CLASSICAL ARTICLES IN COLOR

On The Relationship of Four-Color Theory to Three-Color Theory

Erwin Schrödinger

Über das Verhältnis der Vierfarben- zur Dreifarbentheorie, Sitzungsber. Kaiserl. Akad. Wiss. Wien [*IIa*] **134,** 471–490 (1925).

Translation by National Translation Center

Commentary by Qasim Zaidi

Erwin Schrödinger (1887–1961) is best known for his equation which provides a description of behavior at the quantum level of molecules, atoms, and subatomic particles. In a different domain, his book "What is Life?", has been described as a major inspiration by some of the founders of molecular biology. During time that he spent in Zurich, he wrote (to my knowledge) four articles on color vision. Two of these were published in translation by MacAdam.[1] The present translation of a third article was obtained from the National Translation Center in 1982 (the identity of the translator was not available). I have made a small number of modifications to the translation in the interest of clarity. This paper was a major influence on the work of the color theorist Deane Judd, both in his analyses of opponent-color theories[2] and in the specification of the CIE 2° primaries.[3]

In this paper, Schrödinger shows that the representation of lights on the basis of opponent-color theory is an affine transformation of the representation on the basis of trichromatic theory, if certain assumptions are valid about the combination of the three cone signals into the two opponent and one achromatic system.[4] As I will show below, this geometrical equivalence in turn makes it possible to test the combination assumptions critically. The failure of some of these tests has led to a refinement of our understanding of the opponent stage.

Given the sophistication of his other papers, the treatment of trichromatic theory in this paper is surprisingly inadequate. Instead of building on Maxwell's[5] work which established the three dimensionality of metameric color-matches, Schrödinger treats trichromacy as a basis for the phenomenological appearance of colors. However, this shortcoming has no effect on the subsequent analysis which uses König's Fundamentals that are based on trichromatic color matching functions (CMF) and the assumption that cone spectral sensitivities are linear combinations, with constant coefficients, of the CMF's.

Schrödinger shows that Hering[6] type valence curves are equivalent to straight lines through achromatic white in any chromaticity diagram. The essential insight was that since color matching functions are a linear transform of cone spectral sensitivities, if, as proposed by von Kries,[7] valence curves could be identified with linear combinations of cone spectra, then valences should also be a linear transform of color matching functions. The explicit derivation is easily described in matrix notation. If RG, YB, *and* LD, *the red-green, yellow-blue and light-dark valence associated with a spectral light* λ, *are independent linear combinations of the quanta caught by* L, M, *and* S *cones, then:*

$$\begin{pmatrix} RG \\ YB \\ LD \end{pmatrix} = \begin{pmatrix} l_{RG} m_{RG} s_{RG} \\ l_{YB} m_{YB} s_{YB} \\ l_{LD} m_{LD} s_{LD} \end{pmatrix} \begin{pmatrix} L \\ M \\ S \end{pmatrix}. \quad (1)$$

Since L, M, *and* S *are linear combinations of color matching functions,* R, G, B, *the valences can be described as:*

$$\begin{pmatrix} RG \\ YB \\ LD \end{pmatrix} = \begin{pmatrix} r_{RG} g_{RG} b_{RG} \\ r_{YB} g_{YB} b_{YB} \\ r_{LD} g_{LD} b_{LD} \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (2)$$

where

$$\begin{pmatrix} r_{RG} g_{RG} b_{RG} \\ r_{YB} g_{YB} b_{YB} \\ r_{LD} g_{LD} b_{LD} \end{pmatrix} = \begin{pmatrix} l_{RG} m_{RG} s_{RG} \\ l_{YB} m_{YB} s_{YB} \\ l_{LD} m_{LD} s_{LD} \end{pmatrix} \begin{pmatrix} r_P\, r_D\, r_T \\ g_P\, g_D\, g_T \\ b_P\, b_D\, b_T \end{pmatrix}^{-1}.$$

The coefficients in the inverted matrix are the coordinates of the confusion vectors of congenital dichromats[8] *or equivalently the chromaticity coordinates of dichromatic copunctal points.*[9] *In three-dimensional* RGB *space, lights with zero* RG *valence fall on a plane described by:*

$$RG = r_{RG}R + g_{RG}G + b_{RG}B = 0 \quad (3)$$

In a two-dimensional chromaticity space plotted in barycentric coordinates r *and* b, ($g = 1 - r - b$), *the null plane projects to a straight line given by:*

$$b = \left(\frac{g_{RG}}{g_{RG} - b_{RG}}\right) + \left(\frac{r_{RG} - g_{RG}}{g_{RG} - b_{RG}}\right) r. \quad (4)$$

Similar derivations lead to straight null lines in chromaticity space for YB *and* LD.

Schrödinger estimated the coefficients for RG *and* YB *by using the chromaticity coordinates of psychologically unique hues. Unique blue and unique yellow are hues that are neither reddish nor greenish, and unique red and green are neither yellowish nor bluish. The line passing through unique yellow, achromatic white and unique blue was taken as the null line for* RG. *Similarly, the line passing through unique red, achromatic white and unique green was used to derive the coefficients for* YB. *The coefficients for* LD *were derived from a new kind of line, the alychne, representing the null of Exner's luminosity function. The three valences derived using König's* CMF*'s, when plotted on a wavelength axis, resembled Hering's qualitative valence curves. (Abney*[10] *includes an English translation of Hering's descriptions). In other words, if the opponent-hue mechanisms are linear combinations of cone signals, all the information about valence curves is contained in the location of the four unique hues.*

The relationship between the two theories thus depends critically on the collinearity of the pairs of unique hues with achromatic white. Subsequent measurements have invalidated this assumption. In 1939, Dimmick and Hubbard[11,12] *measured the chromatic locations of the unique hues. Spectral lights of 477, 515, and 583 nm were identified as unique blue, green, and yellow respectively. The line joining unique blue and yellow passed through the reddish side of white. Since all spectral reds have a yellowish tinge, the chromatic location of unique red was ascertained by mixing unique blue with a spectral red to cancel the yellow percept and obtain a color that was neither "yellowish" nor "bluish". Unique red was found to be complementary to 493.6 nm and not to unique green, implying that the line joining unique red to achromatic white was not collinear with the line joining achromatic white to unique green. Consequently, the yellow-blue hue system cannot be a linear combination of cone inputs. For each of the unique hues, Burns et al.*[13] *measured the chromatic locus as saturation was varied while luminance was kept constant. These loci did not fall on straight lines as would be required to generate hue valence curves. If a pair of unique hues, red-green or yellow-blue, is collinear with white, and the less saturated null hues fall on the straight line joining the three null points, then the valence curve for the opponent mechanism can be derived directly from that line. However, if a pair of unique hues is not collinear with white, then a valence curve can not be defined for the corresponding mechanism.*[8,14] *Further work with the hue-cancellation technique*[14,15,16] *has provided additional evidence that opponent mechanisms based on hue judgments are not linear combinations of cone signals.*

