Pergamon

Vision Res. Vol. 34, No. 10, pp. 1279–1299, 1994

0042-6989/94 $7.00 + 0.00

Luminance Noise and the Rapid Determination of Discrimination Ellipses in Colour Deficiency

B. C. REGAN,* J. P. REFFIN, J. D. MOLLON*

Received 16 July 1993; in revised form 1 October 1993

A computer-controlled test of colour vision is described, in which luminance noise and masking contours are used to ensure that the subject's responses depend on chromatic signals. The test avoids the need—common to most computer-controlled tests—to define equiluminance for the individual subject before the colour test itself can be administered. The test achieves a good separation of protan and deutan subjects and reveals the large range of chromatic sensibilities among anomalous trichromats. As a population, dichromats had higher thresholds on the *tritan* axis of the test than did normals. In an extension of the test, full discrimination ellipses were measured for normal and colour-deficient observers. The nature of anomalous trichromacy is discussed and the possibility is raised that hybrid genes, resulting from genetic recombination, may code for incorrectly labelled or functionally impaired molecules.

Colour vision Colour deficiency Discrimination ellipse Dichromacy Anomalous trichromacy Genetics

INTRODUCTION

In the natural world, it is in finding fruit among foliage that the colour-deficient observer is especially handicapped. He is seldom challenged by the juxtaposition of differently coloured but equiluminant surfaces. For it is rare in the natural world that one surface lies in front of another in such a way that both have the same reflectance, both lie at the same angle to the incident illumination, and the nearer throws no shadow on the farther. Rather, the colour-deficient observer is thwarted when he must find an object that is defined by chromaticity but lies in a variegated background where lightness is varying randomly—as will be the case when masking objects present varying angles to the illuminant or where the illuminant is itself dappled (Mollon, 1987, 1989). In one of the earliest reports of colour blindness, Huddart (1777) wrote tellingly of the shoemaker Harris: "Large objects he could see as well as other persons; and even the smaller ones if they were not enveloped in other things, as in the case of cherries among the leaves". And a modern survey found that colour-deficient people frequently experienced difficulty in finding berries or fruits amongst foliage (Steward & Cole, 1989).

The test with which Nature defeats the daltonian teaches us how to construct screening tests of our own. When the first pseudoisochromatic plates were developed in the 1870s, the designers initially attempted to test colour vision with solid figures of one chromaticity on a uniform background of a second chromaticity. It was quickly found that it was impossible to print figure and ground in such a way as to eliminate all edge artefacts. Moreover, a figure and ground that were equally light for one daltonian were not necessarily so for another. The ophthalmologist Stilling solved these problems by two ingenious manoeuvres. Firstly, he broke the target and field into many small patches, each with its own contour; and secondly, instead of attempting to *equate* the lightness of target and field, he varied the lightness of the individual patches (Stilling, 1877). Thus neither edge artefacts nor luminance differences could be used as a cue to discrimination of the target against the field.

Stilling's two manoeuvres have been incorporated into all subsequent sets of pseudoisochromatic plates. It is odd, however, that his principles have not been adopted in most of the computer-controlled tests of colour vision that have been introduced in recent years (e.g. King-Smith, Chioran, Sellers & Alvarez, 1983; Fallowfield & Krauskopf, 1984; Hart, Hartz, Hagen & Clark, 1984; Sellers, Chioran, Dain, Benes, Lubow, Rammohan & King-Smith, 1986; King-Smith, Vingrys & Benes, 1987; Heard, Stone, Gregory & Marmion, 1987; Arden, Gündüz & Perry, 1988; Cuvinot, 1992). In such tests, the subject has typically been required to detect the presence of a coloured spot or grating on an equiluminant field that is presented on a raster display. The great advantage of computer control is that it allows the possibility incorporated in the "chromatophotometer" of Stilling's

*Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England.

contemporary, Chibret—the possibility of dynamically and adaptively varying the chromatic difference of target and field along different directions in colour space (Chibret, 1887). However, if the target is a coloured spot or grating on an equiluminant field, several difficulties confront the designers of computer-controlled tests:

(i) The presence of a chromatic boundary may be revealed by edge artefacts that arise either from the slight misalignment of one or more of the guns of the colour monitor or from the intrinsic pattern of phosphor dots (Vingrys & King-Smith, 1986).
(ii) If the display incorporates chromatic variations at high spatial frequencies, e.g. sharp edges between target and field, then the chromatic aberration of the eye may render colour boundaries visible by introducing unintended luminance contours.
(iii) There are large variations in luminance matches, among normal subjects and especially among colour-deficient subjects.
(iv) Different luminance matches may be needed in different parts of the field, and this may particularly be the case in visual disease.

The first and second of these difficulties can probably be overcome by using chromatic variations of low spatial frequency. But the third requires that a preliminary luminance match be made for each individual subject, in order to ensure that the spots or gratings used in the main test are truly equiluminant for that subject. Such a procedure is time-consuming, both because of the extra time required for the actual measurements, and because of the time required to familiarize the subject with the photometric procedure. Elderly or handicapped patients may not find the photometry easy to perform, and the examiner has no independent criterion for judging the accuracy of their setting. And when the luminance equation has been made, there still remains the fourth difficulty. One solution to these problems is to define equiluminance empirically during the main test, by establishing the axis in stimulus space that gives the highest thresholds when sensitivity is probed in a number of directions in the plane defined by the luminance axis and one chromatic axis (Sellers *et al.*, 1986); but this method is too lengthy if the examiner wishes routinely to probe several different chromatic axes within the equiluminance plane.

Mollon and Reffin have proposed that the disadvantages of computer displays can be sidestepped by introducing spatial and luminance noise into the stimulus, i.e. by forming the target and field from a mosaic of discrete patches that each have their own contour and that vary randomly in luminance. The principles of Stilling and Chibret can then be combined in a single test (Mollon & Reffin, 1989; Reffin, Astell & Mollon, 1991). We describe here the further development of a test of this kind. A mosaic of small patches, varying in luminance and size, is presented on a computer-controlled colour monitor (Fig. 1). A subset of patches forms the target (in the present trials, a C-shape of varying orientation) and the chromatic difference between the target and field is adaptively adjusted during testing. In using a C-shaped target, we follow Schaaff, pupil of Landolt and Physician-Oculist to the Railways of Alsace-Lorraine (Schaaff, 1925).* In the basic form of our test, sensitivity is probed only on protan, deutan and tritan confusion lines (Fig. 2), but we show that it is feasible to extend the procedure to a larger number of directions in colour space, in order to derive discrimination ellipses within a short testing session.

*A similar target is used in the later pseudoisochromatic tests of Ikeda and Ohkuma (Tokyo, 1975) and of Yustova (Moscow, 1992).

METHODS

Apparatus and stimuli

The stimuli were generated by a Sigma Electronics Systems 5688 colour graphics system and were displayed on a slave Barco CD351 monitor. The Sigma system allows the output of each gun of the monitor to be specified with a precision of 8 bits, a resolution that proves adequate for measuring thresholds for normal subjects under our experimental conditions.

The stimulus array consisted of spatially discrete patches of varying size and luminance, presented on a black background (Fig. 1). The luminance of any given patch varied from trial to trial and was randomly assigned to one of six equally spaced and equally probable levels in the range 7.6–17.0 $cd \cdot m^{-2}$. The majority of the patches constituted the field and were of a single chromaticity, which remained constant across trials. A subset of patches formed the target, having the form of a C. The chromaticity of the target patches differed from that of the field patches and was systematically varied in successive presentations. The outer diameter of the C subtended 4.3 deg of visual angle and the inner diameter subtended 2.2 deg; the gap subtended 1 deg. The orientation of the target was varied randomly from trial to trial, with the gap lying at the top, bottom, left or right. The average luminance of the target was always the same as that of the background: instead of attempting to anticipate the spectral sensitivities of individual subjects, we rely on luminance noise to oblige the subject to use chromatic signals to solve the perceptual task.

