Response Functions for Types of Vision According to the Müller Theory

By Deane B. Judd

According to the Müller theory of vision there are three stages in the visual process, an initial photochemical stage, an intermediate chemical stage relating to the chromatic aspect, and a final stage of excitations of the optic-nerve fibers. By taking advantage of recent precise information regarding the metamers characteristic of normal, protanopic, and deuteranopic vision there have been derived the spectral variations of the responses for each stage as functions of wavelength. These response functions account precisely for the same normal metamers as the ICI standard observer, and closely for the same confusions by color-blind observers as the simpler König theory. Furthermore these functions describe chromatic thresholds of the normal eye (Abney, Priest-Brickwedde) as a gradual approach to tritanopic vision as field size and luminance are decreased.

I. Introduction

In the nineteenth century, two rival theories of vision monopolized most of the interest of investigators. One of these is the Young-Helmholtz three-components formulation; the other is the Hering opponent-colors theory. The discovery of the facts of red-green blindness dealt fatal blows to the then current forms of both of these simple theories, though proponents of the respective theories continued to pump a semblance of life into them with wordy battles. The opponent-colors theory in its original simple form can be made to yield but a single form of red-green blindness, that known as deuteranopia. It must overlook the established fact of a second type of red-green blindness, protanopia, in which the luminosity function is deficient in the long-wave portion of the spectrum and in which the chromaticity confusions are consistently different from those of deuteranopia. The three-components theory explains the confusions made by both types perfectly but in its original simple form has to predict that deuteranopic vision consists of mixtures of red and violet and protanopic vision consists of mixtures of green and violet. When cases of unilateral red-green blindness showed consistently that the perceptions of red-green-blind observers have the hues blue and yellow and no others, the original simple three-components formulation became obsolete. Some advocates of this simple theory took refuge in a suggestion by Fick [1] that red-green confusion is the result, not of the nonfunctioning of either the red or the green receptor system, but rather of the two receptor systems having identical photosensitive substances, either that for red (deuteranopia) or that for green (protanopia). By this suggestion, the responses from the red cones combine with those from the green, regardless of the photosensitive substance in either, to give yellow. This combination can take place in the postretinal portion of the nervous system, as emphasized by Hecht [2], for binocular fusion of colors, and it is permissible to assume that it always takes place there even in binary stimulation of one eye alone. From this view it is only a step to the theoretical position originally proposed by Donders [3] and later espoused by König [4], von Kries [5], and Adams [6] that the three-components formulation holds for processes in one stage of the visual mechanism (perhaps the photosensitive-substance stage), while the opponent-colors theory holds for
processes in a later stage (perhaps the optic nerve). This view may be called the stage or “zone” theory of vision. Furthermore, a very able advocate of the opponent-colors theory, G. E. Müller, adopted a theoretical view [7] that, although divergent in detail and elaborated to include an additional stage, was essentially in agreement with the stage theories favored by Donders, König, von Kries, and Adams.

The dotted curve is based on the Müller theory; see section VII. ⋄, Observed Abney; ○, observed Watson; — — —, \( [0.05(a-x) + (0.9Y)]/a^2 \).

It was shown by Abney [8] in 1910, by Priest and Brickwedde [9] in 1926, by Guild [10] in 1928, by Holmes [11] in 1941, by MacAdam [12] in 1942, and probably by others, that the nearly achromatic color of noon sunlight is more confusible with the greenish yellow color of spectrum light at 570 m\( \mu \) by a normal observer than with any other nearby portion of the spectrum, definitely more than with the yellow portion (575 to 585 m\( \mu \)). Figure 1 shows the brightness in foot lamberts found by Abney to be required to produce the perception of a chromatic color noticeably distinct from the achromatic color of light from the carbon arc. Figure 2 shows the Priest-Brickwedde determination of minimum perceptible colorimetric purity. In both of these figures the maximum near 570 m\( \mu \) is outstanding. Similar results were found by Guild, Holmes, and MacAdam. This outstanding maximum might suggest that pigmentation of the eye media of the normal eye absorbs a large fraction of the short-wave (violet) portion of the spectrum, or for small fields it could mean that because of the chromatic aberration of the eye, the short-wave portion of the sunlight spectrum is out of focus and largely lost. But the most likely explanation is that the normal eye, at least in the fovea, has some of the characteristics of a tritanopic eye; a tritanope has a neutral point in the spectrum near 570 m\( \mu \) where the normal observer has this quasi-neutral point. Furthermore, for very small fields subtending 20’ or less, it has been shown by König [4], Willmer [13], Hartridge [15], and Wright [14] that the fovea is tritanopic.

In an attempt to describe the chromaticity sensibility of the normal observer in terms of an approach to tritanopia, Judd [16] derived a transformation of the OSA “excitations” corresponding to an 80-percent dilution of the violet excitation with red and green. Figure 3 shows the resulting excitation curves and Maxwell triangle, and Figure 4 shows how this formulation corresponds with Priest’s data on minimum perceptible colorimetric purity. This formulation corresponds to a theoretical suggestion similar to Fick’s proposal to account for red-green confusion; it suggests that in the fovea the red and green substance from the red and green cones has leaked into the violet cones to a serious degree (80% leakage). A similar degree of success was demonstrated by Hecht [17] in another development of the Young-Helmholtz theory. Both explanations suffer, however, from a failure to permit an account of dichromatic vision, as do various coordinate systems empirically derived to represent in a simple way the facts of chromaticity sensibility [18, 19, 20].
There are, however, two accounts of chromaticity sensibility that do seem also to permit good explanations of dichromat vision, that by Adams [21] and the recent excellent treatment of chromaticity sensibility by Stiles [22]. An account of protanopia and tritanopia by the Adams theory has not yet been worked out in detail.

An outstanding defect of the three-component accounts of chromaticity sensibility is that there is no satisfactory explanation of the primary character of the spectrum in the neighborhood of 475 mμ. Most normal observers (though not all) see this portion of the spectrum as blue, and they see the short-wave extreme as binary in character, a mixture of red and blue. In commenting on this difficulty, it was remarked by Judd [23] in 1932, “The most satisfactory solution yet offered is Müller’s theory which ascribes primacy to both blue and violet, the latter in the retinal processes, and the former in the optic nerve.”

