

FUNDAMENTAL STUDIES OF COLOR VISION FROM 1860 TO 1960

By DEANE B. JUDD

NATIONAL BUREAU OF STANDARDS

What I call fundamental studies of color vision got their start with Maxwell's determination (1860) of the color-matching functions of the normal human eye. Figure 1 (Judd, 1961) compares Maxwell's determinations for himself and one other observer (K) with modern color-matching functions. The open points and the lines drawn through them come from the 1931 CIE standard observer; the points marked otherwise are from Maxwell. Except for the extremes of the spectrum the discrepancies are only slightly more than can be explained by individual differences among observers with normal color vision. These curves, R , G , and B , indicate the radiant flux of spectrum primaries [630, 528, 457 nanometers (nm)] required in an additive mixture to match unit radiant flux for each part of the spectrum in turn. Note that at the red primary the R -curve is unity; the other two, zero. This means simply that one unit of radiant flux at this wavelength color-matches itself, and the same for the other two primaries, of course. Note also that some of the radiant fluxes indicated by the curves are less than zero. To speak of a negative radiant flux makes no sense physically; there is no such thing. But in color matching there is a useful meaning, universally recognized since Maxwell's time, of a negative radiant flux. It means simply that the negative amount of the primary, instead of being added to the other two primaries, is added to the part of the spectrum whose color is being determined. For example, at 500 nm Figure 1 shows that by adding about 0.2 unit of radiant flux of the red spectrum primary (630 nm) to one unit of radiant flux at 500 nm, there is produced a blue-green color that can be matched by about 0.4 unit of the green primary (528 nm) added to about 0.2 unit of the blue primary (457 nm).

The functions $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, $\bar{b}(\lambda)$ not only show how much of the primaries is required to match each part of the spectrum in turn, but they also indicate how much of the primaries is required to match any known combination of the various parts of the spectrum. On this account these functions are known as color-matching functions. By means of these functions, successful predictions may be made of whether any two lights of known spectral distributions $P_{\lambda,1}$ and $P_{\lambda,2}$ will, or will not, look alike. The conditions for color match are:

$$\begin{aligned}\int_0^\infty (P_{\lambda,1} - P_{\lambda,2}) \bar{r}(\lambda) d\lambda &= 0, \\ \int_0^\infty (P_{\lambda,1} - P_{\lambda,2}) \bar{g}(\lambda) d\lambda &= 0, \\ \int_0^\infty (P_{\lambda,1} - P_{\lambda,2}) \bar{b}(\lambda) d\lambda &= 0.\end{aligned}$$

These conditions for color match follow directly from Grassmann's law (1853) of additivity, well verified (König, 1887; v. Kries, 1905) for the macular regions of the retina. They state merely that if for each spectral component of a compound light the equivalent is given in amounts of the three primaries, the sum of these equivalents will color-match the sum of the spectral components making up the compound light. If the conditions for a color match are satisfied because the two lights are physically identical ($P_{\lambda,1} = P_{\lambda,2}$) throughout the visible spectrum, the match is said to be isomeric (following Ostwald's terminology) or spectral (Nimer-

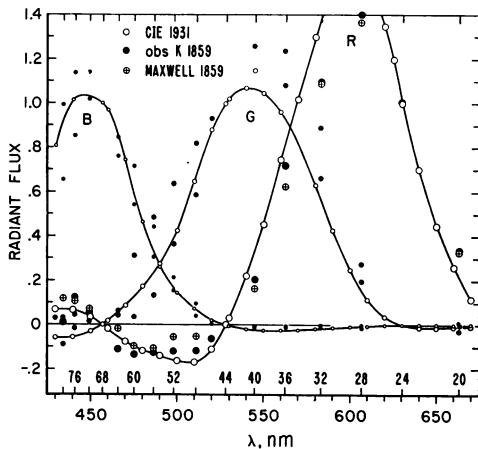


FIG. 1.—Radiant flux of the spectrum primaries, 630.2, 528.1, and 456.9 nm, required in an additive mixture to match unit radiant flux of the spectrum. 1931 CIE standard observer, \circ — \circ ; Maxwell's observer K, \bullet ; Maxwell (observer J), \otimes . Note that Maxwell's data expressed in units of radiant flux are in substantial agreement with the 1931 CIE standard observer. The numbers above the baseline refer to wavelengths designated on Maxwell's instrumental scale, nonlinear because of prismatic dispersion.

off, 1965). In this trivial case, the color-matching functions are not needed. If the two lights have different spectral distributions, the color match is said to be metameric; and in this usual case the color-matching functions are crucial.

Because of this generality of their application, studies yielding color-matching functions might be said to be fundamental. There are, however, two other reasons for regarding them as fundamental.

Types of Color Vision.—Table 1 summarizes the principal types of human color vision (Judd, 1943).

Normal color vision requires three color-matching functions for its definition, as shown in Figure 1. On this account normal color vision is classed as trichromatic. Trichromatic observers can make light-dark, yellow-blue, and red-green discriminations. There are three types of abnormal trichromatic color vision, known as protanomaly, deuteranomaly, and tritanomaly (rare).

Partially color-blind vision requires but two color-matching functions for its definition, and is known as dichromatic vision. There are three types of dichromatic vision, known as protanopia, deuteranopia, and tritanopia. A dichromat can match any color by a suitable additive combination of but two primary lights; and he is able to make discriminations of but two sorts as indicated in Table 1.

Totally color-blind vision requires but one color-matching function for its definition, and is known as monochromatic vision. A monochromat can match any color by a suitable amount of any light taken as a primary simply by adjusting the primary light to the same brightness, because a monochromat cannot make any discrimination other than light-dark.

Since the color-matching functions provide a basis for classifying human vision into the three types, trichromatic, dichromatic, and monochromatic, this is another sense in which we may regard them as fundamental.

Theories of Color Vision.—Major attention between 1860 and 1960 has been devoted to three categories of color-vision theory: three-components theory, opponent-colors theory, and stage theory. All of these theories, and indeed any tenable theory of color vision, must explain in detail why precisely the conditions for color match expressed by color-matching functions hold. The color-matching

TABLE 1
TYPES OF HUMAN COLOR VISION

Types	Response			Maximum luminosity at λ (nm)	Type designation
	Light-dark	Yellow-blue	Red-green		
Trichromatism	Yes	Yes	Yes	555	Normal
	Yes	Yes	Weak	540	Protanomaly
	Yes	Yes	Weak	560	Deutanomaly
	Yes	Weak	Yes	560	Tritanomaly
Dichromatism	Yes	Yes	—	540	Protanopia
	Yes	Yes	—	560	Deutanopia
	Yes	—	Yes	560	Tritanopia
Monochromatism	Yes	—	—	510	Cone blindness
	Yes	—	—	560	—
	Yes	—	—	540	—

functions thus play the role of quantitative boundary conditions for any theory of color vision. This is the major reason for regarding them as fundamental.

Three-components theory: This theory was briefly stated in 1807 by Thomas Young, and was elaborated by Helmholtz about 50 years later. It assumes the existence of three independent response mechanisms in the normal eye: one predominantly sensitive to long-wave light and yielding the response *red*; a second predominantly sensitive to middle-wave light and yielding the response *green*; and a third sensitive to short-wave light and yielding the response *violet*. Young thought that the long-wave and short-wave extremes of the spectrum could excite, respectively, the red and violet responses to the exclusion of the other two. Let us see how color-matching functions yield precisely the response curves implied by this assumption.

