

Color Vision and Colorimetry

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	Colorimetry	13
3.1	CIE 1931 XYZ Space	13
3.2	Chromaticity Diagram and Its Interpretation	13
3.3	Color Cube	17
3.4	HSB/HSL/HSV Space	20
3.5	Display Native Gamut	21
3.6	Color Management	22
3.7	Color Differences and “Perceptually Uniform” Color Spaces	24
4	Post-Receptoral Color Encoding: Opponent Processes	26
4.1	Hue Cancellation Experiment	27
4.2	Light-Dark Mechanism and Luminous Efficiency Function	29
4.3	Neural and Physiological Basis	31
4.4	There are Many Inconvenient Truths	35

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

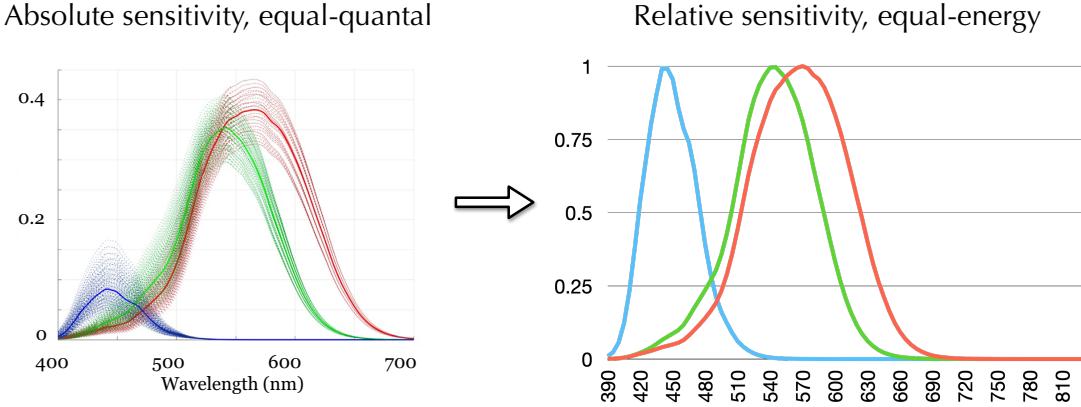


Figure 1: Physiological measurements give us absolute spectral sensitivities at an equal-quantal basis (left), but in color science each cone fundamental function is usually normalized to peak at unity and then converted to an equal-energy form (right). Left: from [Wandell et al. \[2022, Fig. 10a\]](#); 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.

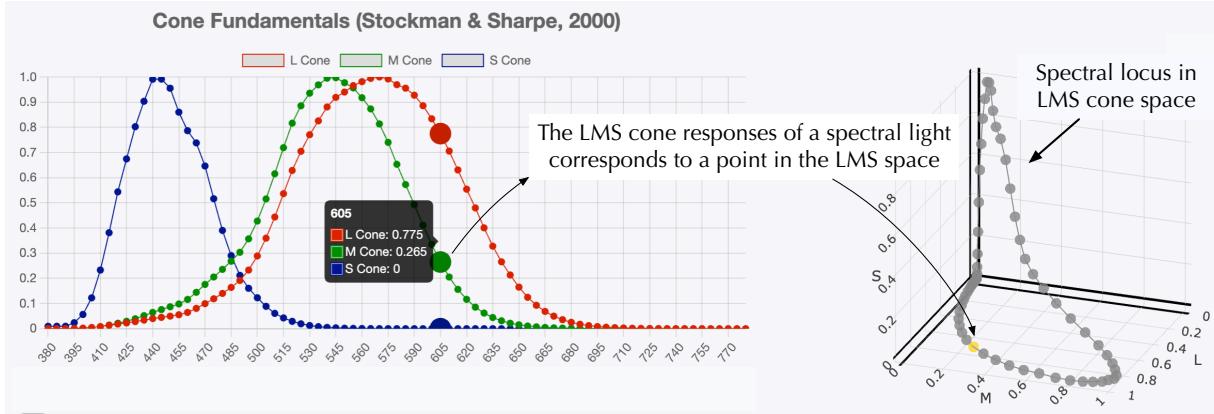


Figure 2: Spectral locus in LMS cone space. From the interactive tutorial in Zhu [2022b].

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 \[2022d\]](#) is an interactive tutorial that visualizes the HVS gamut in the cone space (and others), which you are invite to go through. We will have more to say about the HVS gamut in Chapter 3.2.

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.

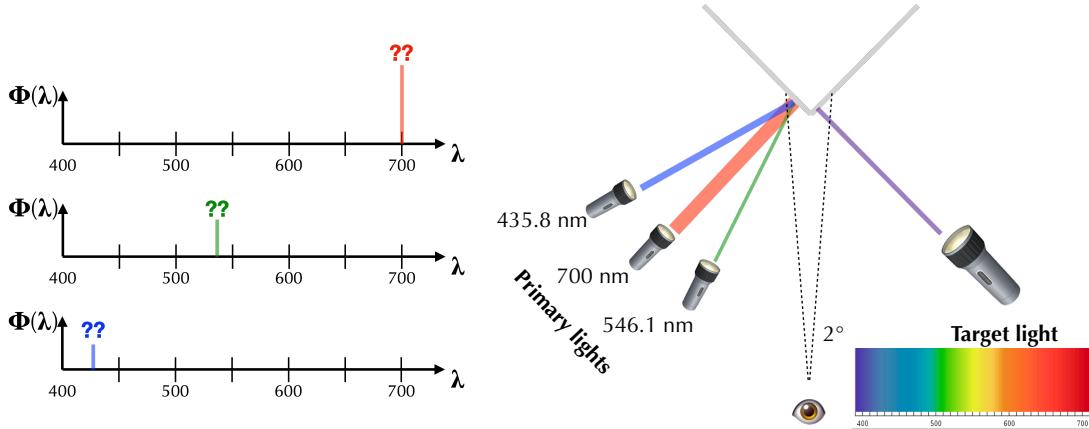


Figure 3: Color matching experiment setup. In CIE 1931 standardization of the experiment, the primary lights are spectral lights at 435.8 nm, 546.1 nm, and 700 nm, and they swept the visible spectrum [380 nm, 780 nm] at a 5-nm interval as the target light. Note that CIE 1931 did not do any actual experiments;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

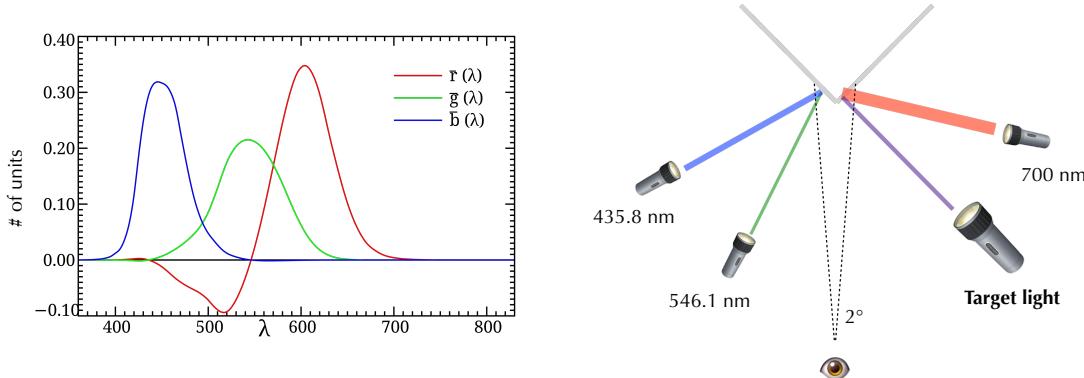


Figure 4: Left: CIE 1931 RGB Color Matching Functions (CMFs); from [Marco Polo \[2007\]](#). The y -axis shows the number of units needed of each primary so that the mixture matches the color at each wavelength (x -axis) on an equal-energy basis. The unit system is so defined that mixing equal amount (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.

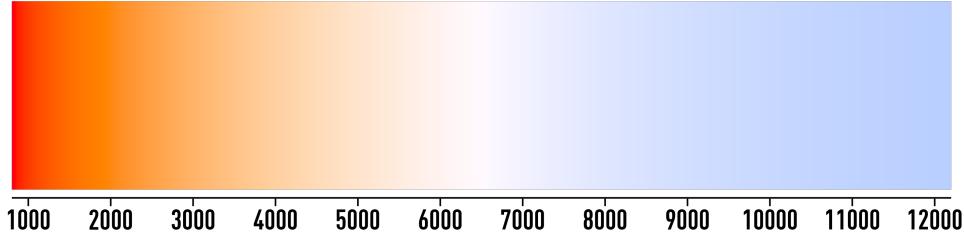


Figure 5: Color from black-body radiation at different temperatures (x -axis; unit: Kelvin). CIE Standard Illuminant D65 approximates the SPD of a noon daylight; its color is similar to that of a 6500K black-body radiation. From [Bhutajata \[2015\]](#).

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 [2022b] 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 .

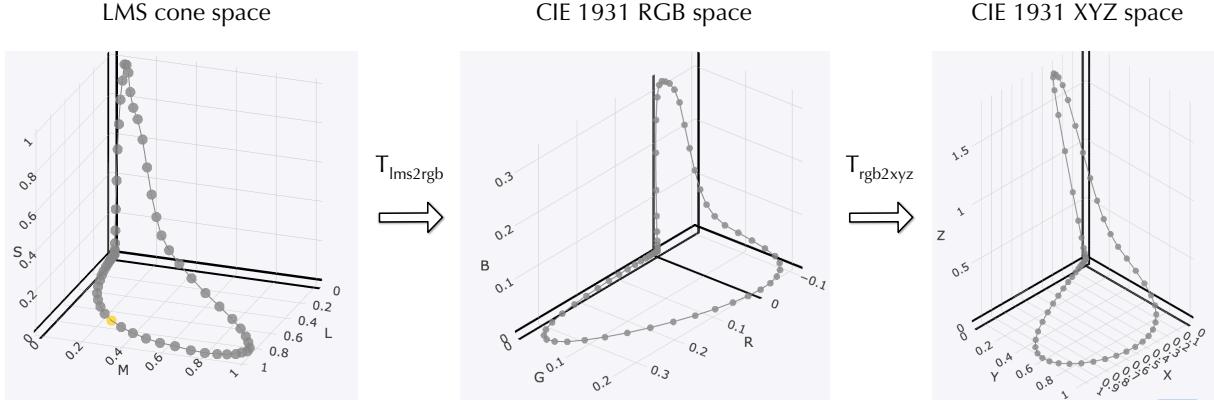


Figure 6: The spectral locus in the LMS cone space, CIE 1931 RGB space, and CIE 1931 XYZ space. The color spaces are a linear transformation away from each other. From the interactive tutorials in Zhu [2022b] and Zhu [2022e].

