Evaluation of single-pigment shift model of anomalous trichromacy*

Joel Pokorny and Vivianne C. Smith

Eye Research Laboratories, The University of Chicago, 930 East 59th Street, Chicago, Illinois 60637

(Received 30 August 1976; revision received 13 May 1977)

The spectral sensitivity of the visual photopigments, the interobserver variability in color judgments, and the spectral locus of unique yellow provide three major problems for accounts of X-chromosomal-linked anomalous trichromacy. According to the single-pigment hypothesis, the primary defect in anomalous trichromacy is a wavelength shift in the peak sensitivity of one of the three visual photopigments. We show that this shift results in reduction of the anomalous trichromat's r-g opponent chromatic channel. The distribution of response variability in Rayleigh equation match widths due to factors other than the spectral characteristics of the photopigments is similar in normal and anomalous trichromats. When normal and anomalous trichromats make hue estimations of sets of stimuli designed to contain similar chromatic characteristics of the photopigments is similar in normal and anomalous trichromats. When normal and anomalous trichromats make hue estimations of sets of stimuli designed to contain similar chromatic characteristics of the photopigments is similar in normal and anomalous trichromats.

I. INTRODUCTION

The congenital X-chromosomal-linked color defects are protanopia, protanomaly, deuteranopia, and deuter-

anomaly. These are the common color defects and account for the 8% to 10% incidence in males of European and North American populations. Modern classification is based on the colorimetric performance of individuals

---

*This work was supported in part by the NSF under Grant No. ENG 75-22625 and in part by the Joint Services Electronics Program under Contract No. F44620-74-C-0056.

---

\[
\begin{array}{|c|c|}
\hline
f(v) & q_2 \\
\hline
0 & 0 \\
3^2 + 6l + 4 - v^2 & a_1(3^2 + 6l + 4 - v^2) \\
(3^2 + 6l + 4 - v^2) & (a_1 + a_2) \\
- 2(3^2 + 6l + 2a_1a_2 + 2a_2) & - a_1(3^2 + 2a_2 + 4v) \\
& + v^2(a_1 + a_2 + a_2) \\
\hline
\end{array}
\]

The lowest-order correction to \( cr = n_0 \) involves the \( m = 2 \) term

\[
ct \approx n_0 - p_2 k^4 ,
\]

or more accurately from (68),

\[
ct \approx \left( n_0 - \frac{q_1}{2h} + O(k^{-2}) \right) \left( n_0 - \frac{q_1}{k} + \frac{q_2}{k^2} + \frac{2}{k^3} \right)^{1/2} .
\]

The \( p_2 \) coefficients for the various profiles are ascertained from (17d), (63), and Table I.

IV. CONCLUSIONS

A previously developed new asymptotic method for analyzing modal propagation in transversely inhomogeneous slab waveguides has here been extended to modal propagation in radially inhomogeneous graded index fibers with analytic refractive index profiles. It has been shown that the method generates term by term the asymptotic expansion derived from the exact solution when the profile function is parabolic. Confidence having been established thereby, the method has been applied to various polynomial profiles that perturb the parabolic shape. In these instances, exact solutions are not available. The new method avoids difficulties associated with modal caustics in other asymptotic schemes, and it provides a new interpretation of guided propagation in terms of evanescent waves. Extension to fibers with step index discontinuities is now being studied, as is the application of noncircular shapes.

ACKNOWLEDGMENT

This work was supported in part by the NSF under Grant No. ENG 75-22625 and in part by the Joint Services Electronics Program under Contract No. F44620-74-C-0056.

---

having these defects. The protanope and deuteranope are dichromats, and require only two primaries for color matching experiments. The protanomalous and deuteranomalous observers are trichromats, and require three primaries.

The color vision of protanomalous and deuteranomalous individuals differs from that of normal trichromats. The anomalous trichromat requires three primaries in spectral color matching but selects different primary ratios from those selected by a normal trichromat. Anomalous trichromats may be further distinguished as either simple anomalous trichromats, who have a moderately narrow matching range, or extreme anomalous trichromats who have a wide matching range that includes the normal match, the anomalous matches, and perhaps one of the primaries.

Pedigree study suggests that these color defects are mediated by two genes each with multiple alleles, one governing a protan sequence of defects, and one a deutan sequence of defects. Jaeger3 suggested that simple and extreme anomalous trichromats each represent a separate allele in the sequence. At the protan locus, the alleles are genes for normal color vision, simple protanomaly, extreme protanomaly, and protanopia. A similar set exists for the deutan locus. In the heterozygote female, with abnormal alleles on both X chromosomes, a dominance pattern among the alleles has been found: the color vision is that of the "least serious" defect. For example, with the allele for simple protanomalous trichromacy on one X chromosome, and the allele for simple or extreme protanomalous trichromacy or protanopia on the other, the color vision is that of simple protanomalous trichromacy.

Analysis of color matching data led von Kries3 to conclude that anomalous trichromacy is an alteration system in which one or more of the visual photopigments differs from normal in its spectral position. This idea led to the generally accepted view that the protan and deutan gene sequences govern the formation of normal and abnormal cone visual photopigments.

There have been two major theoretical approaches: (i) the formation of a normal, or a spectrally shifted (anomalous) or a "missing" (anope) long-wavelength-sensitive (LWS) photopigment is governed by the protan gene, and the formation of a normal or a spectrally shifted (anomalous) or a "missing" (anope) middle-wavelength-sensitive (MWS) photopigment is governed by the deutan gene. We call this the single-pigment hypothesis; and (ii) Jameson and Hurvich6 postulate that the deutan gene regulates a spectral shift of all three cone photopigments to longer wavelengths and the protan gene regulates a spectral shift of all three photopigments to shorter wavelengths. We call this the multiple-pigment hypothesis. Jameson and Hurvich postulated, in addition, that the amount of the shift may vary from individual to individual. In terms of description of stationary color defects as reduction, alteration, and collapse systems,6 the single-pigment hypothesis combines a reduction system for dichromacy with an alteration system for anomalous trichromacy. The multiple-pigment hypothesis postulates an alteration system for both dichromacy and anomalous trichromacy. An additional neural factor (discussed in Sec. III) differentiates dichromatic and trichromatic observers, giving an alteration system with a further modification by a collapse system.