The key to this problem is choice of primaries. In Figure 1 are shown color-matching functions expressed relative to the spectrum primaries (630, 528, and 457 nm) chosen by Maxwell. From Grassmann's law of additivity it can be shown that precisely the same decisions as to color-match, or failure to color-match, can be obtained from functions R' , G' , and B' , as from R , G , and B , provided the new functions are weighted sums of the old:

$$R' = K_{11}R + K_{12}G + K_{13}B,$$

$$G' = K_{21}R + K_{22}G + K_{23}B,$$

$$B' = K_{31}R + K_{32}G + K_{33}B,$$

where K_{11} , K_{12} , . . . , K_{33} are constants that may be chosen arbitrarily provided the determinant differs from zero. It is legitimate to proceed therefore by trial and error to search for constants that will yield response curves, R' , G' , and B' , in accord with Young's hypothesis; that is, curves with no negative values anywhere, a green curve that is zero for as much as possible of the long-wave and short-wave portions of the spectrum, a red curve that is zero for the short-wave portion, and a violet curve that is zero for the long-wave portion. This choice of constants amounts to choice of red and violet primaries at the extremes of the spectrum.

This possibility was clearly explained by Helmholtz. He states, for example (p. 145): "The choice of the three fundamental colours is somewhat arbitrary. . . Young may have been guided by the consideration that the terminal colours of the spectrum seem to have special claims by virtue of their positions." The König-Dieterici

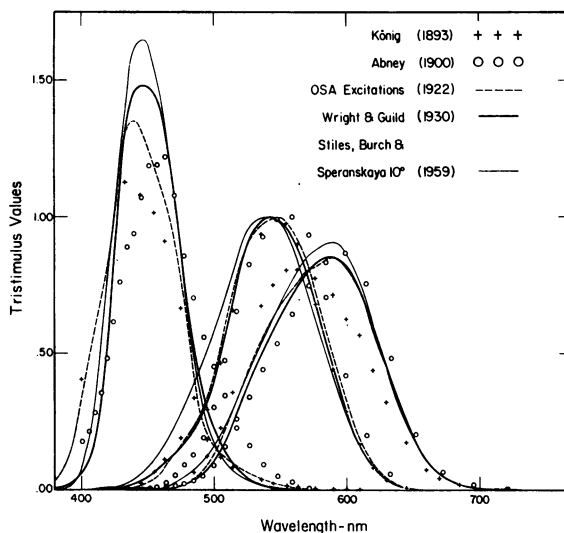


FIG. 2.—Color-matching functions expressed relative to the Young primaries (spectrum red, spectrum violet, and green outside spectrum). By the Young theory these functions represent the spectral absorptances of the retinal photopigments multiplied by the spectral transmittances of the ocular media through which light passes to reach the retina. König (1893), crosses; Abney (1900), circles; OSA excitations (Troland, 1922), dashed lines; CIE standard observer based on Wright and Guild (1930), heavy lines; CIE supplementary observer for 10° viewing fields based on Stiles and Burch (1959), and Speranskaya (1959), light lines.

determinations of the color-matching functions were presented (1893) with Young's choice of primaries under the name "Elementar-Empfindungs-Curven," and Abney's were similarly presented (1900) under the name "sensation curves." This choice of primaries was used by Weaver, who averaged the König and Abney results to produce the "elementary excitation curves" recommended in 1922 by the Committee on Colorimetry, Optical Society of America, under Troland's chairmanship. Figure 2 compares these color-matching functions with those based on the work of Wright (1929-1930) and Guild (1931) recommended in 1931 by the International Commission on Illumination (CIE), and with those of Stiles and Burch, and Speranskaya (1959) for extramacular vision (field subtending 10°) recommended in 1964 by the CIE for large-field colorimetry.

The color-matching functions in Figure 2 all show the amounts of the Young primaries required to color-match unit radiant flux of each part of the spectrum. The scale values for the G' -functions are adjusted to make the maximum be unity; those for the R' - and B' -functions are adjusted to make the areas under the curves equal to that for the G' -function; that is, the so-called equal-energy source (source whose radiant flux per unit wavelength is constant) corresponds for each set of functions to equal amounts of the three primaries. To make the König-Dieterici determinations (shown by crosses) refer to unit amount of spectral radiant flux, it was assumed that König's sunlight was equivalent to sunlight at air mass 2 from the compilation by Parry Moon (1940). Similarly, the Abney data (shown by circles) were adjusted by assuming that the positive crater of the carbon arc used by him was equivalent to the blackbody at $3,800^\circ\text{K}$. It will be noted that Weaver's average of these two sets of color-matching functions (shown by dashed lines) falls

fairly satisfactorily between the crosses and circles; so he must have made a fairly similar adjustment. The Guild-Wright determination (shown by heavy solid lines) corroborates the Weaver average surprisingly well considering the rather large differences between the König and Abney data. The Stiles-Burch-Speranskaya determination (shown by light solid lines) departs from these chiefly in ways that correspond to the use of extramacular retinal regions instead of macular, and this indicates that the extramacular cones have spectral sensitivities not much different from those of foveal cones.

It will be noted that Young's opinion that the long-wave portion of the spectrum can excite the red response to the exclusion of the other two is not contradicted by any of the color-matching functions. These functions, however, do not permit the view that the short-wave extreme can excite the violet response alone; a small red response (about 1/200 of the maximum) has to be admitted.

There is one final remark on Figure 2. If the additional assumption be made that each type of cone contains its own characteristic photosensitive pigment, then these color-matching functions expressed relative to the Young primaries should correspond to the product of the spectral absorptance of the three retinal pigments by the spectral transmittance of the ocular media (principally lens and macula). When Rushton (1957) announced that by measuring the spectral reflectance of the retina both before and after bleaching by strong light, he had discovered the presence of two bleachable pigments, one with peak absorptance at 590 nm, the other at 540 nm, it was very tempting to point to these two color-matching functions peaking very closely at these two wavelengths, and to say that this is striking corroboration of the Young theory. When the spectral transmittances of the ocular media are divided out of these curves, however, the predicted wavelengths of maximum absorptance are found to be somewhat lower than those indicated by Rushton's very difficult technique.

Partial color blindness: Young's theory has a simple built-in explanation of the three forms of dichromatic vision. These forms (protanopia, deuteranopia, and tritanopia) are reduction forms of normal vision. Any dichromat finds any color match set up by an observer of normal color vision with the same ocular pigmentation to be an acceptable match for him also. The simplest explanation for dichromatic vision by the Young theory is to say that a protanopic observer is an otherwise normal observer who has lost the red response. Similarly, deuteranopia corresponds to loss of the green response, and tritanopia, to loss of the violet. Since in each case the remaining two responses are identical with those of normal vision, the dichromatic observer has no way to detect any error in a normal match. A corollary of this explanation, however, is that an observer who has lost the red response would be blind to the long-wave part of the spectrum where the other two mechanisms have zero sensitivity. Protanopes have reduced sensitivity to long-wave light, but not zero. Similarly, tritanopes have reduced sensitivity to short-wave light, but not zero.

