How the CIE 1931 Color-Matching Functions Were Derived from Wright–Guild Data

Hugh S. Fairman,[1]* Michael H. Brill,[2] Henry Hemmendinger[3]

[1]Hemmendinger Color Laboratory
438 Wendover Drive
Princeton, New Jersey 08540

[2]David Sarnoff Research Center
201 Washington Road
Princeton, New Jersey 08540

[3]Hemmendinger Color Laboratory
438 Wendover Drive
Princeton, New Jersey 08540

Received 28 February 1996; accepted 10 June 1996

Abstract: The principles that guided the founders of the CIE 1931 system for colorimetry are examined. The principles are applied to the Wright–Guild experimental determinations of the color mixture data to show in detail how and why each step in the development of the CIE 1931 system for colorimetry came about. These steps are examined in the light of 65 years advanced knowledge of colorimetry. The necessity for each of these principles in the modern world is examined critically to determine whether one might hold to the same principles if the system were being freshly formulated today. © *1997 John Wiley & Sons, Inc. Col Res Appl, 22, 11–23, 1997.*

Key words: CIE; CIE 1931 system; colorimetry; transformation; Wright–Guild data; color-matching; spectrum locus; alychne; primary colors; chromaticity diagram

INTRODUCTION

At 9:30 a.m. on September 18, 1931, a meeting of the Colorimetry Committee of the Commission Internationale de l'Éclairage convened at Trinity College, Cambridge, England.[1] Present were 21 delegates from six countries: France, Germany, Great Britain, Japan, the Netherlands, and the United States. John Guild opened the meeting with apologies that the agendum circulated in advance was by then out of date. He noted, however, that CIE rules then in effect allowed countries four months after the vote for national consideration. Then he presented five revised resolutions to the Committee. Those five resolutions would prove in retrospect to be the most important single event ever to occur in colorimetry, because they would set the colorist's agenda for the next 65 years, and into the foreseeable future. In fact, resolutions of any sort could have been adopted only if the two English-speaking nations could come to agreement upon them. This was because in the period after the First World War, with the devastation of the European continent, colorimetry did not have a very prominent position in Europe. Virtually all seminal writings from this period came either from the United States or from Great Britain. In the week previous to the Cambridge meeting, Guild, T. Smith (his deputy at the National Physical Laboratory), and Irwin G. Priest (Chief of the Colorimetry Division of the U.S. National Bureau of Standards) hammered out the differences between the views of the delegates of these two nations. This agreement set the stage for the presentation of the revised resolutions.

First Resolution

The first of the resolutions offered to the 1931 meeting defined the color-matching functions of the soon-to-be-

* Correspondence to: Mr. Hugh S. Fairman

adopted standard observer in terms of Guild's spectral primaries centered on wavelengths 435.8, 546.1, and 700 nm. Guild approached the problem from the viewpoint of a standardization engineer. In his mind, the adopted primaries had to be producible with national-standardizing-laboratory accuracy. The first two wavelengths were mercury excitation lines, and the last named wavelength occurred at a location in the human vision system where the hue of spectral lights was unchanging with wavelength. Slight inaccuracy in production of the wavelength of this spectral primary in a visual colorimeter, it was reasoned, would introduce no error at all.

Guild was further motivated by his view that colorimetry would in the future take place, as it did then in his laboratory, on a visual colorimeter. Accordingly, the primaries had to be producible. He largely failed to foresee the potential for control of object color by spectrophotometry, with indirect colorimetry by calculation. This further motivated him to insist upon spectral primaries to define the standard observer.

The color-matching functions thus adopted are those we know now as the CIE 1931 *RGB* data. They appear in CIE Publication 15.2 today with every digit presented in 1931 intact.[2] The data were an amalgamation of functions derived experimentally by Wright in 1929 and by Guild himself earlier, but not published until 1931. Wright published the chromaticity coordinates of his data in 1929.[3] He then published the color-matching functions themselves a year later.[4] In this article, Wright also included the chromaticity coordinates of the spectrum transformed to Guild's spectral primaries. Guild found such good agreement with functions previously developed by him, but as yet unpublished, that he published the amalgamated data with the recommendation that they be adopted as a standard observer for colorimetry.[5]

Second Resolution

The second resolution presented that day in 1931 defined the standard-illuminant spectral power distributions we know today as Illuminants A, B, and C. Illuminants A and C survive intact to the present. It was, however, recommended by the CIE in 1963 that Illuminant C be supplemented by a series of daylight illuminants.

Third Resolution

The third resolution defined the conditions for illumination and viewing for the measurement of surface colors. The conditions specified in 1931 were illumination 45° from the specimen normal and viewing at 0° from this normal. This was the one resolution with which the delegates from United States still disagreed by the time of the meeting, as the General Electric–Hardy Spectrophotometer, development of which was then being completed in the U.S., illuminated normally and viewed diffusely using an integrating sphere.

Fourth Resolution

The fourth resolution was an arcane statement of the principles by which the CIE reference primaries X, Y, and Z should be chosen. Wright, who was there, tells us that hardly anyone in the room understood the resolution, and it was adopted without discussion.[1] It had mostly to do with assuring that, among the various ways in which a colorimetric specification might be made, the trichromatic method was adopted. This eliminated the monochromatic-plus-white method, as well as several proprietary methods then being used, which were thought by metrologists to be merely derivatives of the trichromatic method. As such, it was innocuous enough that it could be adopted without discussion.

Fifth Resolution

What people in the room could understand about the selection of the new CIE reference primaries was contained in resolution five, which was a statement of the color-matching equations relating the newly adopted *RGB* color-matching functions, of Resolution (1), to the *XYZ* system for colorimetry. These equations contained the transformation coefficients, as we know them today, for transformation from *XYZ* to *RGB*.

Also contained in resolution five was a table of chromaticity coordinates at 5-nm intervals of the spectrum in terms of x and y, and a table of what are now called color-matching functions in terms of $\bar{x}$, $\bar{y}$, and $\bar{z}$. Smith, Guild, and Priest were said to have stayed up most of the night before the Cambridge meeting calculating these data for presentation at the meeting. The alternative they faced was the postponement of international agreement for three more years.

Purpose of This Article

The system for colorimetry adopted by the CIE in 1931, with some major additions and some minor modifications, has served well for many decades, and is so widely used throughout the world that major changes should not be made except for very good cause. It was, however, recognized at an early date that some visual observations were inconsistent with the predictions of the CIE system, especially with regard to luminance.[6,7,18] The problem has been under study by Project Committee 49 of the Inter-Society Color Council since 1992. Concern with the problems raised by the recent work of Thornton motivates the presentation of this review.[8]

We have, however, attempted to restrict our review to the understanding of the colorimetrists of 1931; we have not set out to reconcile the 1931 definitions based on a more modern understanding of the distinction between *colorimetric* definitions and *appearance* characteristics.

This article will examine, to the extent that we can reconstruct the discussions of 1931, what led the partici-

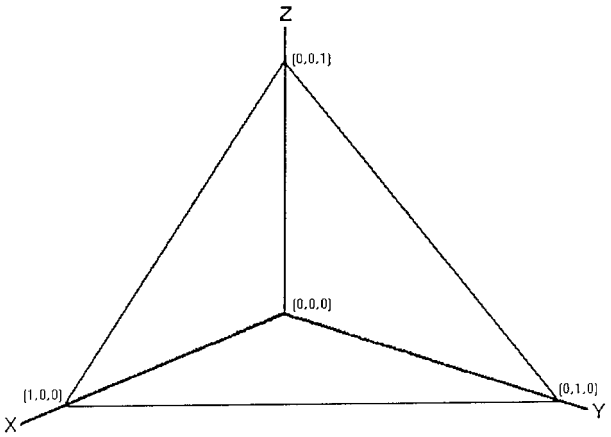

FIG. 1. (a) The framework of the CIE *XYZ* three-space. Superposed on the three coordinate axes is the Maxwell triangle, an equilateral triangle defined by $X + Y + Z = 1$.

pants from the English-speaking nations to agree in the week before September 18th on the transformation coefficients in Resolution (5). In doing so, we accept the Wright–Guild data of Resolution (1). We assume that if the experiments leading to these data were to be repeated by others using the best equipment and techniques available to them at that time, the results would be the same within the experimental variance. In examining the principles that led to the transformation coefficients, we discuss whether or not these principles would be adopted if the system were to be newly formulated today, with current knowledge and resources.

Viewed from the vantage point of 1995, it is noteworthy that the delegates who formulated the 1931 resolutions exhibited an admirable readiness to submerge their individual priorities for the sake of reaching agreement by consensus. The concepts embodied in the 1931 agreement were in large part first enunciated only within the previous decade.