The impetus for the single-pigment hypothesis came from consideration of the color matching performance and the pedigree analysis which postulates two (one protan, one deutan) sets of multiple alleles, each with a dominance sequence governing formation of the visual photopigments. The multiple-pigment hypothesis was derived from color perception data and does not address the findings of the pedigree studies.

Although there is a complete description of a model of color vision in anomalous trichromacy using the multiple-pigment hypothesis,4,6 there has been no general model of color vision in anomalous trichromacy using the single-pigment hypothesis. It has been suggested4 that the single-pigment hypothesis cannot fully explain color vision of anomalous trichromats. Criticism has centered on two issues: the variability of chromatic judgments made by anomalous trichromats5 and modeling of the (r-g) opponent channel.4 Our purpose in this paper is to evaluate the single-pigment hypothesis as a general model of anomalous color vision.

II. CONE VISUAL PHOTOPIGMENTS IN X-CHROMOSOMAL-LINKED COLOR DEFECT

Since both the single-pigment hypothesis and the multiple-pigment hypothesis involve dichromacy as well as anomalous trichromacy, it is important to review studies of dichromacy as well as anomalous trichromacy for those that might differentiate the two hypotheses. The single-pigment hypothesis applied to dichromacy states that the protanope shows evidence of only one visual photopigment active for wavelengths longer than 520 nm, a photopigment identical to the normal MWS visual photopigment; and that the deuteranope shows evidence of only one visual photopigment active for wavelengths longer than 520 nm, a photopigment identical to the normal LWS visual photopigment. Protanope, deuteranope, and normal all have a similar short-wavelength-sensitive (SWS) visual photopigment since the gene series do not affect the SWS visual pigments in this theory. The multiple pigment hypothesis applied to dichromacy states that the protanope will have three classes of visual photopigment. All classes demonstrate a (yoked) spectral shift to shorter wavelengths in comparison with the normal visual photopigments. The deuteranope also has three classes of visual photopigments, and all demonstrate a yoked spectral shift to longer wavelengths. A third class of observer is the "neuteronope,"8 a term suggested by Hurvich to describe the Aitken-Leber-Fick fusion deuteranope. This dichromat has three classes of visual pigments, which are identical to those of normal trichromats. According to the multiple-pigment hypothesis, dichromacy is said to result from "neural loss."6

A weakness of the single-pigment hypothesis is the difficulty of explaining "missing cone photopigments."
The meaning of “missing” is not specified. Some workers postulate that there is a replacement of the “missing” photopigment by the remaining normal photopigments in dichromacy. A variation of this theory has recently been developed by Piantanida. Others (e.g., Vos and Walraven) postulate that the “missing” photopigment is simply absent. This issue is not resolved, and there are no unambiguous experimental data to support replacement as opposed to absence.

Partial bleaching studies demand different predictions of dichromatic behavior for the single- and multiple-pigment hypotheses. According to the single-pigment hypothesis, the spectral sensitivity of the dichromat, being determined by only one visual photopigment, will remain unchanged in spectral position above 520 nm when measured following chromatic adaptation. Spectral sensitivity measured at bleaching levels should show only a slight narrowing of spectral shape in accordance with the theory of self-screening of a single visual photopigment. On the other hand, according to the multiple-pigment hypothesis, where two classes of photopigments are active above 520 nm, chromatic adaptation should change the shape of the spectral sensitivity curve of dichromats just as it does in normals. The fact is that chromatic adaptation does not change the position or shape of dichromatic spectral sensitivity curves; this finding supports the single-pigment hypothesis. Hurvich argued, however, that partial bleaching studies in dichromats or anomalous trichromats do not yield definitive evidence for or against either hypothesis. He points out that anomalous trichromats who are known to have two classes of photopigment active above 520 nm show minimal changes in the shape of their spectral sensitivity curve after chromatic bleach. According to the single-pigment hypothesis, the failure of the partial bleaching technique in anomalous trichromats occurs because their visual photopigments are spectrally very close together. According to the multiple-pigment hypothesis, failure of the partial bleaching technique in anomalous trichromats and dichromats is caused by a lack of sensitivity of the method. Thus, adherents of either hypothesis bring arguments to support their views.

Large-field matches made by X-chromosomal-linked dichromats are difficult to interpret within the multiple-pigment theory. Dichromats accept a full range of red-green mixtures on the anomaloscope with a 2° field. This behavior is of course a diagnostic criterion for dichromacy. The single-pigment hypothesis attributes this behavior to the fact that only one visual photopigment is active above 520 nm. The multiple-pigment hypothesis explains the behavior as due to reduced discrimination by an opponent chromatic channel which mediates perception of redness-greenness. As field size is increased, chromatic discrimination improves for all observers. We found that dichromats make unique matches with an 8° field. At least two visual photopigments must be active in the red-green region of the spectrum for the 8° field. According to the multiple-pigment hypothesis, the unique match occurring in protanopes by virtue of improved discrimination would correspond in its midpoint to the large-field matches of protanomals; similarly, in the case of deuteranopes, the midpoint is given by the large-field matches of deuteranomals; and in the case of “neuteranopes,” the correspondence is with the large-field matches of normals. Contrary to this expectation, we demonstrated that the red-green matches made by the dichromats are mediated by two visual photopigments, the one mediating the 2° match and a visual photopigment with the spectral sensitivity of rhodopsin. The improved discrimination available by use of the large field does not reveal evidence for the additional classes of cone photopigment demanded by the multiple-pigment hypothesis. The large-field matches offer no difficulty to the single-pigment hypothesis.