Schrödinger's mathematical analysis of opponent mechanisms is more general than his estimation procedure which was restricted to hue judgments. Linear postreceptoral mechanisms can be identified with straight null lines in chromaticity space as a general rule, and the coefficients can be estimated on the basis of a variety of measurements.[17,18] *A different approach to second-stage mechanisms was taken by Krauskopf et al.*[19] *who directly searched for three linear second-stage mechanisms defined with reference to cone signals rather than unique hues. As the result of selective desensitization experiments, the null lines for the opponent-mechanisms were identified as a constant S cone line and a constant L & M cone line. Electrophysiological measurements have shown that these lines correspond to null lines for the two classes of chromatic cells in Macaque lateral geniculate nucleus.*[20]

This paper also introduces the remarkable invention of the alychne. The alychne provided a graphical means of conceiving of lights of equal brightness. Schrödinger used Exner's brightness data. A standard luminosity function, $V(\lambda)$ *had been adopted by the CIE in 1924, based on a mixture of measurements involving flicker photometry and direct brightness matching.*[21] *However, the differences between comparisons made by the two procedures have become clearer,*[22] *and flicker photometry is now used to derive an additive metric for luminance. From the location of the alychne in the chromaticity diagram, Schrödinger theorized that S-cones do not contribute to* *luminance. This position has been supported by the evidence that S-cone signals do not contribute to flicker nulls,[23,24] except under exceptional circumstances.[25] The concept of the alychne has also been influential in practical color work. In the CIE[3] all-positive color-diagram, two of the primaries,* X *and* Z, *were placed on the alychne, i.e. defined to have zero luminance. All the luminance information is carried by* y(λ), *which is thus identical to the CIE* V(λ) *function.*

The paper concludes with a thought experiment about the evolutionary sequence of the color sense. Schrodinger speculated that the yellow-blue color system evolved earlier than the red-green system. This sequence is consistent with the molecular analysis that shows that whereas the S-cone gene is similar in basic structure to the human rhodopsin gene, it is very different from the genes for the long-wavelength sensitive pigments. The L and M-cone opsin genes have a very high degree of sequence homology, indicating a more recent differentiation.[26,27]

In this commentary, the treatment of developments in opponent-color theory has purposely been kept terse. Judd[2] summarized developments up to 1950, and a number of readings[28–33] can provide more details. Readers may also want to compare Schrödinger's opponent-axis color space (Fig 3) to later theoretically based color representations[34,35,36,20], and notice the elegant treatment of the luminance relationships between complementary colors[37] facilitated by this space.

QASIM ZAIDI
Department of Psychology
Columbia University
New York, NY 10027

REFERENCES

1. E. Schrödinger, Thresholds of Color Differences in Muller-Pouillets Lehrbuch der Physik, 2nd edition, Vol. 2, Lehre von der strahlenden Energie (Optik), Part 1 (1926). E. Schrödinger, Grundlinien einer Theorie der Farbenmetrik im Tagassehen. *Ann. Physik,* **63,** 397–447, 481–520 (1920). Translations in D. L. MacAdam, Sources of Color Science. M.I.T. Press, Cambridge, MA (1970).
2. D. B. Judd, Response functions for types of vision according to the Müller theory, Journal of Research National Bureau of Standards, Washington, DC, 42, 1–16 (1949).
D. B. Judd, Basic correlates of the visual stimulus, In *Handbook of Experimental Psychology, Chapter 22,* 811–867 S.S. Stevens, Ed., John Wiley and Sons, New York (1951).
3. CIE, CIE Proceedings, Cambridge University Press, Cambridge (1931).
4. H. Helmholz, Handbuch der Physiologishen Optik. 2nd edit, L. Voss, Hamburg, (1896).
5. J. C. Maxwell, On the theory of compound colours and the relations of the colours of the spectrum, *Phil. Trans.* **150,** 57–84 (1860). Reprinted *Color Res. Appl.* **18,** 270–287 (1993).
6. E. Hering, Zur Lehre vom Lichtsinne, Carl Gerold's Sohn, Wien (1878).
7. J. von Kries, Die Gesichtsempfindungen, in *Handb. d. Physiol. des Menschen.,* W. Nagel, Ed., 3, 102–282 (1905).
8. Q. Zaidi, Parallel and serial connections between human color mechanisms, in *Applications of Parallel Processing in Vision,* J. Brannan, Ed., North Holland, New York, 227–259 (1992).
9. G. Wyszecki and W. S. Stiles, Color Science, Wiley, New York (1982).
10. W. deW. Abney, Colour Vision, S. Low, Marston and Co., London (1895).
11. F. L. Dimmick, and M. R. Hubbard, The spectral location of psychologically unique yellow, green and blue. *Am. J. Psychol.* **52,** 242–254 (1939).
12. F. L. Dimmick and M. R. Hubbard, The spectral components of psychologically unique red, *Am. J. Psychol.* **52,** 348–353 (1939).
13. S. A. Burns, A. E. Elsner, J. Pokorny, and V. C. Smith, The Abney effect chromaticity coordinates of unique and other constant hues. *Vision Res.* **24,** 479–489 (1984).
14. J. Pokorny, V. C. Smith, S. A. Burns, A. Elsner, and Q. Zaidi, *Modeling Blue-Yellow Opponency,* Proc. 4th Int. Congr. Int. Color Assoc. (AIC), Berlin (1981).
15. J. Larimer, D. H. Krantz, and C. M. Cicerone, Opponent-process additivity. I. Yellow/blue equilibria and nonlinear models. *Vision Res.* **18,** 723–731 (1975).
16. M. Ikeda, and M. Ayama, Additivity of opponent chromatic valence. *Vision Res.* **20,** 995–999 (1980).
17. S. L. Guth, and H. R. Lodge, Heterochromatic additivity, foveal spectral sensitivity, and a new color model. *J. Opt. Soc. Am.* **63,** 450–462 (1973).
18. S. L. Guth, R. W. Massof, and T. Benzschawel, Vector model for normal and dichromatic color vision. *J. Opt. Soc. Am.* **70,** 197–212 (1980).
19. J. Krauskopf, D. R. Williams, and D. M. Heely, The cardinal directions of color space, *Vision Res.* **22,** 1123–1131 (1982).
20. A. M. Derrington, J. Krauskopf, and P. Lennie, Chromatic mechanisms in lateral geniculate nucleus of macaque, *J. Physiol.* **357,** 241–265.
21. K. S. Gibson, The relative visibility function in *Proc. of 1924 Session of Commission Internationale de l'*Eclairage, 232–238, Cambridge Univ Press, Cambridge.
22. R. M. Boynton, *Human Color Vision.* Holt, Rinehart and Winston, New York (1979).
23. V. C. Smith and J. Pokorny, Spectral sensitivity of the foveal cone pigments between 400 and 500-nm, *Vision Res.* **15,** 161–171 (1975).
24. A. Eisner and D. I. A. MacLeod, Blue-sensitive cones do not contribute to luminance, *J. Opt. Soc. Am.* **70,** 121–123 (1980).
25. A. Stockman, D. I. A. MacLeod, and D. D. DePriest, The temporal properties of the human short-wave photoreceptors and their associated pathways, *Vision Res.* **31,** 161–171 (1991).
26. J. Nathans, D. Thomas, and D. S. Hogness, Molecular genetics of human color vision: The genes encoding the blue, green, and red pigments, *Science* **232,** 193–202 (1986).
27. J. Nathans, T. Piantanida, R. L. Eddy and T. B. Shows, Molecular genetics of inherited variation of human color vision, *Science* **232,** 203–232 (1986).
28. F. C. Donders, Uber Farbensysteme. Archives of Ophthalmology, **27,** (1) 155–223 (1881).
29. E. Q. Adams, A theory of color vision. *Psychological Review,* 30, 56 (1923). X-Z planes in the 1931 ICI system of colorimetry. *J. Opt. Soc. Am.* **32,** 168–173 (1942).
30. G. E. Müller, Darstellung und Erklarung der verschiedenen Typen der Farbenblindheit. Vandenhoek and Ruprecht, Göttingen. (1924) Uber die Farbenempfindungen. *Z. Psychol., Erganzungsb,* **17,** 18 (1930).
31. D. Jameson, and L. M. Hurvich, Some quantitative aspects of an opponent-colors theory, I. Chromatic responses and spectral saturation. *J. Opt. Soc. Am.* **45,** 546–552 (1955).
32. L. M. Hurvich, and D. Jameson, Further development of a quantified opponent-colors theory, in *Visual Problems of Colour,* **2,** 693–723, NPL Symposium No. 8, Her Majesty's Stationary Office, London (1958).
33. C. R. Ingling, Jr., The spectral sensitivity of the opponent color channels, *Vision Res.* **17,** 1083–1089 (1977).