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 $\mathbf{T}_{lms2rgb}$. One way to intuitively interpret this transformation is to look at how the spectral locus is transformed between the two spaces. 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. The cone space and the CIE 1931 RGB space give two ways to represent the color of a light Φ . So it is natural to ask: 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 $\mathbf{T}_{lms2rgb}$:

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

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.

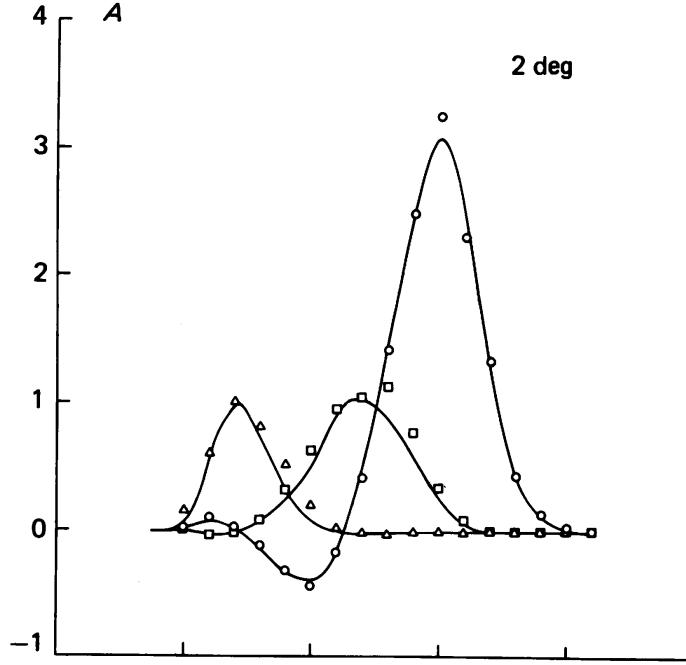


Figure 7: Smooth curves are the CMFs from [Stiles and Burch \[1955\]](#), which uses a different set of primaries and white point than those used in the CIE 1931 RGB CMFs. The markers are the predicted CMFs based on the cone fundamentals measured from macaques. From [Baylor et al. \[1987\]](#), Fig.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

[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.

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.

3 Colorimetry

Colorimetry is about quantitatively studying color, a subjective experience. Not until we can put our experience to numbers can we rigorously study colors. We have seen two ways to geometrically interpret a color as a point in a three-dimensional space: the cone space and the CIE 1931 RGB space. We will study a few other ways to quantitatively analyze colors in this chapter.

3.1 CIE 1931 XYZ Space

There are two slight inconveniences with the CIE 1931 RGB color space. First, it depends on the exact primary colors (and reference white) you choose. Second, there are also inevitable going to be colors that can be “produced” only by using negative amount of the primaries, no matter what primaries you choose. While mathematically and physically rigorous, it is not quite intuitive. So CIE in 1931 wanted to standardize a color space that 1) can be used as a “common language” (without having to laboriously specify what the primaries are used every time you say “the RGB color space”) and that 2) all human visible colors are produced by mixing non-negative amount of the primaries. That color space is called the **CIE 1931 XYZ** color space, sometimes referred to simply as the XYZ color space.

You might be wondering: isn’t the LMS cone space a color space that satisfies the two conditions above? The cone space is tied intrinsically to the HVS, so it does not vary (significantly) in population. It is also a color space where all colors are expressed using positive amounts of the primaries (cone responses). These are all true, but remember the cone fundamentals were not reliably available back in 1931 (Chapter 2.3).

[Fairman et al. \[1997\]](#), [Brill \[1998\]](#), and [Service \[2016, Sec. 4\]](#) describe the process and the (sometimes rather arbitrary) design decisions that went into turning the 1931 RGB space into the 1931 XYZ space. [Zhu \[2022e\]](#) is an interactive tutorial that walks through the math.

The bottom line is that the transformation from the RGB to the XYZ space is another linear transformation. Figure 6 shows how the spectral locus is transformed from the RGB to the XYZ space, governed by the matrix $\mathbf{T}_{rgb2xyz}$. We can see that in the RGB space the spectral locus enters negative octants but it stays entirely within the all-positive, first octant in the XYZ space. The transformation also gives a new set of CMFs in the XYZ space.

3.2 Chromaticity Diagram and Its Interpretation

How do a color that is mixed from 1:2:4 units of RGB primaries and a color that is mixed from a 2:4:8 units of the primaries relate? The amount of a primary is directly proportional to the power of that primary, so the second color can be obtained by doubling the power of each primary in the first color. Similarly, halving the power of each primary in the second color gets us the first color. Intuitively, lights that have the same primary quantity ratio have the same “objective color quality” while differing in the intensity.

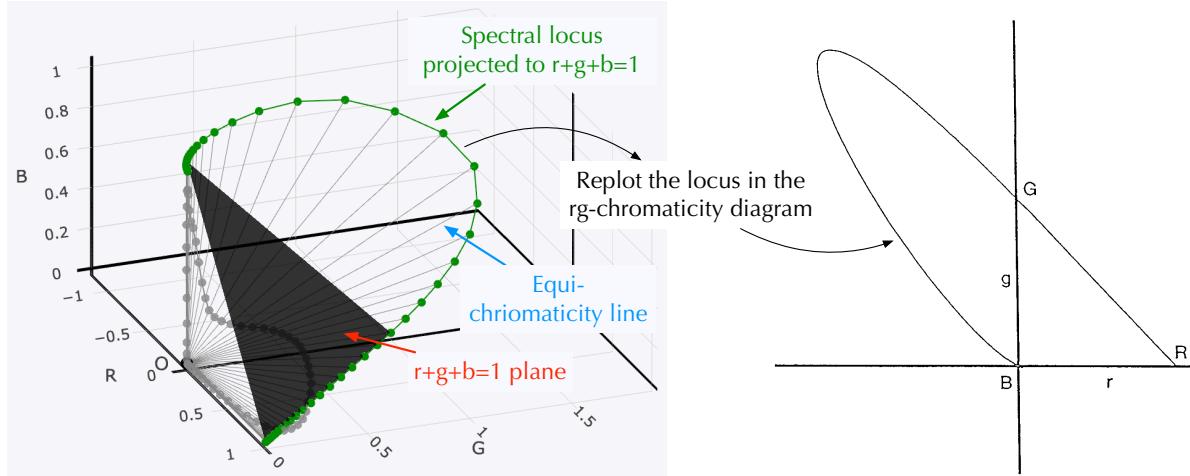


Figure 8: Visualization of the CIE 1931 RGB space and its rg-chromaticity diagram. Left: the transformation from a $[R, G, B]$ color to its $[r, g, b]$ chromaticity is a perspective projection to the $r + g + b = 1$ plane. Each line that go through the origin is an “equi-chromaticity” line, in that all the colors on that line have the same chromaticity. We use the CIE 1931 RGB color space for illustration here, but the same idea applies to other color spaces as well, e.g., the CIE 1931 XYZ space. From the interactive tutorial in Zhu [2022a]. Right: visualization of the spectral locus in CIE 1931 RGB space; from Fairman et al. [1997, Fig. 2].

Chromaticity is the Result of a Perspective Projection

More formally, we can calculate the primary ratio $r : g : b$ of a color and then normalize the ratio such that $r + g + b = 1$ (100%). The so-calculated r, g, b values of a color are called the **(RGB) chromaticity** values of that color. Mathematically, the chromaticity of a color defined in a RGB space is calculated from its absolute quantity by:

$$r = \frac{R}{R + G + B} \quad (10a)$$

$$g = \frac{G}{R + G + B} \quad (10b)$$

$$b = \frac{B}{R + G + B} \quad (10c)$$

Geometrically, going from the RGB values of a color to the rg chromaticity is equivalent to a *perspective projection*, where we project a $[R, G, B]$ point through the origin to the $r + g + b = 1$ plane. The left panel in Figure 8 visualizes this projection. Each line that go through the origin is an “equi-chromaticity” line, in that all the colors on that line have the same chromaticity. The spectral locus is so projected to the $r + g + b = 1$ plane. Since there are only two degrees of freedom in chromaticity, we can visualize the chromaticity in a two-dimensional space, and