Evidence from the study of anomalous trichromacy tends to yield weak support for the single-pigment hypothesis. First, psychophysical evidence against one aspect of the multiple-pigment hypothesis for anomalous trichromacy may be found in the data of Schmidt, namely, the corollary assumption of a graded series of spectral shifts. Schmidt made careful measurements of the ratios of green to red (g/r) required in a match of yellow for populations of normals, protanomalous, and deuteranomalous trichromats. The green-red ratios of these groups showed three distinct populations. When these data are replotted as percentiles as a function of k log 10 g/r, they have similar variance. Analysis of the data of Waaler and Willis and Parnsworth leads to the same result. Schmidt suggested the source of variability to be common in the three groups. Such data are not indicative of a continuous spectral range of pigment maxima in anomalous trichromacy. Pokorny and Smith determined that the normal range of variability of inert ocular pigments accounts for two-thirds of the variation of green-red ratios in normal trichromats. The remaining one-third was due to receptoral variation—either pigment shift or optical density. These calculations reduce the possible role of pigment perturbation in accounting for interobserver variability in green/red match midpoints.

Second, psychophysical evidence for the single-pigment hypothesis in anomalous trichromats rests on the observation that the dichromat accepts the matches of both the normal trichromat and the corresponding anomalous trichromat. This behavior is required under the single-pigment hypothesis, since the dichromat’s visual photopigment is presumed identical to one of the visual photopigments of the normal and the corresponding anomalous trichromat. That is, the dichromat has a reduction form of the trichromatic color vision found in the normal and in anomalous observers. According to the multiple-pigment hypothesis, protanopes and deuteranopes may or may not reject matches of corresponding anomalous trichromats but should reject matches of normal trichromats. Neuteranopes should accept only the matches of normal trichromats. Quantitative data, published by Mitchell and Rushton and by Rushton, Powell, and White, showed that protanopes accept protanomalous and normal trichromatic matches, and that deuteranopes accept the deuteranomalous and normal trichro-
matic matches. Despite these generalizations, a given protanope may or may not accept the color matches made by an individual protanomalous trichromat or an individual normal trichromat. Procedures in which an individual is asked to accept or reject a different individual's color match are called confrontation experiments. Acceptance of another's match in confrontation may fail because of individual differences in inert ocular pigments combined with receptor variation. On the other hand, acceptance may occur because of poor chromatic discrimination. Confrontation experiments are inconclusive; they have been used to support and to refute both single- and multiple-pigment hypotheses.

Attempts to derive the spectral sensitivity curves for the photopigments of anomalous individuals have primarily been made within the framework of the single-pigment hypothesis. A comparison of some estimates is shown in Fig. 1 for the anomalous photopigment of deuteranomalous and protanomalous trichromats. For protanomalous observers, the long-wavelength slope of the theoretical anomalous visual photopigment (L'WS) lies 7 nm from the MWS photopigment postulated for both normal and anomalous trichromats. For deuteranomalous observers, the theoretical anomalous photopigment (M'WS) lies 6 nm from the long-wavelength-sensitive photopigment postulated for both normal and anomalous trichromats. Given the considerable spread in the experimental estimates, the theoretical functions provide a general fit to the data.

The Schouten hypothesis is a special case of the single-pigment hypothesis in which the shifted visual photopigment of anomalous individuals have primarily been made within the framework of the single-pigment hypothesis. A comparison of some estimates is shown in Fig. 1 for the anomalous photopigment of deuteranomalous and protanomalous trichromats. For protanomalous observers, the long-wavelength slope of the theoretical anomalous visual photopigment (L'WS) lies 7 nm from the MWS photopigment postulated for both normal and anomalous trichromats. For deuteranomalous observers, the theoretical anomalous photopigment (M'WS) lies 6 nm from the long-wavelength-sensitive photopigment postulated for both normal and anomalous trichromats. Given the considerable spread in the experimental estimates, the theoretical functions provide a general fit to the data.

The accumulated evidence does support the single-pigment hypothesis for anomalous trichromacy although, as we have emphasized, much of the evidence is weak; and there is not good agreement on the exact spectral location of the photopigments of anomalous trichromats. In the remainder of this paper, for modeling purposes, we use the solid lines of Fig. 1 as estimates of the anomalous (shifted) photopigments of anomalous trichromats in our evaluation of the single-pigment hypothesis.

III. DISCRIMINATION LOSS

Large interobserver variability in results with colorimetry, with wavelength and saturation discrimination, and with hue naming is characteristic of individuals having anomalous color vision. Wavelength and saturation discrimination functions of individual anomalous trichromats vary from those appearing close to normal to those appearing close to dichromatic. Interobserver variability occurs also in the matching ranges on the anomaloscope. Anomaloscope ranges, the total errors on the Farnsworth Munsell 100-hue test, and the shape of the wavelength discrimination function are correlated, suggesting that all reflect a common source of discrimination loss in each individual. The enlarged variability refers to judgments of hue and hue balance. Mitchell and Rushton, and Rushton, Powell, and White stated that the precision of brightness discrimination made by normal and anomalous trichromats is similar. Helve's analysis of Farnsworth Munsell 100-hue data shows that the chromatic discrimination loss experienced by anomalous trichromats is primarily...
of judgments of the redness-greenness component of color fields.