**Figure 3.** The response functions and Maxwell triangle corresponding to a three-components explanation (Judd) of the data shown in figure 2.

The dotted line represents a part of the data shown in figure 2; the solid line is based on a three-components explanation (Judd, see fig. 3) of the tendency of the normal eye under these conditions to make tritanopic confusions. The three “distribution” curves which give the mixture diagram shown.

As a prerequisite to a quantitative explanation in terms of the Müller theory for the confusibility of sunlight with the spectrum at 570 mμ and for chromaticity sensibility generally, there must be derived the colorimetric coordinate systems corresponding to the two additional stages, the retinal and the optic nerve stages, of the Müller theory. They have so far been described only qualitatively, or at least semiquantitatively. Since the theory has been adjusted to correspond qualitatively with the facts of colorblindness, and since those facts have recently become known quantitatively (chiefly through the work of Pitt [33]), it is now possible to evaluate these coordinate systems, and so lay the ground work for a possible explanation of chromaticity sensibility based on the Müller theory.

**II. Formulation for Normal Vision**

According to the Müller theory, light stimuli can elicit three different primary sensitizing processes (P-processes) in the cone mechanism, whose strengths are determined according to wavelength of the incident radiant energy according to functions similar to those defining the three components of the Young-Helmholtz theory. The P-process is aroused by the spectral region 475

---

²Acknowledgment is made to Michael J. Zigler, Department of Psychology, Wellesley College, who kindly supplied a very helpful translation into English of these parts of Müller’s discussion of color-blindness [7].
m m up to the long-wave visible extreme. The $P_1$-process is aroused by the spectral region between the long-wave end stretch ($770$ m m to a wavelength greater than $655$ m m) and the short-wave end stretch ($380$ m m to a wavelength less than $450$ m m). The $P_2$-process is aroused by the spectral region between $540$ m m and the short-wave visible extreme. From this description, the distribution curves of the $P$-processes are seen to resemble closely the OSA excitation curves [24]. The best modern evaluation of these distribution curves based upon the ICI standard observer [25] is to be obtained [26] by the following transformation:

$$
\begin{align*}
P_1 &= 3.1956X + 2.4478Y - 0.6434Z \\
P_2 &= -2.5455X + 7.0492Y + 0.4963Z \\
P_3 &= 0.0000X + 0.0000Y + 5.0000Z
\end{align*}
$$

which are graphed in figure 5. The reverse transformation [27] is given by

$$
\begin{align*}
X &= 0.24513P_1 - 0.08512P_2 + 0.03999P_3 \\
Y &= 0.08852P_1 + 0.11112P_2 + 0.00036P_3 \\
Z &= 0.00000P_1 + 0.00000P_2 + 0.20000P_3
\end{align*}
$$

Except for the small secondary maximum of the $P_1$ curve in the neighborhood of $430$ m m, these functions conform essentially to the description by Müller of the primary sensitizing processes. These processes contribute immediately to excitation of a black-white "substance" of the optic-nerve fibers in the sense of producing white. They also act upon certain assumed chromatic sensory substances in the cones of the normal eye as follows: the $P_1$-process acts to produce a major yellowish red ($yR$) process in a yellowish red-bluish green ($yR-bG$) substance and a minor greenish yellow ($gY$) process in a greenish yellow-reddish blue ($gY-rB$) substance. The $P_2$-process acts to produce a major bluish green ($bG$) process in the $yR-bG$ substance and a minor greenish yellow ($gY$) process in the $gY-rB$ substance. The $P_3$-process acts to produce a reddish blue ($rB$) process in the $gY-rB$ substance. It is further assumed that the processes within each of these chromatic sensory substances are antagonistic so that a $yR$-process cancels completely a $bG$-process of equal strength; and a $gY$-process may cancel completely an $rB$-process. The wavelengths at which stimulation by homogeneous radiant energy would produce these cancellations are between $560$ and $570$ m m for $yR$ to cancel $bG$, and near $495$ m m for $gY$ to cancel $rB$.

From this description it would seem that the transformation from the amounts of the primary sensitizing processes to the amounts of the chromatic sensory processes might take on the simple form:

$$
\begin{align*}
yR &= -bG = a_1P_1 - a_2P_2 \\
gY &= -rB = b_1P_1 + b_2P_2 - b_3P_3
\end{align*}
$$

where $a_1$, $a_2$, $b_1$, $b_2$, and $b_3$ are constants greater than zero, with $a_1$ greater than $b_1$, and $a_2$ greater than $b_2$.

In addition to the primary sensitizing processes ($P_1$, $P_2$, $P_3$) and the chromatic sensory processes ($yR-bG$, $gY-rB$), there are six different excitations of the optic nerve ($w$, $s$, $r$, $g$, $y$, $b$), which correlate with the introspectively pure white, black, red, green, yellow, and blue sensations, respectively. The chromatic excitations are assumed to arise from the chromatic sensory processes alone, the $yR$-process arousing a major $r$-excitation and a minor $y$-excitation as indicated by the notation $yR$. Similar major and minor excitations are aroused by the $bG$, $gY$, and $rB$ processes. The white-excitation of the optic nerve comes from the immediate effect of the primary sensitizing processes ($P_1$, $P_2$, $P_3$) to which secondary contributions from the $yR$- and $gY$-processes are added. The black-excitation comes chiefly by induction [7a, p. 85], from a white
surrounding field or from a white preexposure field, but a secondary contribution comes from the $bG$ and $rB$ processes. Like the chromatic sensory processes, the four chromatic excitations of the optic nerve make antagonistic pairs, an amount of $r$-excitation cancelling a like amount of $g$-excitation, and the same cancellations for $y$- and $b$-excitation. For stimulation by homogeneous radiant energy, the one cancellation occurs at the wavelength for arousing unitary yellow, which is probably between 575 and 582 m,u under usual observing conditions for an average normal observer. A second $(r,g)$-cancellation occurs at the wavelength for arousing unitary blue, and a $(y,b)$-cancellation occurs at the wavelength for unitary green.