Maxwell was among the first to recognize that dichromatism is a reduced form of normal trichromatism, and he described a method by which to determine precisely the normal color corresponding to the missing primary process. In a letter of June 4, 1855, to G. Wilson, Maxwell wrote: "The mathematical expression of the difference between colour-blind and ordinary vision is that colour to the former is a

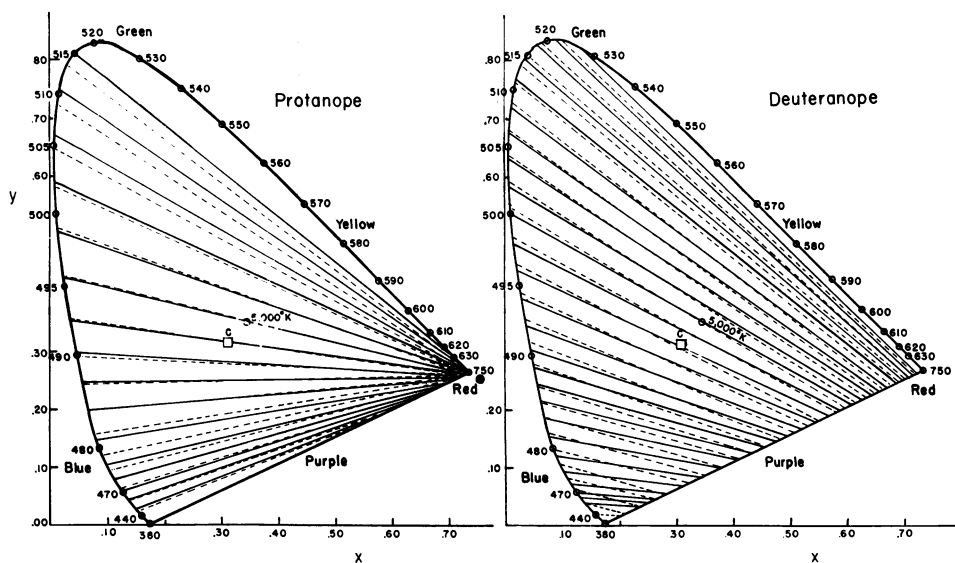


FIG. 3.—Chromaticity confusions of the protanope and deuteranope shown on the (x,y) -chromaticity diagram of the 1931 CIE colorimetric coordinate system from determinations by Pitt (1935). The points along each dotted line correspond to chromaticities found by Pitt to be indistinguishable in dichromatic vision. The solid lines intersect at a single point: protanope at $x = 0.753$, $y = 0.247$; deuteranope at $x = 1.000$, $y = 0.000$.

function of two independent variables, but to an ordinary eye, of three; and that the relation of the two kinds of vision is not arbitrary, but indicates the absence of a determinate sensation, depending perhaps upon some undiscovered structure or organic arrangement, which forms one-third of the apparatus by which we receive sensations of colour. . . . If we find two combinations of colours which appear identical to a colour-blind person, and mark their positions on the triangle of colours, then the straight line passing through these points will pass through all points corresponding to other colours, which, to such a person, appear identical with the first two. We may in the same way find other lines passing through the series of colours which appear alike to the colour-blind. All these lines either pass through one point or are parallel, according to the standard colours which we have assumed, and the other arbitrary assumptions we may have made. Knowing this law of colour-blind vision, we may predict any number of equations which will be true for eyes having this defect."

Figure 3 shows on the CIE (x,y) -chromaticity diagram by dotted lines the chromaticity confusions found by Pitt (1935) for six protanopes (on the left), and for six deuteranopes (on the right). The protanopic confusion lines are seen to intersect at very nearly one point as stated by Maxwell, and indicate the color for whose perception protanopes have no mechanism. The deuteranopic confusion lines are not as consistent, but the average intersection point is at about $x = 1.10$, $y = -0.10$. The solid lines intersect at $x = 1.00$, $y = 0.00$, which may be taken as an upper y -limit for Pitts' six deuteranopes. By basically the same method, though different in detail, König (1892) also evaluated the intersection points for confusion lines by two protanopes and one deuteranope. Similar evaluations for confusion lines were made by Farnsworth (1955) for one tritanope, and by Thomson and Wright (1953)

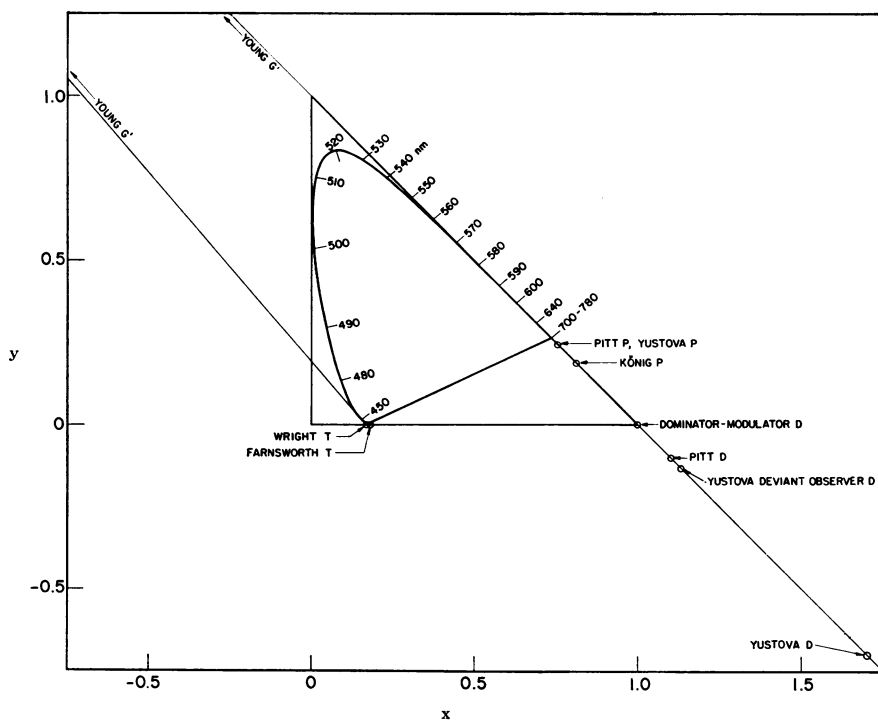


FIG. 4.—Intersection points of the chromaticity-confusion lines for the three types of dichromatic vision: protanopic, *P*; deuteranopic, *D*; and tritanopic, *T*, shown on the (x,y) -chromaticity diagram of the 1931 CIE colorimetric coordinate system. Note that the two determinations (Wright, 1952; Farnsworth, 1955) of the *T*-point fall close to $x = 0.17$, $y = 0.00$, and that the two recent determinations (Pitt, 1935; Yustova, 1953) of the *P*-point fall close to $x = 0.75$, $y = 0.25$. The determinations of the *D*-point (Pitt, 1935; Nuberg and Yustova, 1955), on the other hand, spread over a considerable range. The dominator-modulator theory implies a *D*-point at $x = 1.00$, $y = 0.00$. Young's theory in its original form implies that the deuteranope chromaticity confusion lines are nearly parallel on the (x,y) -chromaticity diagram (see the two straight lines labeled *Young G'*).

for seven tritanopes. Figure 4 shows the locations of the intersection points of the chromaticity-confusion lines for protanopes (*P*), deuteranopes (*D*), and tritanopes (*T*) from the studies mentioned above. König's deuteranope yields an intersection point off the graph.