Jameson and Hurvich\textsuperscript{4} and Jameson\textsuperscript{5} present a theoretical analysis of the interobserver variability of match widths found in anomalous trichromats. They postulate that variation in Rayleigh equation match widths for normal and anomalous trichromats may be modeled by variation of the weighting constant on an $\alpha - \gamma$ opponent channel. Their model thus includes two postulates: (i) the sets of three photopigments characteristic of normal trichromatic vision are shifted on the frequency axis, and (ii) there is independent reduction of neural activity in opponent channels. The adoption of two postulates allows prediction of match widths, both for those occasional anomalous trichromats who show superior discrimination and for those who show reduced discriminatory ability. Conversely, the model adds an additional, novel, and unexplained problem: to explain the distribution of match widths, it must be assumed that, as a population, anomalous trichromats have a reduction of neural activity compared with normal trichromats and that the neural factor in anomalous trichromats is governed by the multiple allele sequence. How do the two postulated independent factors (pigment shift and neural reduction) arrive at covariance in an entity presumed governed by a single gene?

An alternate question may be asked: is there evidence of discrimination loss in the neural channels of anomalous trichromats when the single pigment hypothesis is assumed? We\textsuperscript{37} and others\textsuperscript{20,26,37} have suggested that the single-pigment hypothesis predicts an average loss of chromatic discrimination of redness-greenness in anomalous trichromacy. Normal trichromats also share interobserver variability in their FM 100-hue test scores\textsuperscript{35,36,38,39} and Rayleigh match widths on the Nagel anomaloscope.\textsuperscript{18,36,40} We can compare the statistical distributions of chromatic discrimination in normal and anomalous trichromats. We must therefore ask: does the discrimination loss predicted by the single-pigment hypothesis account for interobserver variability of anomalous trichromats, allowing for similar statistical variation of neural processing in normal and anomalous trichromats? The first analytic method is a comparison of normal and anomalous Rayleigh equation match widths in terms of the relative information available to a chromatic processing channel. The second method is hue naming in normal trichromats, using a spectrum whose relative chromatic content is similar to that available (for the whole long-wavelength spectrum) to anomalous trichromats.

Analysis of Rayleigh match widths

The distributions of match widths in groups of observers differ markedly.\textsuperscript{18,36,40,41} Generally, anomalous trichromats show larger match widths and greater interobserver variability than do normal trichromats. We recomputed these match widths in terms of the effective chromatic signals generated by theoretical sets of photopigments for each class of observer and then compared the recomputed distributions.

We required match width data for an unselected population of normal, protanomalous, and deuteranomalous trichromats obtained on an anomaloscope with narrow bandwidth spectral primaries. Such data have been reported by Willis and Farnsworth\textsuperscript{18} for a population of Naval recruits using a Nagel (model 2) anomaloscope.

Conversion of the match width data to express the effective chromatic signal generated by the underlying visual photopigments is accomplished by first transforming the raw data to a new form, called analytical modes, where the red and green primary energies are matched either for protanopic sensitivity (protan mode) or for deuteranopic sensitivity (deutan mode). The analytical modes were first proposed by Baker and Rushton\textsuperscript{12} and were expanded by Mitchell and Rushton.\textsuperscript{18} In their experiments, a protanope set the $R$ and $G$ primaries equal in luminance for the protan mode, and a deuteranope set the $R$ and $G$ primaries equal in luminance for the deutan mode.\textsuperscript{12,42}

We have previously shown\textsuperscript{37} that theoretical equations may be written for the anomaloscope in analytical mode. For normal trichromats in the deutan mode:

$$\frac{SM_L}{SL'} = \frac{SM_L}{SL} - \frac{SM_G}{SL} q + \frac{SM_G}{SL} ;$$

and in the protan mode:

$$\frac{SL_L}{SL'} = \frac{SL_L}{SM} - \frac{SL_G}{SM} q + \frac{SL_G}{SM} .$$

In these equations, $q$ represents the fraction of $R$ primary at the match to the $Y$ test color, $SM$ represents the theoretical sensitivity of the MWS visual photopigment which occurs in the normal trichromat, the protanomalous trichromat, and the protanope, and $SL$ represents the theoretical sensitivity of the LWS visual photopigment which is assumed to occur in the normal trichromat, the deuteranomalous trichromat, and the deuteranope. Similar equations may be written for the visual photopigments of anomalous trichromats. In this case, $SM$ is replaced by $SM'$ in Eq. (1) and $SL$ by $SL'$ in Eq. (2) to give theoretical equations for the deuteranomalous in the deutan mode and the protanomalous in the protan mode.

The advantage of the modes is that the pigments sensitivity ratios ($SM/SL$ for normals, $SM'/SL$ for deuteranomalous in the deutan mode, and $SL'/SM$ for normals, $SL'/SM$ for protanomalous in the protan mode) are shown to be linear functions of $q$, the fraction of long wave primary in the match. This statement is illustrated in Fig. 2. Figure 2(a) shows the theoretical calculated values of $q$ for three values of $SM/SL$ for the normal, and $SM'/SL$ for the deuteranomalous trichromat: namely, at the wavelength of the $G$ primary where $q$ is zero, at the wavelength of the Rayleigh equation match, and at the wavelength of the $R$ primary where $q$ is one. At the Rayleigh wavelength, $q$ takes on values of 0.57 for normals and 0.25 for deuteranomals (i.e., normals have a greater fraction red in their match than deuteranomals). The plotted values of $q$ would represent the theoretical match midpoint value. The three values of $q$ are collinear, as demanded by Eq. (1). Figure 2(b) shows similar calcu-
lutions for normal and protanomalous trichromats in the protan mode. In this case, the value of \( q \) is higher for protanomals than normals at the Rayleigh wavelength.

In Fig. 2, the normal slopes are steeper than the corresponding anomalous slopes by a factor of 2.5:1 for the deutan mode and 10:1 for the protan mode; that is, a normal match width \( \Delta q \) of one unit or a deuteranomalous match width \( \Delta q \) of 2.5 units, would reflect a similar effective chromatic signal for the normal as for the deuteranomalous individual.