Procedure

The subject viewed the display binocularly from a distance of 2.4 m. The spatial arrangement of the four response buttons corresponded to the four possible positions of the gap in the target C, and the subject's task was simply to press the appropriate button within 4 sec of the onset of the display. Having found in earlier trials that many untrained subjects, in particular elderly or visually impaired patients, were reluctant to guess when completely uncertain as to the position of the gap, the standard instructions for the present trials advised the subject not to respond at all on such presentations. The

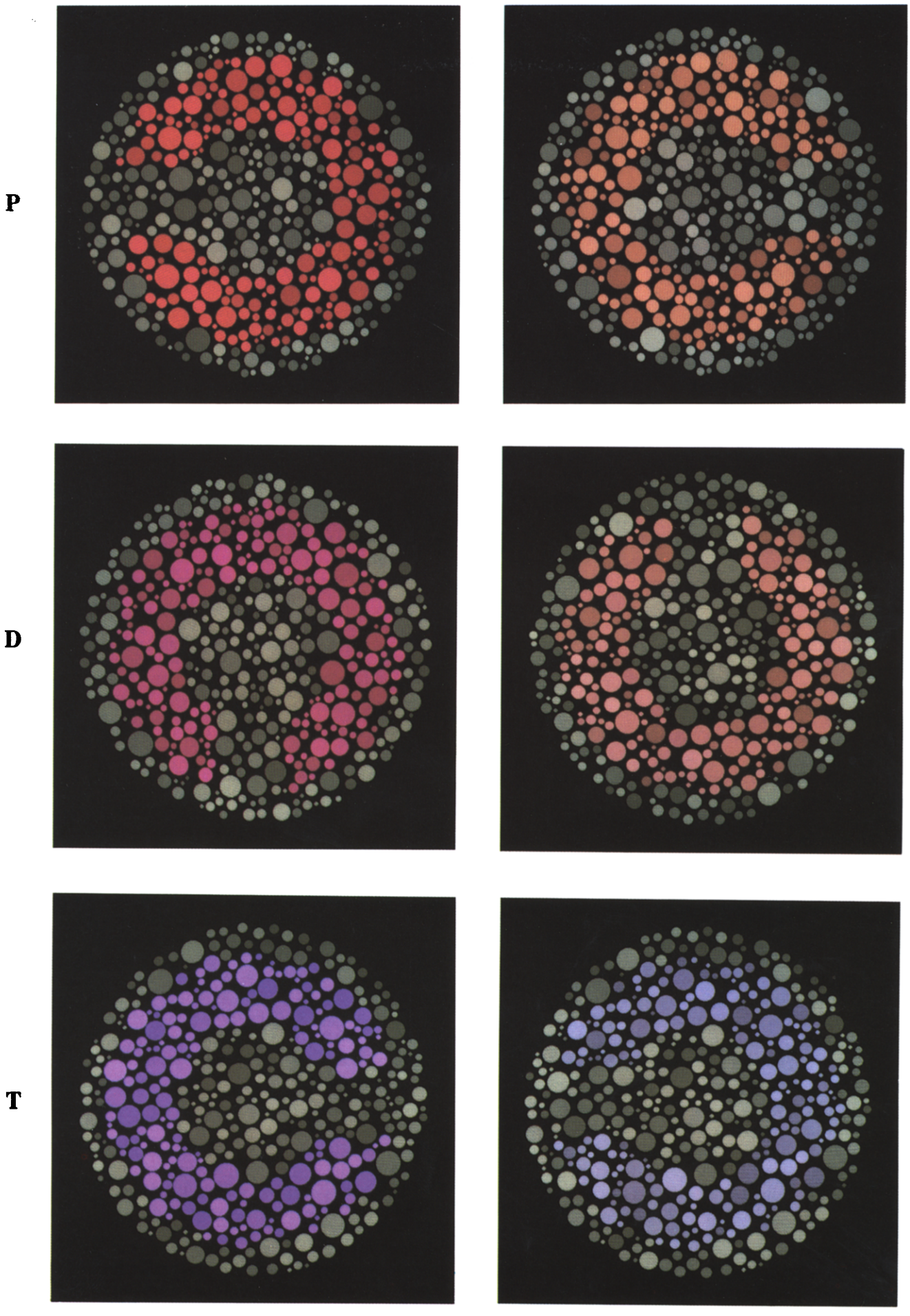

FIGURE 1. Typical stimuli used in the basic test. The top, middle and lower panels show protan, deutan and tritan test stimuli respectively. The left-hand stimulus of each pair is saturated and would be shown near the beginning of the test; the less saturated stimulus on the right represents a more difficult discrimination. Owing to the limitations of the printing process, this figure should not be taken to reproduce exactly the chromaticities or luminances used in our test.

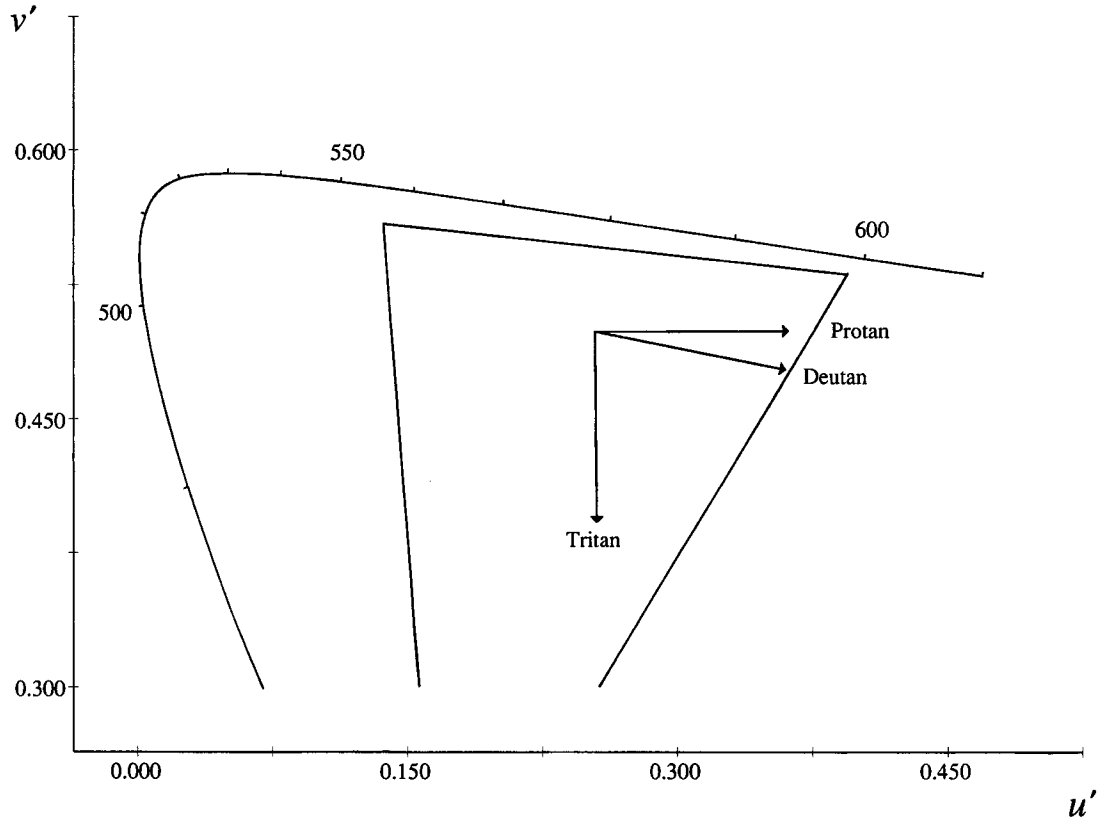

FIGURE 2. A section of the u', v' colour diagram showing the three directions probed in the basic test. The triangle represents the gamut of colours that could be reproduced by the monitor, and the arrows indicate the largest excursions made in each of the three directions away from the chromaticity of the field.

computer treated non-responses and incorrect responses as equivalent.

Chromatic sensitivity was measured along different axes of colour space (Fig. 2), using a staircase procedure (Cornsweet, 1962). To calculate the protan, deutan and tritan confusion lines, we used here the following dichromatic copunctal points: $x = 0.747$, $y = 0.253$; $x = 1.40$, $y = -0.40$; $x = 0.171$, $y = 0$. (These values were adopted on the basis of earlier experience with the test and their validity was tested empirically in the course of the present experiments. See below.) The host computer maintained a separate staircase for each direction being tested. In the basic test, trials on each of the three confusion lines were randomly interleaved. The chromatic difference between target and field along a given line was adaptively increased or decreased according to the subject's performance on the previous presentation on that line. Testing on any one staircase was terminated after 11 reversals, or after the subject had five times failed to detect the most saturated stimulus that the display could generate in that direction. In the former case, the mean of the last six reversals was taken as the threshold estimate for the direction being tested; in the latter case, the threshold was assigned an arbitrary maximum value. A small subset of presentations, intermingled with test presentations, served as control trials: on these trials, a target was presented at maximal saturation and at a chromaticity far from any dichromatic confusion line. The control trials served to detect malingering or misunderstanding, and also to give the subject occasional clear cases when he or she was near threshold.

Similar procedures were used in more elaborate tests (see below), in which a large number of directions were probed. In these cases, the program tested a pair of directions, randomly interleaved, before moving on to a second, randomly chosen, pair of directions.