THE FORMULATING PRINCIPLES FOR THE TRANSFORMATION FROM *RGB* TO *XYZ*

The formulating principles upon which Smith, Guild, and Priest came to agreement in the days leading up to the Cambridge meeting may be summarized as follows:

a. The validity of Grassmann's Laws was assumed.
b. The ratios of the luminance factors of R, G, and B at wavelengths 700, 546.1, and 435.8 nm were to be 1:4.5907:0.0601.[9] As will be seen, this choice emerged from the least-square best fit of the color-matching functions to the 1924 $V(\lambda)$ function.
c. All coordinates of all real stimuli were to be positive.
d. The units of the primaries were to be chosen so that the chromaticity coordinates of the equal-energy spectrum would be equal to each other. This principle applies to all three chromaticity coordinates. Because only two of these chromaticity coordinates are independent, we sometimes overlook that there are actually three such coordinates.
e. Subject to (b), (c), and (d) above, the chromaticities of the reference stimuli were to be chosen so that the area occupied by chromaticities of real stimuli would be the maximum area in the right triangle nearest the origin in the unit positive quadrant. The triangle occupied by these real chromaticities is then bounded at its apices by the chromaticities of the primaries at (1, 0), (0, 1), and (0, 0) for X, Y, and Z respectively. (See Fig. 1)
f. Finally, Judd had noted that there was an opportunity to set the long-wave end of the $\bar{z}$-function to all zeroes,[10] thus diminishing the labor of calculation of tristimulus values for heterochromatic stimuli, in an era when that was done utilizing mechanical calculators. Judd's proposal was adopted as a formulating principle.

Formulating Principles Detailed

Grassmann's Laws

Although it is not often stated explicitly, Grassmann's law of additivity underlies the linear transformations of colorimetry, as was pointed out by Judd in 1930.[10] It is instructive in this connection to reread a footnote in

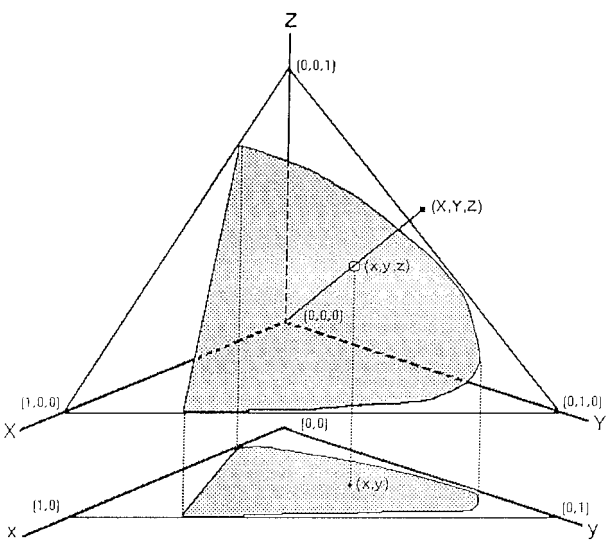

FIG. 1. (b) The location of the chromaticity plane in *XYZ* space. Two sequential projections are required to project the points in this space to the chromaticity plane. The first is a central projection of points (X, Y, Z) from the origin to the Maxwell triangle. The second is a projection of the points of the first projection parallel to the Z axis and onto the XY plane, now called by convention the xy plane. (A stereoscopic view of the tristimulus values of the spectrum in three-space has been published by E. Q. Adams.)[21]

Judd's article, in the context of his reference to Newton's work:[11]

> We find here a statement of Newton's ''law of color mixture.'' Although this law makes no explicit mention of additive distribution curves, and does not directly state that contributions from different sources to a given primary color process are to be combined by addition, still these concepts are contained implicitly in the law. Our indebtedness to Newton in this respect was acknowledged by J. C. Maxwell (Sci. Papers, Cambridge, p. 149; 1890), who was among the first to state explicitly and use these concepts. Probably Grassman [sic] (Pogg. Ann. **89,** pp. 69–84; 1853) first stated them explicitly.

Luminance Coefficients

Judd[10] had formalized the proposal, originally made by Schrödinger, that one of the color-matching functions could, by careful selection of the location of the other two primaries, be made to have a luminance coefficient of unity with the other two primaries having luminance coefficients of zero. This would in theory make the color-matching function associated with that primary an exact copy of the 1924 $V(\lambda)$ curve. Thus, what was currently being done as an ancillary calculation of the luminance of the light under consideration would be built into the colorimetric calculations, again diminishing the burden of calculation. This could be accomplished, Judd realized, by placing the other two primaries on the so-called alychne, meaning the lightless line, whose points have zero luminance.

There was an overwhelming consideration apart from the convenience of calculation. If what was at that time considered the standard brightness function were not incorporated into the standard color functions, then two colors might match in all respects [that is, be a complete color match in, say, dominant wavelength, luminance, and purity (to use the concepts of the day] but differ in standard brightness. Such an apparent contradiction could not be tolerated in 1931; only in later years was a fuller understanding reached regarding the distinction between luminance and brightness.

All Positive Values of the Spectral Tristimulus Values

It may have been harder to agree on the formulating principle that all coordinates of real stimuli have positive values. By real stimuli are meant any stimuli the coordinates of which lie within the convex region of the chromaticity diagram bounded by the spectrum locus and the line connecting its extremities. It was, after all, those stimuli for which one would be cross-multiplying and summing seemingly endless columns of numbers. Any simplification would decrease human error. Priest felt particularly strongly about this item,

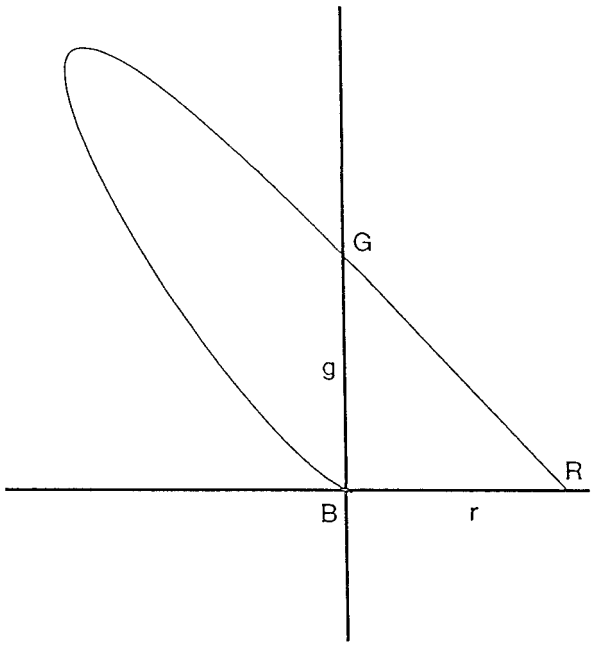

FIG. 2. The CIE 1931 *r*, *g* chromaticity diagram. The Wright-Guild amalgamated data are expressed on *r*- and *g*-axes with the primaries $\mathcal{R}$, $\mathcal{G}$, and $\mathcal{B}$ located at the apices of a right triangle (*RGB*). The triangle is located in the unit-positive first quadrant, but some of the spectrum locus extends outside the triangle bounded by the primaries, because some color-matching data are negative.

as did most officials in the U.S. Priest reported so to Guild by letter in July,[1] just two months before they were to meet. This ran counter to a British interest that the standard observer be described by Guild's selected primaries and by the mean Wright–Guild experimental data. The eventual solution must, therefore, have been a great compromise: the Wright–Guild data as represented by the Guild primaries would define the standard observer in accord with Resolution (1), and the system would be transformed to all positive color-matching functions in accord with Resolution (5). Both parties would realize their objectives.

It had been known for a long time that the conditions of transformation that would lead to a system of all-positive values was one in which the lines joining the new primaries in the chromaticity diagram of the old system passed completely outside the spectrum locus and purple boundary.[13] Looking at Fig. 2, one can see those portions of the spectrum locus that pass outside the lines joining the primaries *R*, *G*, and *B*. Those spectral tristimulus values have at least one negative value, and, indeed, it may be seen from Fig. 2 why no system of real primaries can fail to have some negative values.