Raw match settings for any anomaloscope can be transformed algebraically to analytical form using the subject's anomalous quotient and a value for \( q_a \) predicted by a set of MWS and LWS visual photopigments for normal trichromats. We weight the raw score \( R/G \) mixture to give \( q_a \) at the normal match. For the Nagel model 2, the equation to convert from the \( R/G \) mixture to \( q \) is

\[
q = \frac{a(R/G \text{ mixture})}{a(R/G \text{ mixture}) + (75 - R/G \text{ mixture})},
\]

where \( a \) is a weighting function to convert the \( R/G_a \) mixture set by normals to the value \( q_a \) predicted by the normal visual photopigments, and 0 to 75 is the range of \( R/G \) mixture values on the instrument. The transformation is algebraically identical to that devised by Willis and Farnsworth, who pointed out that the anomalous quotients remain invariant. Figure 3 shows the relation of the transformed score \( q \) to the raw score (the \( R/G \) mixture) for the Nagel model 2. Figure 3(a) shows the calculation for the deutan mode, while 3(b) shows the calculation for the protan mode.

A feature of the Nagel model 2 instrument was that the average normal match could be set to occur either at an \( R/G \) mixture scale value of 40 or at a mixture of 60. To move the match to 40 a neutral filter was placed in front of the 536 nm primary. Schmidt reported that the average normal match width was smaller by a factor of 0.6 for model 2 anomaloscopes set to give a normal match at 60 rather than at 40. A test of our analysis is thus whether we can predict this finding. Figure 3 shows two sets of calculations for \( q \) as a function of \( R/G \) mixture scale, one for the normal match at 40 and one for the normal match at 60. When the Nagel model 2 is set to give a normal match midpoint of 40 by placing a neutral filter in front of the 536 nm primary, the transformed score \( q \) (shown as squares) for the deutan mode [Fig. 3(a)] is, to a first approximation, linearly related to the Nagel \( R/G \) mixture scale, and match width is independent of match midpoint. This finding reflects the fact that the Nagel anomaloscope is designed to have equal luminance at the \( R \) and \( G \) primaries for normal trichromats, and is set virtually in the deutan mode. When the instrument is set to give a normal match midpoint of 60 (circles), there is slight bowing.

For a normal trichromat, a fixed \( \Delta q \) represents a constant change in the SM/SL pigment sensitivity ratio. The predicted match width on the \( R/G \) mixture scale for a fixed \( \Delta q \) is smaller when the normal match is set at 60 than when it is set at 40 by a factor of 0.6 in accordance with Schmidt's report. This agreement gives us confidence that our method of analysis is correct and appropriate. We attribute the difference in average normal match widths reported by Schmidt to the manner of "setting up" the anomaloscope.

In the protan mode [Fig. 3(b)], there is severe nonlinearity in the function relating \( q \) to \( R/G \) mixture scale. The match widths may vary with match midpoint. These nonlinearities reflect the setting of the Nagel anomaloscope to have approximately equal luminance for normal trichromats.

The Willis and Farnsworth raw scores were transformed algebraically to give \( q \), the match settings in the appropriate analytical modes. We recalculated data for normal and simple deuteranomalous trichromats in the deutan mode and normal and simple protanomalous trichromats in the protan mode. For example, in the deutan mode, data for normal and deuteranomalous trichromats were first transformed into \( q \) scores. Then, for
FIG. 4. Block histograms of the frequency of match widths in transformed score, $\Delta q$ (panel a), and the frequency of the corresponding changes in theoretical pigment sensitivity ratios (panel b). Data are shown for 40 normal (left side) and 56 deuteranomalous (right side) trichromats. Computations are described in the text. Associated values of $\chi^2$ are:

panel (a), $\chi^2(11) = 70.66; \{p: \chi^2(11) > 70.66 > 0.001\}$

panel (b), $\chi^2(8) = 9.8; \{p: \chi^2(8) > 9.8 < 0.75\}$

each subject, theoretical pigment sensitivity ratios for the extremes of the matching range ($q_{\text{max}}$ and $q_{\text{min}}$) were calculated from Eq. (1). The data were then retabulated grouping observers first according to their transformed match width ($\Delta q$) and then according to the corresponding change in the pigment sensitivity ratio ($\Delta[SM/SL]$ for normal and $\Delta[SM'/SL]$ for deuteranomalous trichromats). A similar analysis of the Willis and Farnsworth data$^{18}$ for normal and protanomalous trichromats was performed in the protan mode.

Block histograms compare the results for normal and deuteranomalous trichromats in the deutan mode (Fig. 4) and for normal and protanomalous trichromats in the protan mode (Fig. 5). The frequency of occurrence of transformed match ranges ($\Delta q$) is shown in Figs. 4(a) and 5(a) and the frequency of occurrence of theoretical changes in pigment sensitivity ratios in Figs. 4(b) and 5(b). When block histograms are made of the match ranges, the normal and deuteranomalous trichromats show significantly different distributions ($p < 0.001$). When pigment sensitivity ratios are computed, the widths of the deuteranomalous trichromats are multiplied by a factor of about 0.4 (Fig. 2). The distributions of the change in pigment sensitivity ratio for the extremes of the match width yield similar histograms for normal and anomalous trichromats. The $\chi^2$ test for normal and deuteranomalous trichromats indicates that these two distributions are not significantly different ($p < 0.75$); they may come from the same parent population. Similar trends are evident for the comparison of normal and protanomalous trichromats (Fig. 5) but the small number of protanomalous observers precludes use of the $\chi^2$ test.

From this analysis we conclude that the single-pigment hypothesis for anomalous trichromacy adequately explains the poorer than normal discrimination of this class of observer typified by the acceptable match range on the anomaloscope. Within the population, some anomalous trichromats may show discriminations as sensitive as some normal trichromats. Neural or other factors which contribute to variability in discrimination are similar in both anomalous and normal trichromats.