An important issue in the design of an adaptive clinical test of this kind is the unit in which staircases are computed and thresholds are expressed. Although the familiar CIE (1931) x, y chromaticity diagram is appropriate for specifying the chromaticities used, equal distance in different parts of the diagram correspond to very different perceptual differences (Wright, 1941; MacAdam, 1942) and it would be inappropriate to specify, say, the step-size of a staircase in those units. In the measurements reported here, we have computed staircases, and expressed results, in units of the CIE u', v' diagram, which is (for a normal observer) an approximation to a uniform colour diagram, and which has the advantage of being a linear transformation of the x, y diagram (Wyszecki & Stiles, 1982). Experience with the test may demonstrate that slightly different units are appropriate for a clinical version of the test, but the discrimination ellipses we obtain from normal observers (see below) suggest that the u', v' diagram is reasonably appropriate for representing chromaticity thresholds.

The chromaticities and gamma functions of the monitor primaries were measured with a Minolta chromameter and were incorporated in the software for stimulus

generation. The additivity of the guns was assessed by measuring 31 chromaticities used in the psychophysical tests, a single patch being measured in a typical field of patches. The average discrepancy between computed and measured chromaticity was 0.005 units in x, y space.

Subjects

Normal and colour-deficient subjects were recruited by advertisement or by personal contact. Phenotypes given below refer to the classification obtained at the Nagel anomaloscope (Model I; Schmidt and Haensch, Berlin). We define as "dichromats" those who accept a match at both extremes of the Nagel scale, and as "extreme anomalous trichromats" those whose matching range includes one end of the Nagel scale but not the other. Ancillary information was provided by the Farnsworth–Munsell 100-hue test and the Ishihara plates (10th edn), administered under Illuminant C at approx. 100 lx.

RESULTS

Basic test

In the basic form of the test, intended for clinical screening, the subject's sensitivity is probed along three axes of colour space, corresponding to protan, deutan and tritan axes (Fig. 2). In the present trials, the field chromaticity was fixed at $u' = 0.254$, $v' = 0.499$ ($x = 0.413$, $y = 0.360$). Subjects took approx. 4 min to do the complete test.

Results were obtained for 48 colour-deficient subjects and 41 colour normals. Mean ages were 25.0 yr for dichromats, 26.7 yr for anomalous trichromats, and 23.5 yr for normals. The graphs of Fig. 3 show, for individual subjects, the threshold excursion along the protan line plotted against the threshold excursion on the deutan line. The units of the two axes are distances in u', v' space multiplied by 10^3.

Figure 3(c) shows the results for the normal observers. No normal subject scores more than 15 on either axis.

Figure 3(a) shows data for the 20 subjects who were identified as dichromats on the Nagel anomaloscope, i.e. who accepted a match of the 589 nm primary to both the red (666 nm) and the green (546 nm) primary. With one exception, all the dichromats exhibit a maximum score on at least one axis of the test; and there is a complete separation of protanopes and deuteranopes.

Figure 3(b) shows the results for those subjects identified as anomalous trichromats on the Nagel anomaloscope. Here the striking feature is the large range of discrimination abilities exhibited by different subjects: there are several protanomalous and deuteranomalous subjects who fail to detect the maximum excursion on one of the confusion lines, and conversely there is one deuteranomalous and one protanomalous subject whose scores lie within the normal range. The

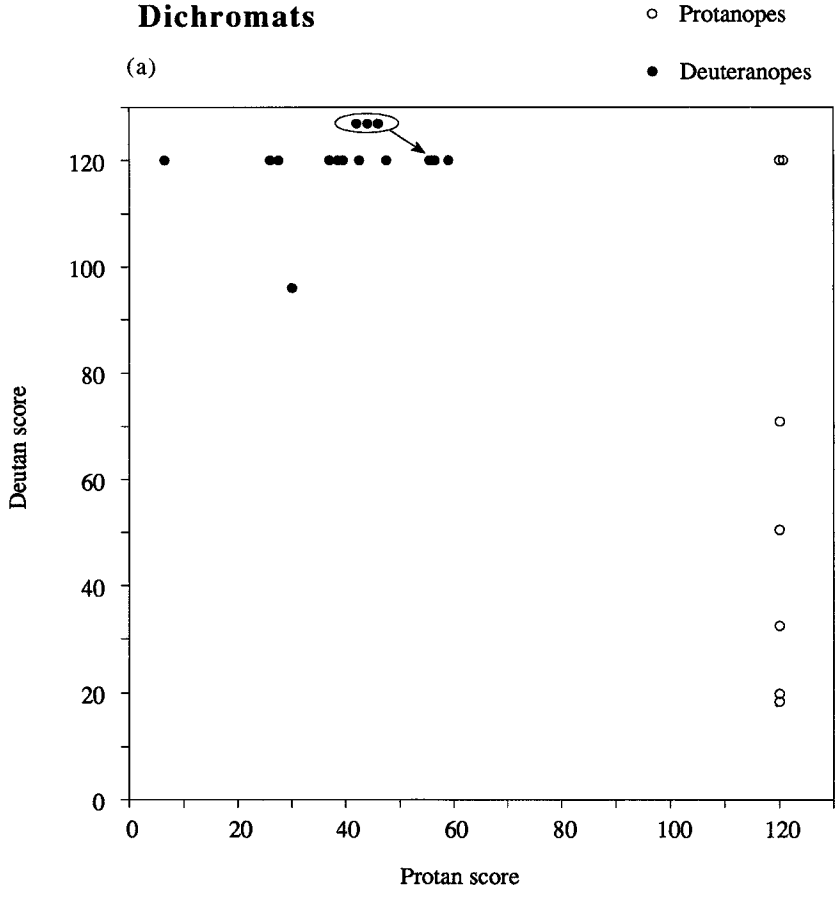

FIGURE 3(a). Scores for dichromatic observers on the protan and deutan lines of the basic test.

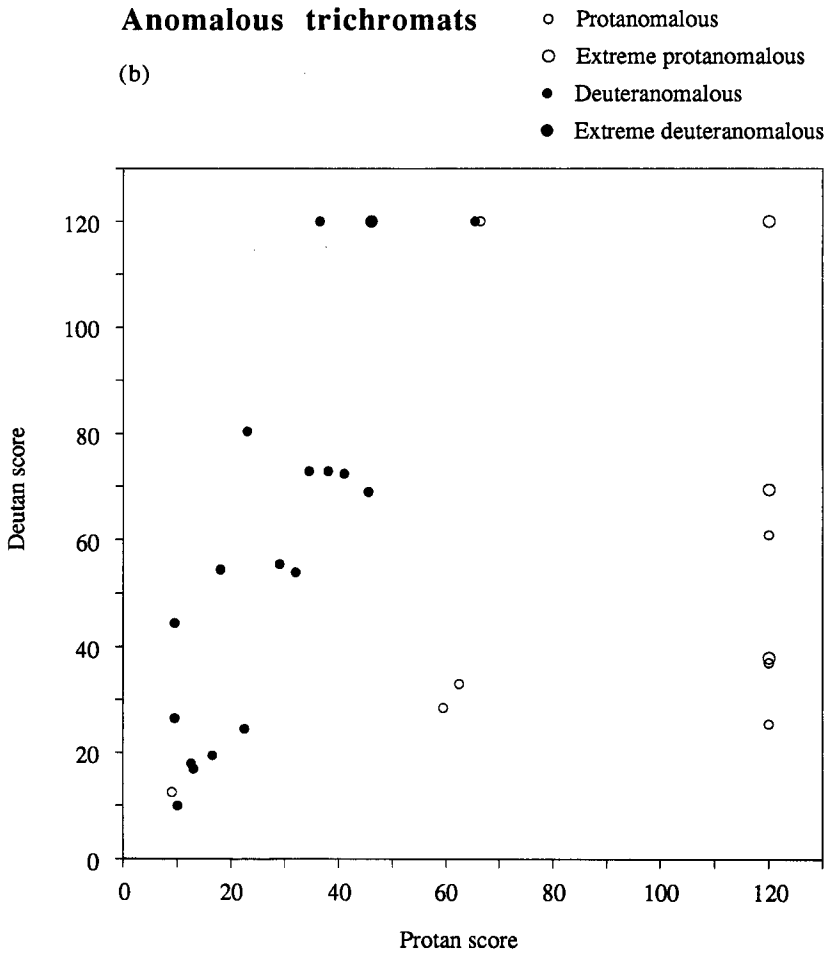

FIGURE 3(b). Scores on the basic test for anomalous trichromats.

huge variation in the discrimination abilities of anomalous trichromats is one of the classical mysteries of this field (Köllner, 1915; Pokorny, Smith, Verriest & Pinckers, 1979) and is discussed further below. That some anomalous trichromats reach the maximum score on one axis of the test is not mysterious: the confusion axis tested by the Nagel anomaloscope offers an effective gamut greater than that available on a colour raster display.