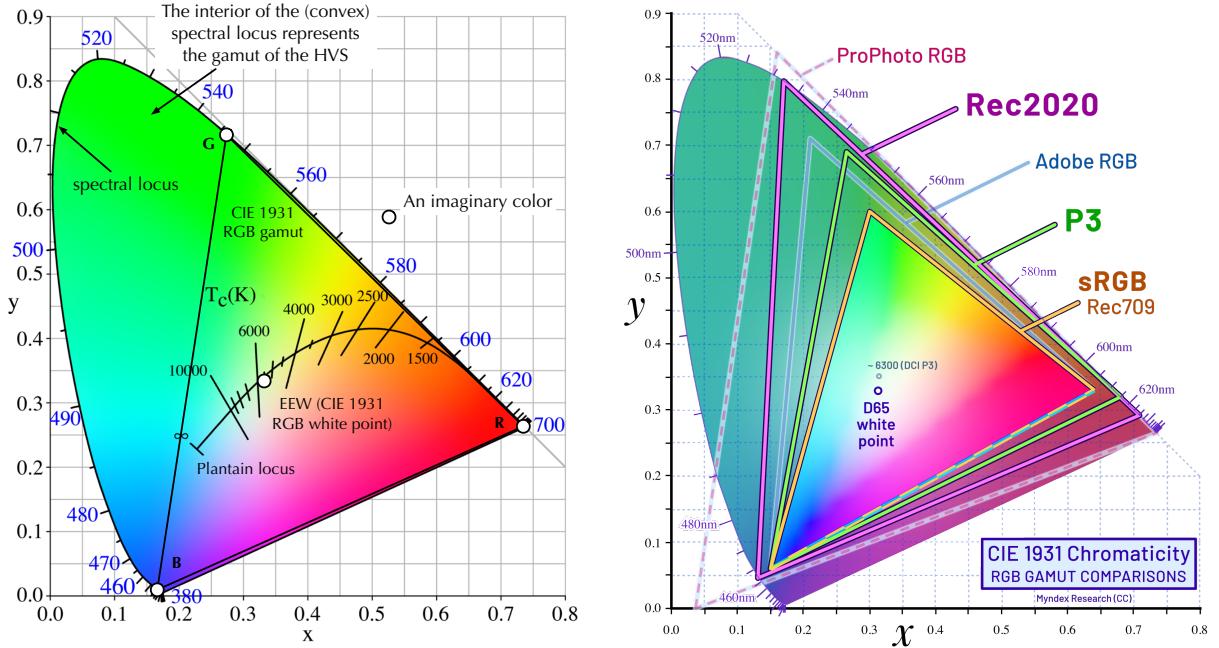


Figure 9: Left: The gamut and spectral locus of the CIE 1931 RGB space visualized in the xy-chromaticity diagram; adapted from [User:PAR \[2012\]](#). The Planckian locus is shown for the reference too. A point outside the (convex) spectral locus is an imaginary color. Right: comparison of different color spaces in the xy-chromaticity diagram; from [Myndex \[2022\]](#). A color space's chromaticity gamut is a triangle; a color outside the triangle cannot be physically produced in that color space.

usually the r and g coordinates are used. The right panel in Figure 8 shows the spectral locus in the rg-chromaticity diagram.

xy-Chromaticity Diagram and Its Interpretation

Of course we can do the same if a color is defined in the XYZ space or the LMS cone space, and we omit the trivial math here. The left panel in Figure 9 shows the xy-chromaticity diagram. It is obtained by first converting from the XYZ space to the xyz space and then plot only the x and y axes (z is implicit in that $x + y + z = 1$). The horseshoe curve is the spectral locus. For the reference, we also show the three primary lights and the white point of the CIE 1931 RGB color space as well as the Planckian locus, which shows the chromaticities of the black-body radiation at different temperatures (Figure 5).

We can make a few general observations. First, the triangle in the diagram represents the chromaticity values of all the colors that can be produced by mixing different amounts of the three colors whose chromaticities are the vertices of the triangle. That is, given three colors $[R_1, G_1, B_1], [R_2, G_2, B_2], [R_3, G_3, B_3]$ and their chromaticity coordinates $\mathbf{c}_1 = [\frac{R_1}{R_1+G_1+B_1}], \mathbf{c}_2 = [\frac{R_2}{R_1+G_1+B_1}], \mathbf{c}_3 = [\frac{R_3}{R_1+G_1+B_1}]$, the triangle formed by the points $(x_{\mathbf{c}_1}, y_{\mathbf{c}_1}), (x_{\mathbf{c}_2}, y_{\mathbf{c}_2}), (x_{\mathbf{c}_3}, y_{\mathbf{c}_3})$ contains all the colors that can be produced by mixing the three colors in the ratios R_1, R_2, R_3 , G_1, G_2, G_3 , and B_1, B_2, B_3 .

$[\frac{R_2}{R_2+G_2+B_2}]$, and $\mathbf{c}_3 = [\frac{R_3}{R_3+G_3+B_3}]$, we can show if we mix these colors to form a color C, $[\alpha R_1 + \beta R_2 + \gamma R_3, \alpha G_1 + \beta G_2 + \gamma G_3, \alpha B_1 + \beta B_2 + \gamma B_3]$ (α, β, γ are the contributions of the primary colors), C's chromaticity is necessarily inside the triangle $\Delta\mathbf{c}_1\mathbf{c}_2\mathbf{c}_3$. So the triangle $\Delta\mathbf{RGB}$ represents the chromaticities that can be physically produced by the CIE 1931 RGB primary lights. We call that the **chromaticity gamut** of the color space, or sometimes simply the gamut of the color space, but we should keep in mind that the actual gamut of a color space is always a three-dimensional concept.

Second, we can extend from mixing three colors to mixing arbitrary number of colors and show that the interior of the spectral locus represents the chromaticities of all the colors that humans can see, i.e., the gamut of the HVS. This is true because the shape of the spectral locus is convex, so connecting any two points (i.e., mixing two colors) on or inside the locus will never go beyond the locus. By extension, a positive linear combination of any points on or inside the locus will always stay inside the locus. A natural implication is that any point outside the spectral locus represents an imaginary color, since that point can never be constructed by a positive linear combination of points on or inside the spectral locus.

Third, the right panel in Figure 9 shows the gamut of a few common color spaces. The sRGB color space is the most commonly used color space; virtually every single display supports it and an image, by default, is encoded in the sRGB format. We will have more to say about displays and image encoding later. Observe how small the sRGB gamut is: it covers about 35% of the HVS gamut. P3 is a more wider gamut that is supported in many new displays. Rec.2020 is an even wider gamut that is yet to be widely supported; it is 72% larger than the sRGB gamut and 37% larger than the P3 gamut. ProPhotoRGB contains colors that are beyond the HVS gamut, so to produce all the real colors in the ProPhotoRGB space we will need more than three primary lights. It is mostly used in Adobe Lightroom and Adobe Camera RAW software. They both deal with RAW images before they are encoded in a format that is displayable. We will talk about RAW imaging and processing later in the class.

Finally, no display can produce all the colors that humans can see. No matter where you choose to place the primary colors in the chromaticity diagram and how many primaries you choose, the resulting gamut will never completely cover the entire HVS gamut as long as the primary colors are real colors (i.e., on or inside the spectral locus) and you have finite number of them. This is again because the spectral locus is convex. For this reason, do not trust the colors in any xy-chromaticity diagram: the undisplayable colors are approximated by in-gamut, displayable colors. This is called gamut mapping, which we will discuss in Chapter 3.6.

HVS Gamut

We can systematically sample the chromaticities in the chromaticity diagram to visualize how the HVS gamut looks like. Figure 10 visualizes the HVS gamut in both the XYZ space and the xy-chromaticity diagram. Comparing the two, you can see how a selected set of colors in highlighted the XYZ space map to a curve in the xy-chromaticity diagram.

There are of course many ways you can sample the chromaticities to get a good coverage of the HVS gamut, and [Zhu \[2022d\]](#) is an interactive tutorial that talks about this in detail (you can also see how the HVS gamut looks like in different color spaces). A common way seems

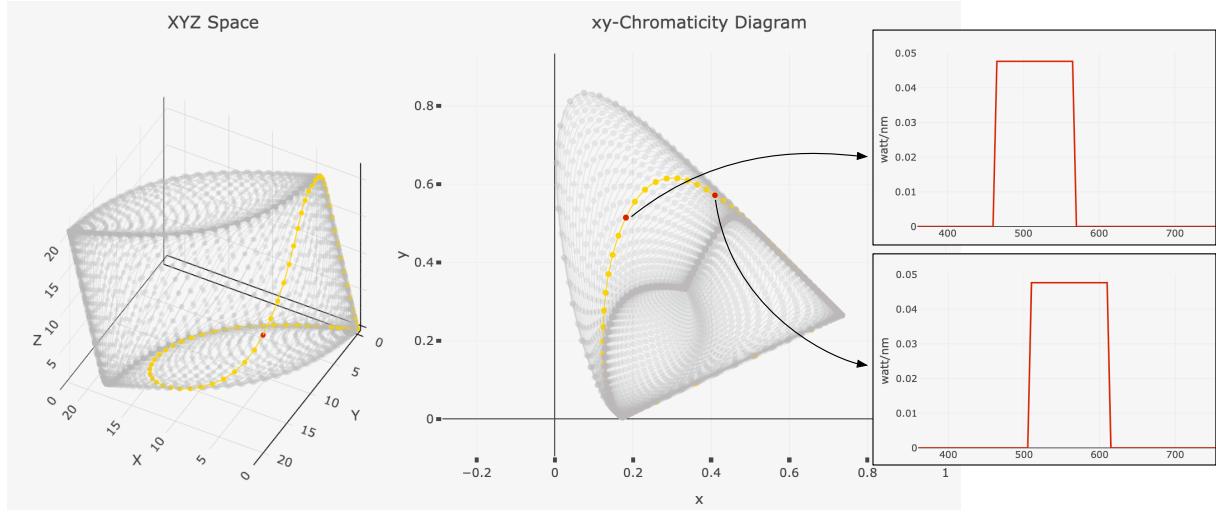


Figure 10: HVS gamut visualized in the XYZ space and in the xy-chromaticity diagram. We systematically sample the chromaticities in the chromaticity diagram using square pulses as the light SPDs (insets on the right). From the interactive tutorial in [Zhu \[2022d\]](#).

to be to generate SPDs that are square pulses with equal peaks (see the insets on the right), which will guarantee that you do not repeatedly sample the same chromaticity point. This is what the popular Python package Colour [[NumFOCUS](#)] does, but nothing prevents you from using a different method, as explored in [Zhu \[2022d\]](#). Of course, the actual HVS gamut has no boundary: we can indefinitely grow the gamut by simply scaling up the light power.