Yustova (1949, 1953) carried out studies of dichromatic vision, and in 1955 Nuberg and Yustova published an extensive determination of intersection points of chromaticity-confusion lines for four protanopes and 12 deuteranopes. They used a long-overlooked method, also originated by Maxwell (1860), which yields the intersection point from a single pair of colors confused by the dichromat. The method is to obtain the tristimulus values set by the dichromatic observer to match any arbitrarily chosen color and to subtract them from the tristimulus values of the same color set by an observer having normal trichromatic vision. The differences specify the color for the perception of which no mechanism exists in the dichromatic observer (Judd, 1964). The Yustova results are also shown in Figure 4. It will be noted that the point found by Yustova for four protanopes agrees perfectly with that found by Pitt for six protanopes. Only one of the 12 deuter-

anopes, however, agrees with Pitt's *D*-point. The other 11 deuteranopes yield intersection points between Pitt's *D*-point and König's early determination for one deuteranope.

It must be pointed out that the location of the intersection points at various places along the line $z = 0$ in the CIE coordinate system correlates with the wavelength of the maximum of the long-wave response curve for the corresponding dichromat. This long-wave response curve is essentially the dichromatic luminosity function because the contribution of the short-wave response function to luminosity, if not precisely zero, differs negligibly from zero. For the intersection point at $x = 0.75$, $y = 0.25$, as for protanopes, the dichromatic luminosity function peaks at about 540 nm. For $x = 1.00$, $y = 0.00$, as in the dominator-modulator theory (to be discussed presently), the luminosity function is identical with that for normal vision, and peaks at 555 nm. For $x = 1.10$, $y = -0.10$, the peak is at about 560 nm, as for Pitt's six deuteranopes. But for Yustova's 11 deuteranopes yielding $x = 1.7$, $y = -0.7$, the peak is at 573 nm. The luminosity curves found by Hsia and Graham (1957) average to a curve with a peak at about 575 nm in good agreement with the Yustova determination. It should be recalled, also, that Willmer (1949) announced the discovery of two distinct types of deuteranopia, one with normal luminosity function, the other with luminosity function shifted toward the long-wave end of the spectrum compared to normal. If it is true, as announced by Willmer, that there are two types of deuteranope, then perhaps all of the *D*-points shown on Figure 4 are correct. By this view, all of Pitt's six deuteranopes, one of Hsia and Graham's six deuteranopes, and one of Yustova's 12 deuteranopes should be classed as having essentially normal luminosity functions; the remainder of the deuteranopes studied have luminosity functions shifted from the normal position to the long-wave end of the spectrum.

If the color-matching functions are expressed relative to the *P*-, *D*-, and *T*-points, they may be interpreted as the responses that three different kinds of cones would have to have to explain the three types of dichromatic vision. Figure 5 shows such response curves derived from the CIE 1931 color-matching functions in which the *D*-point is taken at $x = 1.00$, $y = 0.00$. By this choice, which is barely admissible from Pitt's determination, and inadmissible from Yustova's, deuteranopic luminosity is taken as identical to normal luminosity. The obvious difference between these response curves and those expressed relative to Young's primaries is that the long-wave curve peaks, not at 590 nm, but at 555 nm; and indeed this curve is precisely the normal luminosity function. The other important, though not obvious, difference is that the middle-wave curve does not drop all the way to zero at either end of the spectrum until the other curves do. The short-wave curve remains unchanged. The usual explanation of the differences between these two sets of response curves is that those referred to the Young primaries reflect the spectral absorptances of photosensitive retinal pigments, while those expressed relative to the intersection points of the chromaticity-confusion lines for the three types of dichromatic vision reflect the responses of three types of cones. The fact that the short-wave curves of the two sets are identical is taken to mean that the cones giving violet or blue signals contain only the short-wave-absorbing pigment. The fact that the middle- and long-wave curves expressing cone response are broader than those expressing absorptance of photosensitive pigments may be taken to mean that

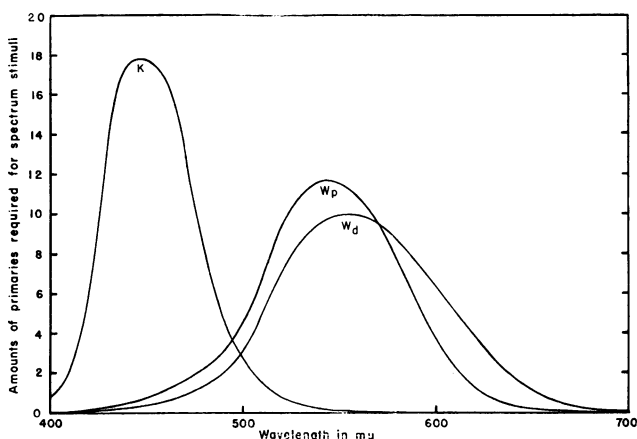


FIG. 5.—Color-matching functions for the 1931 CIE standard observer expressed relative to the primaries implied by the König (1897) tentative form of dominator-modulator theory. The red primary is placed at the protanopic intersection point ($x = 0.753, y = 0.247$); the violet primary is placed at the tritanopic intersection point ($x = 0.18, y = 0.00$); and the green primary is placed at the deutanopic intersection point ($x = 1.00, y = 0.00$) implied by the dominator-modulator theory. By this theory the W_d curve represents for normal color vision both the luminous-efficiency function and the red response function; for deutanopic vision it is both the luminous efficiency function and the “warm” response function. Protanopic vision does not have this function. Similarly, the W_p curve represents for protanopic vision both the luminous-efficiency function and the “warm” response function. Finally, the K curve represents the response function for the violet modulator in normal vision and for the “cold” modulator for protanopic and deutanopic vision. Tritanopic vision does not have this function.

the cones giving green signals contain, in addition to middle-wave-absorbing pigment, slight amounts of the other two pigments, and the cones giving red signals contain about equal amounts of middle-wave- and long-wave-absorbing pigments, with perhaps a slight admixture of short-wave-absorbing pigment. This interpretation is based on the idea that the cause of dichromatic vision is a “loss” of one of the three components of normal vision. Another well-explored interpretation of the relation of dichromatic vision to normal trichromatic vision is that proposed by Leber (1869) and Fick (1879). They said that deuteranopia, for example, might be explained by assuming that both the cones signaling red and those signaling green alike contain the same admixture of middle-wave- and long-wave-absorbing pigments. This idea leads to the same cone-response curve as the “loss” hypothesis, but has the advantage that by this view the dichromat could never experience red separate from green, but would always see what unilateral dichromats testify they see: either blue or yellow (Judd, 1948).

The set of response functions in Figure 5 accords with the particular variety of three-components theory tentatively championed by König following his study of tritanopia in 1897. For example, v. Kries in a note specially prepared in 1924 for insertion in the English edition of Helmholtz' *Physiological Optics*, states (p. 412): “König long ago called attention incidentally to the curious fact that the luminosity values of the colours . . . turn out to be very nearly the same function of wavelength for deuteranopes and persons with normal vision. Since in the case of deuteranopes light of long wave-lengths has no effect except on the red component, we must suppose also that this action depends on the wave-length in the same way. But

then the further result is that the distribution of luminosity is not affected, or at least almost inappreciably affected, by the addition of the green component by which the deuteranopic visual organ is converted into a normal organ; in other words, that in the case of normal vision also the luminosity goes practically hand in hand with the action of the red component." König was, I believe, the first among three-components theorists to suggest the possibility that the production of the brightness sensation might be entrusted to a single one of the three receptor mechanisms, with the other two mechanisms having the function of modulating this response to produce the chromatic aspects of vision. After Granit's work on retinas of various animals with the microelectrode indicated (1943) that these retinas contained a preponderance of brightness sensors, such theories of color vision came to be known as dominator-modulator theories.

