuteranopic **confusion line**, as all colors on that line look the same to a Deuteranope. The left panel in Figure 18 plots two such confusion lines for Deutanopia, although it is easy to see that there are infinitely many confusion lines.

The confusion lines are more commonly visualized in the CIE 1931 xy-chromaticity diagram, as illustrated in the three diagrams in Figure 18 to the right. We will discuss the xy diagram in the colorimetry lecture soon, but briefly, the xy diagram is a *perspective projection* from the LMS space that provides a useful 2D representation of colors by discarding the luminance dimension. Therefore, parallel confusion lines in the LMS space converge in the xy diagram.

Not all colors on a dichromatic confusion line are confusing to the corresponding anomalous trichromatic individuals. From a modeling perspective, a typical approach is to restrict the confusing colors to a small segment on a confusion line [Flatla et al., 2015]. Figure 19 illustrates a model to reason about this using Protanomaly as an example. C1 and C2 are two colors that differ only in the L cone response. For simplicity, Figure 19 shows only the L and M dimensions. Since the L and M cone fundamentals are very similar, the L and M cone responses are similar too (the extreme case being L = M). So C1 and C2 are mapped to C1' and C2' in a Protanomalous color space, where both colors are pulled toward the L=M line. C1' and C2' are never going to exactly coincide, because the L and M cone fundamentals do not exactly overlap, but when C1' and C2' are sufficient close, they can still be perceptually undiscriminable. This is not unique to CVD individuals: even for normal trichromats the color discrimination is not perfect, as has been demonstrated extensively through psychophysics [MacAdam, 1942], which we will discuss in the Colorimetry chapter soon.

Isochromes

We know that all colors on the same confusion line are perceptually the same, but still we do not know, for all those colors on a confusion line, exactly what is the color a dichromate sees. The key to answer this question is the notion of **isochromes**, which are colors perceived correctly by a dichromate. The question is how do we find the isochromes?

It is impossible to find isochromes by simply querying trichromats and dichromates. Imagine

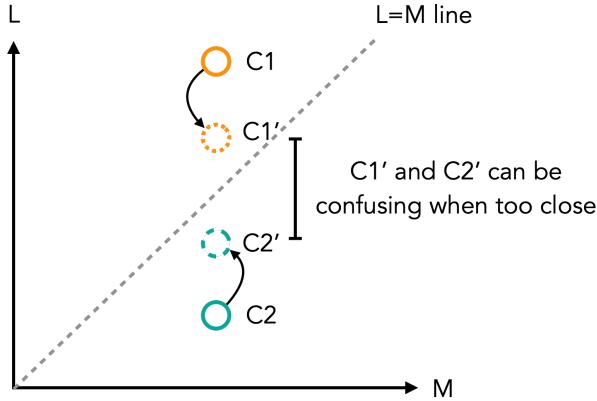


Figure 19: In a Protanomalous color space, the L and M cone fundamentals are very similar, so the L and M cone excitations are similar. C1 and C2 are colors in a trichromatic color space. They are mapped to C1' and C2' in a Protanomalous color space, where both are moved closer to the L=M line (while keeping the M-axis unchanged, since the M cone fundamental is unaffected). When C1' and C2' are sufficiently close, they become confusing.

we have a normal trichromat and a protanope look at a color; even if they have the same color sensation, how would they communicate with each other about it? You might be tempted to find isochromes by asking a dichromate whether two colors appear the same.

Remarkably, there is an exceedingly rare form of CVD called **unilateral dichromacy**, where an individual has one dichromatic eye and another trichromatic eye. Color matching between the two eyes by a unilateral dichromate would allow us to identify isochromes, assuming of course that the dichromatic eye and the trichromatic eye are similar to those of a “normal” dichromatic and trichromatic eye, respectively. This is a remarkable luck we get from nature; without unilateral dichromats we might never be able to quantitatively study dichromats’ color vision. There are a handful of studies reported on unilateral dichromates. [Judd \[1949\]](#) meticulously summarized data from prior studies, where only 8 had quantitative data that were useful. [Sloan and Wollach \[1948\]](#), [Graham and Hsia \[1958\]](#), and [MacLeod and Lennie \[1976\]](#) reported results for unilateral Protanopes/Deuteranopes, while [Alpern et al. \[1983\]](#) reported results for a unilateral Tritanope.

Such studies show that monochromatic lights at 475 nm and 575 nm are isochromes for protanopes and deuteranopes (i.e., no significant difference between these two types when it comes to isochromes, but of course their confusion lines are different), and for tritanopes isochromes are found at 485 nm and 660 nm.