3.3 Color Cube

The various color spaces we have been discussing are great, but they do not seem to be the sort of color spaces we use in everyday software when specifying colors. By far the most common way in practical applications to specify colors is by using a **color cube**, where you can specify the primary values (usually R, G, and B) of a color, each an integer between 0 and 255. What exactly are the colors that can be represented by such a color cube? How is it related to the color gamut we have discussed, and how do we construct a color cube? These are questions explored in the interactive tutorial [[Zhu, 2022d](#)], which you are invited to go through. Figure 11 illustrates the idea, and we will give a brief summary of the main steps.

- From Chapter 2.3, we know that a color space is defined by its three primary colors and the white point, which you get to choose when building your own color cube. The left panel shows one such choice, which happens to be what is used by the sRGB color space.
- Knowing these four points uniquely defines the shape of a parallelepiped in the XYZ space (middle panel). The space inside the parallelepiped corresponds to actual colors that can be produced by using the primary colors.

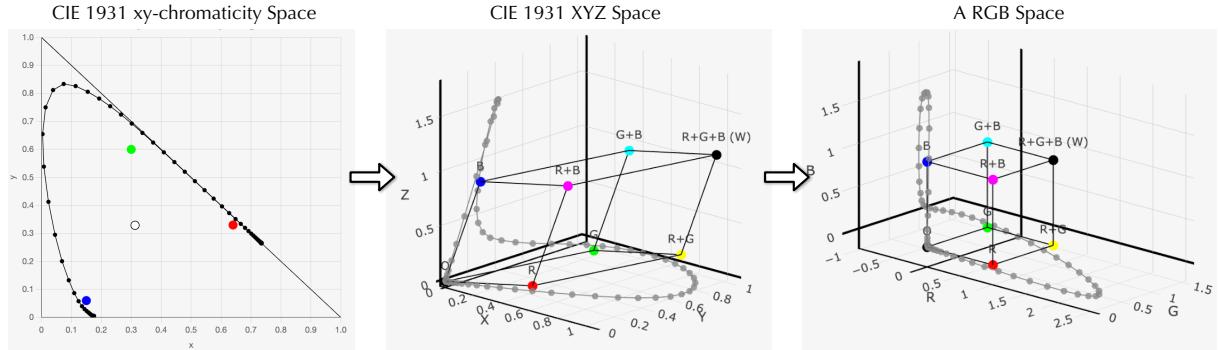


Figure 11: Pick the primary colors (which usually are termed R, G, and B, because they usually are red-ish, green-ish, and blue-ish) and the white point in the xy-chromaticity space (left panel) and then construct a color cube from them (right panel). Note how the spectral locus is now positioned in the constructed RGB space. From the interactive tutorial in [Zhu \[2022c\]](#).

Note that at this point we know only the relative shape, but not the absolute scale, of the parallelepiped: we can uniformly scale the power of the primary colors and white point, which will not change their chromaticity values but will expand or shrink the parallelepiped. The convention is to set the Y value of white to be 1 and normalize everything else accordingly, but of course the actual luminance of white (and any other color) depends on the actual device used.

- Now we turn the parallelepiped to a cube that is positioned between $[0, 1]$ in all three directions (right panel). The white point in the XYZ space will be $[1, 1, 1]$ in the color cube, signifying that white is produced from equal units of the three primary colors. This amounts to a linear transformation from the XYZ space.

Note also how the spectral locus is now positioned in the RGB space: part of the locus (and by extension the HVS gamut) is now outside the RGB cube, showing that there exists real colors (i.e., inside the HVS gamut) that cannot be produced by the choice of the primary colors. This is consistent with our gamut interpretation in the chromaticity diagram (Figure 9).

Color Quantization and Gamma

We get a cube now, but we are not done yet. The cube is a continuous solid between $[0, 0, 0]$ and $[1, 1, 1]$, but the digital representation of a color is discrete and finite, so we have to quantize the solid. Assuming we have, say, 8 bits (i.e., 256 discrete levels) to represent the contribution of each primary color, the question is how to allocate the 256 levels to the $[0, 1]$ range.

So far the contribution of a primary color is linearly correlated with the power of the primary: doubling the contribution of a primary requires doubling the power of the corresponding light. Therefore, a uniform allocation of the bits would mean uniformly quantizing the light power

range. As we have seen in the photoreceptor chapter, the electrical response of a photoreceptor is not linearly proportional to the light power (even though the amount of photon absorption and pigment excitation are!); the response incrementally saturates as the light power increases. As a result, the perceptual brightness level also gradually saturates with the light power. Therefore, uniforming quantizing the power range would lead to a *non-uniform* quantization of the brightness range, which is what we ideally want in order to best use the limited bit budget.

To uniformly quantize the brightness levels, a common method is to first model the brightness level (B) as a power-law function of the raw channel value ($v \in [0, 1]$) by $B = v^{1/2.2}$ and then quantize B uniformly. The constant factor 2.2 is called the **gamma** of the system. For instance, a red-channel value of 0.5 would translate to $\lfloor 0.5^{1/2.2} \times 255 \rfloor = 186$ in an 8-bit encoding. The relationship between B and v is called the Opto-Electronic Transfer Function (**OETF**). OETF is usually performed by an imaging system such as a camera, which turns optical signals (luminance) to electrical signals (bits in a color space).

Note that the gamma-based OETF does not model the actual relationship between perceived brightness and light luminance, but it is a close engineering hack. The behavioral brightness perception is largely accounted for by the photoreceptor/RGC response to light intensity. As we discussed in the photoreceptor chapter, the relationship between the electrical response of a photoreceptor and the light intensity is usually modeled by a (generalized) Michaelis equation, which incrementally saturates and exhibits a diminish return, just like a power-law function using a gamma.

The sRGB color space [Anderson et al., 1996] slightly modifies this OETF to avoid numerical issues when v is small. The sRGB standard uses a linear scaling when v is very small³ and adjust the gamma to be 2.4 so that the overall quantization function approximates a uniform power-law function with a gamma of 2.2.

There are two caveats here. First, v is *proportional* to luminance L , but is *not* exactly L , so the same v will result in different L s on different displays that differ in their peak luminance. So encoding B as a power-law function of v does not mean the OETF actually models the correct relationship between B and L . That is why the sRGB standard specifies the peak luminance of the display (white point) as 80 cd/m². Second, light adaptation (a later topic) will also play a role, since the HVS responds to contrasts over the mean illuminance, rather than absolute illuminance, and the mean illuminance vary largely across viewing environments. The sRGB standard also specifies that the mean illuminance level of the viewing environment to be 64 lux. When viewing an sRGB image on an actual display, both condition are rarely met, so take all these with a huge grain of salt.

Color Spaces are Linearly Related in Luminance-Linear Space

Different color spaces might use different gammas and quantization schemes. In the end, the discrete RGB values are usually not linearly related to luminance. We can remove gamma (gamma decoding) to go back to a luminance-linear space from the discrete RGB values. Once

³This makes sense given our understanding in the photoreceptor chapter that the receptor's electrical response is approximately linear against the light luminance when the luminance is very low.

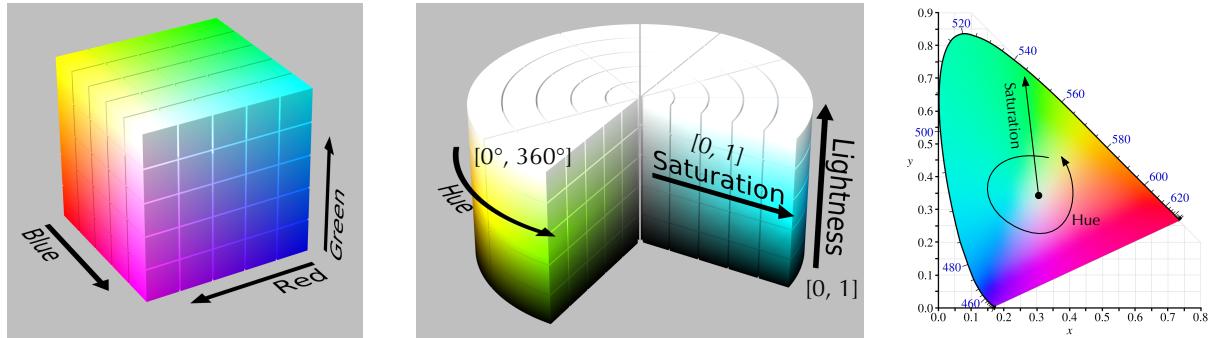


Figure 12: We can represent an RGB color cube (left) using cylindrical coordinates. One such representation is the HSL color space (right), where hue, saturation, and lightness have intuitive interpretations. Hue and saturation also have intuitive interpretations in the CIE 1931 xy-chromaticity diagram, which normalizes luminance so lightness information is absent. Left: from [SharkD \[2010b\]](#). Middle: adapted from [SharkD \[2010a\]](#). Right: from [BenRG \[2009\]](#).

in a luminance-linear space, different color spaces are simply a linear transformation away from each other. The transformation matrix can be calculated based on the primary colors and the white point of the two color spaces. We will omit the math here, but to get an idea just go back to Figure 11; two color spaces having different primary colors and white points will end up being two different parallelepipeds. The transformation matrix transforms one parallelepiped to the other.

3.4 HSB/HSL/HSV Space

A color cube is one way to represent a (RGB) color space. Another common way to represent an RGB color space is to use a cylindrical-coordinate representation. There are two such representations, HSL (Hue, Saturation, and Lightness) and HSV (Hue, Saturation, and Value), which is also called HSB (B for Brightness). These are not new color spaces; they have the exactly same gamut as the corresponding RGB color space. They are just different ways to organize colors in a color space; instead of using three-dimensional coordinates to represent a color as in a color cube, they use cylindrical coordinates.

Figure 12 compares a typical color cube (left) and its HSL representation (right). We omit the transformation math here, but one can imagine how we turn the white point in a color cube to the top plane, the black point to the bottom plane, and expand everything else so that a cube surface morphs into a cylindrical surface. The three dimensions in an HLS space are: hue, saturation, and lightness. Very informally, hue represents subjectively different colors (red, orange, yellow, etc.), saturation represents how much white a color has (a color with a higher saturation means it is more “pure”), and lightness represents the brightness. In this sense, hue and saturation also find their interpretations in the CIE 1931 xy-chromaticity diagram (right), where a color closer to the spectral locus has a higher saturation (and colors closer to white-ish colors are *desaturated*) and the spectral locus cycles through different hues. Lightness is not

concerned with in the chromaticity diagram, which normalizes the color intensity.