For self-luminous areas with a neutral surrounding field the $s$-excitation acts as negative $w$-excitation; however, they do not cancel, but combine to give gray. From this description it would appear that the excitations of the optic nerve could be found for the normal observer by the following transformations:

$$
\begin{align*}
    r &= -g = c_1 yR + c_2 rB = -c_1 bG - c_2 gY \\
    y &= -b = d_1 gY + d_2 yR = -d_1 rB - d_2 bG \\
    w &= e_1 P_1 + e_2 P_2 + e_3 P_3 + e_4 gY + e_5 bG \\
    s &= e_4 bG + e_5 rB
\end{align*}
$$

where the luminance of the area is given by the difference, $w - s$, between the white and black excitation, and the symbols with subscripts represent constants evaluated so far only by the conditions that $c_1$ is greater than $d_2$, and $d_1$ is greater than $c_2$.

Equation 3 is similar to those set up by Schrödinger [28] in accord with the theoretical views of von Kries [5], and by Adams [21] in accord with his own theory [6]. Adams has, moreover, pointed out the advantages and theoretical plausibility of the view that the various stages are not linearly connected. For simplicity in the present derivation, attention will be confined to the assumption represented in eq 2 and 3 that the connection is linear and homogeneous.

**III. Dichromatic Vision**

According to the Müller theory, protanopia corresponds to the failure of the $(yR, bG)$-chromatic substance; and on this account, it is called by him outer red-green blindness. Thus, for protanopia, $yR = bG = 0$, and we may write from eq 3:

$$
\begin{align*}
    r_p &= -g_p = c_2 rB = -c_2 gY \\
    y_p &= -b_p = d_2 gY = -d_2 rB \\
    w_p &= e_1 P_1 + e_2 P_2 + e_3 P_3 + e_4 gY \\
    s_p &= e_2 rB
\end{align*}
$$

From eq 4 one might think that the Müller theory predicts for protanopia simply the sensations of black and white plus the two chromatic sensations greenish yellow and reddish blue, since a given amount of blue excitation is always bound up intrinsically with the same minor red excitation. This is indeed the simplest prediction from the formulation and corresponds fairly well with reports of unilaterally protanopic observers. Müller, however, points out that although failure of the $(yR, bG)$ chromatic process is sufficient to produce the symptoms of protanopia completely, this failure could be accompanied by a failure of some of the chromatic processes in the optic nerve, which combination of circumstances could give rise to an observer having protanopic vision by all tests actually sensing only the hues yellow and blue, or even only the hues red and green. Such observers could be distinguished from each other only if one eye had trichromatic, and the other, protanopic vision.

Deuteranopia, on the other hand, is ascribed by Müller to failure of the $(r,g)$-sense of the optic nerve and is called inner red-green blindness. Thus, for deuteranopia, $r_d = -g_d = 0$, and we may write from eq 3:

$$
\begin{align*}
    y_d &= -b_d = d_1 gY + d_2 yR \\
    w_d &= e_1 P_1 + e_2 P_2 + e_3 P_3 + e_4 gY \\
    s_d &= e_1 bG + e_2 rB
\end{align*}
$$

From eq 5 it is plain that the predicted sensations of deuteranopes must be black, white, yellow, and blue; there are no alternatives.

Tritanopia, like protanopia, is ascribed to a retinal defect. It corresponds to failure of the $(gY, rB)$-chromatic substance and is called outer yellow-blue blindness. Thus, for tritanopia, $gY = rB = 0$, and we may write from eq 3:
\[ \begin{align*}
  r_i &= -g_i = c_y R = -c_y b G, \\
  y_i &= -b_i = d_y y R = -d_y b G, \\
  w_i &= e_i p_1 + e_2 p_2 + e_3 p_3 + e_4 y R, \\
  e_i &= e_i b G
\end{align*} \] (6)

The sensations of tritanopes are seen to be predicted as black, white, and either yellow and blue, or red and green, or some fixed combination such as yellowish red and bluish green. The latter hues correspond to the simplest prediction, since the \((yR, bG)\) chromatic sensory process is unaffected. These hues agree well with the reports of tritanopes who have acquired the defect through a disease of the retina.

**IV. Evaluation of the Constants**

Response functions for normal and dichromatic vision according to the Müller theory can be evaluated from eq 1, 2, and 3, for all three stages of excitation, provided the 14 constants of eq 2 and 3 be evaluated. For the normal mechanism adapted to a stimulus yielding an achromatic color, the stimuli for the unitary hues, red, yellow, green, and blue, are known \([29, 30]\) within limits; these stimuli must excite only the respective \(r\), \(y\), \(g\), and \(b\)-processes of the optic nerve. The stimulus yielding the achromatic color, itself, must cause the chromatic sensory processes \(yR\), \(gY\), \(bG\), and \(rB\) to vanish, and also reduce to zero the chromatic \(r\), \(y\), \(g\), and \(b\)-processes of the optic nerve. The colors confused with gray by the typical protanope, deuteranope, and tritanope must conform to the \(yR\)-process, the \(g\)-process, the \(rB\)-process, and their complements, respectively. The difference between the spectral luminosity function of the typical protanope and the same function for the normal observer must be a constant fraction of the \((yR, bG)\)-process in order to conform to Müller’s proposal; and similarly the difference between normal and tritanopic luminosity must be a constant fraction of the \((gY, rB)\)-process.

There are many more than 14 conditions to be satisfied, including several that are set down by Müller only in qualitative terms, and some that relate to sensibility to chromaticity differences. In evaluating these constants, the best determined conditions (marked by asterisks in the next sections) have been satisfied perfectly; and from the resulting excitation curves it may be seen to what extent the less well-determined conditions are satisfied. For example, of the data for the normal stimuli for the unitary hues, only those for unitary yellow have been used, since they were specifically mentioned by Müller as indicating the stimulus to be between 575 and 582 \(\mu\). Unitary blue has necessarily to be taken as the complement of unitary yellow relative to the stimulus for white or gray; and unitary red and green were taken as the confusion colors for typical deuteranopia.