Brettel et al. Model

There have been two main ways to build a model for dichromatic color vision: that of phenomenological nature such as [Brettel et al. \[1997\]](#); [Viénot et al. \[1999\]](#); [Meyer and Greenberg \[1988\]](#) and that based on first principles of the visual pathway such as [Jiang et al. \[2016\]](#); [Rodriguez-Pardo](#)

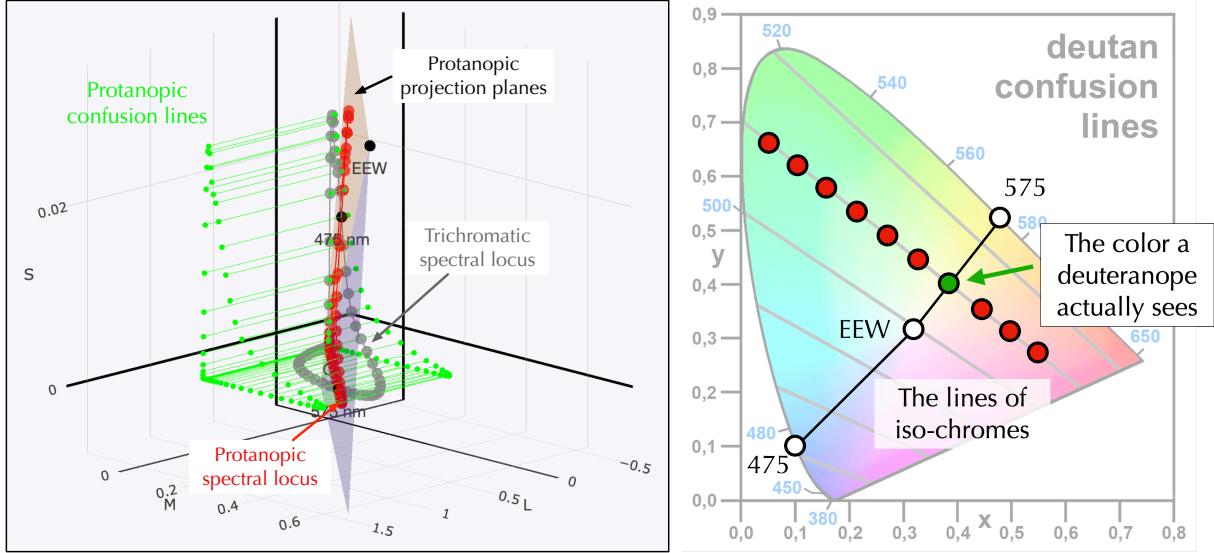


Figure 20: Left: isochrome planes and Protanopia confusion lines in LMS cone space; from Zhu [2022b]. According to Brettel et al. [1997] model, the intersection between the isochrome planes and a confusion line is the color a dichromat actually sees for all the colors on that confusion line. Right; the isochrome lines and Deutanopia confusion lines in the xy-diagram; adapted from Curran919 [2022].

and Sharma [2011]. We will primarily focus on the phenomenological model described in Brettel et al. [1997]. Like all other models, this model is not perfect. It makes assumptions that are not experimentally validated on unilateral dichromats, but it is a popular model that seems to work well in practice. Zhu [2022b] discusses cases where this model falls apart.

Brettel et al. [1997] assumes that Equal-Energy White (EEW) is also an isochrome. In fact, they assume that the entire plane that contains EEW, 475 nm, 575 nm, and Black is an isochrome plane, on which all colors are isochromes for deuteranopes and protanopes. Similarly, the plane that contains EEW, 485 nm, 660 nm, and Black is an isochrome plane for tritanopes. Two caveats must be noted here. First, this is not validated against unilateral dichromats; it is just an assumption. Second, the isochrome “plane” is not an actual plane. It is more like two half-planes that share a border. For Protanopia and Deutanopia, the two half-planes are almost parallel so that they look like part of one plane. The left panel in Figure 20 shows the two half planes (with distinct colors) for Protanopia.

Assuming that the isochrome plane assumption by Brettel et al. [1997] is true, we can then reason about the colors that dichromats actually see: the intersection between the isochrome planes and a confusion line is the color a dichromat actually sees for all the colors on that confusion line. The left panel in Figure 20 visualizes this, where the trichromatic spectral locus is projected to the isochrome planes along the direction of the confusion lines. The resulting locus lies completely on the isochrome planes and represents how the spectral colors will actually

be perceived by a Protanope.

The two isochrome half-planes become two line segments in the xy-diagram. The right panel in Figure 20 shows how the [Brettel et al. \[1997\]](#) model predicts a Deuteranopia's color perception in the xy-diagram.

Modeling anomalous trichromacy is “easy” as long as we know the new set of cone fundamentals. Assuming their subsequent neural processing is the same as that of normal trichromacy, we can then calculate the cone responses in a anomalous trichromatic color space given any light spectrum.

5.2 CVD Assistive Technologies

There are many assistive techniques attempting to enhance the color vision of CVD individuals. [Zhu et al. \[2024, Sec. 3\]](#) provides a good review.

Re-Coloring Helps Color Discrimination

By far the most commonly used technique is called “re-coloring”; it is available on most Operating Systems in laptops, PCs, and smartphones. The idea is to apply a (usually linear) transformation to colors in an image (colloquially referred to as color filters) so that initially confusing colors become distinct (i.e., no longer on a confusion line).

The main limitation of re-coloring is that, while initially confusing colors might be distinguishable after a transformation, these colors will inevitably be confused with others. Fundamentally, a dichromat's color vision is still two-dimensional, and the confusion lines are still there. Usually the transformation is designed so that a set of ecologically relevant confusing colors (colors that we commonly encounter in everyday life and are important to discriminate, e.g., red flowers and green leaves) become distinct.