Our most curious result concerns the performance of red–green dichromats on the *tritan* axis of our test. The tritan axis probes the phylogenetically ancient subsystem of colour vision (Mollon & Jordan, 1988), which (in a protanope or deuteranope) depends on comparing the signal of the short-wave cones with that of either the middle- or the long-wave cones. It is commonly assumed that this subsystem is of normal sensitivity in the red–green dichromat, although systematic tests are seldom made. In Fig. 4 we compare the distribution of tritan scores for red–green dichromats (a) and normals (c): as a population, the dichromats have higher scores and the difference is highly significant by a Mann–Whitney test ($U = 132.5$, $P < 0.0001$). Although the dichromats include many young adults, no dichromat scores less than the median score for normals. The difference between dichromats and normals is significant for both protanopes and deuteranopes and the two types do not differ significantly from each other. Anomalous trichromats [Fig. 4(b)] are intermediate between dichromats and normals: as a population they score just significantly worse on the tritan axis than do the normals ($U = 363.5$, $P < 0.05$).

Discrimination at angles close to the confusion lines

As a check on the suitability of our copunctal points, we ran a subset of colour-deficient subjects on an extended version of the test, in which discrimination was probed along a fan of 16 closely-spaced axes in the neighbourhood of the protan and deutan confusion lines. We had two additional reasons for making these measurements. First, it is important for the test designer to know how much scatter there is in the confusion lines of a population of colour-deficient subjects; and detailed data on this issue are not readily available. Second, we were interested in whether the separation of deutans from protans (or dichromats from anomals) might be enhanced in later versions of our test by slightly splaying the probe lines relative to the true confusion lines—as appears to have been done, intentionally or unintentionally, by the designers of the HRR pseudoisochromatic plates (Lakowski, 1966, Fig. 6).

The results of these tests are shown in Fig. 5. Data are

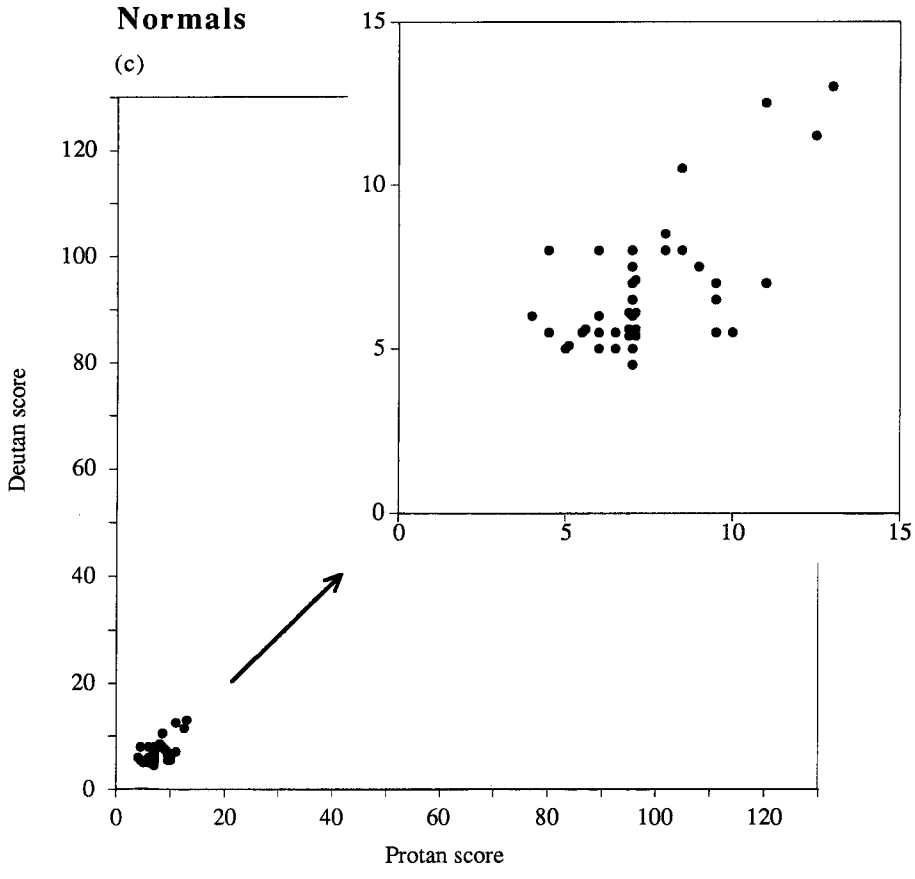

FIGURE 3(c). Scores on the protan and deutan lines of the basic test for 41 colour normals.

shown separately for dichromatic ($n = 18$), extreme anomalous ($n = 4$), and anomalous subjects ($n = 21$). The abscissa of each panel represents the angle in u', v' space along which sensitivity was probed, and the ordinate represents the mean threshold for that axis expressed as distance in u', v' space multiplied by 10^3. The histogram in each panel indicates the distribution, across individual subjects, of the axis on which the maximal threshold is recorded. (If the subject reached the maximal score on more than one axis, the average angle was plotted.) We refer to this angle as the "pessimum angle".

The results of Fig. 5 show that the protan and deutan probe lines used in our basic test are close to the true confusion lines. The data for dichromats and extreme anomals suggest that a very small rotation of the deutan probe line might be needed to bring it into coincidence with the empirical maximum, but the correction needed is of the order of only one degree.

A second conclusion to be drawn from Fig. 5 is that there is only limited scatter in the pessimum angles for the most severely deficient groups (dichromats and extreme anomals). This finding is the more striking in that our observers came to the laboratory untrained and completed these measurements within 1 hr. The results for anomalous trichromats [Fig. 5(c)] when averaged within the protan and deutan classes, reflect the functions obtained from the corresponding groups of dichromat. However, it was already clear from Fig. 3 (above) that anomals vary enormously in the degree to which their thresholds are elevated along the confusion line that bears the label of their phenotype. In the present test, several anomals exhibited virtually flat functions, and the absence of any systematic peak in such functions largely accounts for the increased spread of pessimum angles seen in the histogram for the anomals.

We have raised above the question of whether the basic test would achieve a better separation of protans and deutans if the probe lines were in fact splayed away from the true confusion lines. For each angle used in the present experiment, we have used the Mann–Whitney U-statistic to quantify the efficiency with which the protan and deutan groups are separated. For dichromats, the minimal U-values (i.e. most efficient separations) in fact coincide with the two estimated confusion lines; for anomalous trichromats, a minimum in the U-values is found at the protan line, but a second minimum lies at an angle 4 deg less than that of the deutan line. Given these findings and given the narrow range over which dichromats exhibit a maximal error score [Fig. 5 (a)], we conclude that only very small adjustments are desirable to the probe lines used in the basic test.

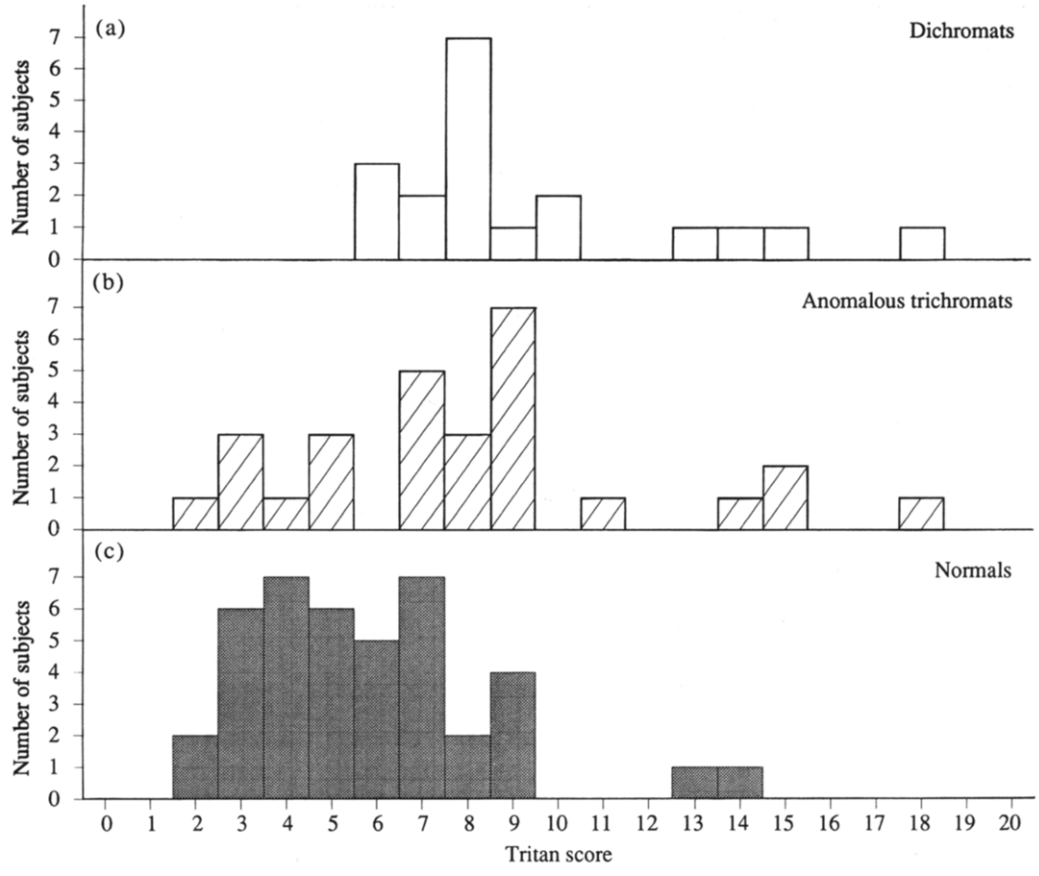

FIGURE 4. Distributions of scores on the tritan axis of the basic test for 41 colour normals, 28 anomalous trichromats and 20 dichromats.