To summarize the 1960 status of three-components theories, there was one for the photosensitive pigments of the retina, and there were several cone-response theories, including dominator-modulator theories, depending on the relationship assumed between deuteranopic and normal vision. All of these theories assume that yellow is produced by the sum of red and green responses and that white is produced by the sum of equal amounts of red, green, and violet responses; but none of them so far discussed here provide any explanation for these assumptions. The Ladd-Franklin theory (1892) is a skillful attempt to formulate a three-components theory overcoming this defect. Its outstanding weaknesses are its failure to account for tritanopia and its implication that the blue process contributes to luminosity for protanopic vision, but subtracts from it for normal and deuteranopic vision.

Opponent-colors theory: This theory was proposed and explained in detail by Ewald Hering in 1878. It is based on an analysis of sensations of color rather than of the stimuli required to evoke them. It assumes that there are six independent unitary colors (red, yellow, green, blue, white, and black), no one of which partakes of any other; that is, for example, yellow is a basic color in its own right, not the product of combining red with green. The Hering theory tacitly assumes that light is absorbed in the receptors by photopigments, that this absorption starts activity in the rest of the visual system, and that this activity is directly responsible for the colors we see. This activity is not found in six separate systems, but in three opposing pairs of processes: black-white, yellow-blue, and red-green. Black and white blend to produce gray, but equal amounts of yellow and blue, and of red and green cancel to zero.

This formulation of human vision supplies a specific model for the colors perceived, which the three-components theory does not. It has been very fruitful in suggesting many researches in such aspects of vision as chromatic adaptation, the influence of the surround on the perceived color, Bezold-Brüche phenomenon, and dependence of color perception on luminance of test field and surround; and it has provided a framework within which the results of such studies can be clearly stated and analyzed.

Just as is true for the various three-components theories, so also do color-matching functions provide quantitative information regarding the dependence of the opponent-color processes on wavelength. It is necessary only to express the color-matching functions relative to the primaries implied by the special assumptions

of the particular form of opponent-colors theory under consideration. This quantitative information will be presented under stage theories of vision.

Stage theories of human vision: It has already been pointed out that there are three-components theories for two stages in the visual process: the retinal photopigment stage and the cone-response stage. It was suggested by v. Kries (1905) that the Young-Helmholtz three-components theory may be taken as valid at the receptor level but that the signals from the receptors are so processed that at some later stage the opponent-colors theory of Hering applies. Schrödinger (1925) derived the response curves for the two stages implied by the v. Kries proposal.

A somewhat similar two-stage theory has been used by Hurvich and Jameson (1951 on) in their comprehensive and detailed studies of color vision. The opponent-colors theory has probably received at their hands its most successful quantification and application.

There has been considerable uncertainty as to precisely at what stage of the visual process the signals from the receptors can be said to be organized in opponent colors, whether this be in the retina, the optic nerve, or the occipital lobe of the cortex. Adherents of the opponent-colors theory have been greatly encouraged by the work of Svaetichin and MacNichol (1953-1961) on shallow-water fish with the microelectrode technique. They obtained two types of response curves from the bipolar-cell layer of the fish retina. Both types of curves changed polarity depending on the wavelength of the light introduced into the retina. One corresponded in spectral location and shape to a yellow-blue process of an opponent-colors theory, the other to a red-green process. They suggested that the *Y-B* and *R-G* spectral response curves are expressions of signals delivered by *Y-B* and *R-G* cones, the chromoreceptors which form the basic mechanism for the color vision of this fish. This result for fish vision suggests that the signals in human vision might be organized according to opponent colors in the retinal stage immediately following the receptors themselves.

Müller three-stage theory: This first three-stage theory was developed by G. E. Müller between 1920 and 1930. The three stages are: photopigment stage, cone-response stage, and optic-nerve stage. The photopigment stage follows the three-components theory based on the Young primaries (see Fig. 6, upper-left quadrant). The cone-response stage follows an opponent-colors form with yellowish-red opposing bluish-green, and greenish-yellow opposing reddish-blue; see upper-right quadrant of Figure 6. These response curves (Judd, 1949) derived from the CIE color-matching functions are strikingly similar to those found by Svaetichin and MacNichol from the retinas of shallow-water fish. The optic-nerve stage is the opponent-colors formulation of Hering with red opposing green, and blue opposing yellow (see lower-right quadrant). This is close to the formulation used in the second stage by Schrödinger and by Hurvich and Jameson. The details of Müller's ingenious explanation of all three principal forms of dichromatic vision in terms of these three stages need not be explored here. Some hint of them, however, is shown in the lower-left quadrant. Müller assumed that the photopigment stage contributes directly to luminosity in such proportions as yields closely the protanopic luminosity function (see W_p); then for normal vision the $\eta R-bG$ process makes another contribution of such size as to make up the difference between the normal and protanopic luminosity functions. Protanopia is explained by the loss of the

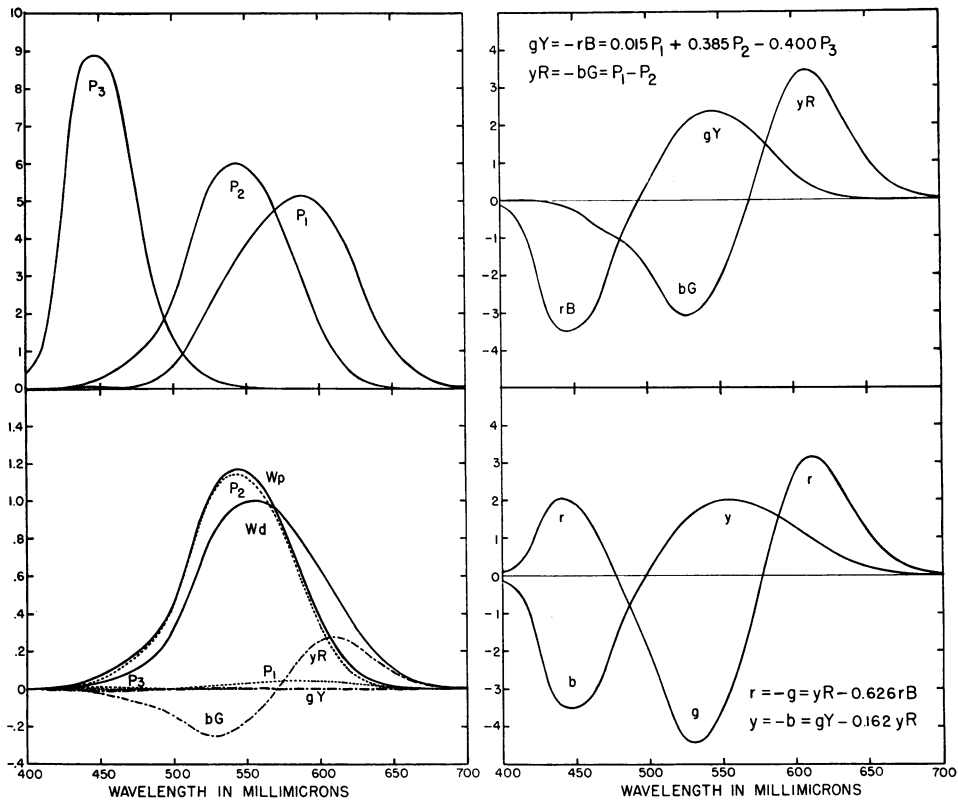


FIG. 6.—Response functions of the Müller stage theory evaluated by means of the 1931 CIE standard observer. *Upper-left* quadrant shows the response functions at the photopigment stage (same as Fig. 2, heavy lines). *Upper-right* quadrant shows the opponent-color response functions at the receptor stage. *Lower-right* quadrant shows the opponent-color response functions at the optic-nerve stage. These optic-nerve response functions also correspond to the Hering opponent-colors theory, and to the optic-nerve stage of the Adams theory. The *lower-left* quadrant shows how the Müller theory explains the luminous efficiency W_d function as the sum of contributions from the P_1, P_2, P_3 processes of the photopigment stage and from the yR and gY processes of the receptor stage. By this theory protanopia is a defect of the receptor stage whereby the yR ($= -bG$) process is lost. This loss explains both the inability of the protanopic observer to distinguish red from green, and the shortening of the luminous-efficiency function at the long-wave end whereby the W_d curve is supplanted by the W_p curve. Deuteranopia, on the other hand, is explained as a defect in the optic-nerve stage whereby the r ($= -g$) process is lost. This loss explains the inability of the deuteranopic observer to distinguish red from green, but it implies that the luminous-efficiency function of the deuteranopic observer must be identical to that for the normal observer.