**1. Hueless Point and Unitary Yellow**

For an achromatic color, both the chromatic processes of the optic nerve and the chromatic sensory processes must cancel to zero. It is known \([31]\) that stimulation of the normal eye by a source of equal energy results under ordinary conditions of observation in a closely achromatic, or hueless color. For such a source, \(X = Y = Z\), and from eq 1, \(P_1 = P_2 = P_3\); hence we may write with sufficient accuracy from eq 2

\[ gY = b_1 P_1 + b_2 P_2 - b_3 P_3 = 0, \]

whence:

\[ \frac{a_1}{a_2} = 1.000; \] (7)*

and similarly

\[ yR = a_1 P_1 - a_2 P_2 = 0, \]

whence we find

\[ b_1 + b_2 = b_3. \] (8)*

By setting \(r = g = y = b = 0\) in eq 3, and substituting eq 7 and 8, the expressions vanish, and no further relation is found.

For the \(r-g\) cancellation point between 575 and 582 \(\mu\), we may take somewhat arbitrarily for simplicity the crossing point of \(X\) and \(Y\) at 578.1 \(\mu\). For stimulation by homogeneous energy of this wavelength we may write

\[ r = -g = c_y R + c_x R B = 0 \]

\[ = c_1 (a_1 P_1 - a_2 P_2) - c_2 (b_1 P_1 + b_2 P_2 - b_3 P_3) = 0 \]

\[ = (a_1 c_1 - b_1 c_2) (3.1956 X + 2.4478 Y - 0.6434 Z) - \]

\[ (a_2 c_1 + b_2 c_2) (-2.5455 X + 7.0492 Y + 0.4963 Z) + \]

\[ b_3 c_1 (5.0000 Z) = 0 \]

By substituting \(X = Y = 0.4996, Z = 0.0008\), which refer closely to 578.1 \(\mu\), we obtain
2. Dichromatic Copunctal Points

The chromaticity confusions of dichromatic vision may be represented on the Maxwell triangle by families of straight lines, all lines of one family intersecting at a common point, known as the copunctal point [32]. These copunctal points have been evaluated from recent determinations, chiefly by Fitt [33], and expressed in terms of the ICI standard coordinate system. They embody the essential information regarding the chromaticity confusions of dichromats and lead to a convenient expression of four conditions affecting the unknown constants.

According to the Müller theory, protanopia corresponds to the failure of the \((gY, bG)\)-chromatic substance. For certain stimuli the \((gY/rB)\)-process is also reduced to zero. These stimuli are the equal-energy stimulus and all of those confused with it by the protanope. These chromaticity confusions are indicated on the Maxwell triangle by a straight line passing through the equal-energy point, and for every point on this line, \(gY = -rB = 0\). In particular, since all of the chromaticity confusion lines pass through a single point, these conditions must hold for the protanopic copunctal point defined by \(x = X/(X+Y+Z) = 0.747\), \(y = Y/(X+Y+Z) = 0.253\), \(z = Z/(X+Y+Z) = 0\). Hence we may write from eq 2

\[
\begin{align*}
&b_1P_1 + b_2P_2 - b_3P_3 = \\
&b_1(3.1956X + 2.4478Y - 0.6434Z) + \\
&b_2(-2.5455X + 7.0492Y + 0.4963Z) - \\
&b_3(5.0000Z) = 0;
\end{align*}
\]

whence we find

\[
b_1/b_2 = 0.0393. \tag{10}\]

Similarly, from the less well-determined tritanopic copunctal point, \(x = 0.18\), \(y = 0.00\), \(z = 0.82\), we may write

\[
yR = -bG = a_1P_1 - a_2P_2 = 0, \\
\]

\[
= a_1(3.1956X + 2.4478Y - 0.6434Z) - \\
\]

\[
a_2(-2.5455X + 7.0492Y + 0.4963Z) = 0; \
\]

whence we find \(a_1/a_2 = -0.0512/0.0476 = -1.07\), a value contradictory to eq 7 based upon the achromatic stimulus for the normal observer, which is much more reliably established than the tritanopic copunctal point. By setting \(a_1 = a_2\) in conformity with eq 7, we find that a condition for the tritanopic copunctal point by the Müller formulation is that \(P_1 = P_2\), which from eq 1 is equivalent to \(y = 1.987x - 0.327\). This line intersects the \(x\)-axis of the chromaticity diagram at \(x = 0.165\), \(y = 0.000\), which is in as good agreement with the facts as the approximate evaluation of the copunctal point \((x = 0.18, y = 0.00)\) estimated with the help of the König theory [32]. The latter accords with Müller’s view that tritanopia is characterized by a single neutral point in the spectrum (near 570 m\(\mu\)); the former places a second neutral point near the violet extreme of the spectrum (430 m\(\mu\)) and arises from application of the Müller theory to the properties of the standard observer. As already noted, \(P_1\) evaluated by this means has a small secondary maximum in the neighborhood of 430 m\(\mu\), causing it to cross the \(P_2\)-curve at this point, a result quite unanticipated by Müller. Actual reports of tritanopic vision are fairly well divided in this respect; for example, the cases reported by König [4] and Köllner [34] yielded a single neutral point; those by Collin and Nagel [35] and Piper [36] had neutral points or areas in the neighborhood of 430 m\(\mu\). Willmer and Wright [14] found an indication of such a neutral region for small fields in the normal fovea, and Pitt [37] considers this to be typical of tritanopia. The difference in chromaticity between 430 m\(\mu\) and the short-wave end of the spectrum is small, and it is possible that individual variations among tritanopes can account for the slight discrepancy in the report between no short-wave neutral point and one near 430 m\(\mu\). On the other hand, it seems to be fairly frequent that tritanopes have ocular media pigmented heavily with brown pigment, and it is also possible that this pigmentation would cause the spectrum to become invisible to many tritanopes at a wavelength greater than 430 m\(\mu\) as in a case reported by Farnsworth [38]. The formulation could be made to accord strictly with the Müller view on this point by choosing constants in eq 1, so that the representation of \(P_2\) is everywhere higher than \(P_1\) for wavelengths less than 500 m\(\mu\), such as that found by Stiles [22], but such a wavelength dis-
tribution of $P_2$ is itself contrary to Müller's view. It seems more useful to proceed with an account of tritanopia involving a second neutral point near 430 m.

Take therefore $a_1 = a_2$ as in eq 7.