Hue estimation

Smith, Pokorny, and Swartley$^{34}$ described hue estimations made by five deuteranomalous trichromats using restricted response categories "red," "green," "yellow," "blue," and "white." They stated that the response category "blue" was the same for normals and deuteranomals. This response occurred minimally or not at all above 500 nm. The deuteranomals showed greater intersubject variability than normals in using response categories "red," "yellow," and "green," a phenomenon that does not occur in normal trichromats.
The deuteranomalous trichromats did not always need three independent color names to describe the spectrum above 510 nm. In deuteranomalous trichromats, the Bezold-Brücke hue shift measured by hue estimation was exaggerated compared to that of the normal trichromats (also reported by Inuma), with a disproportionate increase in the use of the yellow category at higher luminances. Smith, Pokorny, and Swartley emphasized that these findings presented several inconsistencies with available theory.

However, the variability of the data may reflect the reduction of chromatic signals in anomalous trichromats. To investigate this possibility, we required normals to estimate the hues of a set of stimuli of greatly reduced chromatic range. To derive such a set of stimuli, theoretical assumptions concerning the neural organization beyond the photopigment level are necessary. In this section, we rely on reported measurements of the spectral locus of unique yellow for each class of trichromat (a plausible model for correctly predicting such shifts using the single-pigment hypothesis is discussed later in Sec. IV.) Pigment sensitivity ratios were calculated for theoretical pigments of normal, deuteranomalous, and protanomalous trichromats. We arbitrarily adjusted the heights to cross at reported values of unique yellow for each class of observer.

The deviations of these ratios from unity express the relative information available to a chromatic processing channel. For the normal trichromat, the ratio set at unity at 580 nm rises rapidly as a function of wavelength and levels off only after both pigments have reached their limiting gradients above 680 nm. In contrast, the ratios rise slowly in the anomalous trichromat, showing little change as a function of wavelength. The ratios are set at unity at 590 nm for deuteranomalous and at 570 nm for protanomalous trichromats. The change in the ratios for the deuteranomalous photopigments for the 150 nm spectral range from 530 to 680 nm compares to a spectral range for normals of 26 nm, which represents wavelengths 571–597 nm (Fig. 6(b)). We believe that the signals to the r-g chromatic channels in normals and deuteranomals are proportional to these ratios, and that a spectrum with equivalent anomalous ratios can be calculated for normal observers. Thus, we view the deuteranomalous trichromat as having the r-g discrimination capacity for wavelengths 530 to 680 nm equivalent to a reduced spectrum for normal trichromats including wavelengths 571 to 597 nm. A reduced spectrum may also be made for the ratios of protanomalous trichromats. It has a width of 44 nm comprising wavelengths 570 to 614 nm. This approach is limited to the spectral region where the chromatic contribution of a b-y chromatic channel is reasonably constant (λ > 510 nm).

Normal observers (naive to the purpose of the experiment, as was the person recording the data) were asked to assign color names “green,” “yellow,” and “orange” by percent, or the color names “green,” “yellow,” and “red” by percent as a function of wavelength. The stimuli were 12 min in diameter and were flashed for 12 ms. We chose wavelengths between 571 and 597 nm which had equivalent pigment sensitivity ratios to the wavelengths at 20 nm intervals between 510 and 690 nm for deuteranomals. The method was that used in the study of deuteranomalous hue estimation.

The choice of the color name “orange” or “red” proved arbitrary. We chose “orange” since the reduced spectrum extended only to wavelengths usually identified as orange by normal trichromats. We thought that the name “orange” might allow better differentiation of the spectral region. In fact, we found our observers chose one of two approaches. Some observers showed no differences in hue estimation whether using “red” or “orange” as the third chromatic name. The others showed a consistent difference in “orange” and “red”; the “red” response function was displaced to longer wavelengths (5–10 nm) as occurs on full spectrum hue estimation when both response categories are allowed. In addition, all observers showed some anchoring effects, that is, the shortest wavelength was reported as containing more “green,” the longest as more “red” or “orange” than occurs in their full spectrum naming.

The data for a low level of retinal illuminance, 25 tads, are shown in Fig. 7. On the left are shown the chromatic response functions for six deuteranomalous trichromats reproduced from Smith, Pokorny, and Swartley. On the right are the functions obtained with the reduced spectrum on four normals using response categories “green,” “yellow,” and “orange.” For normals, the upper scale shows the reduced spectrum (i.e., the actual wavelengths used), while the lower scale shows the equivalent spectrum for deuteranomalous trichromats. The distributions for both classes of observers are shown for the response “green” in Fig. 7(a) and for “yellow” in Fig. 7(b). Figure 7(c) shows “orange” for normal, compared with “red” for deuteranomalous tri-
FIG. 7. Chromatic response functions for six deuteranomalous (left hand side) and four normal (right hand side) trichromats as a function of wavelength at 25 tds. Upper scale shows the actual wavelengths presented to the normal trichromats. Lower scale shows the actual wavelengths presented to the deuteranomalous and “equivalent” wavelengths for the normal trichromats. Panel (a) shows data for the response “green,” panel (b) for the “yellow” for both classes of observer. Panel (c) shows data for the response “red” for deuteranomalous and “orange” for normal trichromats.

In Fig. 8, similar data at a higher level (250 tds) are shown. Average data are compared, for the two levels, in Fig. 9. The shape and variability of the normal responses on the reduced spectrum are comparable to the responses of the deuteranomalous trichromats. Figure 10 shows a comparison of the average data at the two light levels. Data for deuteranomalous individuals are shown in Fig. 10(a), and for normals on the reduced spectrum in Fig. 10(b). When run with a reduced spectrum, the normal trichromats show the same exaggerated effect of luminance change as do the chromats.
deuteranomalous trichromats. The actual neural processing probably does not differ between normal and anomalous trichromats.