34. Y. Le Grand, Le seuils differentials de coleurs dans la theorie de Young. *Rev Opt Theo Inst* **28,** 261–278 (1949).
35. R. W. Rodieck, The Vertebrate Retina, Principles of Structure and Function. W. H. Freeman and Co., San Francisco (1973).
36. D. I. A. MacLeod and R. M. Boynton, Chromaticity diagram showing cone excitation by stimuli of equal luminance. *J. Opt. Soc. Am. A* **69,** 1183–1186 (1979).
37. D. L. MacAdam, Photometric relationships between complementary colors, *J. Opt. Soc. of Am.* **28,** 103–111 (1938).

I.

As is well known, two apparently very different conceptions of the color continuum confront each other at present, the three-color theory, usually linked with the name of Helmholtz, and the four-color theory, emphatically defended by Hering though its origin goes back to Aubert. The former, in keeping with the indisputable three-dimensionality of the color continuum, claims that every color is to be regarded as a mixture of three basic colors, a basic red, a basic green and a basic blue (or violet). This is in contradiction to the direct perception of the naive observer, that pure yellow is also a psychologically homogeneous color sensation in addition to red, green and blue, in which with the best of will he cannot perceive a mixture of equal quantities of red and green, as he should according to the three-color theory. Likewise the perception of pure white does not seem in the least related to the colors named, but is psychologically homogeneous and not to be further analyzed, whereas according to the three-color theory it is produced by mixing equal quantities of all the three basic colors.

On the other hand, the four-color theory adheres closely to the psychological color arrangement. According to Hering, every individual color, in addition to a "white valence," possesses two colored valences of which the first is called the "red-green valence" and is either red or green for a certain color, whereas the second, called the "blue-yellow valence," is either blue or yellow. That corresponds entirely to the psychological observation that, on the one hand, the perceptions of red with green and blue with yellow are absolutely incompatible, whereas, on the other hand, every color perception is classified according to its color tone between one of the two colors of the former and one of the two colors of the second "opponent-color pair." An increase in the white valence corresponds to a brightening and, if the color valences are unchanged, at the same time to a whitening (desaturating); a decrease in the white valence, to a darkening and blackening. However, addition of color valence to a pure white valence also changes the brightness, in so far as red and yellow have a specific "brightening" effect, and green and blue a specific "darkening" effect.

This theory, which led Hering to a very definite physiological-chemical interpretation of the visual process, still has many adherents, especially among the psychologists. Others object that it is scarcely compatible with the empirical three-dimensionality of the color continuum, if only because of a superfluous increase in variables, and for that reason no quantitative examination on colors can be adequately expressed in its terminology. von Kries took a conciliatory position in his so-called zone theory.[1] According to it, the three-color conception applies to the physiological process on the retina, and the four-color conception, on the other hand, to a more centripetal "zone" of the visual organ, which would explain its closer connection to the psychological color arrangement.

To me this von Kries interpretation of the theory seems very probable. What I wish to show in the following, is entirely independent of the deeper conception of the physiological substratum of the visual process. It is a question of the mere determination that, purely superficially, the relationship between the two theories—the three and the four-color theory—may be regarded as extremely simple, i.e. as a mere transformation of the variables. From the purely mathematical standpoint, the true facts are not particularly intricate, but to my knowledge they have never been expounded with entire clarity and certainly have never been recognized by many, otherwise the discussion would have taken a different course. The evident numerical contrast, which seems substantial because of the notation selected by two theories, may also have contributed to the ambiguity.

Let us consider the conventional color triangle of the Helmholtz theory, Fig. 1. In it the three basic color portions of a color, which we shall designate in short as their coordinates x_1, x_2, x_3, geometrically constitute the projective or triangle coordinates of the point corresponding to the color in question, the center of gravity of the triangle being selected as the "unit point" (in accordance with projective geometry [geometric stipulation]). However, whereas the coordinates of projective geometry are only significant as proportional numbers, the color coordinates further have the absolute meaning that their sum $x_1 + x_2 + x_3$ indicates the mass which must be attributed to the color point in order to find the result of a color mixture according to the known center of gravity construction. A further, by no means necessary, stipulation on the units of the standard lights to be selected places white in the center of gravity of the triangle (physico-physiological stipulation).

The three-color coordinates of the spectral lights in the diffraction spectrum of the earth's sunlight, plotted as functions of the wave lengths, are called the "basic perception curves" or fundamentals.[2] The color points in question—for a spectrum of any desired distribution of intensity—lie on the dashed "spectral curve" in Fig. 1.