By the same argument given for protanopia, we may set $y = -b = 0$, for the deuteranopic copunctal point defined by $Y = Z = 0$. For eq 1, 2, and 3 we may write

$$y = -b = d_1 g Y + d_2 y R = d_1 (b_1 P_1 + b_2 P_2 - b_3 P_3) + d_2 (a_1 P_1 - a_2 P_2) = 0,$$

$$= (b_1 d_1 + a_1 d_2) (3.1956X + 2.4478Y - 0.6434Z) + (b_2 d_1 - a_2 d_2) (-2.5455X + 7.0492Y + 4.963Z) - b_3 d_1 (5.0000Z);$$

whence we find, since $Y = Z = 0$,

$$3.1956 (b_1 d_1 + a_1 d_2) - 2.5455 (b_2 d_1 - a_2 d_2) = 0.$$

(11)*

Two other conditions may be derived from the protanopic and deuteranopic copunctal points. It has been shown [32] that the deuteranopic copunctal point lies on the deuteranopic achlyne, that is, the line on the Maxwell triangle associated with zero deuteranopic luminosity; and the protanopic copunctal point lies on the protanopic achlyne. Hence for the respective copunctal points we may set $(w - s)_d$ and $(w - s)_p$ equal to zero, and from eq 1, 2, and 3 there are found:

$$e_1 + 0.0393 e_2 + 2 e_3 (b_1 - 0.0393 b_2) = 0,$$

and

$$3.1956 (e_1 + 2 e_4 a_1 + 2 e_5 b_1) - 2.5455 (e_2 - 2 e_4 a_2 + 2 e_5 b_2) = 0.$$

(13)*

3. Dichromatic Luminosity Functions

It has been shown that the luminosity functions of red-green-blind observers can be expressed as functions of $X$, $Y$, and $Z$ [39] by eq 14 and 15, in which $W_p$ is the protanopic luminosity and $W_d$, the deuteranopic

$$W_p = 0.460 X + 1.359 Y + 0.101 Z,$$

$$W_d = Y.$$

From eq 1a, these functions may be written in terms of the primary processes, $P_1$, $P_2$, and $P_3$:

$$W_p = 0.00754 P_1 + 0.19017 P_2 + 0.00229 P_3,$$

(14a)

$$W_d = 0.08852 P_1 + 0.11112 P_2 + 0.00036 P_3,$$

(15a)

But from eq 2, 4, and 5, we may write

$$W_p = (w - s)_p = (e_1 + 2 a_1 e_4 + 2 b_1 e_5) P_1 + (e_2 - 2 a_2 e_4 + 2 b_2 e_5) P_2 + (e_3 - 2 b_3 e_5) P_3,$$

(14b)

$$W_d = (w - s)_d = (e_1 + 2 a_1 e_4 + 2 b_1 e_5) P_1 + (e_2 - 2 a_2 e_4 + 2 b_2 e_5) P_2 + (e_3 - 2 b_3 e_5) P_3.$$

(15b)

By equating the coefficients of $P_1$, $P_2$, and $P_3$ in eq 14a and 14b, we obtain three additional conditions to be satisfied by the constants:

$$e_1 + 2 b_1 e_5 = 0.00754,$$

(16)

$$e_2 + 2 b_2 e_5 = 0.19017,$$

(17)

$$e_3 - 2 b_3 e_5 = 0.00229.$$

(18)

Similarly, by equating the coefficients in eq 15a and 15b we obtain three more conditions:

$$e_1 + 2 a_1 e_4 + 2 b_1 e_5 = 0.08852,$$

(19)

$$e_2 - 2 a_2 e_4 + 2 b_2 e_5 = 0.11112,$$

(20)*

$$e_3 - 2 b_3 e_5 = 0.00036.$$

(21)*

Equations 16 to 21 are not entirely independent of the conditions previously found, nor are they all congruent. From eq 16 and 17 there may be derived eq 12; and from eq 19 and 20 there may be derived eq 13. Furthermore, eq 7, 10, 12, 19, and 20 combine to give eq 16 and 17. Equations 18 and 21 are contradictory. Since it is an essential part of the Müller theory that the deuteranopic luminosity function be the same as the normal, we must accept eq 21 and reject eq 18. This choice will prevent $(w - s)_p$ in eq 14b from being exactly equal to $W_p$ in eq 14. It remains to be found whether the resulting evaluation of $(w - s)_p$ is as representative of available data on the protanopic luminosity function as is $W_p$. Thus, we have obtained only two additional independent conditions from protanopic and deuteranopic luminosity functions, eq 20 and 21.

If it be assumed for the moment that tritanopic luminosity is the same as deuteranopic and nor-
nal luminosity, as it may well be judging from the available information [7a, pp 53 and 102; 34, 36, 40, 41, 42, 43], then from eq 2 and 6 we obtain eq 22:

\[(w - s) = (w - s)_1 = e_1 P_1 + e_2 P_2 + e_3 P_3 + 2e_4 (a_1 P_1 - a_2 P_2). \quad (22)\]

By comparing the coefficients of \(P_1, P_2,\) and \(P_3\) in eq 15b and 22 we see from each of the three comparisons that \(e_5\) must be zero; that is, there can be no darkening effect from the \(gY\)-process, such as implied by eq 3. The Müller theory thus cannot abide having equality between tritanopic and normal luminosity. Müller was well aware that his theory required the tritanopic luminosity function to be different from normal and remarks [7a, p. 63] “In regard to spectral luminosity distribution in tritanopia, there must be, if no complications exist, because of the absence of the \(w^\prime\) value of the \(gY\)-process, a decrease in the luminosity of yellow and yellowish lights in comparison to normal. ... Unfortunately there have been up to now no investigations of the spectral luminosity characteristic of tritanopia.” Since the present purpose is to find the coordinate systems implied by the Müller theory, we must disregard the rather inconclusive indications that there is no difference between tritanopic and normal luminosity; hence no attention can be paid to eq 22 in evaluating the constants, and \(e_5\) must be given a positive, though small value. Take arbitrarily, then:

\[e_5 = 0.03 e_4. \quad (23)^*\]

### 4. Chromaticity Sensibility and Theory

We may now take stock of the conditions that must be satisfied by the 14 constants:

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Resulting condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hueless point</td>
<td>Eq 7 and 8</td>
</tr>
<tr>
<td>Unitary yellow</td>
<td>Eq 9</td>
</tr>
<tr>
<td>Protagonopic copunctal point</td>
<td>Eq 10 and 12</td>
</tr>
<tr>
<td>Deuteranopic copunctal point</td>
<td>Eq 11 and 13</td>
</tr>
<tr>
<td>Dichromatic luminosity</td>
<td>Eq 20, 21, and 23.</td>
</tr>
</tbody>
</table>

These 10 equations have been marked with asterisks to show that they were used in the derivation of the constants.