You can imagine what the benefit of using an HSL/HSB color space is. It is more intuitive to pick colors in these color representations since the three dimensions have intuitive interpretations that better align with how we describe colors in our everyday language. So we can more easily reason about how a color will change if we vary a dimension. In contrast, it is sometimes hard to predict how a color will change when we, say, increase the red channel by 10. I almost exclusively use the HSL/HSB space when picking colors in graphing software.

3.5 Display Native Gamut

The display has a native color space. Each display pixel is implemented by (usually) three sub-pixels, each of which has an implementation-specific SPD and acts as a primary light. The retina then spatially integrates the lights from the three sub-pixels, i.e., mixing the three primary colors. We can individually control the luminance of each sub-pixel and, by extension, the actual color of the mixed pixel. The luminance can be controlled by 1) the duty cycle of a pixel through Pulse Width Modulation (PWM), 2) the current supply to each sub-pixel, or 3) the voltage supply to each sub-pixel. The luminance is strictly linear with respect to the drive signal in the first case, approximately linear in the second case, and non-linear in the third case [Miller, 2019, p. 112]. The mapping from the electrical drive signal strength to the luminance level is usually called the Electro-Optical Transfer Function (**EOTF**).

The display's native color space is mostly like not exactly sRGB or any standard color space. The primary colors (and the white point) depend on the emission spectrum of each sub-pixel, which in turn depends on the material used. For instance, inorganic LEDs have a narrower emission spectra than the organic LEDs [Huang et al., 2020], so they tend to be able to generate more saturated colors and, thus, the resulting display gamut is wider. One has to balance multiple trade-offs in a display design, such as invariance of chromaticity vs. luminance, lifetime, power consumption, and cost, so it is difficult to tune the pixel spectra *just* so that the colors precisely match that of a standard.

Field Sequential Displays (FSD) rely on the temporal integration of our visual system to create different colors. The most common example of a FSD is modern Digital Light Processing (DLP) projectors. We will discuss display implementation technologies later in the class. For now, we will focus on the color space of a display regardless of how the colors are produced.

As an example, Figure 13 shows the the sub-pixels images of the green primary colors in the P3 and sRGB color space as displayed on a 4th-generation iPad Pro. We can make a few observations. First, the emission patterns of P3 green and sRGB green are different. The P3 green is more “pure”, where the red and blue sub-pixels are contributing very little, whereas the sRGB green requires noticeable contribution from the red sub-pixels. This is not surprising because the P3 green is much more saturated (closer to spectral colors) as the sRGB green, as shown in the right figure in Figure 9. The actual contributions of red sub-pixels in sRGB green as seen by my eye are not as strong as seen in this iPhone-taken image; the image signal processing pipeline in the iPhone definitely has introduced its artifacts.

Second, even for the P3 green, there are still some contributions from the red sub-pixels. This suggests that the native display gamut is different from (and larger than) P3. This makes

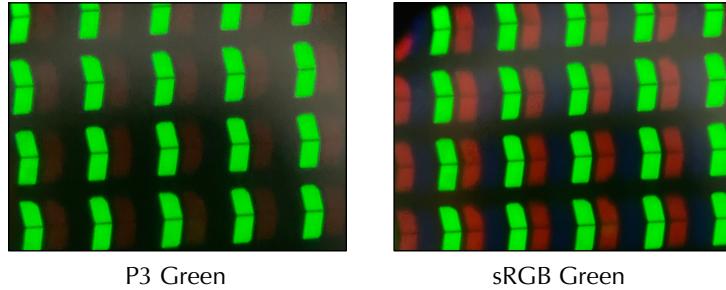


Figure 13: Microscope-magnified subpixel images of P3 green and sRGB green primary (both are $[0, 255, 0]$ in their respective color spaces) on a 4th-generation iPad Pro taken from an iPhone 12 Pro (whose image signal processing chain introduces color inaccuracies; the red sub-pixel contributions to the sRGB green are not as strong when seen by naked eye). As a side note, you can also see that when the image is focused on the green sub-pixels, the red (and blue) sub-pixels are out of focus, a result of chromatic aberration.

sense: for a display to support a particular color space, say, the P3 space, the display’s native color space must be no smaller than the P3 space.

3.6 Color Management

An end-to-end workflow might involve multiple color spaces, and it is important to correctly translate colors between color spaces to retain color accuracy. For instance, you might edit a photo encoded in the P3 color space, save the photo in a file and share it your friend who will view the image on a display that supports only the sRGB color space.

Multiple color spaces are involved here. The image is first encoded in P3 space, and then will have to re-interpreted in the sRGB space. A color, say, $[10, 20, 30]$ encoded in the P3 color space is not the same color as the sRGB color $[10, 20, 30]$, so we must correctly translate a color encoded in the source color space to the destination color space. A critical issue in this transformation is that the P3 color space has a larger gamut than that of sRGB, so there will necessarily be colors in the photon that will never be accurately reproduced on your friend’s display — what do we do with these colors? Each display also has its own native color space, and an sRGB/P3 image will have to be transformed to the display native space for display. Fundamentally, if we want to display a, say, P3-encoded image, the display’s native gamut must be no smaller than P3.

Taking care of all these is part of **color management**, whose goal is to maintain a consistent color appearance throughout the workflow that might involve wildly different devices. It requires a collaboration between every single piece that touches color in the workflow: the image file must specify what color space its pixel colors are encoded in (called a *profile*), the software that manipulates image content must correctly read and interpret the profile and perform the necessary transformation, potentially through APIs exposed by the Operating System (OS), and the display firmware and drive must communicate with the OS what sort of color spaces they

support. [Giorgianni and Madden \[2009\]](#) and [Sharma \[2018\]](#) are two excellent references for color management. We will describe the key issues here.

Converting Pixel Colors to Drive Signals

When opening and viewing an image encoded in, say, sRGB on a display, a few transformations have to happen [[Miller, 2019](#), Chap. 7.1]. Since the display's native color space is mostly like not exactly sRGB or any standard color space, we must correctly translate a color encoded in the sRGB space to the display's space. A color [10, 20, 30] encoded in sRGB is not the same color as [10, 20, 30] in the display's color space. This transformation is done in two steps.

First, the image file ideally has metadata that tells us what color space its pixel colors are encoded in or, better, the transformation matrix from the image's color space to a device-independent color space, say the CIE XYZ space. The way to describe such information has been standardized by International Color Consortium (ICC) in what is called the *ICC profile* [[International Color Consortium, 2019](#)]. We can embed an ICC profile in common image file formats such as JPEG. Second, the display itself also has to report its native color space. To do that, modern displays usually come with an ICC profile that describes how to transform from the CIE XYZ space to the display native space. Now when the Operating System gets the image file, it would first transform the sRGB colors to the XYZ space using the ICC profile in the image and then transform the colors in the XYZ to the display native space using the display profile⁴. You can see that the XYZ space here serves to connect the input color space and the output color space. ICC calls such a space a Profile Connection Space (PCS).

The transformation from the XYZ space to the display native space is necessarily linear. To calculate the transformation matrix, we will first measure the chromaticity values of the display native primary colors and the white point offline [[Balasubramanian, 2003](#)]. Then we take the exactly the same steps as described in Chapter 3.3: we are essentially creating a color cube for the display ([1, 1, 1] represents the display white point, i.e., when all the sub-pixels emit maximum luminance, etc.).

After this transformation, we have obtained a set of luminance-linear, analog (between [0, 1]) color values in the display's native color space. The next step is to turn the real-valued colors to discrete values (drive signals) that can be sent to the display to control the luminance of each sub-pixel. Ideally, we want 255 (assuming 8 bits) to produce maximum luminance and 0 to produce minimum luminance. Depending on how the display adjusts its luminance (by PWM, current, or voltage), the drive signal vs. luminance relationship, i.e., EOTF, may or may not be linear. Either way, we can offline calibrate an EOTF look-up table (or regress a function), from which we can then map a desired luminance level to a discrete value.

What is the desired luminance level for a pixel? It would be *amazing* if your display can reproduce the scene luminance, but that is unlikely, because the real world has a much higher, orders of magnitude higher, dynamic range (DR) than that of a display. A main challenge in imaging and display, thus, is **tone mapping**, which is concerned with mapping a high-dynamic-

⁴While in the XYZ space we usually perform an additional transformation so that sRGB white becomes the white point in the display space. This is called *chromatic adaptation*, which we will discuss later in the class.

range scene to a low-dynamic-range display. This mapping can be described by a Opto-Optical Transfer Function (**OOTF**). Both the OETF of an imaging system and the EOTF of a display participate in the OOTF, and if the product of OETF and EOTF is not the desired OOTF, one would need to implement an Electro-Electrical Transfer Function (**EETF**) as part of the image processing pipeline to reach the desired OOTF. Tone mapping is the focus of extensive research [Reinhard, 2010; Mantiuk et al., 2015].

Gamut Mapping

When viewing a P3-encoded image on a display whose gamut is smaller, e.g., similar to that of sRGB, the colors might not be accurately reproduced. The best thing we can do is to approximate an out-of-gamut color with a in-gamut color to minimize the color error. This is called **gamut mapping**. Morovič [2008] and Glassner [1995, Chap. 3.6] describe the basic algorithms with the former being more recent and comprehensive.

The simplest strategy would be to simply clamp out-of-range values, so a color of [12, 200, 300] would become [12, 200, 255]. Clearly, other than being extremely simply to implement this strategy would introduce large color reproduction errors. ICC has defined four **rendering intents**, each of which corresponds to a gamut mapping algorithm (vaguely worded and the implementation detail might vary). For instance, the *Absolute* rendering intent leaves all the in-gamut colors unchanged but maps the out-of-gamut colors to the boundary of the color gamut. The *Perceptual* rendering intent can be implemented by uniformly projecting all the colors to the white point so that all the colors are in-gamut. You can imagine that while this maintains the relative color appearance between colors (which the Absolute rendering intent fails at), but it would also change in-gamut colors that could have been accurately rendered!