So the formulating principles that all values should be positive and minimum space should be wasted could be met by causing the sides of the triangle join-

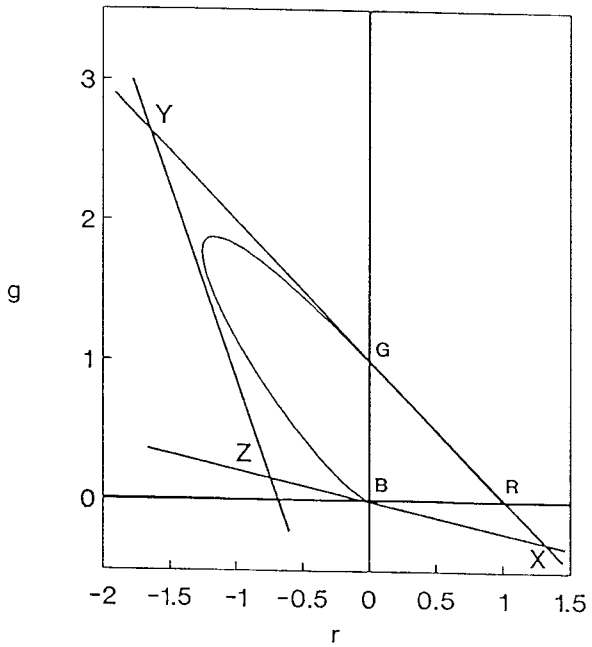

FIG. 3. The CIE 1931 *r*, *g* chromaticity diagram. Boundaries of a new triangle, which will form the legs of the transformed *XYZ* unit-positive, first-quadrant triangle, are identified with the locations of the $\mathcal{XYZ}$ primaries. The *r*, *g* chromaticity values of these $\mathcal{XYZ}$ primaries will be utilized to calculate the transformation coefficients from the old *RGB* system to the new *XYZ* system.

ing the three primaries of the *XYZ* system to pass just slightly outside, but not too far from, the spectrum locus in the *RGB* chromaticity diagram of Fig. 2. Such a triangle is shown in Fig. 3, and we must now define the equations of the triangle and the coordinates of its vertices.

Additional Zeroes on the Long-Wave End of the $\bar{z}$-Function

Choice of the red primary at 700 nm in the Guild primary system provided the useful property that all the values of the blue-function at the long-wave end beyond 660 nm were zero. This was useful, because, while calculating tristimulus values, one need not calculate the contribution of this function beyond 660 nm, saving considerable work in the days of hand calculation. When it was noticed that the spectrum locus of the *RGB* system was, to the precision necessary, a straight line above about 650 nm, plans were made to choose the transformation such that the line $x + y = 1$ coincided with this straight-line part of the spectrum locus. This would, in turn, assure the maximum number of zeroes in the transformed $\bar{z}$-function.

Units of the Primaries to Be Chosen on the Basis of the Equal-Energy Spectrum

In the British system for colorimetry then in use, the NPL primaries were normalized to the NPL white, which was most similar to the new CIE Standard Illuminant B in color temperature and spectral power distribution.[9] The British were accustomed to specifying colorimetry by four so-called cardinal stimuli, three primaries and a white, for normalization of the resulting units in what was called the unit color equation. Thus, Resolution (5) gave four transformation equations: three for transforming the values of *X*, *Y*, and *Z* to Guild's primaries and the fourth to Illuminant B, as was British practice. Apparently unnoticed was that the fourth was redundant. It was there for information only, because the transformation was completely specified by the choice of units in the first three equations; those units were chosen so that the chromaticity coordinates of the equal-energy spectrum were equal to each other, and hence equal to 0.3333.

Priorities

The priorities that emerged from the agreements held dear that the luminance coefficients of the primaries should reproduce the 1924 $V(\lambda)$ curve in the $\bar{y}$-function. That all coordinates of the color-matching functions be positive was a fundamental principle, apparently won at the negotiating table in exchange for specifying the standard observer in Wright–Guild data with Guild's primaries. That the units of the primaries were to be normalized to place the equal-energy spectrum in the center of the chromaticity diagram was noncontroversial as the Wright–Guild data were already so normalized. The balance of the formulating principles (zeroes on the end of one function, minimum wasted area) would be convenient if they could be incorporated. Judd had dealt abstractly with the transformations;[10] now it was time to calculate the transformation coefficients for the agreed real set of data.

STATEMENT OF THE PROBLEM

The 1931 founders arrived at their system of colorimetry through a logical progression. To effect this progression, they used a formalism whose primitives were lights, their superpositions, and their color matches. This traditional mode of thinking had in its favor concreteness, but, as we will show, it involved a number of implicit notational complexities, such as primary lights, superposition operators for the lights, and ''qualified equals signs'' that represent colorimetric (but not spectral) matches between the lights. Furthermore, some of the notation was left ambiguous, a practice that made it less accessible to more recent generations. In the present section, we lay out the 1931 problem in the traditional notation (enhanced for modern readers). For the benefit of readers who are comfortable

with matrix algebra, we also use this more common language in Appendix A to present the same information. Comparison of the two treatments has been very illuminating for the authors, and may also help readers with various backgrounds.

The very first formulating principle we mentioned above was that Grassmann's Laws were implicitly assumed to be valid. Colorists have traditionally used these laws to make algebraic representations of experimental color-matching relationships. Consider the colorimetric equation

$$Q \triangleq R\mathcal{R} \hat{+} G\mathcal{G} \hat{+} B\mathcal{B}. \quad (1)$$

This equation states that some light Q, whose spectral power distribution is unknown and unneeded, but whose color is to be assessed, is matched by amount R of a primary light represented here by the symbol $\mathcal{R}$, added to amount G of the primary represented by $\mathcal{G}$, added to amount B of the $\mathcal{B}$ primary. Each of the primaries represents some light, producible or not. The quantities R, G, and B are the tristimulus values of the match, and are scalar multipliers on the primaries. The primaries carry the dimensions. In the shorthand of Eq. (1), light Q is said to have tristimulus values R, G, and B. More important to the understanding of a colorimetric equation, but less often discussed, are the meanings of the connecting symbols, ($\triangleq$) and ($\hat{+}$). The symbol ($\triangleq$) means that light Q is matched colorimetrically, that is, in chromaticity and brightness, by the mixed light on the right-hand side of Eq. (1). The symbol ($\hat{+}$) means that the individual primary lights are ''additively mixed,'' i.e., superposed.

So we have a light Q, perhaps on one side of a bipartite field of a visual colorimeter, matched by certain quantities of three primaries on the other side of the bipartite field. The two lights, which may be spectrally different, are a complete match for each other, meaning that they match in all perceptual attributes.

Notice that the certain quantities of the primary lights of which we speak are the products $R\mathcal{R}$, $G\mathcal{G}$, and $B\mathcal{B}$, not the tristimulus values alone as we are sometimes told. Further, while the sign ($\hat{+}$) implies superposition of the component lights, the implied multiplication of R times $\mathcal{R}$ is algebraic. It will become important later in this derivation to remember which operations are algebraic and which operations refer to physical superposition.

Suppose color S has tristimulus values R_S, G_S, B_S. Its colorimetric equation would be represented by

$$S \triangleq R_S\mathcal{R} \hat{+} G_S\mathcal{G} \hat{+} B_S\mathcal{B}. \quad (2)$$

Now suppose that color S is physically added to both sides of colorimetric Eq. (1) resulting in

$$Q \hat{+} S \triangleq (R + R_S)\mathcal{R} \hat{+} (G + G_S)\mathcal{G} \hat{+} (B + B_S)\mathcal{B}. \quad (3)$$

Now Grassmann's Law of Additivity has been assumed to write the plus signs connecting the tristimulus values of the two additively mixed colors as algebraic plus-signs, meaning algebraic addition as opposed to superposition, as implied by the sign ($\hat{+}$). One other subtle change has taken place. Although Q and S are both lights, we now refer to them as colors without contradiction. This is because Grassmann's Laws tell us that a *color*, regardless of its spectral composition, added to both sides of a colorimetric equation, preserves the match. Although we have physically added the light S to light Q in the left hand side of Eq. (3), that equation states that the colors of both sides of the equation still match, that the sign ($\triangleq$) is still warranted, and that an algebraic plus sign is warranted between *like* quantities such as the R tristimulus values, or the G tristimulus values. Unlike quantities, such as quantities of different primaries, still require a superposition sign, such as ($\hat{+}$).

Equation (3) is the basis under which we perform what is called colorimetric integration. The tristimulus values of each of the narrow (e.g., 1-nm wide) bands in the visible spectrum are algebraically added to each other in proportion to their relative spectral power in any spectrum to assess a tristimulus value of the entire spectrum.