**Müller Theory of Vision**

There remain four conditions to be set up before the 14 constants can be evaluated. Three of the four conditions refer to the relative sizes of \(a_1\) and \(b_2, c_1\) and \(d_1,\) and \(a_1\) and \(c_1.\)

The first ratio, \(a_1/b_2,\) has to do with the relative sensibility of the eye to yellowish red-bluish green differences on the one hand and greenish yellow-reddish blue differences on the other. Empirical studies on large fields [20, 21] indicate that \(a_1/b_2\) is about 2.5, that is, the normal eye detects \((yR, bG)\)-differences more readily by about a factor of 2.5 than would be judged from the relatively great overlap of \(P_1\) and \(P_2.\) By setting \(e_1/d_1 = 1.0,\) this greater sensibility to \((yR, bG)\)-differences is preserved in the optic-nerve stage.

The ratio of \(a_1/c_1\) has to do with the comparative amounts of the chromatic sensory process and the chromatic excitations of the optic nerve. There does not seem to be any fundamental meaning to this comparison. It has merely to do with a relation between the units expressing the rate of a chemical process in the retinal receptors and those expressing the frequency of the resulting impulses in the fibers of the optic nerve. This ratio may be set arbitrarily, and for simplicity we set \(a_1/c_1 = 1.0.\)

The final condition refers to the size of the arbitrary units in which the chromatic responses are to be expressed; for simplicity take \(a_1 = 1.0.\)

Solution of these 14 equations simultaneously gives the values of the constants:

\[
\begin{align*}
  a_1 &= 1.0000, a_2 = 1.0000, \\
  b_1 &= 0.0151, b_2 = 0.3849, b_3 = 0.40000, \\
  c_1 &= 1.0000, c_2 = 0.6265, \\
  d_1 &= 1.0000, d_2 = 0.1622, \\
  e_1 &= 0.0075, e_2 = 0.1912, e_3 = 0.0013, e_4 = 0.0405, \\
  e_5 &= 0.0012
\end{align*}
\]

It will be noted that, as required by eq 2, \(a_1\) is greater than \(b_1,\) and \(a_2\) is greater than \(b_2.\) Furthermore, as required by eq 3, \(c_1\) is greater than \(d_2,\) and \(d_1\) is greater than \(c_2.\) This correspondence with the Müller description indicates how thorough was his grasp of the facts from purely qualitative data, though probably \(c_2 = 0.6265\) is not as small compared to \(d_1 = 1.0000\) as would be expected from Müller’s designation of \(d_1 gY\) and


c_2gY as a major y-excitation and a minor g-excitation, respectively, resulting from the gY process. The value of c_2 would be reduced somewhat by taking a higher value for the wavelength of the spectrum stimulus for unitary yellow, say 582 m\(\mu\) instead of 578 m\(\mu\); see eq 9.

V. Definition of the Coordinate Systems

We may now insert these constants into eq 2 and 3, and so give explicit definitions of the two new coordinate systems implied in the Müller theory. The coordinate system applying to the chromatic sensory processes of the retinal receptors is defined by eq 2a:

\[
\begin{aligned}
yR &= -bG = P_1 - P_2, \\
gY &= -rB = 0.0151P_1 + 0.3849P_2 - 0.4000P_3.
\end{aligned}
\]

The coordinate system applying to the processes in the optic nerve fibers is defined by eq 3a:

\[
\begin{aligned}
r &= -g = yR + 0.6265rB = -bG - 0.6265gY, \\
y &= -b = gY + 0.1622yR = -rB - 0.1622bG, \\
w &= 0.0075P_1 + 0.1912P_2 + 0.0013P_3 + 0.0405 \\
yY &= 0.0012gY, \\
s &= 0.0405bG + 0.0012rB.
\end{aligned}
\]

These two coordinate systems may also be defined in terms of the standard 1931 ICI coordinate system for colorimetry from eq 1. Equation 2b gives the definition of the colorimetric coordinate system for normal observers corresponding to the luminosity function \(Y\); and eq 2c gives the reverse transformation from this coordinate system to the standard ICI system:

\[
\begin{aligned}
yR &= -bG = 5.741X - 4.601Y - 1.140Z, \\
gY &= -rB = -0.932X + 2.750Y - 1.819Z, \\
Y &= 1.000Y, \\
X &= 0.1581yR - 0.0991gY + Y, \\
Y &= Y, \\
Z &= -0.0810gR - 0.4991gY + Y.
\end{aligned}
\]

Equation 3b gives the definition of the colorimetric coordinate system for normal observers corresponding to the excitations of the optic nerve, and eq 3c gives the reverse transformation from this coordinate system to the standard ICI system:

\[
\begin{aligned}
r &= -g = 6.325X - 6.325Y, \\
y &= -b = 2.004Y - 2.004Z, \\
(\omega-s) &= 1.000Y, \\
X &= 0.1581r + (\omega-s), \\
Y &= (\omega-s), \\
Z &= -0.4991y + (\omega-s).
\end{aligned}
\]

Table 1 gives the response functions of the normal, protanopic, deuteronopic, and tritanopic types of vision derived from the ICI standard observer according to the Müller theory (eq 1, 2b, 3b, 4, 5, and 6).

The very simple transformation equations between the chromatic excitations of the optic nerve according to the Müller theory and the standard 1931 ICI coordinate system for colorimetry arise, of course, from the fact that the \(X\)-primary of the ICI system corresponds to a stimulus for unitary red, and the \(Z\)-primary corresponds to unitary blue. The \(r,y\) \((\omega-s)\) system corresponds to the central stage of the Adams theory [1], and coordinate systems closely resembling that described by eq 3b have been used by Adams with considerable success to explain chromaticity spacing for large fields, chiefly studies of the spacing of the Munsell colors [1].