Re-coloring can also be done optically. Many commercially available glasses for CVD individuals, such as those from EnChroma and VINO, use identical spectral “notch filters” for both eyes. The filter eliminates light from a narrow spectral band, where the human L and M cone sensitivities overlap the most. Recall that for Protanomalous and Deuteranomalous individuals, their L and M cone fundamentals are closer than normal, so the L and M cone excitations are very similar. The notch filters act to pull the two cone fundamentals apart and amplify the difference between L cone and M cone excitations. Therefore, in principle, the method can enhance color discrimination for anomalous trichromats, but provides no benefit for strict dichromats. In comparative tests, the functional effectiveness of such glasses for anomalous trichromats is also not definitive [\[Gómez-Robledo et al., 2018; Patterson et al., 2022\]](#).

Color Recognition By Reconstructing the Missing Dimension

Re-coloring methods, thus, do not help with color *recognition* and *naming* (“pass me that pink marker” or “look at the person in a red shirt”), which is a common user complaint found in a user study [\[Geddes et al., 2023\]](#). For color recognition, we must restore a three-dimensional color space for dichromats. So the key is to somehow reconstruct the missing dimension.

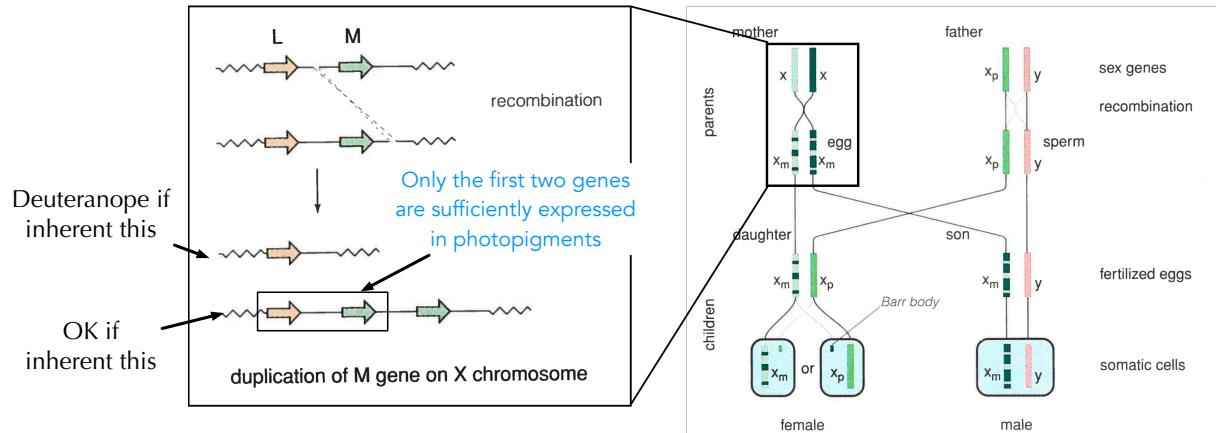


Figure 21: An intergenic that can potentially give rise to Deutanopia. Adapted from [Rodieck \[1998\]](#), p. 219–220].

One idea is to introduce binocular color disparity, where the stimuli are differentially altered for the two eyes. The idea was originated by [Maxwell \[1857\]](#) in the very early phase of the scientific study of color perception, and then later revived by [Cornsweet \[1970\] page?](#). Maxwell conjectured that the disparity across the two eyes would essentially introduce a new dimension of perception, which would augment the existing 2D percept of a dichromat providing 3D color perception. [Knoblauch and McMahon \[1995\]](#) shows that the improved color discrimination afforded by binocular filters might have limited use for Protanopes and Deutanopes.

[Zhu et al. \[2024\]](#) introduces a smartphone App to reconstruct the missing dimension through temporally modulating colors. The idea is that as a user swipes the finger in the App, they apply a color-space transformation such that originally confusing colors undergo distinct color shifts. The combination of the initial 2D color precept with the induced temporal shifts reconstructs a new 3D space for the user. By spending time interacting with our system, users then learn to associate different color names in this new 3D space, thereby recognizing colors.

5.3 Genetic Basis of CVD

L and M cone genes are on the X chromosome. Because the L and M genes are created from a duplication, they are tandemly arrayed (spatially adjacent in a head-to-tail manner) and, thus, are subject to crossovers during recombination of meiosis, which might lead to deletion of a whole gene or hybrid genes in a X chromosome. This is the genetic basis of Protanopia, Deutanopia, Protanomaly, and Deutanomaly. [Onishi et al. \[2002\]](#) shows that only two in a sample of over 3000 macaque monkeys (an old-world monkey) have CVD, must lower than then CVD rate in humans, indicating that the crossovers might be recent [[Lee, 2008](#)].