3.7 Color Differences and “Perceptually Uniform” Color Spaces

In many practical applications we need to calculate color differences. For instance, an image synthesis algorithm might want to be minimize the color difference in the synthesized image and some form of “ground truth”; a display’s color reproduction might not be 100% accurate so we want to quantitatively compare the quality of different displays by measuring the color difference (compared to the colors to be reproduced) each introduces. Fortunately, once we put colors into a three-dimensional coordinate system, calculating color differences becomes natural: the distance between two colors gives a measure of the difference between the two colors.

However, for the Euclidean distance to be a good measure, we must be sure that the distance is proportional to the perceptual color difference. How do we quantify the perceptual color difference? Practically there are not many cases we need to quantify large color differences. What is more important is to quantify small color differences. So a typical approach is to estimate the **Just Noticeable Difference** (JND) threshold of a color. For a given reference color, we can use a threshold-detection psychophysical paradigm to estimate the set of colors that are just noticeably different from the reference color. These experiments are called *color discrimination* tests.

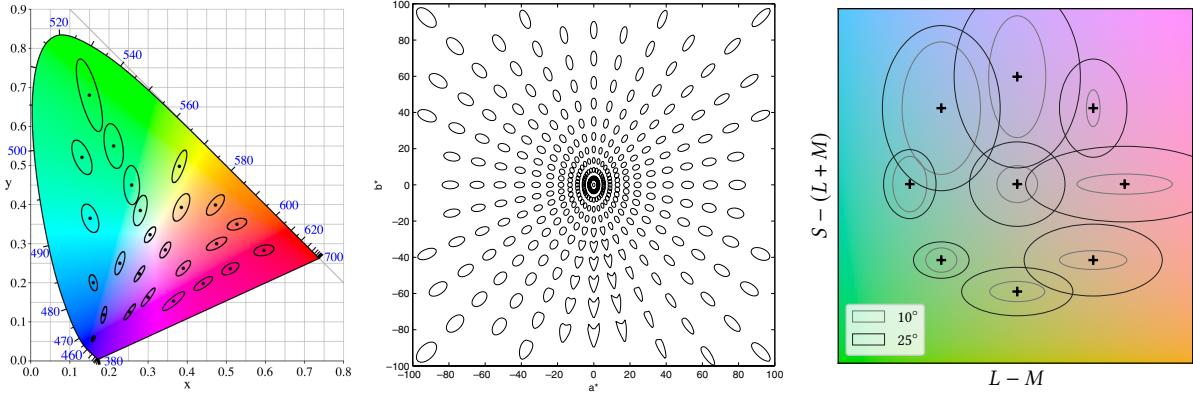


Figure 14: Left: MacAdam ellipses (measured at 2° eccentricity) plotted in the xy-chromaticity diagram (the ellipse sizes are magnified 10 times to be more visible); from [Anonymous \[2009\]](#). Each ellipse is an iso-discrimination contour, within which all the colors are non-discriminable from the center, reference color. Middle: Iso-discrimination contours corresponding to a $\Delta E_{00} = 1.0$ in the $a^* - b^*$ plane in the CIELAB space; from [Sharma \[2003, Fig. 1.18\]](#). Right: A set of MacAdam ellipses in the (chromatic plane of the) DKL space [[Derrington et al., 1984](#)] (also see Chapter 4.3) under two eccentricities; from [Duinkharjav et al. \[2022, Fig. 4b\]](#).

A color space is said to be “perceptually uniform” if the JND measure is the same regardless of where the reference color is in the color space. If so, the Euclidean distance is a good measure of perceptual color differences. Unfortunately, the common CIE XYZ space is not perceptually uniform. This is demonstrated in the seminar work by [MacAdam \[1942, 1943\]](#) (MacAdam did the work while working for Eastman Kodak at Rochester and he later was an adjunct professor at University of Rochester)⁵. He measured the thresholds in the CIE 1931 xy-chromaticity space for a set of colors. He found that the thresholds for a reference color fit an ellipse-shaped contour. Within an iso-discrimination ellipse all the colors are non-discriminable with respect to the center, reference color. A modern rendition of his results is shown in Figure 14 (left panel); the ellipse sizes are magnified ten times to be visible. Of course the actual iso-discrimination contour would be a 3D solid (ellipsoid) in the XYZ space; the ellipses in the xy-chromaticity diagram are projections of the ellipsoids.

We can see that not only the iso-discrimination contours (ellipses) are not circles, their shapes also vary significantly across the gamut, indicating that the XYZ space is not perceptually uniform. Quite a few attempts have been made to transform the XYZ space to a more perceptually uniform space. The two common ones are the CIE 1976 $L^*u^*v^*$ (CIELUV) space and the (more widely used) CIE 1976 $L^*a^*b^*$ (CIELAB) space, both of which are non-linear transformations from the XYZ space.

The so-called CIE Delta E 1976 color difference metric (ΔE_{ab}^*) is defined as the Euclidean distance in the CIELAB space. If CIELAB is truly perceptually uniform (as far as color dis-

⁵He did not use a direct threshold-detection strategy, but indirectly estimated the thresholds using variations in color matching experiments.

crimination is concerned), ΔE_{ab}^* being 1.0 would mean a JND. However, this is not true either. CIE has since recommended a new, much more involved, and non-Euclidean measure in the CIELAB space, called the Delta E 2000 metric (ΔE_{00}), to better achieve better perceptual uniformity [Sharma et al., 2005]. The middle panel in Figure 14 shows iso-discrimination contours corresponding to $\Delta E_{00} = 1.0$ in the $a^* - b^*$ plane in the CIELAB space. If the ΔE_{00} is to be considered a perceptually uniform color difference metric in the CIELAB space, the CIELAB space itself must not be perceptually uniform given the varying contour shapes throughout the space.

MacAdam's original data were collected at 2° eccentricity. Given that the visual acuity reduces as the eccentricity increases, it is only natural that the iso-discrimination ellipses expand in size as the eccentricity. Duinkharjav et al. [2022] measures the ellipses under different eccentricities. The right panel in Figure 14 compares the results between 10° and 25° . Not surprisingly, the ellipses are larger in the latter.

4 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-

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.

4.1 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 15 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.

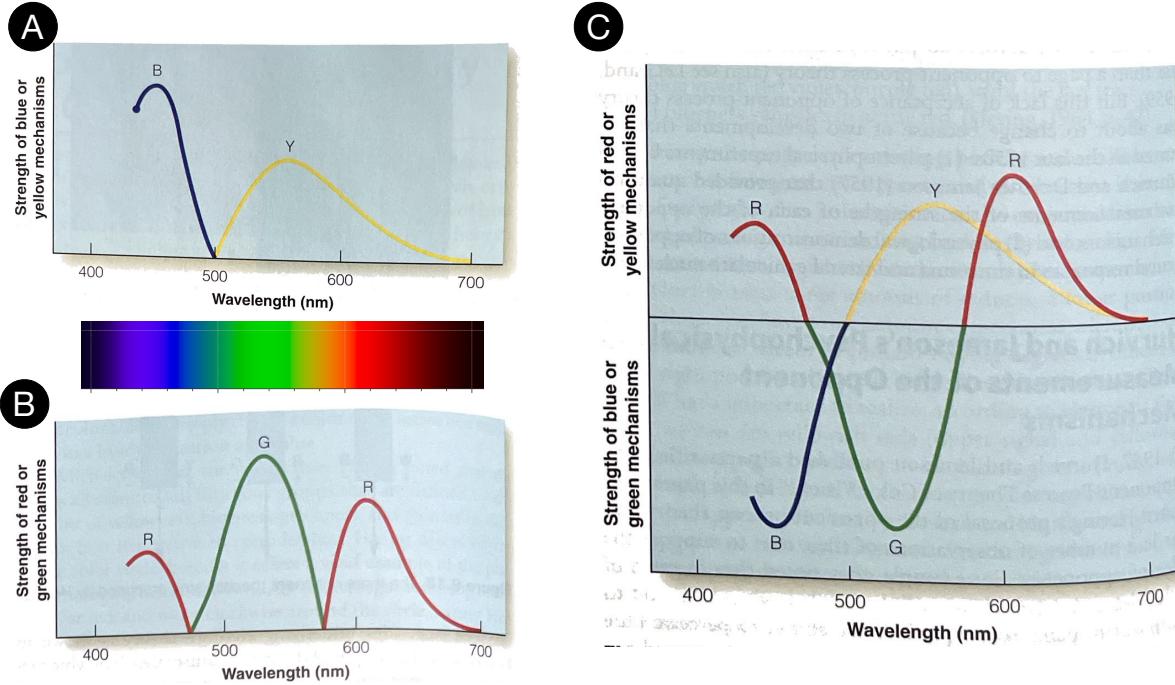


Figure 15: Measurements from the hue cancellation experiment in [Jameson and Hurvich \[1955\]](#). (A) the Blue-Yellow measurement; the y -axis shows the intensity of the Yellow/Blue cancellation light, i.e., the relative strength of the “Blue-ness” and “Yellow-ness” in the test light. (B) the Red-Green 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 15 **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 15 **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.