Derivation of discrimination ellipses

Encouraged by the results of the preceding experiment, we asked whether our method could be used to obtain full discrimination ellipses from untrained subjects. It was in France that the tradition first arose of characterizing colour deficiency by the extent of the achromatic region of a chromaticity diagram. Galezowski, who first measured saturation thresholds in acquired eye diseases, was explicitly influenced by the colour space of Chevreul (Galezowski, 1868); and his pupil Chibret introduced the chromatophotometer as a

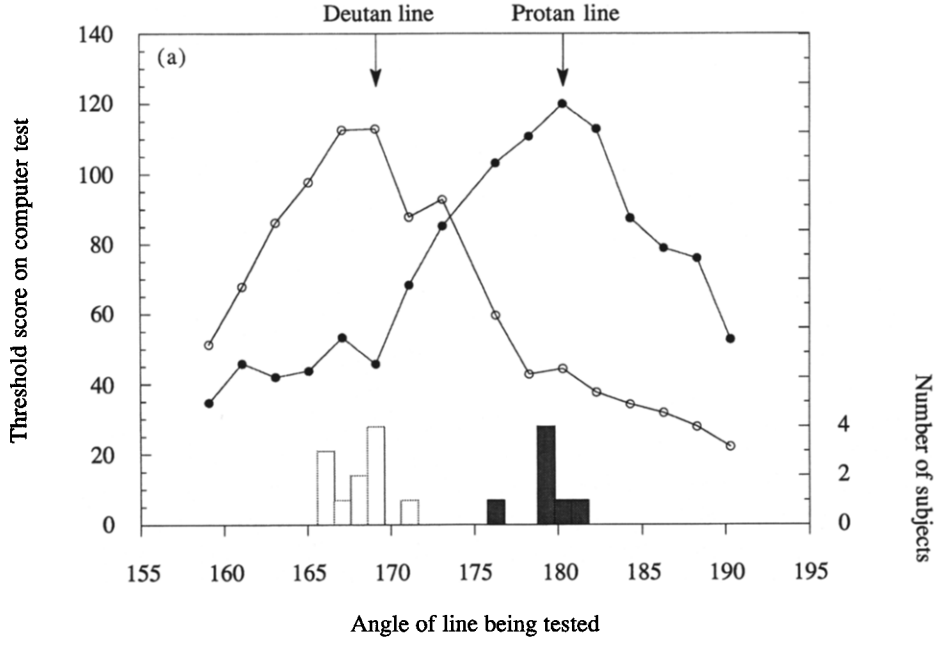

FIGURE 5(a). *Caption on facing page.*

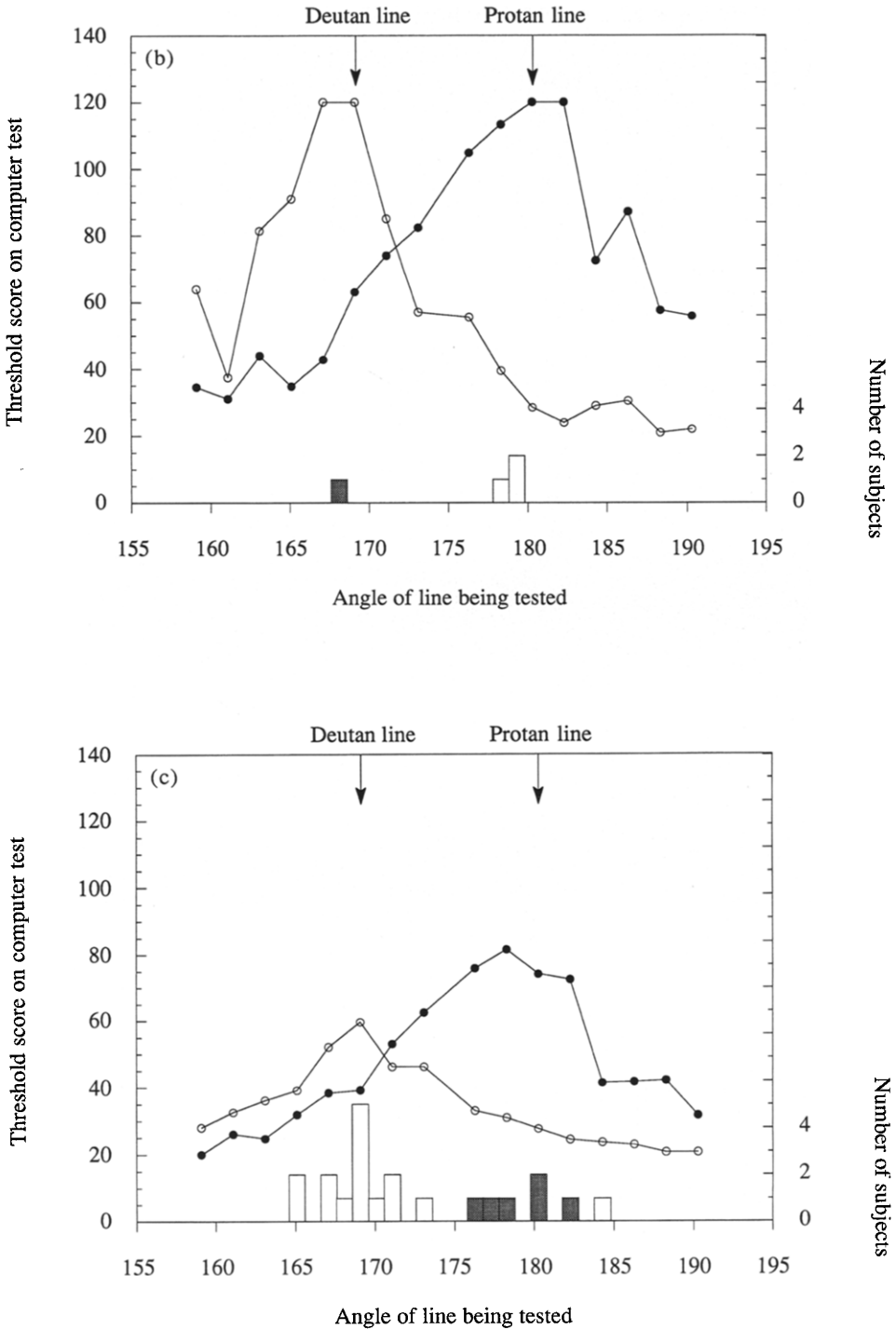

FIGURE 5. Distribution of the direction of worst discrimination, in u', v' space, for colour-deficient observers. (a) Results from 18 dichromats; (b) results from four extreme anomalous trichromats; (c) results from 21 simple anomalous trichromats.

device for establishing the subject's threshold along any axis of colour space (Chibret, 1887). But it was possibly Bruno Kolbe of St Petersburg who first graphically represented thresholds in a quantitative colour diagram, providing, in effect, discrimination ellipses for "red–green" and "blue–yellow" forms of colour deficiency (Kolbe, 1881). His diagram is reproduced in Fig. 6.

In our own experiment we probed the subjects' chromatic discrimination along 20 directions in colour space. The angles of these 20 probe lines were spaced at 18-deg intervals in the u', v' diagram. In most cases, measurements were obtained for three different values of the field chromaticity, i.e. of the origin from which colour discrimination was probed. The chromaticity coordinates of the three fields are listed in Table 1 and were chosen to lie approximately on the same tritan confusion line.