yR - bG process, and it is clear that such a loss, in addition to accounting for inability of the protanope to make these red-green chromatic discriminations, also endows him with a luminosity function peaking correctly at about 540 nm.

Adams three-stage theory: Elliot Q. Adams proposed a three-stage theory of human color vision in 1923, and applied it in 1942 rather successfully to prediction of the perceptual size of color differences. The first and last stages correspond to the Young primaries and to the Hering opponent colors, respectively, just as in the Müller theory. The Adams second stage is a three-components formulation and, surprisingly enough, corresponds precisely to the primaries of the CIE system. Figure 7 shows the CIE color-matching functions in their official form. Adams

takes these without change to be the cone-response functions. The central function, which is the normal luminosity function, refers to brightness-perceiving cones, and a portion of their signals is supposed to be transmitted as nonlinear transforms (Munsell value function from Newhall, 1943) of luminance directly to the optic nerve as white signals. The similarly nonlinear signals from the red cones are partially inhibited by another portion of the white signal, and the difference is transmitted to the optic nerve in the form of the Hering red-green response. Similarly, the nonlinear signals from the blue cones are inhibited by the remainder of the signals from the white cones, and the difference is transmitted to the optic nerve in the form of the Hering blue-yellow response. The outstanding weakness of the Adams three-stage theory is that it provides no built-in explanation of two of the three forms of dichromatic vision. It is noteworthy, however, for two reasons. In the first place, the second and third stages are strictly dominator-modulator in type. The Adams theory antedates the coining of the phrase "dominator-modulator" by 20 years. In the second place, it is the first theory of color vision to take explicit account of the possibility that the frequency of firing of the nerve leading from the cones may not be linear with the rate at which radiant flux is absorbed within the cone, but is likely to be a diminishing function of it, such as the logarithm. This ends the summary of the principal theories of human color vision, and the demonstration that color-matching functions are basic to their quantification.

Chromatic Adaptation.—Another method of studying human color vision, less fundamental perhaps, than the information to be derived from setting of color matches, is to submit the eye to pre-exposure of various known kinds, and to study the differences in response so caused. It has been known since studies by v. Kries and König (about 1890) that such pre-exposures of whatever wavelength composition and over a wide range of retinal illuminances [up to 10,000 trolands according to Wright (1936)] have no influence on color matches set up for foveal vision. This "persistence of optical matches," as v. Kries phrased it, means that the color-matching functions whose determination has already been summarized apply to a wide range of chromatic adaptations. All of the theories mentioned above provide for this fact.

Helmholtz (1860) early indicated the connection conceived by him between chromatic adaptation and the three independent mechanisms of Young's theory by saying (p. 235): "... the nervous substance in question is less sensitive to new reacting light falling on it than the rest of the retina that was not previously stimulated... Thus [p. 240] an eye which has been acted on by yellow light, say, is thereafter in a condition in which the blue components of white light affect it more than yellow does. Accordingly, the effect of fatiguing the retina is not uniformly extended to every kind of stimulation, but chiefly to stimulation similar to the primary stimulation. This fact is explained very simply by Young's assumption of three different kinds of sensory nerves for the different colours. For since coloured light does not stimulate these three kinds of nerves all to the same extent, different degrees of fatigue must also be the result of different degrees of stimulation. When the eye has been exposed to red, then the red-sensitive nerves are strongly stimulated and much fatigued; whereas the green-sensitive and violet-sensitive nerves are feebly stimulated and not much fatigued. If afterwards white light falls on the eye, the green-sensitive and violet-sensitive nerves will be relatively more af-

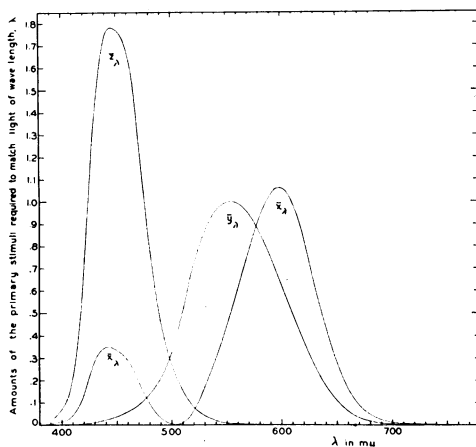


FIG. 7.—Color-matching functions for the 1931 CIE standard observer expressed in the CIE colorimetric coordinate system. The primaries are red (\bar{x}_λ), green (\bar{y}_λ), and blue (\bar{z}_λ), all outside the spectrum. The \bar{y}_λ -curve is identical to the photopic luminous-efficiency function for normal color vision, and this choice of \bar{y}_λ -curve ensures that the \bar{x}_λ - and \bar{z}_λ -curves are unassociated with luminosity. These color-matching functions apply to the receptor stage of the Adams theory (1923, 1942) which for this stage has the dominator-modulator form.

fects by it than the red-sensitive nerves; and hence the impression of blue-green, which is complementary to red, will predominate in the sensation." This explanation of the qualitative changes caused by local chromatic adaptation is quite successful, and equally good predictions are possible by the opponent-colors theory.

The predictions can, however, be subjected to a quantitative test. Suppose one retinal region (one eye, say) be maintained in a state of adaptation to daylight, and another retinal region (the other eye, say) be pre-exposed to green light. All stimuli presented to the green-adapted portion of the retina will appear more purplish than when presented to the daylight-adapted eye, and a measure of this adaptive color shift may be obtained by allowing the observer to alter the light stimulating the daylight-adapted region until he finds a stimulus yielding a color preception equal to that from the green-adapted region. Helmholtz' ideas on chromatic adaptation were quantitatively formulated by v. Kries, as follows: If R , G , V specify the color of any stimulus for neutral adaptation, then for any chromatic adaptation, the resulting color perception of this stimulus will be similarly specified by R' , G' , V' , where

$$R' = S_r R, \quad G' = S_g G, \quad V' = S_v V,$$

where S_r , S_g , S_v are coefficients specifying the state of chromatic adaptation of the eye following the pre-exposure. This formulation is known as the v. Kries coefficient law (v. Kries, 1905; p. 211).