Given the *RGB* system, the CIE wanted to define a new system *XYZ*, characterized by primary lights $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ (producible or not) that satisfied the formulating principles. The *RGB* and *XYZ* systems are formally connected by three colorimetric equations describing matches of the new primaries with mixtures of the old primaries:

$$\begin{aligned} \mathcal{X} &\triangleq a_{11}\mathcal{R} \hat{+} a_{21}\mathcal{G} \hat{+} a_{31}\mathcal{B} \\ \mathcal{Y} &\triangleq a_{12}\mathcal{R} \hat{+} a_{22}\mathcal{G} \hat{+} a_{32}\mathcal{B} \\ \mathcal{Z} &\triangleq a_{13}\mathcal{R} \hat{+} a_{23}\mathcal{G} \hat{+} a_{33}\mathcal{B}. \end{aligned} \quad (4)$$

Here the tristimulus values of the *XYZ* primaries in the *RGB* system are denoted as a_{ij}'s as shorthand, but they are the familiar tristimulus values of Eq. (1).

We know that in the *XYZ* system any color Q is matched in the notation of the colorimetric equation by

$$Q \triangleq X\mathcal{X} \hat{+} Y\mathcal{Y} \hat{+} Z\mathcal{Z}. \quad (5)$$

Substituting Eq. (4) into Eq. (5) with slight rearrangement for convenience of expression, we obtain

$$\begin{aligned} Q \triangleq (a_{11}\mathcal{R} + a_{21}\mathcal{G} + a_{31}\mathcal{B})X \hat{+} (a_{12}\mathcal{R} + a_{22}\mathcal{G} + a_{32}\mathcal{B})Y \\ \hat{+} (a_{13}\mathcal{R} + a_{23}\mathcal{G} + a_{33}\mathcal{B})Z. \end{aligned} \quad (6)$$

By removing parentheses and collecting terms in a different order, we obtain

$$\begin{aligned} Q \triangleq (a_{11}X + a_{12}Y + a_{13}Z)\mathcal{R} \hat{+} (a_{21}X + a_{22}Y + a_{23}Z)\mathcal{G} \\ \hat{+} (a_{31}X + a_{32}Y + a_{33}Z)\mathcal{B}. \end{aligned} \quad (7)$$

Now the parenthetical expressions are the tristimulus values of Eq. (1), and it follows that

$$\begin{aligned} R &= a_{11}X + a_{12}Y + a_{13}Z \\ G &= a_{21}X + a_{22}Y + a_{23}Z \\ B &= a_{31}X + a_{32}Y + a_{33}Z, \end{aligned} \quad (8)$$

and all superposition mixing signs are removed from the expression in favor of purely algebraic operators. Thus, the transformation is carried out among the tristimulus values of the spectra in the two sets of primaries, not among the primaries themselves. Equation (8) is the equation in its most general form applying to the tristimulus values of spectral and nonspectral lights, and even to points outside the gamut of the three-space bounded by the origin, the spectrum locus, and the purple boundary. These latter colors we call imaginary.

While Eq. (8) states *RGB* in terms of *XYZ*, the inverse of that equation states *XYZ* in terms of *RGB* as follows:

$$X = b_{11}R + b_{12}G + b_{13}B$$
$$Y = b_{21}R + b_{22}G + b_{23}B \qquad (9)$$
$$Z = b_{31}R + b_{32}G + b_{33}B.$$

By solving Eq. (8) for R, G, and B, one retrieves the values of b_{ij} in Eq. (9) from the known values of a_{ij} in Eq. (8).

If we can find three points in *RGB* space for which we know the tristimulus values in both *RGB* and *XYZ* space, we can substitute these values into Eq. (8). That will leave Eq. (8) with the only algebraic unknowns being the a_{ij}. That is, Eq. (8) will now define a_{ij} in terms of numeric values. We have only to invert the a_{ij} matrix to obtain the b_{ij} matrix, which contains the unknowns sought in the generalized problem.

The above is sufficient to solve the problem, but there is one thing more we might want to do. Generally large values of a_{ij} lead to small values of b_{ij} in the same sense that the reciprocals of large numbers are small numbers. In the end we might want to redefine the primary units so as to rescale b_{ij} to place them in the same numeric range as a_{ij}, or for some other matter of convenience such as we shall see later in the article.

Figure 1(a) depicts the *XYZ* three-space. In Fig. 1(b), the grey region in the upper portion of the figure is the projection from the origin of the real tristimulus values of this space to the Maxwell triangle whose vertices are the unit vectors of the space. This Maxwell triangle then is the plane $X + Y + Z = 1$ in *XYZ* space, but note that this plane is not yet a standard CIE chromaticity diagram. To make it such a chromaticity diagram, one more projection is necessary. That projection is parallel to the Z axis and projects the points in the upper grey region to points in the *XY* plane, which we now relabel *xy* by the convention that lowercase symbols represent chromaticities.

We seek three points in *RGB* space for which we know the tristimulus values of the colors in both *RGB* and *XYZ* space. We seek these in order to substitute these tristimulus values into Eq. (8) to calculate the unknown b_{ij}'s. It is apparent that we are able to work in the chromaticity diagram alone and then make the inverse projection to the Maxwell unit vector triangle. That will put us in tristimulus space, which is what is required to substitute into Eq. (8). We can derive a result in chromaticity space and represent it on the plane $X + Y + Z = 1$ in tristimulus space.

Algebraically the representation is

$$X = x$$
$$Y = y \qquad (10)$$
$$Z = 1 - x - y,$$

where *XYZ* are tristimulus values and x and y are chromaticity coordinates. A correlate to the above diagram and algebraic expressions holds for *RGB* space.

We will work in the rg chromaticity diagram, and use the formulating principles to identify three points with known chromaticities, and, hence, tristimulus values known in the sense described above. The three known points are the $\mathcal{XYZ}$ primaries. To specify these points we apply the formulating principles in the chromaticity diagram.

IMPLEMENTING THE SOLUTION

Incorporation of the $V(\lambda)$ Curve into the Transformation Coefficients

The *RGB* primaries of Resolution (1) were assigned the luminance coefficients of 1:4.5907:0.0601.[9] These coefficients were the least-square fit of the linear combination of $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ to $V(\lambda)$. Any light having chromaticity coordinates r, g, b would, therefore, have luminance proportional to

$$0.17697r + 0.81240g + 0.01063b, \qquad (11)$$

where the numerical coefficients are the luminance coefficients of the primaries normalized to unity. That expression set equal to zero defines the locus of points having luminance of zero, or the alychne, as follows:

$$0.17697r + 0.81240g + 0.01063b = 0. \qquad (12)$$

But b is dependent in that $b = 1 - r - g$, which may be substituted into Eq. (12). Thus, in the rg chromaticity diagram of Fig. 1, the straight line defined by

$$0.16634r + 0.80177g + 0.01063 = 0 \qquad (13)$$

defines the alychne and, choosing $\mathcal{Y}$ as the primary to carry the luminance function, $\mathcal{X}$ and $\mathcal{Z}$ must have their location on that line. That conveys to $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ luminance coefficients of 0:1:0, respectively, which satisfies Judd's 1930 condition of formulating principle (b) enunciated in Ref. 10.

The coefficients of Eq. (11) are calculated by least-square fit of the *RGB* data to the $V(\lambda)$ curve by a technique also originated by Judd for colorimetric purposes.[12] In this article, written while still a Master's degree candidate in the Department of Physics at Ohio State, and published while he was a Ph. D. candidate at Cornell, Judd outlines the method of calculation of the least-square

fit in a way that was adaptable to the mechanical calculators of the day. It was Judd's second published article on colorimetry and his fourth overall.

We re-calculated Judd's coefficients by matrix inversion using a digital computer with the identical results to the five digits of significance given in 1931.

The first line, XZ, which will connect the chromaticities of the X and Z primaries, is in place in Fig. 3.

All Coordinates of the System to Be Positive

As has been mentioned, the requirement of positive coordinates for all producible lights implies that the spectrum locus lies within the triangle XYZ (connecting the primaries $\mathcal{XYZ}$) in the rg chromaticity diagram. It might seem that the XY line in Fig. 3 could be chosen to coincide with the RG line in Fig. 2. However, this cannot quite be done, because the color-matching function $\bar{b}(\lambda)$ is very slightly negative[14] for wavelengths between the $\mathcal{G}$ primary and 670 nm, hence the spectrum locus must cross the RG line. The CIE indeed chose line XY to pass through $\mathcal{R}$. However, we have not been able to find a detailed discussion of how the slope was selected, so we advance the following scenario as plausible and consistent with the 1931 CIE system.