The similarity between these sets of data for normal (reduced spectrum) and anomalous trichromats (full spectrum) suggests that some measured effects in which anomalous trichromats show exaggerated changes in color perception compared to normals may reflect the fact that large wavelength differences are signaled as small changes in the anomalous trichromat’s r–g chromatic processing channel. As an example, heightened color contrast can be accounted for in this manner, as can the magnitude (but not the direction) of the Stiles-Crawford effect of the second kind (hue change).

IV. RED-GREEN CHROMATIC CHANNEL IN ANOMALOUS TRICHROMATS

Our final concern centered on the nature of the r–g opponent chromatic channel in anomalous trichromats.

A number of investigators noted that the spectral locus of unique yellow (a hue containing neither greenness nor redness) is displaced to shorter wavelengths for protanomalous trichromats and to longer wavelengths for deuteranomalous trichromats than its locus (about 575 nm) for normal trichromats.

Hurvich and Jameson stated that the single-pigment model does not yield acceptable predictions for the unique yellow locus for anomalous trichromats. In their model, the presumed photopigment sensitivities of normal trichromats are weighted so that they cross near the unique yellow locus. Hurvich and Jameson maintain these weighting factors when they model anomalous trichromacy, and, using the multiple-pigment hypothesis, obtain correct predictions for the loci for unique yellow in anomalous trichromats. However, when they apply their technique to the single-pigment hypothesis, they find that predicted loci for unique yellow are displaced to longer wavelengths for both protanomalous (PA) and deuteranomalous (DA) trichromats (610–620 nm for PA,

FIG. 9. Comparison of average data of normal (dashed) and deuteranomalous (solid lines). The abscissa shows actual wavelengths for deuteranomalous and equivalent wavelengths for normal trichromats. Panel (a) shows data at 250 tds; panel (b) at 25 tds.

FIG. 10. Comparison of results of 25 tds (solid) and 250 tds (dashed). The abscissa is as described for Figure 9. Panel (a) shows data for deuteranomalous, panel (b) for normal trichromats.
In order to derive a weighting constant, we assume that there exists a physical white that is equivalent to physiological white (that stimulus which produces no activity in chromatic opponent channels). As an example, Hurvich\textsuperscript{6} has suggested that a white of 5500 K color temperature is neutral in human vision. Physiological white for normal trichromats may not in fact lie on the Planckian locus. However, we do assume that for either the blue-yellow opponent channel or the red-green opponent channel there will be a white on the Planckian locus which is equivalent to the achromatic point of the channel.

Given this assumption, the r-g opponent is calculated by weighting the theoretical pigment sensitivity curves by factors inversely proportional to the area of the sensitivity curve calculated for the appropriate energy distribution. We have used the equal energy spectrum as a convenient approximation to 5500 K white. For each visual photopigment, the weighting factor is reciprocal to the summed area for an equal-energy spectrum using discrete approximations. Full spectrum quantal spectral sensitivities for normal and anomalous photopigments were derived using methods described by Smith, Pokorny, and Starr\textsuperscript{12} and converted to equal energy field sensitivities using methods described by Wyszecki and Stiles.\textsuperscript{13} Figure 11 shows the three r-g channels. For the normal trichromat, the r-g channel was calculated as

\begin{equation}
N(r-g) = 10 \frac{S_{LWS}}{\Sigma_i S_{LWS}} - 10 \frac{S_{MWS}}{\Sigma_i S_{MWS}} \tag{4}
\end{equation}

where $S_{MWS}$ and $S_{LWS}$ are the equal energy field sensitivities for the normal middle- and long-wavelength photopigments. The factor 10 is an arbitrary weighting factor to keep maxima of the weighted sensitivity curves near unity. The r-g opponent channels for deuteranomalous and protanomalous were calculated in a similar fashion. The equation for deuteranomalous trichromats is

\begin{equation}
DA(r-g) = 10 \frac{S_{MWS}}{\Sigma_i S_{MWS}} - 10 \frac{S_{LWS}}{\Sigma_i S_{LWS}} \tag{5}
\end{equation}

and for protanomalous trichromats is

\begin{equation}
PA(r-g) = 10 \frac{S_{LWS}}{\Sigma_i S_{LWS}} - 10 \frac{S_{MWS}}{\Sigma_i S_{MWS}} \tag{6}
\end{equation}

Figure 11 shows that the strength of the theoretical r-g chromatic opponent is reduced for the two classes of anomalous observer compared with the strength of the r-g chromatic opponent for the normal trichromat. The crossovers are the predicted equal energy neutral points. They occur at 569 nm for the normal r-g opponent, at 588 nm for the deuteranomalous r-g opponent, and at 562 nm for the protanomalous r-g opponent. These neutral points are in close agreement with the reported equal energy saturation minimum of normal trichromats and with the secondary relative minima of some anomalous trichromatic subjects.\textsuperscript{24} The crossovers would predict unique yellow at or near these lo-

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure11.png}
\caption{Predicted r-g chromatic opponents. Symbols: V, deuteranomalous; o, normal, and A, protanomalous trichromats.}
\end{figure}

670–680 nm for DA). Although the original choice of weighting factor was arbitrarily chosen to give the desired locus for unique yellow in normal trichromats,\textsuperscript{4,5} Hurvich and Jameson\textsuperscript{51} conclude that the result negates the single-pigment hypothesis.

Only recently have attempts been made to relate psychophysical opponents to presumed visual pigment spectra and to the electrophysiological data from macaque monkeys. There are no "objective" methods for setting pigment curve heights to give chromatic opponents, nor is there a physiological interpretation of the weighting factors. Weighting factors are chosen to give the best possible set of predictions.