In fitting ellipses to the resulting data, we forced the centre of the ellipse to coincide with the chromaticity of the field patches of the array. The ellipse was then fitted by minimizing the sum of squares of the log distances between the ellipse and the data points, the distances

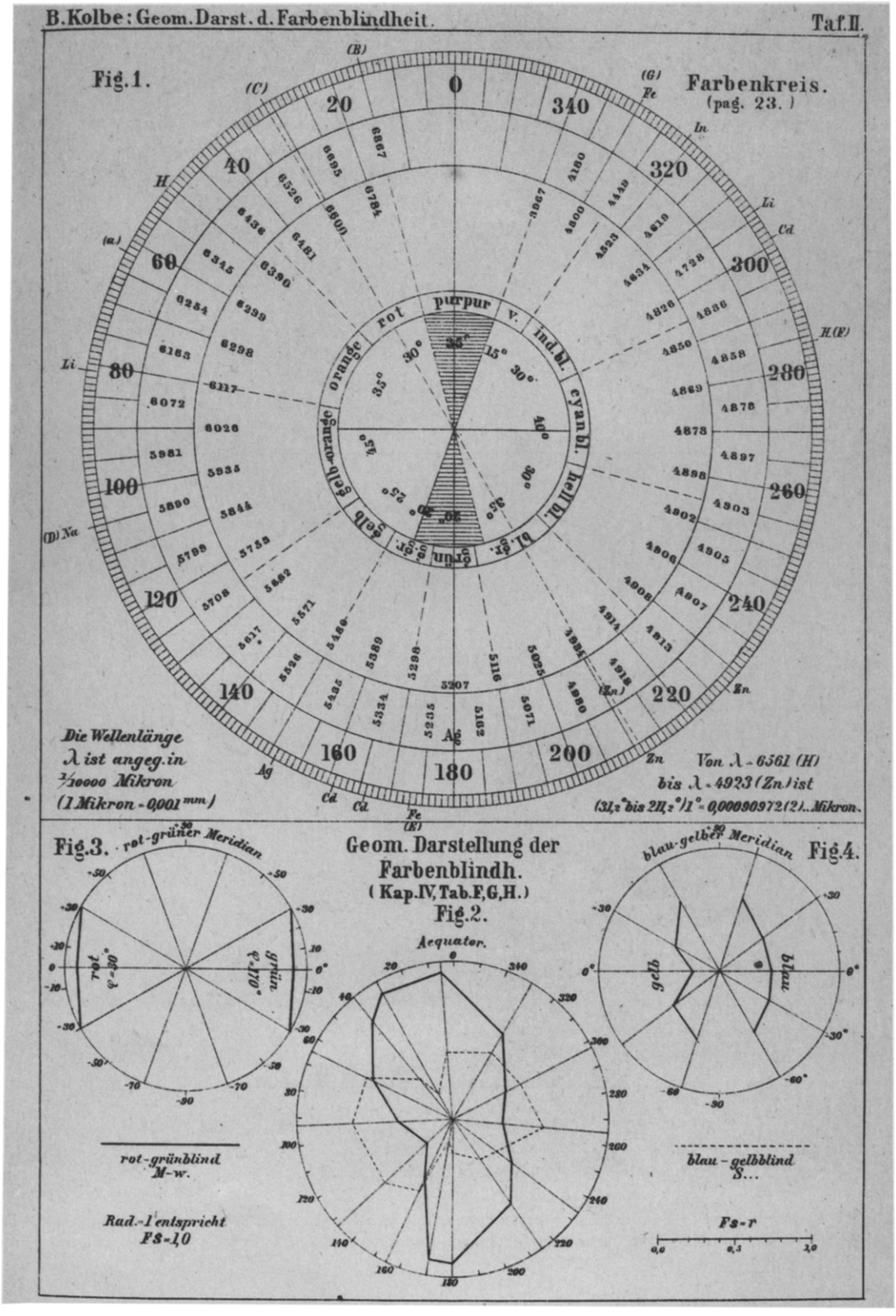

FIGURE 6. A reproduction of what are possibly the first published discrimination ellipses for daltonians, taken from a monograph by Bruno Kolbe published in St Petersburg in 1881. The upper panel shows Kolbe's colour circle, calibrated in wavelengths. The central figure ("Fig. 2") in the lower panel shows, in the same colour diagram, the achromatic regions for a "red–green blind" subject (solid line) and a "blue-yellow blind" subject (broken line). A graduated stimulus was generated by spinning a cone painted with black, white and coloured sections. A sliding aperture exposed an increasingly saturated stimulus until the subject identified the colour.

being measured along radial lines that corresponded to the probe lines. In directions where the subject could not discriminate the maximal available chromaticity difference, the data point was arbitrarily placed at the point where the probe line intersected the limiting gamut of our colour display.

We emphasize that we fit ellipses to our data simply as a consistent way of operationally summarizing the

results from a range of very different phenotypes. For a true dichromat, of course, it is strictly inappropriate to fit an ellipse: the empirical thresholds ought to lie along two confusion lines that represent the minimal detectable increment or decrement in the ratio of short- to long-wave cone signals, and the length of the fitted ellipse will depend on how exactly one of our 20 probe directions coincides with the confusion line that passes through the field chromaticity. For anomalous trichromats a different qualification must be made: their thresholds, when plotted in the u', v' diagram of the normal observer, are valid only for the particular primaries used in the present test.

For comparison with the daltonian subjects, we show in Fig. 7 the discrimination ellipses obtained from five colour-normal observers, all young adults. Nagel midpoints and ranges, and 100-hue error scores, are listed in Table 2. Expressed in the u', v' diagram, the three ellipses for a given subject are approximately constant in size (the lengths of the long axes are listed in Table 2) and show relatively small departures from circularity (the average ratio of long to short axes is 1.56). These features confirm that the units of the u', v' diagram were reasonable ones to adopt in a practical test, although in fact the present results suggest that the older (CIE 1960) u, v diagram would give more nearly circular discrimination ellipses under our experimental conditions. The relationship between the two uniform chromaticity diagrams is $u' = u$ and $v' = 1.5v$. So the effect of using the u, v diagram would be to compress our ellipses in the vertical direction. [Such a compression also appears appropriate in the case of the results of Cavonius, Müller and Mollon (1990) for supra-threshold colour differences.]

All the normal subjects do exhibit a systematic shift in the directions of the long axes of the discrimination ellipses: for the upper two ellipses the angles are similar and are <90 deg, whereas the lowermost ellipse, which lies in the purple region of the chromaticity diagram, is consistently oriented in a tritan direction (Table 2). The female subject SH reports that her father is colour deficient and on the OSCAR test (Estévez, Spekreijse, van Dalen & Verduyn Lunel, 1983) she exhibits a spectral sensitivity known to be characteristic of protan carriers (Mollon, 1987). It may or may not be significant that ellipses 1 and 2 are oriented at a lower angle for her than for the other colour-normal observers.

TABLE 1. Chromaticity coordinates used for testing protans, deutans and normals (upper three rows) and for testing a tritanope (lower three rows)

	x, y (CIE 1931)	u', v' (CIE 1976)
Ellipse 1	0.403, 0.442	0.215, 0.531
Ellipse 2	0.350, 0.340	0.219, 0.480
Ellipse 3	0.302, 0.247	0.225, 0.415
Ellipse 1	0.297, 0.368	0.174, 0.485
Ellipse 2	0.350, 0.340	0.219, 0.480
Ellipse 3	0.409, 0.309	0.278, 0.472

The coordinates given are the field chromaticities that served as the origin for the 20 target vectors used in determining each ellipse.

In Fig. 8 we show the results obtained from five protan subjects, one protanopic (PC), one extreme protanomalous (MD), and the remainder protanomalous. In all cases the long axes of the ellipses converge approximately towards the protan copunctal point; but there is a large variation in the lengths of the long axes. For the simple protanomalous observers the central ellipse varies in its axis ratio from 7.7 to 2.9 (Table 2).

Figure 9 shows the results obtained from two deuteranopic and three extreme deuteranomalous subjects. In all these cases, the ellipses extend to the gamut available from the colour monitor, and the long axes of the ellipses are directed towards the deutan copunctal point. The two oldest subjects (KB, EF) also exhibit a clearly elevated threshold in the tritan direction.

Figures 10 and 11 show discrimination ellipses for ten deuteranomalous subjects and illustrate the varied range of phenotypes that pass under this name. The eight subjects CHI, AJM, TS, CHO, JSB, CJT, NFT and GM exhibit ellipses of different extents but all pointing in an essentially deutan direction. There is a sevenfold variation in the length of the central ellipse.

For the remaining two deuteranomalous subjects, GRC and PJ, both young adults, the long axes of the ellipses are not extended in a deutan direction, but a large difference is seen in the absolute level of discrimination. PJ (a professional artist) is an example of the rare phenotype called *Minimalanomaltrichromat* by Vierling (1935): he passes the Ishihara test but makes matches between 20 and 21 on the Schmidt and Haensch (Model I Nagel) anomaloscope. It is interesting to note that—in the u', v' diagram—he shows one axis ratio that is closer to unity than that for any subject. Moreover, the long axes for all three ellipses for PJ are smaller than the corresponding lengths for any of the normal subjects in our sample (Table 2), although this may reflect the fact that he had had more practice on our basic test than had other subjects.