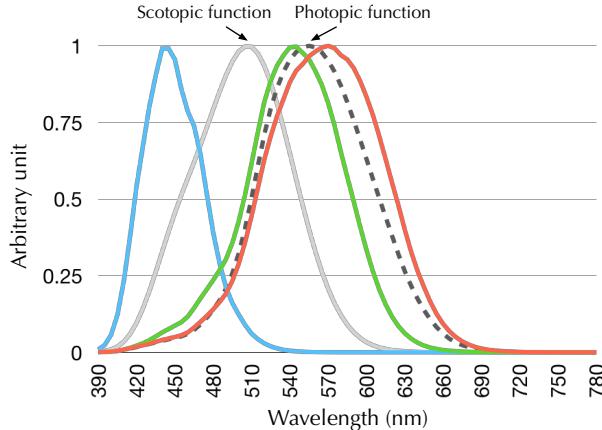


Figure 16: The grey solid curve is the scotopic luminous efficiency function (CIE 1951 standard; based on [Wald \[1945\]](#) and [Crawford \[1949\]](#)). The grey dashed curve is the photopic luminous efficiency function (CIE 2008 “physiologically-relevant” 2-deg function; based on [Sharpe et al. \[2005, 2011\]](#)) The other three curves are the cone fundamentals, shown for the reference.

4.2 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 16 shows a modern version of the photopic LEF (the so-called CIE 2008 “physiologically-relevant” 2-deg function)⁶.

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,

⁶In 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.

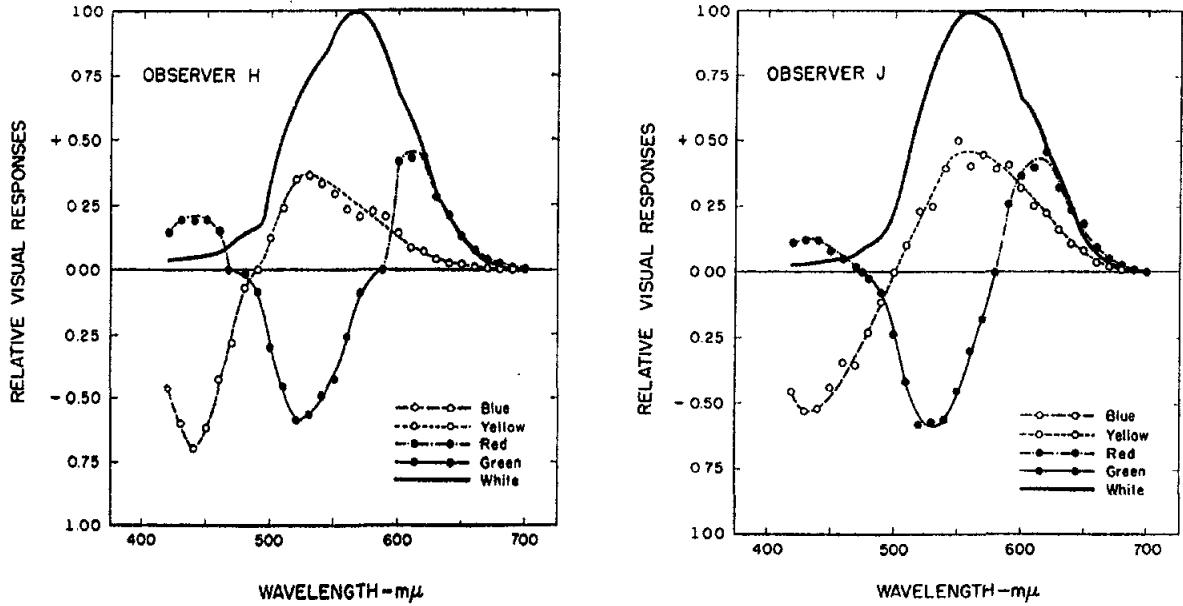


Figure 17: The solid curve is the white-black measurement, indicating the amount of whiteness in a light across the spectrum. The white-black curve in theory matches the luminous efficiency function. The two plots are for two participants. From [Hurvich and Jameson \[1957, Fig. 4\]](#).

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 16 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 17 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 4.4 and discuss how this space relates to the colorimetric spaces we have discussed so far.

4.3 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 18 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 18, 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.

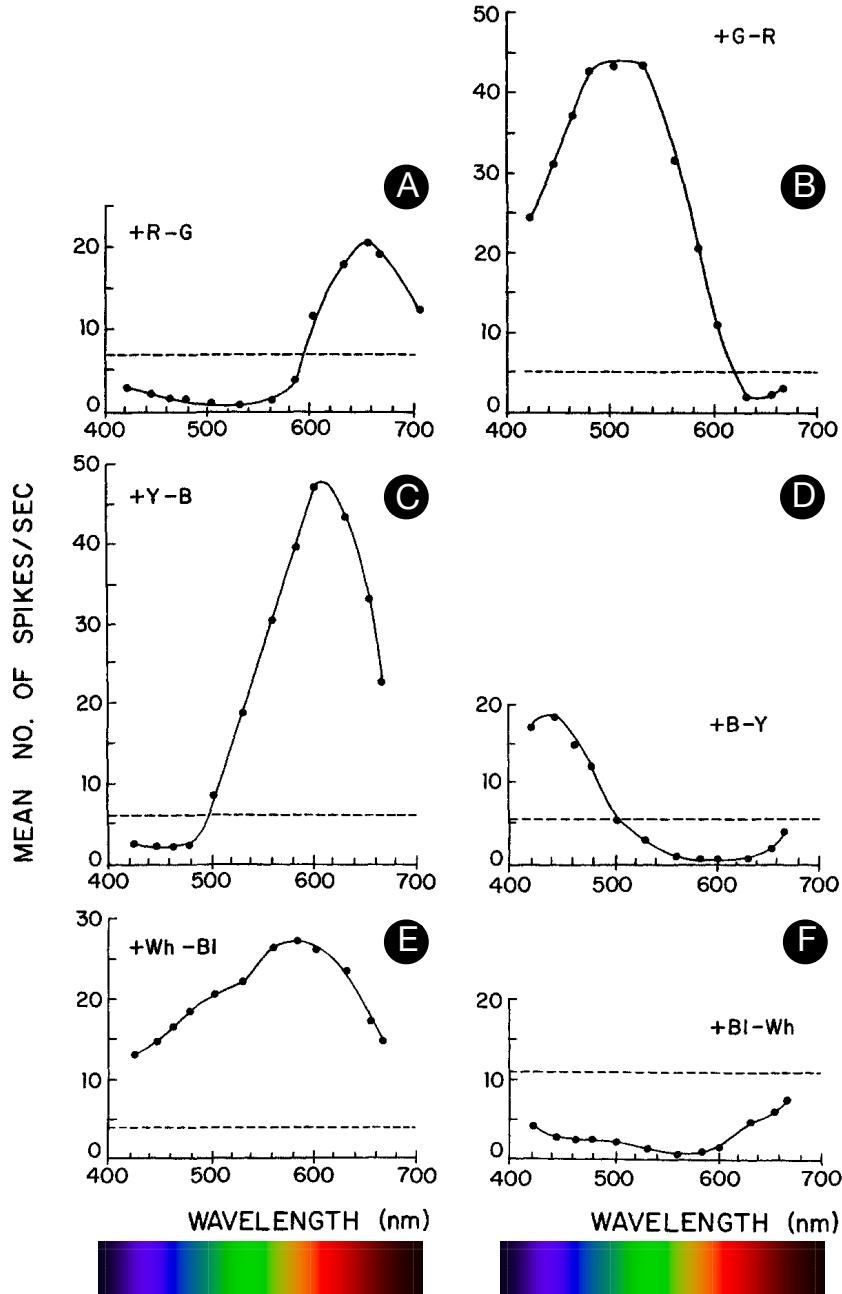


Figure 18: Responses of six typical classes of LGN neurons to incremental flashes of varying wavelengths. y -axis shows the spikes/second under spectral lights of equal energy. Each curve represents a particular energy level. (A): these cells are excited (activity exceeds the spontaneous firing rate) 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. \[1966, Fig. 9–12, 15–16\]](#).

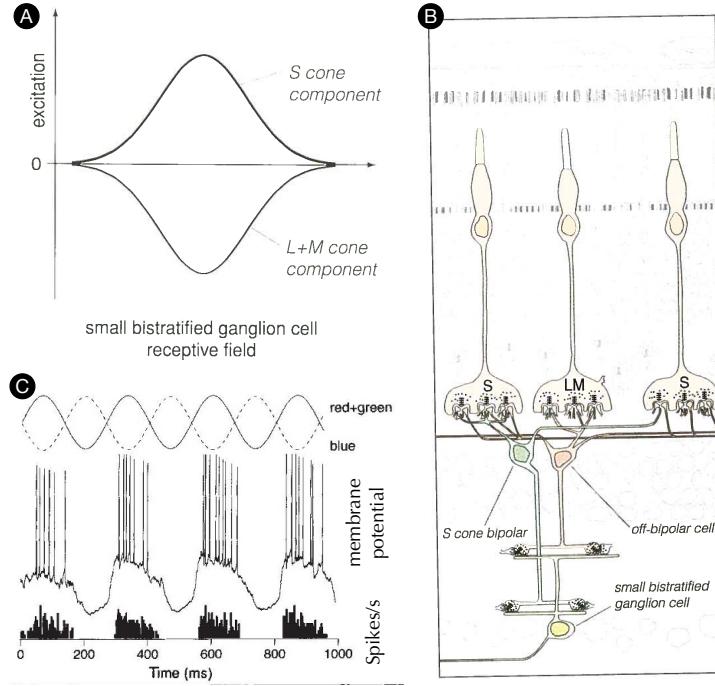


Figure 19: The small bistratified RGCs might be the substrate for the Y-B pathway. A: illustration of the receptive field structure of a small bistratified RGC, which is S-on and L/M-off (there are also S-off and L/M-on ones); from [Rodieck \[1998\]](#), p. 348. B: a small bistratified RGC receives excitatory 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 18 C.

Figure 19 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 18 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 18 E; otherwise we see a tuning curve like Figure 18 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 4.2), 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.