The v. Kries coefficient law will work to some degree or other even if the coordinate system within which R , G , V are evaluated departs rather widely from that based on the true cone responses, but it should work perfectly only for the true cone responses. Studies of chromatic adaptation can therefore be used to check the accuracy of response curves derived by some other means, such as by reference to the protanopic, deuteranopic, and tritanopic intersection points. Studies by Wright (1934), Walters (1942), MacAdam (1956), Brewer (1954), Wassef (1955), and Burnham, Evans, and Newhall (1957) have indicated that the v. Kries coefficient law is at least a good first approximation to the facts, but the primaries yielding the closest approximation to correct predictions do not agree with the dichromatic intersection points. The P - and T -points are found in some studies to

be reasonably successful, but the best third primary is extremely variable. For some studies (MacAdam, 1956) it falls within the range of D -points, but for others, far from that range. MacAdam (1958) has suggested a nonlinear extension of the v. Kries law for improved agreement.

A more rewarding line of attack is that by Stiles (1949, 1959) who determined for spectrum colors the least amount required to be added to the center of a large field of another spectrum color to be just detectable. The large field, of course, controls the chromatic adaptation of the eye of the observer, and may render two of the receptor mechanisms relatively insensitive, in which case the thresholds found should be inversely related to the spectral sensitivity of the third mechanism. If the wavelength of the large field is such as to reduce chiefly the sensitivity of but one of the receptor mechanisms, then the threshold for a given test stimulus will still be related chiefly to the sensitivity of the most sensitive of the remaining two. Only if the sensitivities of the two mechanisms are nearly equal does the interpretation of the increment-threshold data depend on an assumption regarding the combination of the responses of two mechanisms to produce an increment threshold.

The result of Stiles' researches by the two-color increment-threshold method is in one sense a very gratifying check on the cone-response curves derivable from the color-matching functions by taking the D -point in agreement with the Yustova values: $x = 1.7$, $y = -0.7$. One cone mechanism was found to peak at 440 nm, another at 540 nm, and a third at 575 nm. The shape of the "red" cone response curve found by this method does not, however, agree too well with that derived from color-matching functions. Its maximum is considerably more flat. In still another sense the results are puzzling. Not one but three "blue" mechanisms were found with response curves peaking at 440 nm. Furthermore, the response curves for the "green" and "red" cones changed significantly depending on whether the large field provided high or low retinal illuminances. The response for the "green" cone shifted at high illuminances toward somewhat shorter wavelengths, and that for the "red" cone toward longer, so that the maximum was pushed over to 587 nm, that is nearly in agreement with the "red" curve of the photopigment level. Perhaps this means that the photopigment that does not belong in the cone is less stable than the one that does. Evidence both from studies of the v. Kries coefficient law and from the two-color increment-threshold method suggests that chromatic adaptation may be a somewhat more complicated phenomenon than can be adequately described by any simple form of three-components theory.

Summary.—Between 1860 and 1960 several alternative theoretical possibilities were carefully worked out for human color vision but progress has been handicapped by lack of crucial evidence that might decide between these possibilities.

The color-matching functions determined since Maxwell's time (1860) have been reduced to the primaries previously (1807) proposed by Young, and the modern determinations are shown to be in good agreement. The principal theories of color vision are summarized under the headings: three-components (Young, Helmholtz, König), opponent-colors (Hering), two-stage theories (v. Kries-Schrödinger and Hurvich-Jameson), and three-stage theories (Müller, Adams). The results of quantifying these theories by appeal to modern determinations of color-matching functions is shown. For example, by the Young theory predictions of the spectral absorptance of three retinal photopigments are derived, and from the Helmholtz-

König theories, the responses of three types of retinal receptors responsible for color vision are derived. The various experimentally determined relations between the three principal forms of partial color-blindness and normal color vision are intercompared and checked against the various theories, and a brief review of the results of studies of chromatic adaptation (v. Kries coefficient law; Stiles two-color increment thresholds) and their theoretical implications is given.

Key words: Chromatic adaptation, color blindness, color-matching functions, color vision, deuteranopia, opponent-colors theory, protanopia, three-components theory of color vision, tritanopia.

Abney, W. de W., "The colour sensations in terms of luminosity," *Phil. Trans. Roy. Soc.*, **193**, 259 (1900).

Abney, W. de W., *Researches in Colour Vision and the Trichromatic Theory* (London: Longmans, Green, 1913), p. 241.

Adams, E. Q., "A theory of color vision," *Psychol. Rev.*, **30**, 56 (1923).

Adams, E. Q., "X-Z planes in the 1931 ICI system of colorimetry," *J. Opt. Soc. Am.*, **32**, 168 (1942).

Brewer, W. L., "Fundamental response functions and binocular color matching," *J. Opt. Soc. Am.*, **44**, 207 (1954).

Burnham, R. W., R. M. Evans, and S. M. Newhall, "Prediction of color appearance with different adaptation illuminations," *J. Opt. Soc. Am.*, **47**, 35, 443 (1957).

CIE (Commission Internationale de l'Éclairage) Proc. 8th Session, Cambridge, England, September 1931, pp. 19-29.

CIE (Commission Internationale de l'Éclairage), Proc. 15th Session (Paris, France: Bureau Central de la CIE, 1964), vol. A, p. 113.

Farnsworth, D., "Tritanomalous vision as a threshold function," *Farbe*, **4**, 185 (1955).

Fick, A., "Die Lehre von der Lichtempfindung," *Hermann's Handbuch der Physiologie* (Leipzig: Vogel, 1879), vol. 3, part 1, p. 139.

Granit, R., "A physiological theory of colour perception," *Nature*, **151**, 11 (1943).

Grassmann, H., "Zur Theorie der Farbenmischung," *Poggendorf's Ann.*, **89**, 69 (1853); also *Phil. Mag.* (4), **7**, 254 (1853).

Guild, J., "The colorimetric properties of the spectrum," *Phil. Trans. Roy. Soc. (London)*, **A230** 149 (1931).

von Helmholtz, H., in *Treatise on Physiological Optics*, translated from the 3rd German edition, ed. J. P. C. Southall (The Optical Society of America, 1924), vol. 2, pp. 145, 235, 240, 412.

Hering, E., *Zur Lehre vom Lichtsinn* (Berlin, 1878).

Hsia, Y., and C. H. Graham, "Spectral luminosity curves for protanopic, deuteranopic, and normal subjects," these PROCEEDINGS, **43**, 1011 (1957).

Hurvich, L. M., and D. Jameson, "Some quantitative aspects of an opponent-colors theory, II. Brightness, saturation, and hue in normal and dichromatic vision," *J. Opt. Soc. Am.*, **45**, 602 (1955).

Hurvich, L. M., and D. Jameson, "Some quantitative aspects of an opponent-colors theory, IV. A psychological color specification system," *J. Opt. Soc. Am.*, **46**, 416 (1956).

Hurvich, L. M., and D. Jameson, "An opponent-process theory of color vision," *Psychol. Rev.* **64**, 384 (1957).

Jameson, D., and L. M. Hurvich, "Some quantitative aspects of an opponent-colors theory, I. Chromatic responses and spectral saturation," *J. Opt. Soc. Am.*, **45**, 546 (1955).

Jameson, D., and L. M. Hurvich, "Some quantitative aspects of an opponent-colors theory, III. Changes in brightness, saturation, and hue with chromatic adaptation," *J. Opt. Soc. Am.*, **46**, 405 (1956).

Judd, D. B., "Facts of color-blindness," *J. Opt. Soc. Am.*, **33**, 294 (1943).

Judd, D. B., "Response functions for types of vision according to the Müller theory," *J. Res. Natl. Bur. Std.*, **42**, 1 (1949); RP1946.

Judd, D. B., "Color perceptions of deuteranopic and protanopic observers," *J. Res. Natl. Bur. Std.*, **41**, 247 (1948); RP1922; also (condensed version) *J. Opt. Soc. Am.*, **39**, 252 (1949).