The coordinate system set up by Schrödinger [44] in 1925 resembles closely that defined by eq 3b except that it was not adjusted to correspond to the same balance between \((y,b)\)-excitation and the \((r,g)\)-excitation \((a_1/b_3=2.5, c_1/d_1=1.0)\). Schouten [45] made use of conditions derived from the hueless point (eqs 7 and 8), the deuteronopic neutral point (eq 11) and the unitary yellow point (eq 9) to compute response functions for assumed central \(r\), \(y\), \(g\), and \(b\)-processes. These functions bear a considerable resemblance to \(r\) and \(y\) evaluated from eq 3b. Thus it is seen that the essence of eq 3b is neither new nor confined to the Müller theory; it arises from the opponent-colors theory of Hering and has been used in at least three theoretical studies since 1925.

Figure 6 shows as functions of wavelength \(P_1, P_2, P_3; yR\) and \(gY; r\) and \(y\); and finally in the
lower left quadrant the deuteranopic luminosity according to eq 3b, the protanopic luminosity according to eq 4 together with the luminosity contributions of the chromatic sensory processes $gR$ and $gY$.

VI. Protanopic Luminosity Function

In the ICI system the standard luminosity function is represented by the second function, $Y$; and from eq 3b it may be seen that the Müller

---

Table 1. Response functions of the normal, protanopic, deuteranopic, and tritanopic types of vision derived from the ICI standard observer according to the Müller theory

<table>
<thead>
<tr>
<th>Wave length</th>
<th>Primary sensitizing processes (possessed by all types of vision)</th>
<th>Chromatic sensory processes</th>
<th>Optic-nerve excitation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal vision Protanopic vision Deuteranopic vision Tritanopic vision</td>
<td>$gR + gG$</td>
<td>$gR + gG$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$gR + gG$</td>
<td>$(gR + gG)_s$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$r - q$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>380</td>
<td>32</td>
<td>-13</td>
<td>+8</td>
</tr>
<tr>
<td>390</td>
<td>100</td>
<td>+1</td>
<td>-40</td>
</tr>
<tr>
<td>400</td>
<td>349</td>
<td>+3</td>
<td>-136</td>
</tr>
<tr>
<td>410</td>
<td>1,037</td>
<td>+8</td>
<td>-414</td>
</tr>
<tr>
<td>420</td>
<td>3,228</td>
<td>+17</td>
<td>-1,288</td>
</tr>
<tr>
<td>430</td>
<td>6,928</td>
<td>-3</td>
<td>-2,752</td>
</tr>
<tr>
<td>440</td>
<td>8,736</td>
<td>-97</td>
<td>-3,439</td>
</tr>
</tbody>
</table>

---

*a For protanopic and tritanopic vision the Müller theory does not state rigidly that the optic nerve excitation must follow eq 4 and 6, though this is the simplest prediction. Either $r - q$ or $p - b$, but not both, may be zero. If they are not zero, they have wavelength distributions proportional to the chromatic sensory processes.
theory can be formulated, as he claimed, in such a way that the difference between the \(w\)-excitation and the \(s\)-excitation gives the normal luminosity function. This function has already been shown to be as satisfactory a representation of deuteranopic luminosity as it is for some normal luminosity functions, because the deuteranopic luminosity functions fall within normal limits [32]. In these two respects this formulation of the Müller theory conforms exactly to that previously worked out in accord with the König theory [4, 32]. However, it was noted previously that eq 21 contradicts eq 18; so it remains to be seen whether the prediction of protanopic luminosity by this formulation of the Müller theory is as acceptable as that by the König theory. By inserting the constants in eq 4 it is found that this formulation of the Müller theory requires protanopic luminosity to be given by:

\[
(w-s)_p = 0.0075 P_1 + 0.1921 P_2 + 0.0003 P_3. \tag{4a}
\]

The previous formulation of the König theory yielded the equation:

\[
W_p = 0.0075 P_1 + 0.1902 P_2 + 0.0023 P_3. \tag{14a}
\]

Figure 7 is a plot of these functions adjusted approximately to unit maximum, together with upper and lower limits of available data on luminosity functions of protanopic and protanom-
al observers [32]. It will be seen that these data support both functions about equally well.

VII. Chromatic Thresholds, Normal and Tritanopic

We are now in position to inquire whether the Müller theory offers a basis for explaining the chromatic-threshold data of Abney [8] and Priest Brickwedde [9] referred to earlier. According to the Müller theory the ability of an observer to detect a slight variation in chromaticity from a central chromaticity, such as that of the light from a carbon arc, would depend upon the excitation of the chromatic sensory processes. The amounts of the excitation of these processes corresponding to any color specified in terms \((X, Y, Z)\) of the 1931 ICI standard observer can be found from eq 2b. And, in particular, they have been found for the spectrum colors for unit irradiance and are plotted in the upper right quadrant of figure 6. Both of the above sets of data are given, however, in luminous units (luminance of the field just yielding a chromatic difference from carbon-arc light in Abney’s work, or luminance fraction required to be mixed with sunlight to produce a color just noticeably different from sunlight in the work of Priest and Brickwedde). We should expect to compare with them, therefore, the excitations of the chromatic sensory processes corresponding to the colors of a spectrum of constant luminance; that is, we should expect to find the chromatic thresholds in luminance terms to correspond to the reciprocal of some combination of \(yR/(w-s)\) and \(gY/(w-s)\). The exact form of combination would seem to be expressible in terms of the probability of a chromaticity difference being discriminated as a function of the probabilities of each of the two independent chromatic processes becoming effective considered separately. For the present purpose it is sufficient to take tentatively the combination as the square root of the sum of the squares; that is, assume, for the moment, that the effective chromatic excitation for large fields and high luminance is proportional to: 

\[
[(yR)^2 + (gY)^2]^{1/2}/(w-s)
\]

For experimental conditions, such as restricted angular size of field or low luminance, that make the normal eye respond more or less like a tritanopic eye, the effective chromatic excitation may be assumed to be proportional to 

\[
[(yR)^2 + f^2 (gY)^2]^{1/2}/(w-s),
\]

where \(f\) is the relative effectiveness of the \(gY-\beta B\) process compared to the \(yR-\beta G\) process. In general, we would compare to the chromatic thresholds \(dB/dE\), expressed in luminous terms, the reciprocals of these assumed effective chromatic excitations so as to study the validity of the relation:

\[
dB/dE = k(w-s)/[(yR)^2 + f^2 (gY)^2]^{1/2},
\]

where \(k\) is the constant required to adjust the theoretical function to the units in which the chromatic threshold is expressed.