Figure 12 shows results obtained for the left and right eyes of DC, a female subject aged 43 who had become tritanopic 3 yr earlier. This subject is the patient described by Jordan, Sarkies and Mollon (1990). For the purpose of testing this patient we used three field chromaticities that lie on a line that holds constant, for a normal observer, the excitation of the short-wave cones (Table 1). DC shows remarkable consistency between eyes, and the long axes of her ellipses converge convincingly towards the tritan confusion point. The narrowness of her ellipses deserves note: her loss of the phylogenetically older subsystem of colour vision appears to be unaccompanied by impairment of the L/M subsystem. In this respect, the present observations are consistent with her ability to order without error those regions of the Farnsworth–Munsell 100-hue test that are tangential to protan and deutan confusion lines.

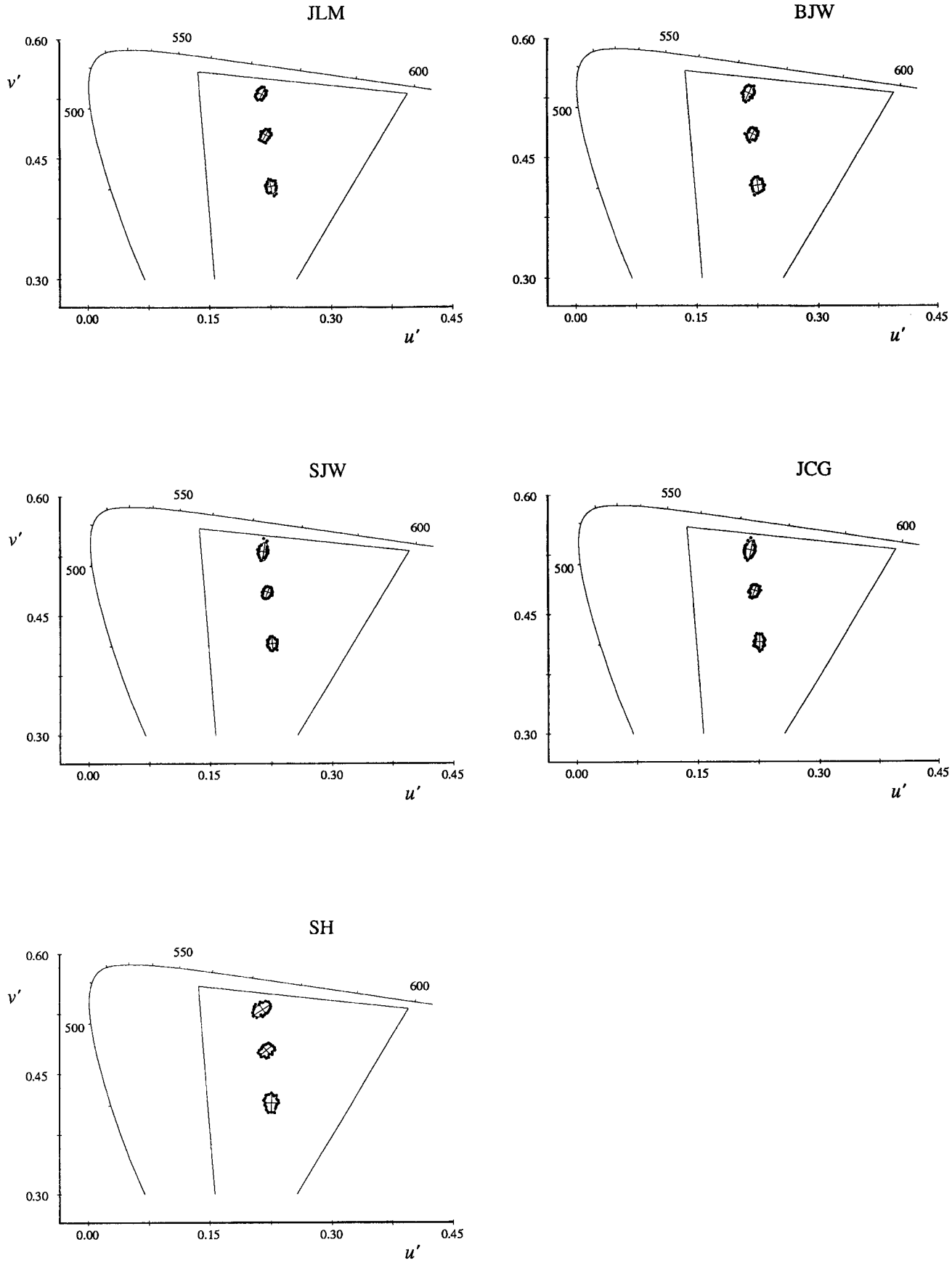

FIGURE 7. Discrimination ellipses obtained for five subjects with normal colour vision.

TABLE 2. Results summarized for subjects for whom full discrimination ellipses were measured

Subject	Classification	Age	Sex	Nagel midpoint	Nagel range	100-hue test score	Ellipse 1			Ellipse 2			Ellipse 3		
							Length	Axis ratio	Angle (deg)	Length	Axis ratio	Angle (deg)	Length	Axis ratio	Angle (deg)
JLM	N	20	F	40.5	4	36	0.0166	1.48	64.0	0.0176	1.55	61.0	0.0192	1.53	99.1
BJW	N	21	M	37.25	1.5	52	0.0210	1.61	64.8	0.0172	1.31	64.4	0.0206	1.34	97.2
SJW	N	19	M	38.75	2.5	NA	0.0236	2.18	81.1	0.0164	1.43	71.2	0.0174	1.44	95.3
JCG	N	21	F	42	2	4	0.0238	1.90	78.2	0.0176	1.46	71.2	0.0210	1.73	90.5
SH	N*	20	F	39.5	5	48	0.0250	1.62	33.7	0.0202	1.40	39.5	0.0236	1.47	89.6
PC	P	27	M	36.5	73	NA	0.1762	10.7	178.8	0.2950	20.1	2.7	0.1050	6.42	10.7
MD	EPA	27	M	32	64	108	0.1350	9.51	178.6	0.1670	12.4	0.0	0.1908	10.8	10.8
HH	PA	31	M	46.5	23	112	0.1950	11.5	176.5	0.1646	7.73	2.2	0.1458	5.13	11.1
DS	PA	28	M	59.5	5	108	0.1000	4.79	178.3	0.0590	3.29	1.9	0.0632	2.73	13.4
JB	PA	15	M	61.5	5	152	0.0850	4.10	178.4	0.0492	2.91	3.5	0.0492	1.78	8.9
CR	D	35	M	36.5	73	NA	0.3160	12.6	169.5	0.2324	9.84	165.2	NA	NA	NA
KB	D	70	M	36.5	73	168	0.2690	6.82	167.5	0.2042	5.80	162.8	0.2040	5.51	157.6
EF	EDA	71	M	20	40	NA	0.3224	8.13	169.5	0.2100	5.06	164.1	0.1844	4.75	162.3
MC	EDA	26	M	33.5	67	152	0.3498	16.9	171.3	0.1672	6.78	165.4	0.1726	5.70	163.8
MT	EDA	18	M	15	30	136	0.1724	8.47	170.8	0.2438	10.5	168.0	0.2044	8.94	166.4
CHI	DA	18	M	13.5	23	132	0.2828	9.40	169.8	0.2572	8.18	166.9	0.1724	5.75	164.6
AJM	DA	56	M	18	14	84	0.1446	6.61	170.0	0.1784	7.68	165.0	0.1324	4.46	161.8
TS	DA	22	M	15.75	5.5	88	0.1090	6.67	167.3	0.1008	4.88	164.4	0.1228	5.73	160.8
CHO	DA	14	M	26.5	1	216	0.0954	5.76	167.7	0.0598	4.13	165.7	0.0952	5.32	161.0
JSB	DA	23	M	20.5	17	NA	0.0692	3.78	170.5	0.0692	3.19	170.1	NA	NA	NA
CJT	DA	15	M	26.5	21	236	0.0614	3.97	170.0	0.0552	3.94	168.0	0.0590	2.91	167.8
NFT	DA	21	M	13†	10	164	0.0494	3.08	163.0	0.0446	2.47	0.3	0.0380	2.57	172.6
GM	DA	15	M	18.75	5.5	52	0.0408	2.71	174.7	0.0384	2.28	170.9	0.0334	1.77	164.3
GRC	DA	23	M	19.5	1	96	0.0362	1.58	173.1	0.0358	1.50	178.6	0.0386	1.08	112.9
PJ	DA‡	28	M	20	2	52	0.0142	1.12	173.7	0.0144	1.41	28.1	0.0172	1.29	106.3
DC (left eye)	T	42	F	41.5	1	208	0.0866	7.22	99.0	0.1510	10.0	93.8	0.1982	9.67	89.0
DC (right eye)	T			41	0	204	0.0760	5.77	98.1	0.2370	16.3	95.0	0.1986	9.82	89.6

Subject classifications: N, normal; P, protanopic; PA, simple protanomalous; EPA, extreme protanomalous; D, deuteranopic; DA, simple deuteranomalous; EDA, extreme deuteranomalous.
NA indicates that a particular score was not available for that subject.
*This subject is thought to be a carrier of a protan form of colour vision deficiency.
†This subject exhibited severe "Umstimmung" on the anomaloscope and consequently the range quoted may be in error.
‡This subject is a "Minimalanomaltrichromat" (Vierling, 1935). See text for a fuller explanation.