Instead of restricting them to the coordinate triangle selected here (which has a deeper meaning according to the Helmholtz theory as a result of measurements on dichromats), the spectral colors and consequently also all the other colors can be referred to any desired other triangle which corresponds to a linear-homogeneous transformation of the coordinates. Naturally this also changes the form of the color matching function (CMF), since every new CMF is a superposition (with certain constant

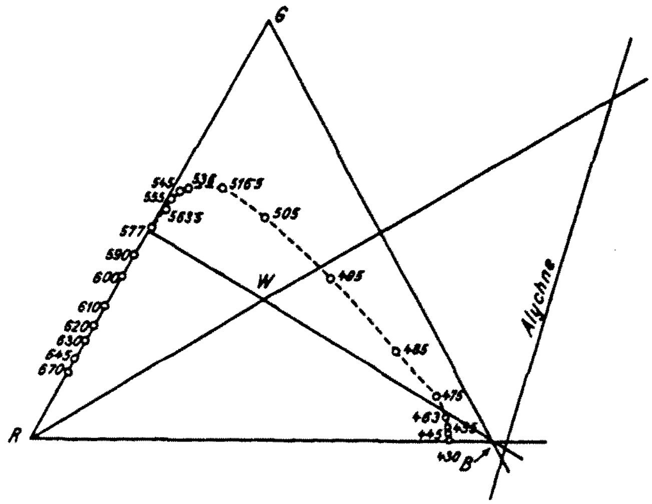

FIG. 1. König color triangle.

coefficients) of the old. It is a question whether the coordinate triangle can be given such a form that the new CMF may be regarded directly as the Hering valence curves—for which, they need only satisfy certain qualitative requirements, since no quantitative information on the valences is known from the Hering theory.

The Hering valence values have the following qualitative course along the spectrum. The long-wave, homogeneous lights from the red end to about λ = 575-nm ("primary yellow") first have an increasing and then a decreasing red valence, and in addition increasing yellow valence. In primary yellow the red valence disappears and passes into the green valence. Up to λ = 495 ("primary green") the yellow valence diminishes while the green valence increases; from there to λ = 472 ("primary blue") there is decreasing green in addition to increasing blue valence. From that point there is again a red valence in place of the green valence; it passes through a secondary maximum and then disappears simultaneously with the blue valence at the violet end. Primary yellow and primary blue are complementary, i.e. in the color triangle, they are on the same straight line through the white point. There is no physically homogeneous "primary red"; it is prepared from extreme spectral red with a small addition of blue, which makes the mixture complementary to primary green. Along the spectrum, the white valence is said to produce the directly observed spectral distribution of brightness, while taking into consideration a certain "brightening" influence of the red and the yellow valences and a "darkening" influence of the green and the blue valences.

The following wave lengths must now be located in the König color triangle or, better still, in the König fundamentals, in which they are represented more accurately as the abscissa of the intersections of two fundamentals in each case: first, that which is complementary to the König basic blue (long-wave intersection of the red and green curve, about λ = 577); second, the wavelength which is complementary to basic red (intersection of the green and the blue curves, about λ = 497); third, the wavelength of basic blue (short-wave intersection of the red and the green curves, about λ = 469). These three wavelengths agree with the Hering primary yellow, primary green and primary blue within the limit of accuracy with which, on the one hand, the intersection of the fundamentals can be determined according to various methods,[3] and on the other, of the Hering primary color for the determination of which the psychological observation of a normally adjusted eye is used, e.g.: this yellow is pure yellow with neither a reddish nor a greenish tinge. Naturally this in the cause of very approximate agreement of the Hering primary red with the König basic red, so that we can say: The Hering primary colors agree in color tone with the König basic red and basic blue and their complement whereas the basic green (the wavelength of which, by the way, is very close the complement of the basic red), plays no marked role in the Hering theory.[4]

The connecting lines passing through the white points to the two complementary pairs of colors are the straight lines RW and BW of Fig. 1. If they are included in the sides of a new coordinate triangle, the geometric significance of the projective coordinates indicates very clearly that the two new "CMF" attributed to them will show the correct change in signs for interpretation of the positive ordinates of the one (attributed to BW) as red valences, and their negative ordinates as green valences according to Hering, and likewise the positive ordinates of the other new CMF (attributed to RW) as yellow valences, and their negative ordinates as blue valences. That follows from the fact that every projective coordinate of a point is proportional to a perpendicular line drawn from this point to one of the sides of the basic triangle, and the sign changes when the point passes from one end to the other of the side of the triangle in question.

For the present the third side of the triangle must still be selected. It is a question of how it should be placed in order that the third new CMF may be regarded as white valence according to Hering.

In order to clarify the selection of this third side, we shall insert an intermediate step. First let us select it in such a way that the distribution of brightening in the diffraction spectrum of the sun is the third new CMF. That is possible and extremely simple from the comprehensive studies of Franz Exner[5] on the calculation of the brightness of any color from the (König) fundamentals. According to Exner, the brightness is expressed homogeneously and linearly by the basic perception portions x_1, x_2, x_3:

$$h = \alpha x_1 + \beta x_2 + \gamma x_3.$$

For the three coefficients which are only to be regarded as proportional numbers, Exner found the values:

$$\alpha = 1 \quad \beta = 0.756 \quad \gamma = 0.024. \tag{1}$$

Evidently, it is only necessary to select for the third side of the new triangle the straight line, which, referred to the König triangle, satisfies the equation:

$$\alpha x_1 + \beta x_2 + \gamma x_3 = 0. \qquad (2)$$

The third of the new coordinates is then proportional to this linear term, and consequently, to the brightness. This straight line (2) is constructed in Fig. 1 and, for reasons to be explained directly, is called alychne ("lightless").

In the König triangle the straight lines RW and BW have the equations:

$$x_3 - x_2 = 0$$
$$x_2 - x_1 = 0. \qquad (3)$$

The conversion equations for transformation to the chromaticity triangle formed from the three straight lines just mentioned, therefore, have the following form:

$$x'_1 = a(x_3 - x_2)$$
$$x'_2 = b(x_2 - x_1)$$
$$x'_3 = c(\alpha x_1 + \beta x_2 + \gamma x_3). \qquad (4)$$

Here a, b and c are at first arbitrary. It is convenient to determine them from the requirement that

$$x'_1 + x'_2 + x'_3 \equiv x_1 + x_2 + x_3 \qquad (5)$$

must be identical. Thereby the mass with which each color enters into a center of gravity construction is preserved, and the barycentric coordinates x'_i for the new color triangle apply to the color continuum drawn in Fig. 1, without requiring any distortion. Requirement (5) leads to:

$$a = \frac{\alpha + \beta - 2\gamma}{\alpha + \beta + \gamma},$$
$$b = \frac{2\alpha - \beta - \gamma}{\alpha + \beta + \gamma},$$
$$c = \frac{3}{\alpha + \beta + \gamma}. \qquad (6)$$

With (1) substituted in (4), this gives the numerical equations:

$$x'_1 = 0.960\,(x_3 - x_2)$$
$$x'_2 = 0.685\,(x_2 - x_1)$$
$$x'_3 = 1.685 x_1 + 1.274 x_2 + 0.040 x_3. \qquad (4')$$

Before we discuss the representation developed in this way, we wish to add some more remarks about the interesting Exner straight line, Equation (2), which we called alychne.