- Judd, D. B., "Maxwell and modern colorimetry," *J. Phot. Sci.*, **9**, 341 (1961).
- Judd, D. B., "Relation between normal trichromatic vision and dichromatic vision representing a reduced form of normal vision," *Acta Chrom.*, **1**, 89 (1964).
- König, A., and C. Dieterici, "Die Grundempfindungen in normalen und anomalen Farbensystem und ihre Intensitätsvertheilung im Spectrum," *Z. Psychol. Physiol. Sinnesorg.*, **4**, 231 (1892); König, A., *Gesammelte Abhandlungen* (Leipzig: Barth, 1903), p. 214.
- König, A., "Ueber 'Blaublindeheit'," *Sitzber. Akad. Wiss. Berlin* (July 8, 1897), p. 718; *Gesammelte Abhandlungen* (Leipzig: Barth, 1903), p. 396.
- König, A., "Die Abhängigkeit der Farben- und Helligkeitsgleichungen von der absoluten Intensität," *Sitzber. Akad. Wiss. Berlin* (July 29, 1897), p. 871; *Gesammelte Abhandlungen* (Leipzig: Barth, 1903), pp. 416-429.
- v. Kries, J., and W. A. Nagel, "Über den Einfluss von Lichtstärke und Adaptation," *Z. Psychol. Physiol. Sinnesorg.* **12x**, 1 (1896).
- v. Kries, J., and W. A. Nagel, "Validity of Newton's law of color mixture," *Z. Psychol. Physiol. Sinnesorg.*, **23**, 161 (1900).
- v. Kries, J., "Die Gesichtsempfindungen," in *Handbuch der Physiologie des Menschen*, ed. W. A. Nagel (Vienna: Braunschweig, 1905), vol. 3, pp. 209, 211.
- Ladd-Franklin, C., "Eine neue Theorie der Lichtempfindungen," *Z. Psychol. Physiol. Sinnesorg.*, **4**, 211 (1892).
- Ladd-Franklin, C., "Colour and Colour Theories" (New York: Harcourt-Brace, 1929), pp. 193-197.
- Leber, Th., "Ueber das Vorkommen von Anomalien des Farbensinnes bei Krankheiten des Auges, nebst Bemerkungen über einige Formen von Amblyopie," *Arch. Ophthalmol.* (3), **15**, 26 (1869).
- MacAdam, D. L., "Chromatic adaptation," *J. Opt. Soc. Am.*, **46**, 500 (1956).
- MacAdam, D. L., "Beat-frequency hypothesis of colour perception," Paper 25, NPL Symposium on Colour, September 1957 (London: Her Majesty's Stationery Office, 1958), vol. 2, p. 577.
- MacNichol, E. F., Jr., L. Macpherson, and G. Svaetichin, "Studies on spectral response curves from the fish retina," Paper 39, NPL Symposium on Colour, September 1957 (London: Her Majesty's Stationery Office, 1958), vol. 2, p. 529.
- MacNichol, E. F., Jr., M. L. Wolbarsht, and H. G. Wagner, "Electrophysiological evidence for a mechanism of color vision in the goldfish," in *Light and Life* (Baltimore: The Johns Hopkins Press, 1961), p. 795.
- Maxwell, J. C., "On the theory of colours in relation to colour-blindness," Letter of Jan. 4, 1855, to G. Wilson, Suppl. to Wilson, G., *Researches on Colour-Blindness* (Edinburgh: Sutherland-Knox, 1855).
- Maxwell, J. C., "On the theory of compound colours and the relations of the colours of the spectrum," *Proc. Roy. Soc. (London)*, **10**, 404, 484 (1860); *Phil. Mag.* (4), **21**, 141 (1860); *Scientific Papers*, pp. 149, 410 (Cambridge, 1890).
- Moon, P., "Proposed standard solar-radiation curves for engineering use," *J. Franklin Inst.*, **230**, 583 (1940).
- Newhall, S. M., D. Nickerson, and D. B. Judd, "Final report of the O.S.A. subcommittee on the spacing of the Munsell colors," *J. Opt. Soc. Am.*, **33**, 385 (1943).
- Nimeroff, I., and J. A. Yurow, "Degree of metamerism," *J. Opt. Soc. Am.*, **55**, 185 (1965).
- Nuberg, N. D., and E. N. Yustova, "Investigation of color vision of dichromats," *Studies of the State Optical Institute (Moscow)*, vol. 25, Issue 143, p. 33 (1955).
- Pitt, F. H. G., "Characteristics of dichromatic vision," Medical Research Council, Report of the Committee on the Physiology of Vision, XIV; Special Report Series, No. 200 (London, 1935).
- Rushton, W. A. H., "Physical measurement of cone pigment in the living human eye," *Nature*, **179**, 571 (1957).
- Schrödinger, E., "Ueber das Verhältnis der Vierfarben- zur Dreifarben-theorie," *Sitzber. Akad. Wiss. (Wien)*, IIa, **134**, 471 (1925).
- Speranskaya, N. I., "Determination of the trichromatic coordinates of spectrum colors for 27 normal trichromats," *Opt. i Spektroskopiya*, **7**, 710 (1959).
- Stiles, W. S., "Increment thresholds and the mechanisms of colour vision," *Doc. Ophthalmol.*, **3**, 138 (1949).

Stiles, W. S., and J. M. Burch, "N.P.L. colour-matching investigation: Final Report (1958)," *Opt. Acta*, **6**, 1 (1959).

Stiles, W. S., "Color vision: The approach through increment-threshold sensitivity," these PROCEEDINGS, **45**, 100 (1959).

Svaetichin, G., and E. F. MacNichol, Jr., "Retinal mechanisms for chromatic and achromatic vision," *Ann. N. Y. Acad. Sci.*, **74**, 385 (1958).

Thomson, L. C., and W. D. Wright, "The convergence of the tritanopic confusion loci and the derivation of the fundamental response functions," *J. Opt. Soc. Am.*, **43**, 890 (1953).

Troland, L. T., "Report of committee on colorimetry for 1920-21," *J. Opt. Soc. Am. Rev. Sci. Instr.*, **6**, 527 (1922).

Walters, H. V., "Some experiments on the trichromatic theory of vision," *Proc. Roy. Soc. (London)*, **B131**, 27 (1942).

Wassef, E. G. T., "Application of the binocular matching method to the study of the subjective appearance of surface colours," *Opt. Acta*, **2**, 144 (1955).

Willmer, E. N., "Further observations on the properties of the central fovea in colour-blind and normal subjects," *Brit. J. Physiol.*, **110**, 422 (1949).

Wright, W. D., "A re-determination of the mixture curves of the spectrum," *Trans. Opt. Soc.*, **31**, 201 (1929-1930).

Wright, W. D., "The measurement and analysis of colour adaptation phenomena," *Proc. Roy. Soc. (London)*, **B115**, 49 (1934).

Wright, W. D., "The breakdown of a colour match with high intensities of adaptation," *J. Physiol.*, **87**, 23 (1936).

Wright, W. D., "The characteristics of tritanopia," *J. Opt. Soc. Am.*, **42**, 509 (1952).

Young, Thomas, *Lectures on Natural Philosophy* (1807), vol. 1, p. 440.

Yustova, E. N., "Investigation of color vision of dichromats," *Dokl. Akad. Nauk SSSR*, **65**, 5, 661 (1949); *Probl. Physiol. Opt.*, **8**, 112 (1953).