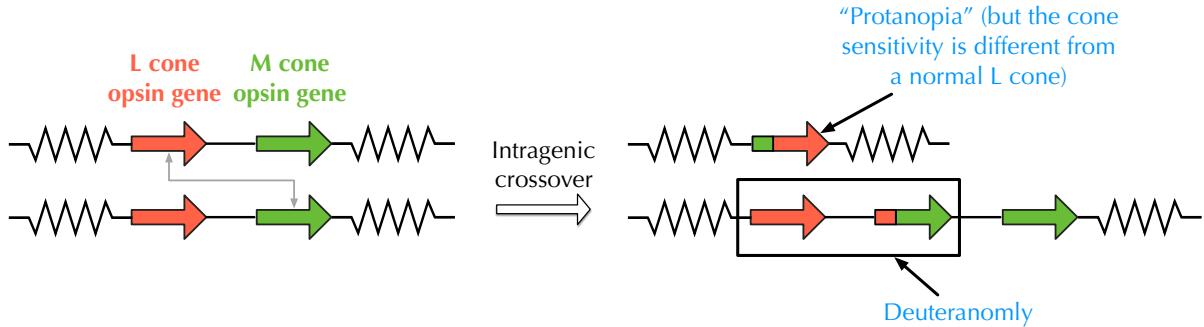


Figure 22: An intragenic crossover that might give rise to Deuteranomaly and “Protanopia”. See text.

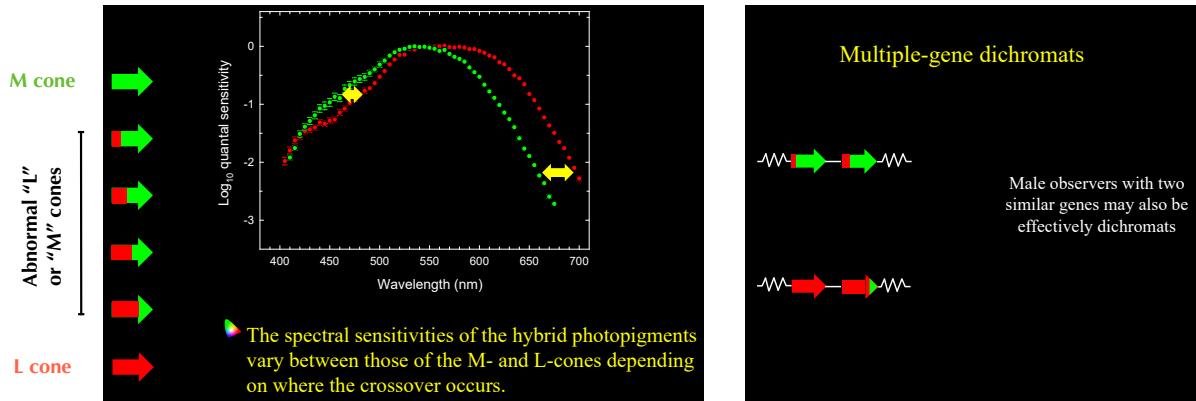


Figure 23: Left: the spectral sensitivities of the hybrid photopigments vary between those of the M- and L-cones depending on where the crossover occurs. Right: if two hybrid genes are sufficiently similar, the individual is effectively dichromatic. Slides credit: Andrew Stockman.

Intergenic Crossovers Give Protanopia and Deuteranopia

Figure 21 illustrates a crossover that can potentially lead to Deuteranopia. The two X chromosomes, due to an **intergenic crossover**, either lose or gain a M gene after recombination. One of the new X chromosome has only the L cone gene, so an individual inheriting that X chromosome from the mother, the individual would get Deuteranopia. Interestingly, while the other X chromosome get an additional M open gene, only the first two genes are sufficiently expressed. The fact that the L and M cone genes are in an X chromosome means that biological females are less vulnerable to Protanopia and Deuteranopia than biological males, simply because there are two X chromosomes in females. Even if one of the inherited X chromosomes has only, say, a L cone gene, the other inherited X chromosome, if normal, can still be sufficiently be expressed to give both L and M cones.

Intragenic Crossovers Give Protanomaly and Deuteranomaly

Intragenic crossover might also occur during recombination, and this would lead to anomalous trichromacy. Figure 22 shows an intragenic crossover that might give rise to Deuteranomaly and “Protanopia”. The second X chromosome after recombination has a normal L opsin gene and another gene that is a mixture of L cone gene and M cone gene. The spectral sensitivity of such a hybrid pigment in-between that of a L cone and a M cone [Sharpe et al., 1999], as illustrated in the left panel in Figure 23. This is the source of anomalous trichromats.

Intergenic Crossovers Can Give Abnormal Protanopia and Deuteranopia

It is interesting to observe that intragenic crossovers can also give a dichromatic vision, although in a somewhat abnormal form. Observe in Figure 22 that the first X chromosome after recombination also has a hybrid gene. If the L cone gene dominates, that gene, when inherited, will be expressed in a pigment that has a spectral sensitivity that is closer to that of a L cone. If you want, you can still say that the inheriting individual has “Protanopia”, but its L cone sensitivity is different from that of a normal L cone.

The right panel in Figure 23 illustrates another scenario where dichromatic vision can arise from intragenic crossovers. If the two hybrid genes in the chromosome are sufficiently similar, they will be expressed in two pigments that are sufficiently similar, and the individual effectively has a dichromatic vision [Sharpe et al., 1999].