This line is the geometric position of colors of vanishing brightness, which at first seems very mystic. Naturally, these colors of vanishing brightness are all virtual; the straight line does not intersect the real color plane. Two such colors are chosen in the representation (4′) as primary colors; the third primary color is white. Actually, these colors have by no means anything mystic about them. The result of adding such a color to a mixture cannot only be stated, but can also be readily demonstrated, e.g. with a color top; more easily than the addition of a König fundamental color. It is merely a matter of changing the color tone (and possibly the saturation) while keeping the brightness constant. For example, if we take two colored papers of the same brightness and mix them in different ratios on a color top, all of these mixtures differ merely in the varying content of a definite color located on the straight line (2). The position of this color is preserved by intersecting the straight line (2) with the straight line which connects positions of the two colored papers.

As may be seen, the alychne passes very close to the König basic blue. That is the result of a low value of the specific brightness γ of the basic blue. The selection of this basic blue, of course, is quite conventional; the spectral standardizations permit far more play than for basic red and basic green, because blindness to blue occurs almost only in seriously diseased visual organs with which it is impossible to experiment accurately and continuously. In any event, such a small displacement of the blue point, which is sufficient to transfer it to the alychne is entirely permissible. The König fundamental would be changed only very imperceptibly thereby; it can even be arranged in such a way that the abscissas of their intersections (which, of course, are also confirmed by other methods),[6] are preserved by shifting the blue point on the straight line BW outward. This change would result in the practical advantage that γ would equal 0, i.e. the brightness of every color could then be calculated from its red and its green content alone, without necessarily taking into consideration the blue content, which even now contributes almost nothing to the brightness. Further, the determination of this small blue coefficient by different observers is so uncertain and varying (Kohlrausch[7] found 0.047, Ives[8] 0.011, both expressed in terms of $\alpha = 1$), that it would not be risky to consider these values as remainders of errors which possibly may be negative occasionally. From the standpoint of the Young-Helmholtz theory, this conception that the perception of blue would, strictly speaking, have the brightness zero, a conception which may practically be regarded as proved by experiments, as already said, would be of considerable importance. The "blue process" of this theory would then probably have to be regarded as being substantially different than the other two, in so far as it only modifies the quality of light perception, but leaves its intensity unchanged. However, we do not wish to make use of this observation for the present, in order that it may not seem as if we wish to or have to distort the results gained by observation.

Let us turn again to Equation (4′). Fig. 2 represents the new "basic perception curves" corresponding to these equations, i.e. the three values x'_i for the spectral lights in

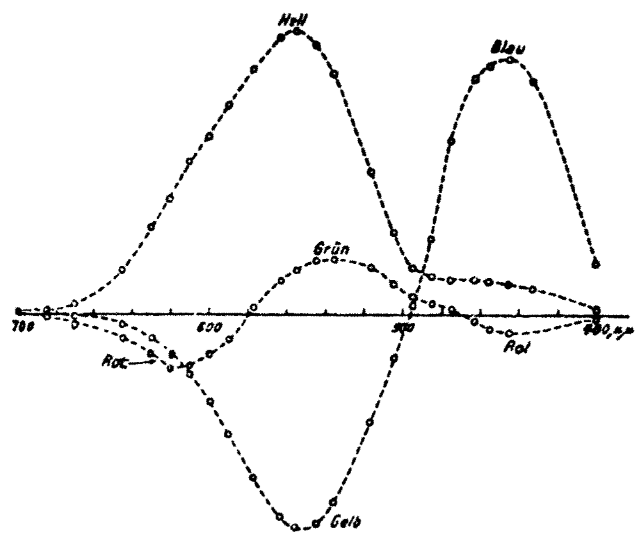

FIG. 2. Hering valence curve, derived from König's fundamentals.

the diffraction spectrum of the sun as functions of the wave lengths. For the x_i the original values of König were used. As was anticipated, the first two curves give a suitable picture of the Hering color valences if the negative ordinates of x'_1 are interpreted as red valences, the positive ordinates of the same curve as green valences, and further the negative ordinates of x'_2 as yellow valences, and the positive ordinates as blue valences. The intersections with the γ-axis are primary yellow, primary green and primary blue, whereas the primary red naturally cannot appear since it does not occur in the spectrum. The third curve (which for lack of space is shown on half the ordinate scale in the figure) simply gives the distribution of brightness, and is exactly identical with that of Franz Exner, i.e., calculated from the original König data by means of his own values of α, β, γ (also used by us).[9] The values of x'_3 cannot as yet be identified with the Hering white valences, because the latter (as Hering's school claims) do not alone determine the brightness, but the color valences, when added are claimed to have a specific brightening or darkening effect. However, it would be easy, and possible in ∞^2 ways, to change the curves by a second linear transformation in such a way that the first two may continue to be regarded as Hering's color valences, and the latter as Hering's white valence; now this could be done by taking into consideration the asserted specific brightening nature of the color valences. In other words, this specific brightening nature of the color valences might be incorporated purely formally. For this purpose it would merely be necessary to write:

$$x''_1 = -Ax'_1$$

$$x''_2 = -Bx'_2$$

$$x''_3 = x'_3 + Ax'_1 + Bx'_2$$

$$\text{Brightness} = x''_1 + x''_2 + x''_3 \, (=x'_3). \qquad (7)$$

A and B are purely arbitrary positive coefficients. Naturally, conversely as before, the positive x''_1 are to be considered as red valences, and the negative x''_1 as green valences; and similarly for x''_2. The brightness is the sum of the color coordinates. Red and yellow have a brightening effect and blue and green a darkening one. (Naturally, the opposite may readily be produced by a negative sign of A and B. There are plenty of advocates of this bizarre conception in the recent literature.)[10]

According to the conception originally advanced by Hering's school, the white valences are given by the twilight values. Accordingly the A and B should be determined in such a way that x''_3 coincides with the curve of the twilight brightnesses, the maximum of which is in the diffraction spectrum of the sun at about $\lambda = 505$. As von Kries emphasized very emphatically,[11] this conception of the twilight values as the pure white valence freed of the influences of the color valences is no longer acceptable; it is refuted by the fact that when the intensity is diminished without rendering perceptible a difference in color tone, certain light mixtures equal in daylight and colorless to him show the deuteranope a Purkinje phenomenon of the enormous amount of 1:100. This fact also prohibits the explanation of the great differences in twilight and daylight brightness to the trichromat through elimination of the color valences. To this argument our present consideration adds another. A closer observation makes it impossible to select the above values A and B in such a way that x''_3 represent the twilight brightness. Otherwise it would have to be possible to construct the twilight curve from the original König fundamental by linear combination with suitable (positive or negative) coefficients. That this is impossible may be recognized almost at first sight, since the maximum of the twilight curve lies at a point where all three fundamentals have relatively small values. I have attempted the adaptation by trying to produce agreement at three points at the crest of the twilight curve and at two points at a medium height. Entirely irrational values were found for the coefficients which produce very high negative values at other points of the superposition curve.