Abney’s data have been found to agree fairly well with eq 24 for \(k=0.0020\) and \(f=0.04\); see dotted curve of figure 1. The course of the experimentally determined function is followed well, except for wavelengths greater than 590 m\(\mu\) where the predicted threshold is considerably lower than that found experimentally. As far as is known, no explanation of these data has previously been suggested. It should be pointed out also that more recent determinations of the chromatic threshold by Purdy [46], and Otero, Plaza, and Casero [47] are quite at variance with these data, and indeed with each other. They show neither the sharp peak at 570 m\(\mu\) nor the decline to small values near 450 m\(\mu\). Needless to say, they are quite unexplainable by the Müller theory. The data by Abney and Watson, how-

Muller Theory of Vision
ever, may be summarized by saying that they conform fairly well to the Müller theory for a retinal region in which the normal \( gY, rB \) process is 96 percent ineffective. Tritanopia corresponds to complete ineffectiveness of this process.

The data of Priest and Brickwedde have been found to agree well with eq 24 for \( k=0.05 \), and \( f=0.5 \); see dotted curve on figure 2. The degree of agreement is quite comparable to that obtained by a coordinate system adjusted empirically to represent such data [18]; see solid curve. In this case, the less complete data by Purdy are in substantial agreement and are also shown. It should be pointed out, however, that these data have been corrected to refer to the standard luminosity function by multiplying them by the ratio of the standard luminosity function to that found by Gibson and Tyndall. It is probable that an improvement in the theoretical account of other psychophysical data by means of the Müller theory would result from revaluation in terms of an observer based on the Gibson-Tyndall experimental mean [18] luminosity function instead of on the standard observer. However, we may say that the Priest-Brickwedde data correspond well to the Müller theory for a retinal region in which the normal \( gY-rB \) process is 50 percent effective.

It is concluded that the Müller theory affords a good explanation of chromatic thresholds in terms of a gradual approach to tritanopic vision. A thorough study of the implications of the Müller theory for chromaticity sensibility of all types such as that carried out by Stiles [22] for the three-components theory would seem to be worth while.

**VIII. Summary and Conclusion**

By taking into account the metamers known to be characteristic of protanopic, deuteranopic, and normal vision as well as data on the stimulus for a neutral color and the stimulus for a color of unitary yellow hue, the spectral variations of the responses for each of the three stages of the Müller theory of vision have been evaluated as functions of wavelength.

These response functions are shown to yield an account of normal, protanopic, and deuteranopic vision that differs in no essential respect from the simpler explanation yielded by the König form of three-components theory. They differ in their explanation of tritanopic vision by requiring the tritanopic luminosity function to be slightly higher in the short-wave end of the spectrum than normal; the three-components explanation requires it to be slightly lower in this part of the spectrum.

The chromatic response functions of the second stage of the Müller theory are shown to lead to a satisfactory and convenient account of the approach to tritanopia exhibited by the normal eye in viewing small fields or fields of luminance near the chromatic threshold.

It is concluded that the qualitative ideas of Müller lead to admissible and consistent coordinate systems. The Müller theory shows how the three-components formulation of Young, Helmholtz, and König (first stage) and the opponent-colors formulation of Hering (third stage) may both be accepted, and the explaining power of both be simultaneously utilized. The intermediate stage is also a promising and powerful theoretical tool. The quantitative consistency of the Müller ideas and the success demonstrated in accounting for tritanopic confusions made by normal observers does not, of course, prove the Müller theory to be completely, or even basically, correct. Alternate explanations are possible. There are important gaps in our knowledge of retinal chemistry and conduction and integration of nerve impulses that, if filled, might disprove the Müller theory and require adoption of an alternate account. Furthermore, several aspects of the Müller explanation, though admissible in the present state of our knowledge, seem implausible and unlikely to be born out by future work. At the very least, however, the Müller theory must be viewed as a forward step, and the coordinate system suggested by the second stage has practical value regardless of any of these future theoretical developments.

**IX. References**

E. G. Muller, Theory of color vision, Psychol. Rev. 30, 56 (1923).

G. E. Müller, Darstellung und Erklärung der verschiedenen Typen der Farbenblindheit (Vandenhoeck-Ruprecht, Göttingen, 1924); G. E. Müller, Uber die Farbenempfindungen, Z. Psychol. Ergänzungsbände 17 and 18 (1930).


E. N. Willmer, Colour of small objects, Nature 153, 774 (June 24, 1944).


D. B. Judd, Sensibility to color change determined from the visual response functions; extension to complete and partial dichromasy, J. Opt. Soc. Am. 16, 115 (1928); Reduction of data on mixture of color stimuli, BS J. Research 4, 515 (1930) RP163.

Seig Hecht, A quantitative formulation of colour vision, Report of a joint discussion on vision, physical and optical societies (Univ. Press, Cambridge, June 1932).


D. B. Judd, Reduction of data on mixture of color stimuli, BS J. Research 4, 515 (1930) RP163.


H. Köllner, Die Störungen des Farbensinnes, p. 48 (Karger, Berlin, 1912).

Collin and A. W. Nagel, Erworbene Tritanopia (Violettblindheit), Z. Sinnesphysiol. 41, 74 (1907).

H. Piper, Beobachtungen an einem Fall von totaler Farbenblindheit des Netzhauteinrums im einen und von Violettblindheit des anderen Auges, Z. Psychol. 38, 155 (1905).


H. Köllner, Erworbene totale farbenblindheit mit Brieft über einen weiteren Fall, Z. Augenh. 21, 193, 309 (1909).


WASHINGTON, July 9, 1948.