Deuteranomaly is the Most Common CVD

Tritanopia and tritanomaly are much rarer than the other forms of CVDs. The gene for S cone opsin is in a nonsex chromosome (chromosome 7), which is not subject to L/M gene crossovers. Gene mutations causing changes in S-cone pigment are much more rare. Among CVDs that are caused by the crossovers, anomalous trichromacy is more common than strict dichromats, and Deuteranomaly is the most common (about 4.9% in caucasians) [Wyszecki and Stiles, 1982, Table 1(5.4.2), p. 464], but the statistics certainly vary across ethnic groups [Birch, 2012].

References

- M Alpern, K Kitahara, and DH Krantz. Perception of colour in unilateral tritanopia. *The Journal of Physiology*, 335(1):683–697, 1983.
- Karl-Heinz Bäuml. A ratio principle for a red/green and a yellow/blue channel? *Perception & psychophysics*, 53(3):338–344, 1993.
- Karl-Heinz Bäuml and Brian A Wandell. Color appearance of mixture gratings. *Vision Research*, 36(18):2849–2864, 1996.
- Denis A Baylor, BJ Nunn, and JL Schnapf. Spectral sensitivity of cones of the monkey macaca fascicularis. *The Journal of Physiology*, 390(1):145–160, 1987.

- Bhutajata. Color temperature black body; CC BY-SA 4.0 license. https://commons.wikimedia.org/wiki/File:Color_temperature_black_body_800-12200K.svg, 2015.
- Jennifer Birch. Worldwide prevalence of red-green color deficiency. *JOSA A*, 29(3):313–320, 2012.
- James K Bowmaker. Evolution of colour vision in vertebrates. *Eye*, 12(3):541–547, 1998.
- James K Bowmaker. Evolution of vertebrate visual pigments. *Vision research*, 48(20):2022–2041, 2008.
- DH Brainard. Cone contrast and opponent modulation color spaces. In *Human Color Vision*, pages 563–579. Optical Society of America, 2 edition, 1996.
- Hans Brettel, Françoise Viénot, and John D Mollon. Computerized simulation of color appearance for dichromats. *Josaa*, 14(10):2647–2655, 1997.
- A Broadbent. Calculation from the original experimental data of the cie 1931 rgb standard observer spectral chromaticity coordinates and color matching functions. *Québec, Canada: Département de génie chimique, Université de Sherbrooke*, pages 1–17, 2008.
- Arthur D Broadbent. A critical review of the development of the cie1931 rgb color-matching functions. *Color Research & Application*, 29(4):267–272, 2004.
- Paul K Brown and George Wald. Visual pigments in single rods and cones of the human retina. *Science*, 144(3614):45–52, 1964.
- Tom Cornsweet. *Visual perception*. Academic press, 1970.
- BH Crawford. The scotopic visibility function. *Proceedings of the Physical Society. Section B*, 62(5):321, 1949.
- Curran919. Color Blind Confusion Lines; CC BY-SA 4.0 license. https://commons.wikimedia.org/wiki/File:Color_Blind_Confusion_Lines.png, 2022.
- Dennis M Dacey and Barry B Lee. The ‘blue-on’ opponent pathway in primate retina originates from a distinct bistratified ganglion cell type. *Nature*, 367(6465):731–735, 1994.
- Herbert JA Dartnall, James K Bowmaker, and John Dixon Mollon. Human visual pigments: microspectrophotometric results from the eyes of seven persons. *Proceedings of the Royal society of London. Series B. Biological sciences*, 220(1218):115–130, 1983.
- RL De Valois, CJ Smith, ST Kitai, and AJ Karoly. Response of single cells in monkey lateral geniculate nucleus to monochromatic light. *Science*, 127(3292):238–239, 1958.
- Russell L De Valois, Israel Abramov, and Gerald H Jacobs. Analysis of response patterns of lgn cells. *JOSA*, 56(7):966–977, 1966.

- Andrew M Derrington, John Krauskopf, and Peter Lennie. Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of physiology*, 357(1):241–265, 1984.
- Russell L DeValois and Karen K DeValois. Spatial vision. 1990.
- Forrest L Dimmick and Margaret R Hubbard. The spectral components of psychologically unique red. *The American Journal of Psychology*, 52(3):348–353, 1939a.
- Forrest L Dimmick and Margaret R Hubbard. The spectral location of psychologically unique yellow, green, and blue. *The American Journal of Psychology*, 52(2):242–254, 1939b.
- Greg D Field, Alexander Sher, Jeffrey L Gauthier, Martin Greschner, Jonathon Shlens, Alan M Litke, and EJ Chichilnisky. Spatial properties and functional organization of small bistratified ganglion cells in primate retina. *Journal of Neuroscience*, 27(48):13261–13272, 2007.
- David R Flatla, Alan R Andrade, Ross D Teviotdale, Dylan L Knowles, and Craig Stewart. Colourid. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015.
- Connor Geddes, David R Flatla, and Ciabhan L Connelly. 30 years of solving the wrong problem: How recolouring tool design fails those with colour vision deficiency. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–13, 2023.
- Andrew S Glassner. *Principles of digital image synthesis*. Elsevier, 1995.
- E. Bruce Goldstein and R. James Brockmole. *Sensation and Perception*. Cengage Learning, 10 edition, 2017.
- Luis Gómez-Robledo, EM Valero, R Huertas, MA Martínez-Domingo, and Javier Hernández-Andrés. Do enchroma glasses improve color vision for colorblind subjects? *Optics express*, 26(22):28693–28703, 2018.
- P Gouras and E Zrenner. Enhancement of luminance flicker by color-opponent mechanisms. *Science*, 205(4406):587–589, 1979.
- Clarence Henry Graham and Yun Hsia. Color defect and color theory: Studies of normal and color-blind persons, including a subject color-blind in one eye but not in the other. *Science*, 127(3300):675–682, 1958.
- John Guild. The colorimetric properties of the spectrum. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 230(681-693):149–187, 1931.
- Ewald Hering. *Zur Lehre vom Lichtsinne: sechs Mittheilungen an die Kaiser. Akad. der Wissenschaften in Wien*. C. Gerold's Sohn, 1878.