For long wavelengths, the magnitude of the b chromaticity coordinate is always less than 0.01 times that of the g coordinate, so a slope of $-100/99$ would be enough to encompass the spectrum locus at the required level of precision. Accordingly, the line joining $\mathcal{X}$ and $\mathcal{Y}$ primaries should be

$$99g = -100(1 - r). \tag{14}$$

It turns out that this line escapes all of the spectrum locus except near 670 nm, where the single nonzero digit of $\bar{b}(\lambda)$ incurs roundoff error anyway. Furthermore, the line is so nearly tangent to the spectrum locus that $\bar{z}(\lambda) = 0$ over a considerable wavelength range.

With the foregoing definition of the line XY, two lines of the three joining the primaries are now in place in Fig. 3, and the intersection of the two, found by solving Eqs. (13) and (14) simultaneously, is

$$r_X = 1.2749, \quad g_X = -0.2777,$$
$$\text{hence} \quad b_X = 0.0028. \tag{15}$$

These are the r_X, g_X, b_X that were as yet unknown to us as we passed through Eq. (8), and were there denoted as a_{ij}, but they are now found. We will continue in this way until we have found the remaining six missing parameters of Eq. (8).

Final Leg of the Encompassing Triangle

Inasmuch as all the important formulating principles are already included, we have a lot of leeway in selecting the location of the final leg YZ of the encompassing triangle. It must pass outside the spectrum locus, which has its r negative extreme near 505 nm. Other than that, we may choose the line to give the most favorable disposition to the spectrum locus in the XYZ system. Choosing an intersection with the alychne at $r = -0.743$, and an intersection with the line connecting X and Y at $r = -1.74$, gives intersections with these lines at

$$r_Y = -1.7400, \quad g_Y = 2.7677,$$
$$\text{hence} \quad b_Y = -0.0277 \tag{16}$$

and

$$r_Z = -0.7430, \quad g_Z = 0.1408,$$
$$\text{hence} \quad b_Z = 1.6022, \tag{17}$$

respectively. This third leg defined by

$$2.6268r + 0.9970g = -1.8113 \tag{18}$$

passes near 505 nm at $r = -1.3422$ when $g = 1.7195$, as it does at 505 nm from the Wright–Guild data. That gives plenty of clearance for the containing line to pass the spectrum locus, the value of r there being -1.3182.

Calculating the Transformation Matrix

Equations (15)–(17) reveal the R, G, B tristimulus values of each of the $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ primaries, subject to as yet unspecified scale factors α, β, γ. In particular, we may write

$$\begin{aligned}
\alpha\mathcal{X} &\triangleq 1.2749\mathcal{R} \mathbin{\dot{-}} 0.2777\mathcal{G} \mathbin{\hat{+}} 0.0028\mathcal{B} \\
\beta\mathcal{Y} &\triangleq \mathbin{\dot{-}}1.7400\mathcal{R} \mathbin{\hat{+}} 2.7677\mathcal{G} \mathbin{\dot{-}} 0.0277\mathcal{B} \\
\gamma\mathcal{Z} &\triangleq \mathbin{\dot{-}}0.7430\mathcal{R} \mathbin{\hat{+}} 0.1408\mathcal{G} \mathbin{\hat{+}} 1.6022\mathcal{B}.
\end{aligned} \tag{19}$$

Equation (19) has the same form as Eq. (4), with $a_{11} = 1.2749/\alpha$, etc. By performing a transpose-inverse on the matrix a_{ij} [exactly the same transformation as that carrying the coefficients of Eq. (4) to Eq. (9)], we can write an equation with the form of Eq. (9):

$$\begin{aligned}
X &= 0.9088\alpha R + 0.5751\alpha G + 0.3709\alpha B \\
Y &= 0.0912\beta R + 0.4187\beta G + 0.0055\beta B \\
Z &= 0.0000\gamma R + 0.0062\gamma G + 0.6236\gamma B.
\end{aligned} \tag{20}$$

Here $b_{11} = 0.9088\alpha$, etc.

The CIE in 1931 chose α, β, γ so that the point $(R, G, B) = (1, 1, 1)$ maps to the point $(X, Y, Z) = (1, 1, 1)$. Hence $1/\alpha = 0.9088 + 0.5751 + 0.3709$, etc. As a result of this scaling choice, the coefficient array b_{ij} in Eq. (9) becomes

$$\begin{matrix}
0.4900 & 0.3100 & 0.2000 \\
0.1769 & 0.8124 & 0.0107 \\
0.0000 & 0.0099 & 0.9901.
\end{matrix}$$

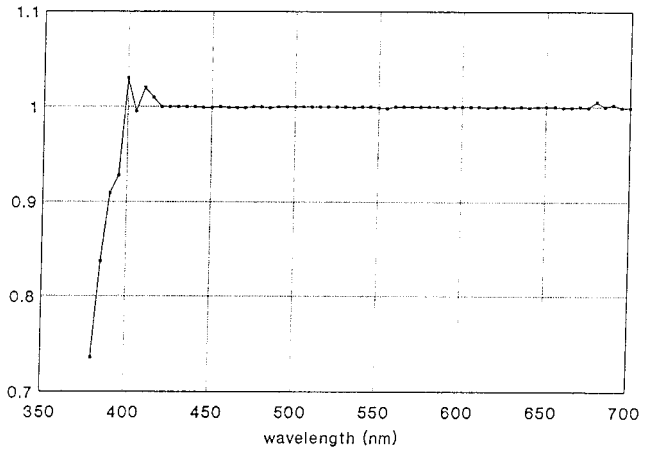

FIG. 4. The ratio $V(\lambda)/\bar{y}_o(\lambda)$ plotted against wavelength. As a test of goodness-of-fit of the rescaled $\bar{y}$-function to the experimental data, the rescaling ratio is plotted at 5-nm intervals. The ratio is very nearly unity except at wavelengths where the absolute value of $V(\lambda)$ is very small.

The luminance coefficients of the second row are known to a higher precision from the least-square fit of *RGB* to $V(\lambda)$ as in Eq. (10). The other transformation equations are likely known to a lower precision than these four digits, so rounding of the first and last row to two digits of precision is justified. The resulting transformation matrix is

$$\begin{matrix} 0.49 & 0.31 & 0.20 \\ 0.17697 & 0.81240 & 0.01063 \\ 0.00 & 0.01 & 0.99, \end{matrix}$$

which was the transformation adopted.[2] The two-digit coefficients were soon appearing in the literature to five digits of precision with three trailing zeroes.[9,14]

Final Adjustment of the Color-Matching Functions Using $V(\lambda)$

One minor problem remained. If these transformation coefficients were applied directly to the tristimulus values of the Wright–Guild data, the $\bar{y}$-function that would emerge would not be exactly $V(\lambda)$, but only a least-square approximation to it. Recall that no linear combination of the R, G, B color-matching functions is exactly equal to $V(\lambda)$. But the CIE deemed it singularly important that $V(\lambda)$ be copied into the $\bar{y}$-function to make the definitions of what the founders considered brightness congruent in photometry and colorimetry. This was done by scaling the resulting color-matching functions [which we call here $\bar{x}_o(\lambda)$, $\bar{y}_o(\lambda)$, $\bar{z}_o(\lambda)$] by the ratio between the photometric luminosity function $V(\lambda)$ and its color-matching function approximation $\bar{y}_o(\lambda)$ derived from the transformation coefficients. The result is a set of modified color-matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$.

$$\begin{aligned} \bar{x}(\lambda) &= \bar{x}_o(\lambda)V(\lambda)/\bar{y}_o(\lambda) \\ \bar{y}(\lambda) &= V(\lambda) \\ \bar{z}(\lambda) &= \bar{z}_o(\lambda)V(\lambda)/\bar{y}_o(\lambda). \end{aligned} \tag{21}$$

The closeness of $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$ to $\bar{x}_o(\lambda)$, $\bar{y}_o(\lambda)$, $\bar{z}_o(\lambda)$ may be assessed by examining the scaling factor $V(\lambda)/\bar{y}_o(\lambda)$ as a function of wavelength. This is done in Figs. 4 and 5, on two different bases, and shows that there is negligible effect incurred by the rescaling.