The only reasonable interpretation of the white valences from the standpoint of the four-color theory is probably that of Kries, i.e. as periphery values.[12] Here it appears that, with recession of the color quality in eccentric vision, practically no change in the brightness takes place. For, when careful attention is given to full adaptation to brightness in determining the peripheral colorless brightness and the macula tinction is taken into consideration in the case of strictly central observation, no distinct difference can be determined between the central or paracentral colored and the peripheral colorless brightnesses.[13] Our representation (4′) in which, in addition to white, two colors "free of brightness" on the alychne are selected as primary colors, exactly corresponds to this. Therefore, there is no reason for carrying out further transformations (7) for the present; the x'_i introduced through (4′) appear as the simplest quantitative expression of the four-color theory.

This expression offers some very unexpected, interesting characteristics. Recently, it has been emphasized by various authors[14] that it is much more suitable for actual constructions in the color triangle to use a right-angled instead of an equilateral triangle, possibly with different scales for the abscissa and the ordinate. As right-angled cartesian coordinates, it would then be simply necessary to plot:

$$x = \frac{x_1'}{x_1' + x_2' + x_3'}$$

$$y = \frac{x_2'}{x_1' + x_2' + x_3'}, \tag{8}$$

whereas the quotient of the third color coordinate is found by the sum of the color coordinates, either algebraically as $1 - x - y$ or geometrically as the distance of the color point from the hypotenuse (drawn through the points (1,0) and (0,1) on the axis). If we do this with the color coordinates x_i', we obtain Fig. 3, in which, in addition to the spectral curve, the original König basic colors are also plotted. At the source is white. According to their sign, the coordinates x, and y give the green, red, blue and yellow content. The distance from the hypotenuse (which hypotenuse we, of course, recognize as alychne) measures the brightness of the unit of the color in question. As will be remembered, we kept this unit—the sum of the coordinates—invariable in the transformation (4′); therefore, it has the same value as that attributed to it by König and is also the mass in Fig. 3 which must be attributed to the color point in making a center of gravity construction. For example, from the figure it is readily possible to answer the common question: How do the brightnesses of two (simple or compound) lights behave in a complementary mixture? For this purpose it is only necessary to divide for each of the two lights its distance

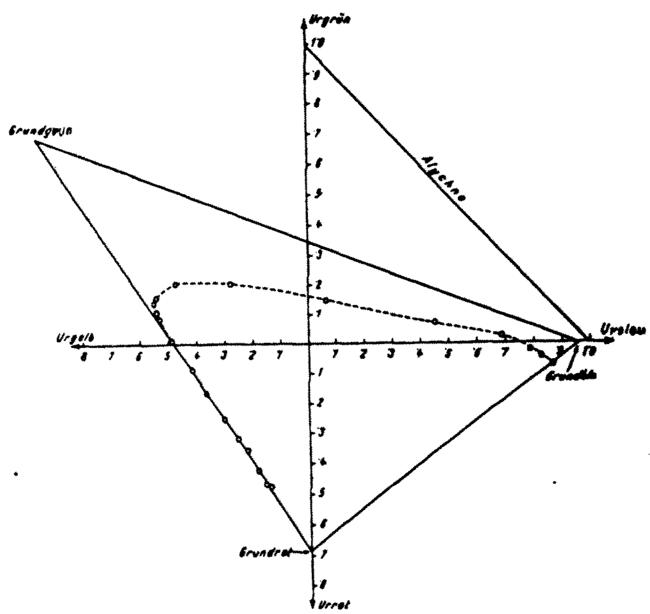

FIG. 3. Hering color valences per unit quantum, plotted as rectangular cartesian coordinates.

from the alychne by its distance from the white point. If two equally bright complementary colors are wanted, they can be found on a conical section one focal point of which is in the white point and the alychne of which is its (corresponding) directrix.

These are all relatively trivial consequences which, by the way, with due alteration of details are already included in König's color triangle when the alychne is constructed from Exner's data. The only simplification produced by the geometric rearrangement is that the two heavy lines (altitudes) of the König triangle which come into consideration are now perpendicular to each other and that the alychne, the construction of which was formerly relatively inconvenient and required the use of the special numerical values α, β, γ, has now become a straight line at 45°. Naturally, it might be expected that the König triangle would lose its simple form, and can only be fitted correctly into the figure with the aid of the numerical values α, β, γ.

Strange to say, that is not the case. In consequence of the chance numerical values of the coefficient in (4′), which in turn depend on the empirical values of the Exner brightness coefficients α, β, γ, the König triangle formed by the primaries red, green, and blue of Fig. 3, the angles of which are naturally constructed correctly according to Equation (4′) as the points which correspond to the triplets (1,0,0), (0,1,0), (0,0,1) of the nonprimed coordinates, proves purely by chance to be rectangular with the right angle at the basic red. Likewise, the short sides of this triangle by chance are in the ratio of $1:\sqrt{2}$. The latter is not a new "chance," but is a necessary consequence of the former when it is considered that the coordinates x, y of the corners of the König triangle according to (4), (5) and (8) in any event must have the form:

$$(0, -b)$$
$$(-a, b)$$
$$(a, 0),$$

in which the particular values of α, β, γ, just result in

$$a = 0.960$$
$$b = 0.685.$$

Once this has been recognized, let us consider that, when there is to be a right angle at basic red, the two right-angled triangles below the x-axis, as partial triangles of the right-angled triangle the hypotenuse of which lies in the x-axis, must be similar to each other and to the latter triangle. (And vice versa, when these two triangles are similar to each other, there is a right angle at the basic red, and they are also similar to the triangle composed of them.) The short sides of these two small triangles (absolute values) are:

$a/2$, b for the left one,

b, a for the right.

The similarity, therefore requires:

$$a = \sqrt{2}.$$

This relationship is fulfilled by the above numerical values to within less than 1%. Then, however, the shorter side (on the right) of the König triangle is:

$$\sqrt{a^2 + b^2} = b\sqrt{3},$$

and its longer side (on the left)

$$2\sqrt{b^2 + a^2/4} = b\sqrt{6}.$$

Therefore, the König triangle is also similar to all those previously mentioned; it also has the ratio of the short sides of $1:\sqrt{2}$, as claimed.

If the short sides of the König triangle are explained as units of measurement of the abscissa and ordinate in the rectangular system of coordinates formed by these short sides, the cartesian coordinates of the color points have the following value with regard to this system of coordinates:

$$\frac{x_3}{x_1 + x_2 + x_3},$$

$$\frac{x_2}{x_1 + x_2 + x_3},$$

i.e. the König blue value, or the König green value by the sum of the coordinates, whereas the corresponding quotient for the König red value is readily found by subtraction of these two values from 1 (or also from the distance from the hypotenuse of the König triangle).