4.4 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 15 C) shows two perceptually neutral colors, as there are two zero-crossings. However, the spectral tuning curve of the R-G neurons (Figure 18 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

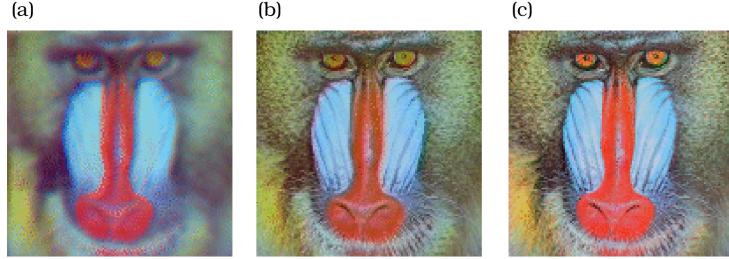


Figure 20: We take an image, decouple it into three channels: luminance, red-green, and blue-yellow. We then spatially blur one of the channels while keeping the other two channels unchanged and then reconstruct the image. Our vision is much more sensitive to spatial blurring in the luminance channel (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 20, 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 green⁷. 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:

$$\begin{bmatrix} Y/B \\ R/G \\ \text{Lum} \end{bmatrix} = T \times \begin{bmatrix} L \\ M \\ S \end{bmatrix} \quad (11)$$

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:

⁷Again, what is considered achromatic depends on the observer's adaptation state; there is no single

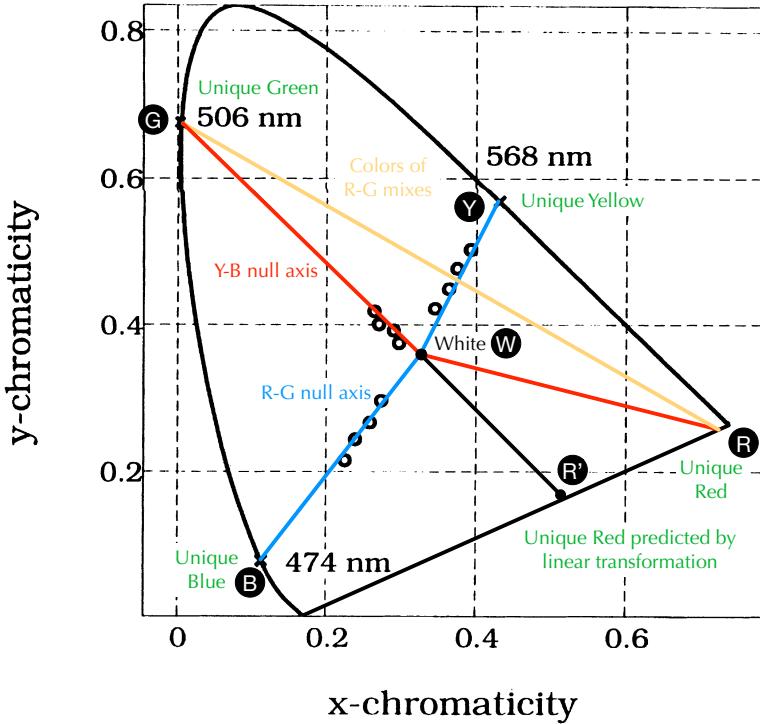


Figure 21: Circles are unique hues derived from psychophysics reported in [Bäuml \[1993\]](#). Fitting lines and extrapolating the lines give us estimations of unique hues that are spectral colors. Three of the four unique spectral hues (blue at 474 nm, green at 506 nm, and yellow at 568 nm) can be accurately 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 \[1996, Fig. 12\]](#).

$$\begin{bmatrix} 0 \\ | \\ | \end{bmatrix} = T \times \begin{bmatrix} L_R \\ M_R \\ S_R \end{bmatrix}, \quad \begin{bmatrix} 0 \\ | \\ | \end{bmatrix} = T \times \begin{bmatrix} L_G \\ M_G \\ S_G \end{bmatrix} \quad (12)$$

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

$$\begin{bmatrix} 0 \\ | \\ | \end{bmatrix} = T \times \begin{bmatrix} aL_R + bL_G \\ aM_R + bM_G \\ aS_R + bS_G \end{bmatrix}, \quad (13)$$

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 21 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 B , green G , and yellow 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 R are complementary to a spectral light at 494 nm; that is, spectral light at 494 nm, white W , and unique red hues should fall on a straight line⁸.

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 B , white W , and yellow 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 R' deviates significantly away from red hue R from actual measurements. If we connect the unique green hue G and unique red hue R' , the line would not cross 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].

References

- Matthew Anderson, Ricardo Motta, Srinivasan Chandrasekar, and Michael Stokes. Proposal for a standard default color space for the internet—srgb. In *Color and imaging conference*, volume 4, pages 238–245. Society of Imaging Science and Technology, 1996.
- Anonymous. MacAdam Ellipses in the CIE1931 xy chromaticity diagram; CC BY-SA 3.0 license. https://commons.wikimedia.org/wiki/File:CIExy1931_MacAdam.png, 2009.
- Raja Balasubramanian. Device characterization. In Gaurav Sharma, editor, *Digital color imaging handbook*, pages 269–384. CRC Press, 2003.
- 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.

⁸Of course, it is conceivable the result might vary in population and depend on the adaptation state (i.e., what is considered white/achromatic).

- 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.
- BenRG. CIE1931 xy chromaticity plot; released into the public domain by the copyright holder. https://commons.wikimedia.org/wiki/File:CIE1931xy_blank.svg, 2009.
- 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.
- DH Brainard. Cone contrast and opponent modulation color spaces. In *Human Color Vision*, pages 563–579. Optical Society of America, 2 edition, 1996.
- Michael H Brill. Erratum: How the cie 1931 color-matching functions were derived from wright-guild data. *Color Research & Application*, 23(4):259–259, 1998.
- 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.
- BH Crawford. The scotopic visibility function. *Proceedings of the Physical Society. Section B*, 62(5):321, 1949.
- 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.
- Budmonde Duinkharjav, Kenneth Chen, Abhishek Tyagi, Jiayi He, Yuhao Zhu, and Qi Sun. Color-perception-guided display power reduction for virtual reality. *ACM Transactions on Graphics (TOG)*, 41(6):1–16, 2022.
- Hugh S Fairman, Michael H Brill, and Henry Hemmendinger. How the cie 1931 color-matching functions were derived from wright-guild data. *Color Research & Application*, 22(1):11–23, 1997.
- 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.
- Edward J Giorgianni and Thomas E Madden. *Digital color management: Encoding solutions*, volume 13. John Wiley & Sons, 2009.
- 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.
- P Gouras and E Zrenner. Enhancement of luminance flicker by color-opponent mechanisms. *Science*, 205(4406):587–589, 1979.
- 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.
- Yuge Huang, En-Lin Hsiang, Ming-Yang Deng, and Shin-Tson Wu. Mini-led, micro-led and oled displays: present status and future perspectives. *Light: Science & Applications*, 9(1): 105, 2020.
- 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.

- International Color Consortium. Specification ICC.2:2019 (Profile version 5.0.0 - iccMAX). <https://color.org/specification/ICC.2-2019.pdf>, 2019.
- 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.
- 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.
- 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.
- 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.
- David L MacAdam. Specification of small chromaticity differences. *Josa*, 33(1):18–26, 1943.
- Rafał Mantiuk, Grzegorz Krawczyk, Dorota Zdrojewska, Radosław Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. High dynamic range imaging. In *Wiley Encyclopedia of Electrical and Electronics Engineering*. Wiley, 2015.
- 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.
- Michael E Miller. *Color in Electronic Display Systems*. Springer, 2019.

- John D Mollon. Introduction: Thomas young and the trichromatic theory of colour vision. In *Normal & defective colour vision*, pages 19—34. 2003.
- Ján Morovič. *Color gamut mapping*. John Wiley & Sons, 2 edition, 2008.
- Myndex. A comparison of RGB gamuts of sRGB, P3, Rec2020, etc. using the CIE1931 chromaticity diagram; CC BY-SA 4.0 license. https://commons.wikimedia.org/wiki/File:CIE1931xy_gamut_comparison_of_sRGB_P3_Rec2020.svg, 2022.
- 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>.
- Allen B Poirson and Brian A Wandell. Appearance of colored patterns: pattern–color separability. *JOSA A*, 10(12):2458–2470, 1993.
- Erik Reinhard. *High Dynamic Range Imaging Acquisition, Display, and Image-Based Lighting*. Morgan Kaufmann Publishers, 2 edition, 2010.
- Robert W Rodieck. *The first steps in seeing*. Sinauer Associates, 1998.
- 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.
- SharkD. HSL color solid cylinder; CC BY-SA 3.0 license. https://commons.wikimedia.org/wiki/File:HSL_color_solid_cylinder_saturation_gray.png, 2010a.
- SharkD. RGB Color Cube; CC BY-SA 3.0 license. https://commons.wikimedia.org/wiki/File:RGB_Cube_Show_lowgamma_cutout_a.png, 2010b.
- Abhay Sharma. *Understanding color management*. John Wiley & Sons, 2018.
- Gaurav Sharma. Color fundamentals for digital imaging. In Gaurav Sharma, editor, *Digital color imaging handbook*, pages 14–127. CRC Press, 2003.

- Gaurav Sharma, Wencheng Wu, and Edul N Dalal. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*, 30(1):21–30, 2005.
- 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.
- 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, 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.
- User:PAR. Plankian locus; released into the public domain by the copyright holder. <https://commons.wikimedia.org/wiki/File:PlankianLocus.png>, 2012.
- 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.
- 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: Chromaticity, Gamut, and the Scary World of Imaginary Colors. <https://horizon-lab.org/colorvis/chromaticity.html>, 2022a.
- Yuhao Zhu. Interative Tutorial: Building a Color Space From Cone Fundamentals. <https://horizon-lab.org/colorvis/cone2cmf.html>, 2022b.
- Yuhao Zhu. Interative Tutorial: Building a Color Cube. <https://horizon-lab.org/colorvis/colorcube.html>, 2022c.
- Yuhao Zhu. Interative Tutorial: Visualizing Human Visual Gamut. <https://horizon-lab.org/colorvis/gamutvis.html>, 2022d.
- Yuhao Zhu. Interative Tutorial: CIE 1931 XYZ Color Space. <https://horizon-lab.org/colorvis/xyz.html>, 2022e.