- Ewald Hering. *Outlines of a theory of the light sense (Translation by Jameson and Hurvich; originally published in 1878)*. Harvard University Press, 1964.
- David M Hunt, Kanwaljit S Dulai, Jill A Cowing, Catherine Julliot, John D Mollon, James K Bowmaker, Wen-Hsiung Li, and David Hewett-Emmett. Molecular evolution of trichromacy in primates. *Vision research*, 38(21):3299–3306, 1998.
- Leo M Hurvich and Dorothea Jameson. Some quantitative aspects of an opponent-colors theory. ii. brightness, saturation, and hue in normal and dichromatic vision. *JOSA*, 45(8):602–616, 1955.
- Leo M Hurvich and Dorothea Jameson. An opponent-process theory of color vision. *Psychological review*, 64(6p1):384, 1957.
- Gerald H Jacobs. Evolution of colour vision in mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1531):2957–2967, 2009.
- Dorothea Jameson and Leo M Hurvich. Some quantitative aspects of an opponent-colors theory. i. chromatic responses and spectral saturation. *JOSA*, 45(7):546–552, 1955.
- Haomiao Jiang, Joyce Farrell, and Brian Wandell. A spectral estimation theory for color appearance matching. *Electronic Imaging*, 28:1–4, 2016.
- Gabriele Jordan, Samir S Deeb, Jenny M Bosten, and John D Mollon. The dimensionality of color vision in carriers of anomalous trichromacy. *Journal of vision*, 10(8):12–12, 2010.
- Deane B Judd. The color perceptions of deuteranopic and protanopic observers. *JOSA*, 39(3):252–256, 1949.
- Deane B Judd. Report of us secretariat committee on colorimetry and artificial daylight. *CIE Proceedings, 1951*, 1:11, 1951.
- Deane B Judd, David L MacAdam, Günter Wyszecki, HW Budde, HR Condit, ST Henderson, and JL Simonds. Spectral distribution of typical daylight as a function of correlated color temperature. *Josa*, 54(8):1031–1040, 1964.
- Kenneth Knoblauch and Matthew J McMahon. Discrimination of binocular color mixtures in dichromacy: evaluation of the maxwell–cornsweet conjecture. *JOSA A*, 12(10):2219–2229, 1995.
- Trevor D Lamb. Evolution of phototransduction, vertebrate photoreceptors and retina. *Progress in retinal and eye research*, 36:52–119, 2013.
- Trevor D Lamb. Why rods and cones? *Eye*, 30(2):179–185, 2016.
- Trevor D Lamb. Evolution of the genes mediating phototransduction in rod and cone photoreceptors. *Progress in retinal and eye research*, 76:100823, 2020.

- Trevor D Lamb. Photoreceptor physiology and evolution: cellular and molecular basis of rod and cone phototransduction. *The Journal of Physiology*, 600(21):4585–4601, 2022.
- Trevor D Lamb, Shaun P Collin, and Edward N Pugh. Evolution of the vertebrate eye: opsins, photoreceptors, retina and eye cup. *Nature Reviews Neuroscience*, 8(12):960–976, 2007.
- James Larimer, Carol M Cicerone, et al. Opponent-process additivity-i: Red/green equilibria. *Vision Research*, 14(11):1127–1140, 1974.
- James Larimer, David H Krantz, and Carol M Cicerone. Opponent process additivity—ii. yellow/blue equilibria and nonlinear models. *Vision research*, 15(6):723–731, 1975.
- Barry B Lee. The evolution of concepts of color vision. *Neurociencias*, 4(4):209, 2008.
- BB Lee, PR Martin, and A Valberg. The physiological basis of heterochromatic flicker photometry demonstrated in the ganglion cells of the macaque retina. *The Journal of physiology*, 404 (1):323–347, 1988.
- Peter Lennie, Joel Pokorny, and Vivianne C Smith. Luminance. *JOSA A*, 10(6):1283–1293, 1993.
- David L MacAdam. Visual sensitivities to color differences in daylight. *Josa*, 32(5):247–274, 1942.
- Donald IA MacLeod and Peter Lennie. Red-green blindness confined to one eye. *Vision research*, 16(7):691–702, 1976.
- Marco Polo. CIE1931 RGB CMF; released into the public domain by the copyright holder. https://commons.wikimedia.org/wiki/File:CIE1931_RGBCMF.svg, 2007.
- WB Marks, William H Dobelle, and Edward F MacNichol Jr. Visual pigments of single primate cones. *Science*, 143(3611):1181–1183, 1964.
- James Clerk Maxwell. Xviii.—experiments on colour, as perceived by the eye, with remarks on colour-blindness. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 21(2):275–298, 1857.
- Gary W Meyer and Donald P Greenberg. Color-defective vision and computer graphics displays. *IEEE Computer Graphics and Applications*, 8(5):28–40, 1988.
- Neda Milić, Dragoljub Novaković, and Branko Milosavljević. Enhancement of image content for observers with colour vision deficiencies. *Color Image and Video Enhancement*, pages 315–343, 2015.
- John D Mollon. Introduction: Thomas young and the trichromatic theory of colour vision. In *Normal & defective colour vision*, pages 19—34. 2003.