Our representation, therefore, gives the following information. It represents in one figure the color valences according to the Helmholtz-König three-color system and according to the Hering four-color system, both in the simplest and most convenient form for practical construction as right-angled cartesian coordinates, in which the Hering system of coordinates simply forms the altitudes of the right-angled König triangle. At the same time it offers the possibility of direct and clear presentation of the specific brightnesses of all lights (simple and compound) according to the measurements of Exner. That this combination is possible at all is due to the empirical value of the specific brightnesses of the basic colors determined by Exner, above all (as we shall see in a moment) on the ratio of the green brightness to the red brightness.

In order to construct our double coordinate system, only one special numerical fact is required after what has been said so far, e.g. the fact that basic blue must be located at $x = 0.960$. This requirement could be eliminated if the Exner γ-value would be exactly 0 instead of 0.024. In that case the basic blue would lie on the alychne, i.e. at $x = 1$. I have already stated above that no observation prohibits this slight displacement of basic blue (with a slight conversion of the red and green curves). I did not follow up this statement before in order not to arouse the suspicion that the simplification produced so far was linked to this step. Even now I shall not follow this course, but an even simpler one. The small persistent γ-value which has resulted from the Exner color disc experiments, does not seem to be a sufficiently certain clue for shifting the basic blue or for how this is to be done, in order to consider it as a brightness color with better approximation than is the case according to its present determination. In practice the slight blue brightnesses which have so far been conscientiously taken into consideration prove to be so inconsequential that it is impossible to confirm in any way that an error is committed when they are simply ignored. The distribution of brightness in the spectrum of the sun calculated by Exner is changed exceedingly slightly by omitting the blue brightness, far less than the difference between the calculated and the observed distribution of brightness at the short-wave end. (In addition the difference is diminished by omitting the blue brightness!) Even though in these most saturated of all blue tones, the contribution of the blue content to the brightness cannot be definitely proved, it can be seen how much less this is the case in other colors. We have already mentioned that Ives found a blue value less than half as great as that of Exner; and we add that, on the other hand, his value of 0.750 for the important green value is entirely in agreement with Exner (0.756); red value always equal to 1, as above.

In short, for the present we feel justified in letting $\gamma = 0$ even without compensating this by a change of fundamental. The numerical coefficients of the equations (4′) then become the following:

$$x_1' = x_3 - x_2$$

$$x_2' = 0.708(x_2 - x_1) = \frac{1}{\sqrt{2}}(x_2 - x_1)$$

$$x_3' = 1.708x_1 + 1.292x_2$$

$$= \left(1 + \frac{1}{\sqrt{2}}\right)x_1 + \left(2 - \frac{1}{\sqrt{2}}\right)x_2. \qquad (9)$$

The simultaneous rectangularity of the two systems of coordinates in Fig. 3 is not only preserved, but if anything, becomes more exact; the $\sqrt{2}$ ratio applies to the new coefficients with even greater approximation:

$$1:0.708 = 1.41 \cdots = \sqrt{2}.$$

If on the basis of (4) and (6) we now study somewhat more closely the question of how this relationship comes about, we find ($\alpha = 1, \gamma = 0$):

$$\frac{2 - \beta}{1 + \beta} = \frac{1}{\sqrt{2}}$$

or calculated:

$$\beta = 5 - 3\sqrt{2} = 0.75736 \cdots \ (\textit{observed}\ 0.756).$$

The small difference from the observed value of the green brightness is due to rounding error.

For drawing the new figure, all of the conversions given in the foregoing (which are very tedious in spite of their simplification, since they must always be carried out numerically for some twenty positions along the spectrum) now become entirely superfluous. The procedure is as follows: The color coordinates of the spectral colors, or any other colors which are of interest, are drawn into a right-angled color triangle scaled in the ratio of $1:\sqrt{2}$, the relative blue value (i.e. divided by the coordinate sum) being plotted on the abscissa and the relative green value on the ordinate. A pair of normals intersecting at right angles, on both of which the distance between the white point and the blue point is considered as the unit, is drawn into this triangle. Finally, the Exner straight line is drawn through the blue point, at an angle of 45° to the two normals. This diagram then expresses the König-Helmholtz and the Hering valences and the brightnesses of the colors in exactly the same way in which this was described with reference to Fig. 3. The difference between the two figures is probably within the accuracy attainable at present of the data used here.

For the present I shall not attempt to answer the question as to whether the very special simplification which the construction of Fig. 4 has undergone through the special Exner-Ives numerical value of the relative green brightness, has a deeper significance for the conception of the visual process. It cannot be denied that Kohlrausch found quite a different numerical value for the green brightness, namely 0.618 (as compared with 0.756) and that Abney[15] also claims that there is a strong individual variation in the coefficient, although the latter

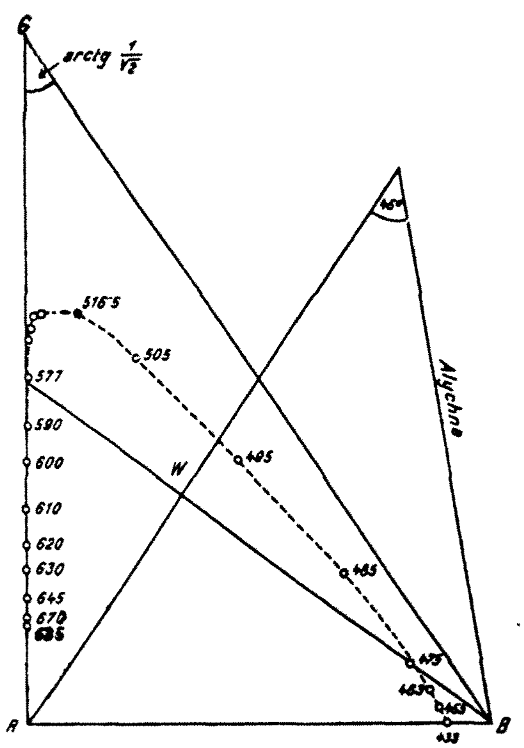

FIG. 4. Direct construction of Fig. 3 (with inconsequential simplification) from the untransformed König data.

used somewhat unsatisfactory methods which led him to the contradictory assertion that the anomalous trichromatism consists merely in an abnormal ratio of green brightness to red brightness. (Actually, from the purely logical standpoint, the brightness coefficients can have no influence on an exact color equation, i.e. when it is found experimentally that two individuals do not accept each others color equations,[16] the adequate theoretic expression of this fact can not consist merely in a difference in the brightness coefficients.)

With a different value of the green coefficient than that of Exner-Ives the construction of Fig. 4 can always be so arranged that the altitudes of the right-angled König triangle intersect each other at right angles by imparting to it the increased height of $\sqrt{2}:1$; however, then the two Hering axes do not have the same scale; consequently, the alychne is no longer a straight line at 45° in the Hering system of coordinates.