Irrespective of the changes incurred on the color-matching functions by Eq. (21), the numerical values of the chromaticities of monochromatic lights remain identical. This can be seen by noting that

$$\bar{x}_o(\lambda)/[(\bar{x}_o(\lambda) + \bar{y}_o(\lambda) + \bar{z}_o(\lambda)] = \bar{x}(\lambda)/[\bar{x}(\lambda) + \bar{y}(\lambda) + \bar{z}(\lambda)] \tag{22}$$

$$\bar{y}_o(\lambda)/[(\bar{x}_o(\lambda) + \bar{y}_o(\lambda) + \bar{z}_o(\lambda)] = \bar{y}(\lambda)/[\bar{x}(\lambda) + \bar{y}(\lambda) + \bar{z}(\lambda)]. \tag{23}$$

Hence the spectrum-locus curve is the same in the scaled and unscaled systems. The invariance of the spectrum-locus curve has also been seen as an encouraging sign in other manipulations of color-matching data, particularly in Wright's normalization of color-matching data to remove effects of lens-yellowing differences among subjects.[15] Readers are cautioned, however, not to confuse invariance of the spectrum-locus curve with invariance of the color-matching functions, and particularly with invariance of chromaticity of lights that are not monochromatic. Clearly, scaling all the color-matching functions by a function of wavelength in general affects color-match

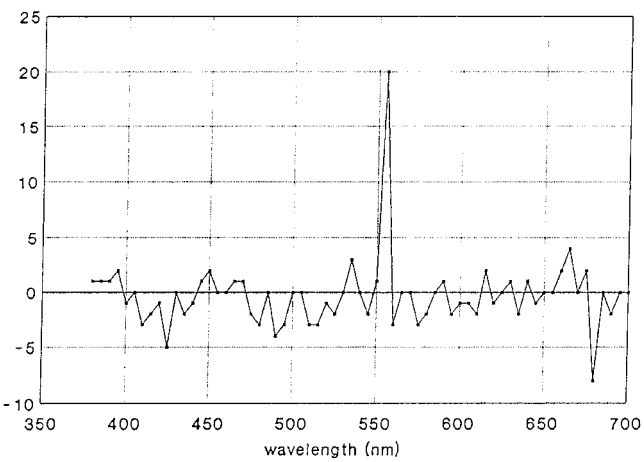

FIG. 5. The absolute difference between $V(\lambda)$ and $\bar{y}_o(\lambda)$ in units of one hundred-thousandths when $V(\lambda)$ is scaled to unity at 555 nm. As an additional test of goodness-of-fit of the rescaled $\bar{y}$-function to the experimental data, the difference is plotted at 5-nm intervals. The worst case, at 555 nm, gives a value of $\bar{y}$ that differs from unity by no more than 0.0002.

predictions, even though it does not much affect them in the case of Eqs. (21).

Since the above explanation of the rescaling follows logically, it is remarkable that none of the relevant literature offers this explanation.[2,9,14,16] Rather, in the explanation provided, both the numerator and denominator of Eqs. (22)–(23) are divided by the sum of $\bar{x}_o(\lambda)$, $\bar{y}_o(\lambda)$, $\bar{z}_o(\lambda)$, converting the terms in both numerator and denominator into chromaticity coordinates. We say it is remarkable, because rescaling using chromaticity coordinates is more difficult although mathematically equivalent.

The explanation given is as follows. Equation (9) has a counterpart in the following transformation of the spectral chromaticity coordinates:

$$
\begin{aligned}
x_o(\lambda) &= \frac{b_{11}r(\lambda) + b_{12}g(\lambda) + b_{13}b(\lambda)}{(b_{11} + b_{21} + b_{31})r(\lambda) + (b_{12} + b_{22} + b_{32})g(\lambda) + (b_{13} + b_{23} + b_{33})b(\lambda)} \\
y_o(\lambda) &= \frac{b_{21}r(\lambda) + b_{22}g(\lambda) + b_{23}b(\lambda)}{(b_{11} + b_{21} + b_{31})r(\lambda) + (b_{12} + b_{22} + b_{32})g(\lambda) + (b_{13} + b_{23} + b_{33})b(\lambda)} \\
z_o(\lambda) &= \frac{b_{31}r(\lambda) + b_{32}g(\lambda) + b_{33}b(\lambda)}{(b_{11} + b_{21} + b_{31})r(\lambda) + (b_{12} + b_{22} + b_{32})g(\lambda) + (b_{13} + b_{23} + b_{33})b(\lambda)}.
\end{aligned}
\tag{24}
$$

Thus, the Wright–Guild data may be transformed from *RGB* chromaticities to *XYZ* chromaticities by Eq. (24). Then the color-matching functions may be derived from the following equations:

$$\bar{x}(\lambda) = x_o(\lambda)V(\lambda)/y_o(\lambda) \tag{25}$$

$$\bar{y}(\lambda) = V(\lambda) \tag{26}$$

$$\bar{z}(\lambda) = z_o(\lambda)V(\lambda)/y_o(\lambda). \tag{27}$$

We may readily see that the rescaling is now in terms of chromaticity coordinates rather than the approximate color-matching functions themselves.

EVALUATION OF THE FORMULATING PRINCIPLES IN TERMS OF MODERN KNOWLEDGE

A final topic is the question of whether the same formulating principles might be incorporated if the system were being formulated for the first time today, with present tools, resources, and knowledge. The answer to this hypothetical question is almost certainly no in every case.

Although luminance was originally defined in terms of flicker photometry (not brightness per se), luminance was said in 1931 to be the correlate of brightness. Judd in 1933 uses the terms luminance and brightness interchangeably,[16] as did MacAdam in 1944.[17] Realizations soon came that brightness was not necessarily additive, a fact that had been discovered by Helmholtz and studied by Kohlrausch, hence the name Helmholtz–Kohlrausch effect for this phenomenon. By 1950, MacAdam had a different posture on this matter, stating, ''It is easy to show that the law of additivity, which necessarily applies to luminance because of its definition, does not apply to brightness.''[6] Judd eight years later described the classic experiment in which a saturated red light is matched for brightness to a saturated green light, and then the two lights are superposed.[18] The resulting yellow light will be judged to be half to three-quarters as bright as either of the saturated lights from which the compound light is composed. If the luminances of the two lights, which are equally bright, are computed, they will be found to be quite different from each other, in spite of their having been matched to each other for the brightness attribute. In any case, if luminance is additive, the resulting luminance of the additive yellow must be equal to the sum of the luminances of the component lights, and thus luminance cannot be a correlate of brightness. It is, therefore, likely that we would, in this day, not require that the standard luminance curve be incorporated into our functions, except for the treatment of spatio-temporal resolution of visual tasks.

The requirement that all values be positive was a requirement stemming from the methods of calculation of the day, and does not offer an important computational advantage to modern digital computers. The same may be said for the introduction of as many zeroes into the functions as was practical.

One reason for normalizing the three functions to place the equal-energy spectrum at the center of the chromaticity diagram was to render the neutral point approximately equi-distant from the spectrum locus at many wavelengths. This in turn would introduce no false differences in the distances from the neutral among colors of approximately equal saturation. Today we know the chromaticity diagram to be so far from a uniform color scale that we would not wish to introduce such considerations, lest they be misinterpreted as representing a uniform color scale.

CONCLUSIONS

We have shown in detail the concepts and methods that led to the CIE 1931 system for colorimetry and particularly the derivation of its transformation from Wright–Guild *RGB* data. We have undertaken to show how the formulating principles of 1931 played an overwhelming role in determining the values of the standard data for colorimetry, which have pervaded the science of colorim-

etry for the past 65 years. We have shown that likely none of these formulating principles would be adopted if the system were formulated from a fresh start today.

APPENDIX A: FORMAL STATEMENT OF THE PROBLEM

The Wright–Guild data gave rise to a set of color-matching functions $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, $\bar{b}(\lambda)$ (λ being visible wavelength) such that the color-matching behavior of any spectral power distribution $S(\lambda)$ can be represented by three coordinates:

$$R(S) = \int \bar{r}(\lambda)S(\lambda)d\lambda$$

$$G(S) = \int \bar{g}(\lambda)S(\lambda)d\lambda \qquad \text{(A1)}$$

$$B(S) = \int \bar{b}(\lambda)S(\lambda)d\lambda.$$

If two lights S_1 and S_2 produce the same values R, G, and B, then they match in color.

The functions $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, $\bar{b}(\lambda)$ emerged from a color-matching experiment with near-monochromatic primary lights at $\lambda_r = 700$ nm, $\lambda_g = 546.1$ nm, and $\lambda_b = 435.8$ nm. Because each primary matches itself within experimental errror, the color-matching experiment tends to yield $\bar{r}(\lambda_r) = \bar{g}(\lambda_g) = \bar{b}(\lambda_b) = 1$, but $\bar{r}$, $\bar{g}$, $\bar{b}$ were subsequently normalized to have the same integrated areas:

$$\int \bar{r}(\lambda)d\lambda = \int \bar{g}(\lambda)d\lambda = \int \bar{b}(\lambda)d\lambda = \int V(\lambda)d\lambda, \qquad \text{(A2)}$$

where $V(\lambda)$ is the CIE 1924 spectral luminous efficiency function normalized so that its peak value is one.