A single-pigment hypothesis can correctly predict spectral positions of unique yellow for normal, protanomalous, and deuteranomalous trichromats. We view trichromacy as a combination of an achromatic (luminance) channel, a blue-yellow (b-y) opponent channel, and a red-green (r-g) opponent channel. Following the scheme set by others\textsuperscript{32,33} for normal trichromacy we let the achromatic be given by a sum of LWS and MWS; the (b-y) opponent by the achromatic opposed to SWS; and the (r-g) opponent by LWS opposed to MWS. Neutral "white" occurs when the b-y channel signals neither "blue" nor "yellow," and the r-g channel signals neither "red" nor "green." Unique yellow occurs when the b-y channel signals "yellow" and the r-g channel signals neither "red" nor "green."
FIG. 12. Predicted neutral points as a function of color temperature. Data were calculated using field sensitivities for a LWS-MWS (tritanopic) opponent chromatic for a normal and a corresponding LWS-M'WS opponent for a deuteranomalous and a L'-WS-MWS opponent for a protanomalous trichromat. Symbols: V, deuteronomalous; O, normal; Δ, protanomalous trichromats.

Our approach leaves unsolved problems for theoretical accounts of normal and anomalous trichromacy. For example, we calculate an (r-g) chromatic opponent using only LWS and MWS cone types. The psychophysical data at threshold can be modeled by this type of opponent with input only from MWS and LWS cones.52,53 Brightness additivity data for violet and green mixtures viewed by normal trichromats are subadditive at threshold but superadditive above threshold.56 The rules governing SWS cone input to the (r-g) chromatic channel are not simple. Ingling, Tsou, and Drum54 have discussed this problem in detail and have postulated a theoretical framework. The (b-y) chromatic opponent is usually modeled by assuming that a sum of MWS and LWS opposes SWS cone types.6°,52,53 Both Hurvich and Jameson57 and Guth and Lodge58 have used the V, curve to represent the y input to a (b-y) chromatic opponent. The V, curve can be expressed as a sum of MWS and LWS, and Z, as SWS.50

The blue-yellow opponent calculated as K( y - z y) would be parallel to our calculation for the (r-g) opponent since the color matching functions are for an equal energy spectrum normalized to equal area. The Hurvich and Jameson57 (b-y) chromatic opponent is of this form. However, the Guth and Lodge58 threshold (b-y) chromatic opponent assumes the form K(l.933 y - z y). Ingling, Tsou, and Drum54 also find it necessary to model the scaling constants for the yellow and blue components separately with varying luminance.

We discuss these difficulties to emphasize that modeling the opponent channels using pigment absorption sensitivities is not a simple business. Our weighting constants were chosen to allow economical modeling of normal and anomalous (r-g) opponent channels within the framework of the single-pigment hypothesis.

V. CONCLUSION

We have evaluated three aspects of the color vision of X-chromosomal-linked anomalous trichromats using the single-pigment hypothesis. We have reviewed evidence on the visual photopigments, and evaluated the chromatic discriminations and the (r-g) chromatic opponent channel of anomalous trichromats within the framework of the single-pigment hypothesis. Although many researchers in genetics and color mixture have postulated the single-pigment hypothesis as an economical explanation of the Rayleigh match data, there has been no general model of the color vision in anomalous trichromacy using the single-pigment hypothesis. Some authors6°,51 have suggested that the single-pigment hypothesis cannot fully explain the color vision of anomalous trichromats.

Our analyses, however, suggest that the single-pigment hypothesis does provide an economical basis for explaining anomalous trichromacy. According to the single-pigment hypothesis, the primary defect in congenital X-chromosomal-linked anomalous trichromacy is a shift in sensitivity in one of the three visual photopigments. As one result of this shift, the inputs to the r-g opponent channel are reduced and the signals are close to the detection capability of the channel. With re-
duced detection capability, the anomalous trichromat's hue judgments seem variable in comparison to the more precise judgments made by normals. By considering the pigment sensitivity ratios, we show that the distribution of response variability in Rayleigh matches due to factors other than the spectral characteristics of the photopigments is similar in anomalous and normal trichromats. In the hue estimation procedure, we show that when a normal trichromat is asked to estimate the hues of a set of reduced spectral stimuli, his discriminations become noisy in the same way as those of anomalous trichromats. We also show that chromatic opponents may be calculated to allow correct predictions of spectral loci for unique yellow for anomalous trichromats.

In conclusion, we believe that the single-pigment hypothesis can provide an adequate framework within which to view the color vision of X-chromosomal-linked anomalous trichromats.

ACKNOWLEDGMENT

Kenneth S. Ginsburg assisted in the collection of the hue estimation data on the normal trichromats.

* Portions of this research were presented at ARVO in 1974 and at OSA in 1976. This work was supported in part by USPH NIH NEI Grant Nos. EY 00901, EY 00652 (J. Pokorny), and EY 00523 (F. W. Newell).


30. We omit the estimates derived by N. V. Lobanova and N. I. Speranskaya ["The spectral sensitivity of the retinal elements of anomalous trichromats," Biophysics 6, 71-78 (1961)], who used M. M. Bongard and M. S. Smirnov's determination of curves for spectral sensitivity of the visual receptors by means of complex curves [see Dokl. Akad. Nauk. SSSR 102, 1111-1114 (1955)]. The technique rests on the assumption that for each of the three visual photopigments there exist pairs of spectral wavelengths to which the photopigment is unresponsive. This assumption can only be realized for the SWS visual photopigment, N. I. Speranskaya and N. V.
Lobanova's estimates "Determination of spectral sensitivity curves of the eyes of normal trichromats," Biophysics 6, 66-70 (1961) of the visual photopigments of normal trichromats show substantial disagreement with the König fundamentals proposed by Vos and Walraven (see Ref. 9) and ourselves (see Ref. 58). The proposed Speranskaya and Lobanova normal receptor I (LWS) is displaced about +5 nm and their normal receptor II (MWS) is -5 nm from the corresponding estimates of Vos and Walraven (Ref. 8) and ourselves (Ref. 58).