With the same reservation regarding a somewhat deeper significance I wish to mention the further striking circumstance that in Fig. 4 the descending branch of the spectral curve coincides with very great accuracy with the hypotenuse of an equilateral, right-angled triangle, the horizontal side of which is RB.

On the other hand, it seems to me of unquestionable importance for the more thorough understanding of the visual process to recognize that the primary blue—whether it be that accepted here or one substituted for it with fully equal justification—is a color free of brightness. If that is correct, then the three "components" of the visual organ are probably not equivalent after all, the conception of the symmetrical role of the three components suggested by the theory is invalidated, and that would have to be taken into consideration in the search for the physiological substratum.

II.

I shall now briefly discuss the relationship of the three-color theory to the four-color theory from still another viewpoint, which, however, I cannot as yet correlate directly to the graphic representations described in the foregoing.[17]

In the introduction it was emphasized as the sharpest contradiction between the two conceptions that, according to the one, white and yellow are almost (white) and entirely (yellow) equivalent as psychologically simple perceptions to the other three colors (red, green, blue), but according to the other white and yellow correspond only to the simplest mixture ratios, namely 1:1:1 (according to definition) and 1:1:0 (according to experiment).

If we now consider the phylogenetic origin of an organ perceiving light, we come to the nearly self-evident assumption that in the primitive beginnings its function was limited to reaction to other radiation—naturally only to the radiation of a limited frequency range; that appears even from the immense difference in the physical effect from the radio waves to the γ-rays. The second stage in the adaptation may be regarded as the one in which within this frequency range the organ commences to react differently qualitatively to different frequencies. A process of development similar to that of the ear—"individually" different reactions for every small frequency range in the finest graduation—is entirely impossible for an organ of light, because under natural conditions it is never exposed to approximately pure frequencies. (If thin layers of luminous gas masses played an important role biologically, this might be different.) Since the prerequisites for the development of an individual power of differentiation of the individual frequencies are lacking, it is apparently most probable that a summary power of differentiation by which a predominance of the short-wave or the long-wave components over the "normal" composition of light (sunlight) is indicated by a special characteristic of perception. This characteristic of perception is the blue-yellow series with neutral white as transition point, the primitive simplicity of which will naturally not be lost because deviations from the norm gradually reach a perceptible emphasis. The state attained is that of dichromatism, which we observe in the normal retina periphery, with the exception of those of partially color blind people, and as it seems, also in many animals (insects).

A second step, entirely analogous to the first, leads to trichromatism. The division due to predominance of short- or long-wave radiation, which formerly affected the entire perceived frequency range, is repeated a second time in the range of the long-wave lights alone. Thus yellow is separated into red and green, as white was previously separated into blue and yellow. Yellow now loses its already established simple color character by a further differentiation, as white did previously. Yellow has the same relation to the color pair red-green that white has to the color pair blue-yellow, namely, it is its neutral transition point.

Although this conception of the successive origin of the color sense cannot be proved by strictly quantitative methods, nevertheless it seems to me to explain the roots of the debate about the "simplicity" of white and yellow, and their role as basic perceptions. White and yellow are true basic perceptions, not recent but ancestral, one from the monochromatic, the other from the dichromatic stage. Of the basic perceptions still "undivided" at present, one (blue) is derived from the dichromatic stage, and the other two (red and green) are of more recent acquisition. This explains why the two latter are still subject to the most disturbances and "setbacks"; it also explains why no disturbances of the perception of blue alone occurs as a physiological anomaly, although it should be just as possible according to the three-component theory; the line of ancestors passed through no such stage of development. With "blue-yellow blindness," "red-green blindness" also occurs regularly; these are cases of inborn color-blindness. Cases have actually been observed in which it seemed to be due not to pure red-vision, but to a degeneration, or, as we prefer to say, an atavism of the "daylight apparatus."[18]

If this conception is correct, we have on the peripheral zones of the normal retina a record of the successive stages of our ancestral methods of vision, the oldest on the outside. Due to the biological importance of foveal vision, it seems entirely comprehensible that new differentiations originate in the fovea and only gradually radiate to the periphery.

ACKNOWLEDGMENTS

Although the foregoing work is of a purely theoretic nature, I wish to express my deep felt thanks to the Stiftung fur wissenschaftliche Forschung at the University of Zurich for their munificence in making it possible for me to undertake the study of some problems on the subject of color.

1. See, for example Nagel's Handbuch der Physiologie de Menschen, Vol. 3, Article by J. v. Kries. The Visual perceptions, p. 269 (Braunschweig, by Vieweg, 1905). In the following this basic article is cited as J. v. Kries, G.E.
2. A. König and C. Dieterici, *Zeitschr. f. Psych. u. Phys. d. Sinnesorgane* **4,** 241–347, 1892.
   A. König, Ges. Abh., Leipzig, J. A. Barth, p. 310, 1903.
3. See these reports (2a) **111,** June 1902 (F. Exner); **115,** p. 39, 1906 (O. Steindler); **122,** p. 1915, 1913 (L. Richtera); **123,** p. 629, 1914 (F. Hauer).
4. A. König and C. Dieterici, l.c. Ges. Abh., p. 317, Note.
5. F. Exner, These reports (2a) **127,** p. 1829, 1918; **129,** p. 27, 1920.
6. See the above citations (F. Exner, O. Steindler, etc.).
7. K. F. Kohlrausch, *Phys. Zeitschr.* **21,** 423, 1920.
8. Journ. Franklin Inst. **195,** p. 23, Jan. 1923 (H. E. Ives).
9. F. Exner, These reports (2a), **129,** p. 40 et seq. Table IV, and Fig. 2, full line curve, 1920. I could not use the correction of the König fundamental which Exner makes further because it results in defective appearance (concavities outward) of the spectral curve in the color diagram.
10. F. W. Fröhlich, Grundzuge einer Lehre vom Licht und Frabensinn; Jena by G. Fischer 1921 (p. 46, 51, 82).
11. J. v. Kries, G.E., p. 192f.
12. J. v. Kries, G.E., p. 203.
13. See the discussion by J. v. Kries, Zeitschr. f. techn. *Physik* **5,** 327, 1924, especially p. 340.
14. *Journ. Opt. Soc. Amer. and Rev. Scient. Inst.* **6,** p. 527, August 1922. -Guild, *Trans. Opt. Soc. London* **26,** 139, 1925.
15. W. Abney, *Proc. Roy. Soc. A.* **83,** 462; **84,** 449, 1910; **86,** 42, 1911; **87,** 326, 1912.
16. Which is the case in normal and anomalous trichromats, as is known.
17. See also *Die Naturwissenschaften,* **12,** 927, 1924.
18. F. Exner, These reports (2a), **131,** 636, 1922.