The first two equalities in Eq. (A2) insured that an equal-energy stimulus, $S_o(\lambda)$ = constant, has equal coordinates (R, G, B). The third equality was a convention whose convenience will soon be clear. The scaling convention in Eq. (A2) does not affect the predictions of Eq. (A1) as to whether lights S_1 and S_2 will match.

Given this starting point, the goal of the CIE in 1931 was to arrive at a new representation of color-matching space that would have attractive computational properties and be independent of any choice of experimental primary lights. Also, the new system would have to be brought into conformance with the 1924 CIE luminance function $V(\lambda)$.

Apart from the incorporation of $V(\lambda)$, and assuming the empirical truth of Grassmann additivity inherent in transformation of primaries as laid down by Ives as early as 1915,[20] the formal problem facing the CIE was to define a linear transformation from (R, G, B) to a new system (X, Y, Z):

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathbf{T}\begin{bmatrix} R \\ G \\ B \end{bmatrix}, \quad \begin{bmatrix} \bar{x}(\lambda) \\ \bar{y}(\lambda) \\ \bar{z}(\lambda) \end{bmatrix} = \mathbf{T}\begin{bmatrix} \bar{r}(\lambda) \\ \bar{g}(\lambda) \\ \bar{b}(\lambda) \end{bmatrix}, \qquad \text{(A3)}$$

where $\mathbf{T}$ is a 3×3 matrix:

$$\mathbf{T} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}. \qquad \text{(A4)}$$

The new functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$ would still embody all the color-matching properties of the Wright–Guild data; if $S_1(\lambda)$ matches $S_2(\lambda)$ by integration with the Wright–Guild $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, $\bar{b}(\lambda)$, then the match, as represented by tristimulus identity, will also occur by $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$ integration.

Finding the matrix $\mathbf{T}$ can be done by finding the coefficients a_{ij} such that

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathbf{T}^{-1}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}. \qquad \text{(A5)}$$

In *RGB*-space, the unit coordinate vectors X, Y, Z are the vectors

$$\begin{bmatrix} R_X \\ G_X \\ B_X \end{bmatrix} = \mathbf{T}^{-1}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix}$$

$$\begin{bmatrix} R_Y \\ G_Y \\ B_Y \end{bmatrix} = \mathbf{T}^{-1}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} \qquad \text{(A6)}$$

$$\begin{bmatrix} R_Z \\ G_Z \\ B_Z \end{bmatrix} = \mathbf{T}^{-1}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \end{bmatrix}.$$

Once one has defined those vectors, one has specified the transformation from *RGB* to *XYZ*.

Three constraints on the a_{ij} emerge from choosing the *XYZ* system to inherit the area-normalization property of Eq. (A2) from the *RGB* system:

$$\int \bar{x}(\lambda)d\lambda = \int \bar{y}(\lambda)d\lambda = \int \bar{z}(\lambda)d\lambda = \int V(\lambda)d\lambda. \qquad \text{(A7)}$$

By inserting the color-matching functions in place of their corresponding tristimulus values in Eq. (A5), integrating both sides over λ, and dividing by a common factor of $\int V(\lambda)d\lambda$, one retrieves the following three linear equations in a_{ij}:

$$a_{11} + a_{12} + a_{13} = 1$$
$$a_{21} + a_{22} + a_{23} = 1 \qquad \text{(A8)}$$
$$a_{31} + a_{32} + a_{33} = 1.$$

To complete the solution for the nine values a_{ij}, it will suffice to find six additional linear equations. This can be done by projecting *RGB* space into what is called the *chromaticity plane* and identifying the X, Y, Z coordinates unit vectors as cardinal points in this plane.

An explanation is needed here. The transformation of tristimulus space (R, G, B) to chromaticity space (r, g) is a combination of *two* projective transformations in sequence as depicted in Fig. 1(b). The first is a central projection of (R, G, B) to the $(1, 1, 1)$ plane with focal point at the origin. This projection is similar to a pin-hole camera perspective projection. The projected points lie on the plane $R + G + B = 1$, and are given by the coordinates

$$r = \frac{R}{R + G + B}, \quad g = \frac{G}{R + G + B},$$
$$b = \frac{B}{R + G + B}. \qquad \text{(A9)}$$

Note that the points in the central projection are still in tristimulus coordinates; that is, there are still three of them. Yet there is a constraint among them: $b = 1 - r - g$. By incorporating that constraint, it is possible to ignore coordinate b, a process that is equivalent to an orthogonal projection on the rg plane parallel to the b-axis. The new two-dimensional representation is known as the chromaticity diagram. Its coordinates are

$$r = \frac{R}{R + G + B}, \quad g = \frac{G}{R + G + B}. \qquad \text{(A10)}$$

Now suppose that we complete the solution for the a_{ij} [begun in Eq. (A8)] using relationships in the chromaticity plane. Further, suppose that we follow the CIE and specify the chromaticities of the unit coordinate vectors of X, Y, Z using Eqs. (A6) and (A8):

$$r_x = \frac{a_{11}}{a_{11} + a_{21} + a_{31}}, \quad g_x = \frac{a_{21}}{a_{11} + a_{21} + a_{31}}$$
$$r_Y = \frac{a_{12}}{a_{12} + a_{22} + a_{32}}, \quad g_Y = \frac{a_{22}}{a_{12} + a_{22} + a_{32}} \qquad \text{(A11)}$$
$$r_Z = \frac{a_{13}}{a_{13} + a_{23} + a_{33}}, \quad g_Z = \frac{a_{23}}{a_{13} + a_{23} + a_{33}}.$$

Specifying the left-hand side of Eq. (A11), which amounts to specifying the vertices of a triangle in rg space, offers the required six equations in a_{ij}. Multiplying Eq. (A11) by its denominators reveals these equations to be linear in a_{ij}.

Equations (A8) and (A11) comprise nine linear equations in the nine unknowns a_{ij}. These were the equations that were solved by the CIE.

It now remains to select the vertices (r_x, g_x), (r_y, g_y), (r_z, g_z) of the chromaticity triangle representing the coordinate unit vectors of X, Y, Z. Since the CIE required all monochromatic lights (and, hence, all lights) to map to the all-positive octant of XYZ space, it follows that the chromaticity triangle should circumscribe the spectrum locus. It would appear that the choice of the vertices is purely a graphical problem. However, there are two constraints on this triangle that do not emerge from examining the triangle's relationship to the spectrum locus. These constraints have to do with the adoption of $V(\lambda)$ as one of the new color-matching functions (the $\bar{y}$-function), and, in fact, emerge from the stipulation that the X and Z coordinate unit vectors fall on the alychne (plane $Y = 0$).

A plane of constant Y is given, using Eqs. (A3)–(A4) by

$$Y = b_{21}R + b_{22}G + b_{23}B. \qquad \text{(A12)}$$

The quantities b_{21}, b_{22}, b_{23} are called luminance coefficients, and are determined by least-square fit of $b_{21}\bar{r}(\lambda) + b_{22}\bar{g}(\lambda) + b_{23}\bar{b}(\lambda)$ to $V(\lambda)$. This means minimizing

$$\int [b_{21}\bar{r}(\lambda) + b_{22}\bar{g}(\lambda) + b_{23}\bar{b}(\lambda) - V(\lambda)]^2 d\lambda, \qquad \text{(A13)}$$

which has the solution

$$\begin{bmatrix} b_{21} \\ b_{22} \\ b_{23} \end{bmatrix} = \begin{bmatrix} \langle \bar{r}^2 \rangle & \langle \bar{r}\bar{g} \rangle & \langle \bar{r}\bar{b} \rangle \\ \langle \bar{r}\bar{g} \rangle & \langle \bar{g}^2 \rangle & \langle \bar{g}\bar{b} \rangle \\ \langle \bar{r}\bar{b} \rangle & \langle \bar{g}\bar{b} \rangle & \langle \bar{b}^2 \rangle \end{bmatrix}^{-1} \begin{bmatrix} \langle \bar{r}V \rangle \\ \langle \bar{g}V \rangle \\ \langle \bar{b}V \rangle \end{bmatrix}, \qquad \text{(A14)}$$

where $\langle\ \rangle = \int d\lambda$.