- Ala Moshedian and Gordon L Fain. Single-photon sensitivity of lamprey rods with cone-like outer segments. *Current Biology*, 25(4):484–487, 2015.
- Jeremy Nathans, Darcy Thomas, and David S Hogness. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science*, 232(4747):193–202, 1986.
- Isaac Newton. *Opticks, or, a treatise of the reflections, refractions, inflections & colours of light*. London: Smith and Walford, 1704.
- NumFOCUS. Colour: open-source Python package for colour science. <https://github.com/colour-science/colour>.
- Toshiyuki Okano, DAISUKE Kojima, Yoshitaka Fukada, Yoshinori Shichida, and Toru Yoshizawa. Primary structures of chicken cone visual pigments: vertebrate rhodopsins have evolved out of cone visual pigments. *Proceedings of the National Academy of Sciences*, 89(13):5932–5936, 1992.
- Toshiyuki Okano, Toru Yoshizawa, and Yoshitaka Fukada. Pinopsin is a chicken pineal photoreceptive molecule. *Nature*, 372(6501):94–97, 1994.
- Akishi Onishi, Satoshi Koike, Miki Ida-Hosonuma, Hiroo Imai, Yoshinori Shichida, Osamu Takenaka, Akitoshi Hanazawa, Hidehiko Komatsu, Akichika Mikami, Shunji Goto, et al. Variations in long-and middle-wavelength-sensitive opsin gene loci in crab-eating monkeys. *Vision Research*, 42(3):281–292, 2002.
- EJ Patterson, RR Mastey, JA Kuchenbecker, Jessica Rowlan, Jay Neitz, Maureen Neitz, and Joseph Carroll. Effects of color-enhancing glasses on color vision in congenital red-green color deficiencies. *Optics Express*, 30(17):31182–31194, 2022.
- Allen B Poirson and Brian A Wandell. Appearance of colored patterns: pattern–color separability. *JOSA A*, 10(12):2458–2470, 1993.
- Robert W Rodieck. *The first steps in seeing*. Sinauer Associates, 1998.
- Carlos Eduardo Rodriguez-Pardo and Gaurav Sharma. Dichromatic color perception in a two stage model: Testing for cone replacement and cone loss models. In *2011 IEEE 10th IVMSP Workshop: Perception and Visual Signal Analysis*, pages 12–17. IEEE, 2011.
- Keita Sato, Takahiro Yamashita, Keiichi Kojima, Kazumi Sakai, Yuki Matsutani, Masataka Yanagawa, Yumiko Yamano, Akimori Wada, Naoyuki Iwabe, Hideyo Ohuchi, et al. Pinopsin evolved as the ancestral dim-light visual opsin in vertebrates. *Communications biology*, 1(1):156, 2018.
- JL Schnapf, TW Kraft, and Denis A Baylor. Spectral sensitivity of human cone photoreceptors. *Nature*, 325(6103):439–441, 1987.