As can be seen from the main text, the least-square solution has a very small residual, so, with high accuracy,

$$V(\lambda) = b_{21}\bar{r}(\lambda) + b_{22}\bar{g}(\lambda) + b_{23}\bar{b}(\lambda). \qquad \text{(A15)}$$

Then, evaluating Eq. (A15) at wavelengths λ_r, λ_g, λ_b of the *RGB* primaries enables a very simple solution for the luminance coefficients:

$$b_{21} = \frac{V(\lambda_r)}{\bar{r}(\lambda_r)}, \quad b_{22} = \frac{V(\lambda_g)}{\bar{g}(\lambda_g)}, \quad b_{23} = \frac{V(\lambda_b)}{\bar{b}(\lambda_b)}. \qquad \text{(A16)}$$

Given these specified luminance coefficients, the alychne constraints are

$$0 = b_{21}R_X + b_{22}G_X + b_{23}B_X$$
$$0 = b_{21}R_Z + b_{22}G_Z + b_{23}B_Z. \qquad \text{(A17)}$$

Dividing the first of these equations by $(R_X + G_X + B_X)$ and the second by $(R_Z + G_Z + B_Z)$ gives

$$0 = b_{21}r_X + b_{22}g_X + b_{23}(1 - r_X - g_X)$$
$$0 = b_{21}r_Z + b_{22}g_Z + b_{23}(1 - r_Z - g_Z). \qquad \text{(A18)}$$

Equations (A18) constrain (r_x, g_x) and (r_z, g_z) to lie on a specified line in rg space. The rest of the triangle-specification problem then becomes a graphical problem, which is described in the main text (see the second and third subsections of the section ''Implementing the Solution'').

Because Eqs. (A17) are homogeneous, it is customary to refer to the luminance coefficients as relative quantities:

$$L_R{:}L_G{:}L_B = 1 : \frac{b_{22}}{b_{21}} : \frac{b_{23}}{b_{21}} = 1 : 4.5907 : 0.0601. \quad \text{(A19)}$$

ACKNOWLEDGMENTS

The authors wish to thank the reviewers for critical insights, which improved the article, and to acknowledge the contribution of the book *The Measurement of Color*[19] in preparing this manuscript. (All four editions are cited, because they are helpful in understanding the evolving understanding of the 1931 CIE System.) Without this material it would have proven impossible to reconstruct the thoughts of 1931. Our thanks, therefore, go to the lone surviving principal player: W. David Wright.

1. W. David Wright, The historical and experimental background to the 1931 CIE system of colorimetry, in *Golden Jubilee of Colour in the CIE*, Proceedings of a symposium held at Imperial College, London, Society of Dyers and Colourists, 1981.
2. Publication CIE No. 15.2, *Colorimetry*, Second Edition, Central Bureau of the Commission Internationale de l'Éclairage, Kegelgasse 23, A1033 Wein, Austria, 1986.
3. W. David Wright, A re-determination of the trichromatic coefficients of the spectral colours, *Trans. Opt. Soc. London* **30,** 141–164 (1928–29).
4. W. David Wright, A re-determination of the mixture curves of the spectrum. *Trans. Opt. Soc. London* **31,** 201–211 (1929–30).
5. John Guild, The colorimetric properties of the spectrum. *Philos. Trans. Roy. Soc. London, Ser. A* **230,** 149–187 (1931).
6. David L. MacAdam, Loci of constant hue and brightness determined with various surrounding colors. *J. Opt. Soc. Am.* **40,** 589–595 (1950).
7. R. E. Harrington, Effect of color temperature on apparent brightness. *J. Opt. Soc. Am.* **44,** 113–116 (1954).
8. William A. Thornton, Toward a more accurate and extensible colorimetry, Part I, *Color Res. Appl.* **17,** 79–122 (1992); Part II, *Color Res. Appl.* **17,** 162–186 (1992); Part III, *Color Res. Appl.* **17,** 240–262 (1992).
9. T. Smith and John Guild, The CIE colorimetric standards and their use. *Trans. Opt. Soc. London* **33,** 18–134 (1932).
10. Deane B. Judd, Reduction of data on mixture of color stimuli. *J. Res. Nat. Bur. Stand.* **4,** 515–548 (1930).
11. Isaac Newton, *Optics: or a Treatise of the Reflexions, Refractions, and Colours of Light*, Sam Smith and Benj. Walford, Printers to the Royal Society, London (1704). Newton's name does not appear on the title page. A facsimile edition was published by Culture et Civilization, 115 Avenue Gabriel Lebon, Bruxelles, in 1966. A reprint based on the 4th. Ed. of 1730, Innys, London, was published in 1952 by Dover, New York. It is this edition to which Judd refers.
12. Deane B. Judd, Chromatic visibility coefficients by the method of least squares. *J. Opt. Soc. Am. and Rev. Sci. Instrum.* **10,** 635–651 (1925).
13. Percival Frost, *Solid Geometry*, MacMillan and Co., London, 1886.
14. Publication CIE No. 15, *Colorimetry*, Central Bureau of the Commission Internationale de l'Éclairage, Paris (1971).
15. W. David Wright, *Researches on Normal and Defective Colour Vision*, Mosby Co., St. Louis, 1947, pp. 126–130.
16. Deane B. Judd, The 1931 ICI standard observer and coordinate system for colorimetry. *J. Opt. Soc. Am.* **23,** 359–374 (1933).
17. David L. MacAdam, On the geometry of color space. *J. Franklin Inst.* **238,** 195–210 (1944).
18. Deane B. Judd, A new look at the measurement of light and color. *Ill. Eng.* **53,** 61–70 (1958).
19. W. David Wright, *The Measurement of Colour*, First edition, 1944; Second edition, 1958; Third edition, 1964, Fourth edition, Adam Hilger, Ltd., London, 1969.
20. H. E. Ives, The transformation of color-mixture equations from one system to another. *J. Franklin Inst.* **180,** 673–701 (1915); **195,** 23–44 (1923).
21. Elliot Q. Adams, *X-Z* planes in the 1931 I. C. I. system of colorimetry. *J. Opt. Soc. Am.* **32,** 168–173 (1942).

ERRATUM

How the CIE 1931 Color-Matching Functions Were Derived from Wright–Guild Data

Michael H. Brill

Sarnoff Corp., CN 5300, Princeton, NJ 08543-5300

In the subject article,[1] there is an error in Eq. (A2) that propagates into Eq. (A8) and into the interpretation of Eq. (A19). This error concerns the CIE's rescaling of the original color-matching functions $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$. Although the scaling indeed equalized the visible-wavelength integrals of these three functions, it is not true that the integrals are equal to $\int V(\lambda)d\lambda$. Rather, the far-right-hand side of Eq. (A2) should be changed to $K\int V(\lambda)d\lambda$, and the right-hand sides of Eq. (A8) should all be K instead of 1. Here, K is given by

$$K = 1/(L_R + L_G + L_B) = 0.1770. \qquad (1)$$

The above value for K is obtained from the observation[2] that $\bar{r}(\lambda)$ was rescaled in such a way that $\bar{r}(\lambda_r) = V(\lambda_r)$, where $\lambda_r = 700$ nm and $V(\lambda)$ is the 1924 CIE luminance function. This rescaling ensures [see Eqs. (A15) and (A16)] that the coefficient b_{21} of $\bar{r}(\lambda)$ in the RGB expansion of $V(\lambda)$ is 1, and hence the luminance coefficients L_R, L_G, L_B defined in Eq. (A19) are identical (not just proportional) to the expansion coefficients b_{21}, b_{22}, b_{23}. Substituting the L's for the b's in Eq. (A15), integrating over visible wavelength, and recalling the equality of the $\bar{r}$, $\bar{g}$, and $\bar{b}$ integrals (Eq. A2), one can then obtain the relation

$$\int \bar{r}(\lambda)d\lambda = K\int V(\lambda)d\lambda, \qquad (2)$$

where K is defined in Eq. (1) above.

In summary, the area under each of the $\bar{r}$, $\bar{g}$, $\bar{b}$ color-matching functions is only 0.1770 times the area under each of the $\bar{x}$, $\bar{y}$, $\bar{z}$ functions, contrary to Appendix A of Ref. 1. This scaling convention is what enables the luminance coefficients to be equal to (not just proportional to) the best-fit coefficients of $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ to $V(\lambda)$. It is also implicit in Hunt's recent discussion.[3]

I thank Robert Risch of IBM Watson Research Center for bringing to my attention the error in Eq. (A2).

1. H. S. Fairman, M. H. Brill, and H. Hemmendinger, How the CIE 1931 color-matching functions were derived from Wright–Guild data, *Color Res. Appl.* **22,** 11–23 (1997).
2. G. Wyszecki and W. S. Stiles, *Color Science*, Second Ed., Wiley, New York, 1982, pp.737, 751.
3. R. W. G. Hunt, The heights of the CIE colour-matching functions, *Color Res. Appl.* **22,**, 335 (1997).