- Erwin Schrödinger. Über das verhältnis der vierfarben-zur dreifarbentheorie. *Sitzungberichte Abt 2a, Mathematik, Astronomie, Physik, Meteorologie und Mechanik Akademie der Wissenschaften in Wien, Mathematisch-Naturwissenschaftliche Klasse*, 134:471–490, 1925.
- Erwin Schrödinger. On the relationship of four-color theory to three-color theory (translation by national translation center; originally published in 1925). *Color Research and Application*, 19(1):37, 1994.
- Phil Service. The wright – guild experiments and the development of the cie 1931 rgb and xyz color spaces. 2016.
- Lindsay T Sharpe, Andrew Stockman, Herbert Jägle, and Jeremy Nathans. Opsin genes, cone photopigments, color vision, and color blindness. *Color vision: From genes to perception*, 351: 3–52, 1999.
- Lindsay T Sharpe, Andrew Stockman, Wolfgang Jagla, and Herbert Jägle. A luminous efficiency function, $v^*(\lambda)$, for daylight adaptation. *Journal of vision*, 5(11):3–3, 2005.
- Lindsay T Sharpe, Andrew Stockman, Wolfgang Jagla, and Herbert Jägle. A luminous efficiency function, $vd65^*(\lambda)$, for daylight adaptation: a correction. *Color Research & Application*, 36 (1):42–46, 2011.
- Steven K Shevell and Paul R Martin. Color opponency: tutorial. *JOSA A*, 34(7):1099–1108, 2017.
- Yoshinori Shichida and Take Matsuyama. Evolution of opsins and phototransduction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1531):2881–2895, 2009.
- MP Simunovic. Colour vision deficiency. *Eye*, 24(5):747–755, 2010.
- Louise L Sloan and Lorraine Wollach. A case of unilateral deutanopia. *JOSA*, 38(6):502–509, 1948.
- Walter Stanley Stiles and Jennifer M Burch. Npl colour-matching investigation: final report (1958). *Optica Acta: International Journal of Optics*, 6(1):1–26, 1959.
- WS Stiles and JM Burch. Interim report to the commission internationale de l'éclairage, zurich, 1955, on the national physical laboratory's investigation of colour-matching (1955). *Optica Acta: International Journal of Optics*, 2(4):168–181, 1955.
- Andrew Stockman and Lindsay T Sharpe. The spectral sensitivities of the middle-and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision research*, 40(13):1711–1737, 2000.
- Andrew Stockman, Lindsay T Sharpe, and Clemens Fach. The spectral sensitivity of the human short-wavelength sensitive cones derived from thresholds and color matches. *Vision research*, 39(17):2901–2927, 1999.

- G Svaetichin. The cone action potential. *Acta Physiol Scand*, 29:565–600, 1953.
- Gunnar Svaetichin. Spectral response curves from single cones. *Acta Physiol Scand*, 39:17–46, 1956.
- Gunnar Svaetichin and Edward F MacNichol Jr. Retinal mechanisms for chromatic and achromatic vision. *Annals of the New York Academy of Sciences*, 74(2):385–404, 1958.
- Yoko Takanaka, Toshiyuki Okano, Masayuki Iigo, and Yoshitaka Fukada. Light-dependent expression of pinopsin gene in chicken pineal gland. *Journal of neurochemistry*, 70(3):908–913, 1998.
- Françoise Viénot, Hans Brettel, and John D Mollon. Digital video colourmaps for checking the legibility of displays by dichromats. *Color Research & Application*, 24(4):243–252, 1999.
- Johannes von Kries. Übersicht der tatsachen, ergebnisse für die theoretische auffassung des sehorgans: Zonentheorie. In *Handbuch der Physiologie des Menschen, Dritter Band: Physiologie der Sinne*, pages 269—274. 1905.
- Johannes J Vos. Colorimetric and photometric properties of a 2 fundamental observer. *Color Research & Application*, 3(3):125–128, 1978.
- George Wald. Human vision and the spectrum. *Science*, 101(2635):653–658, 1945.
- Brian A. Wandell. *Foundations of vision*. Sinauer Associates, 1995.
- Brian A Wandell, David H Brainard, and Nicolas P Cottaris. Visual encoding: Principles and software. *Progress in Brain Research*, 273(1):199–229, 2022.
- Torsten N Wiesel and David H Hubel. Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey. *Journal of neurophysiology*, 29(6):1115–1156, 1966.
- WD Wright. A trichromatic colorimeter with spectral primaries. *Transactions of the Optical Society*, 29(5):225, 1928.
- WD Wright. A re-determination of the mixture curves of the spectrum. *Transactions of the Optical Society*, 31(4):201, 1930.
- William David Wright. A re-determination of the trichromatic coefficients of the spectral colours. *Transactions of the Optical Society*, 30(4):141, 1929.
- Günther Wyszecki and WS Stiles. *Color science: concepts and methods, quantitative data and formulae*. 1982.
- Thomas Young. Ii. the bakerian lecture. on the theory of light and colours. *Philosophical transactions of the Royal Society of London*, (92):12–48, 1802.
- Yuhao Zhu. How the CIE 1931 RGB Color Matching Functions Were Developed from the Initial Color Matching Experiments. <https://yuhaozhu.com/blog/cmf.html>, 2020.

- Yuhao Zhu. Interative Tutorial: Building a Color Space From Cone Fundamentals. <https://horizon-lab.org/colorvis/cone2cmf.html>, 2022a.
- Yuhao Zhu. Interative Tutorial: Understanding and Modeling Color Blindness. <https://horizon-lab.org/colorvis/colorblind.html>, 2022b.
- Yuhao Zhu. Interative Tutorial: Visualizing Human Visual Gamut. <https://horizon-lab.org/colorvis/gamutvis.html>, 2022c.
- Yuhao Zhu. Interative Tutorial: CIE 1931 XYZ Color Space. <https://horizon-lab.org/colorvis/xyz.html>, 2022d.
- Yuhao Zhu, Ethan Chen, Colin Hascup, Yukang Yan, and Gaurav Sharma. Computational trichromacy reconstruction: Empowering the color-vision deficient to recognize colors using augmented reality. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, pages 1–17, 2024.