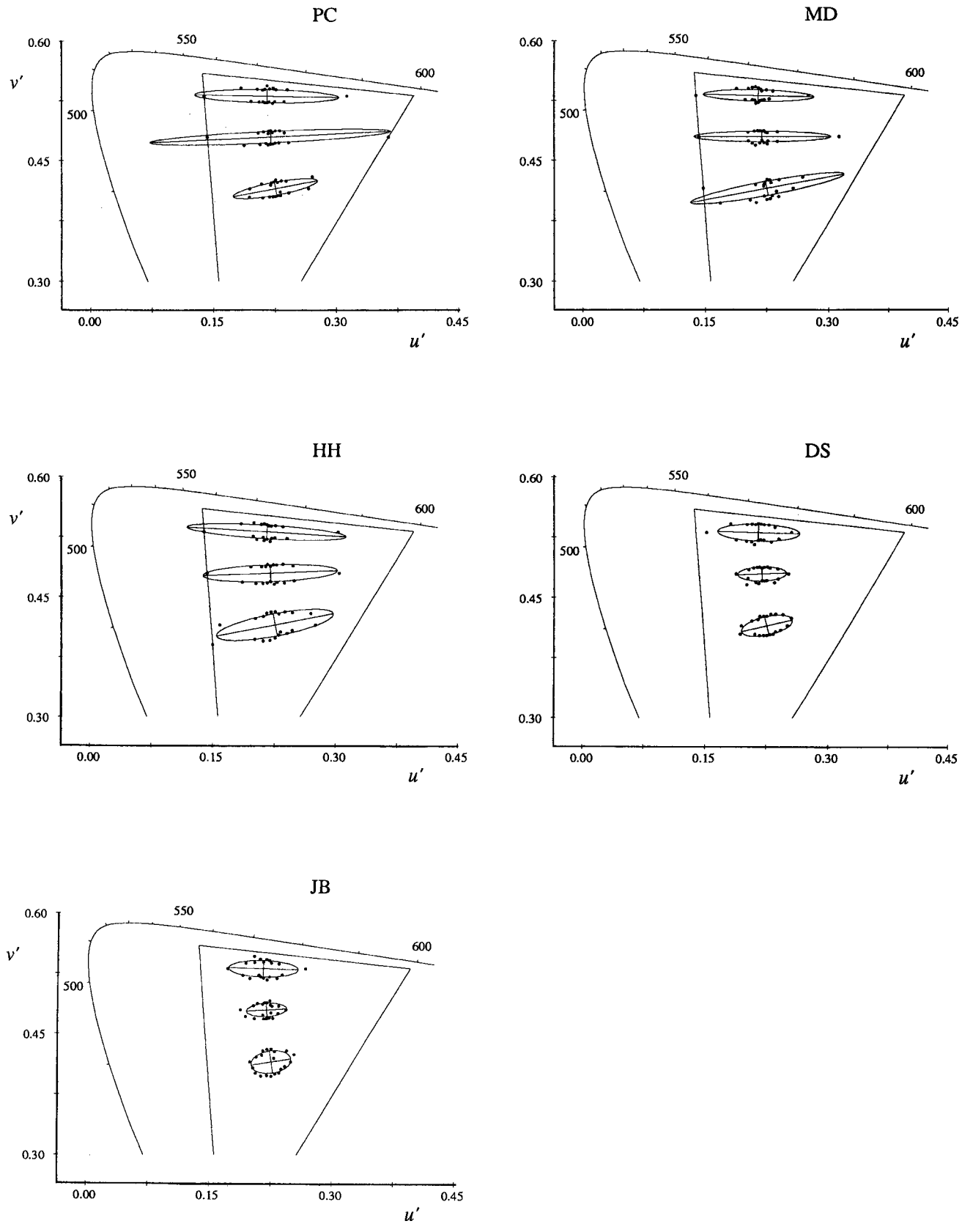

FIGURE 8. Discrimination ellipses obtained from one protanopic, one extreme protanomalous and three simple protanomalous subjects. All ellipses converge towards the protan copunctal point but note the variation in the lengths of the ellipses for the three simple protanomalous subjects.

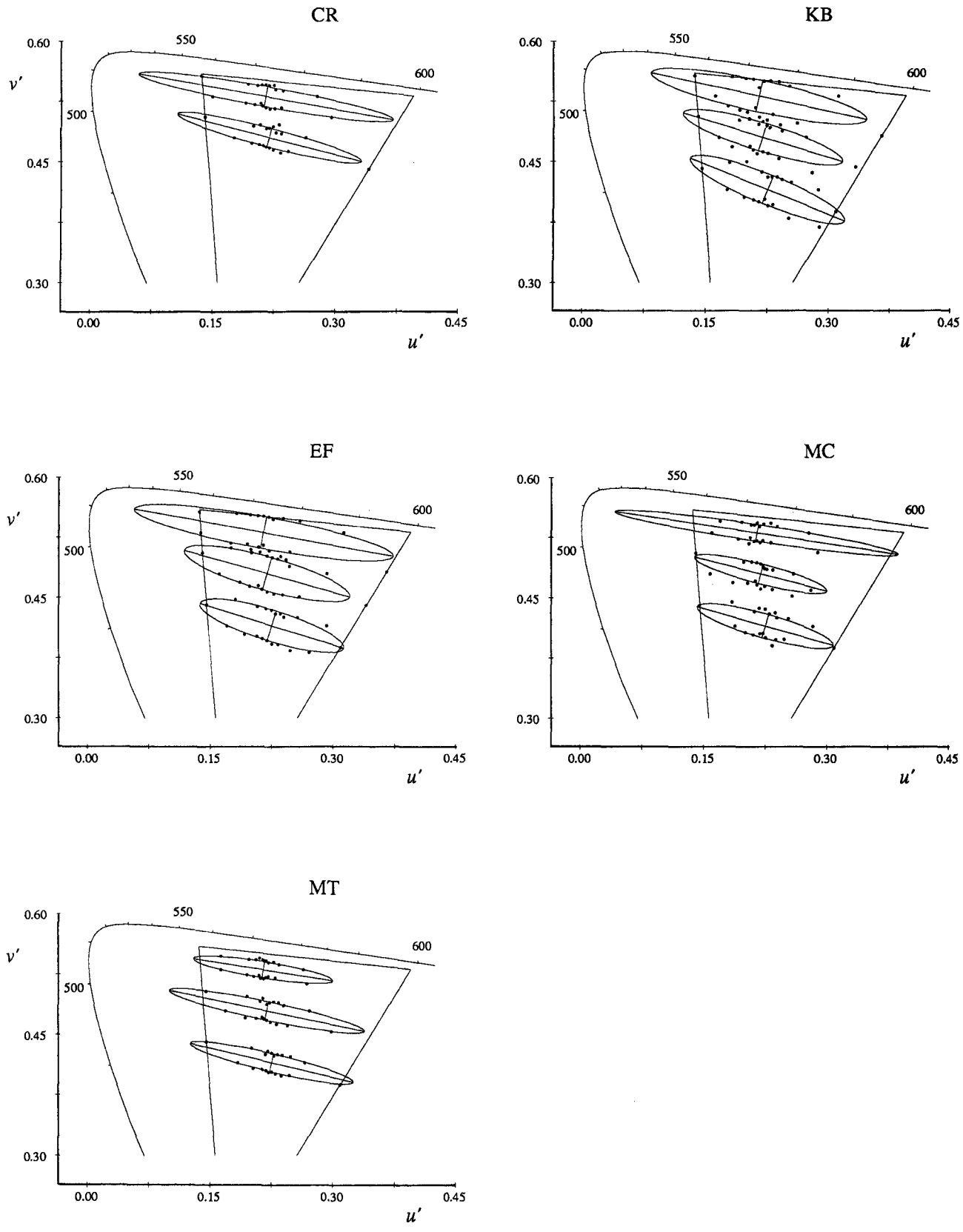

FIGURE 9. Discrimination ellipses obtained for two deuteranopic and three extreme deuteranomalous subjects. Note that in all these cases the ellipses extend to the gamut available from the colour monitor and that their long axes are directed towards the deutan copunctal point.

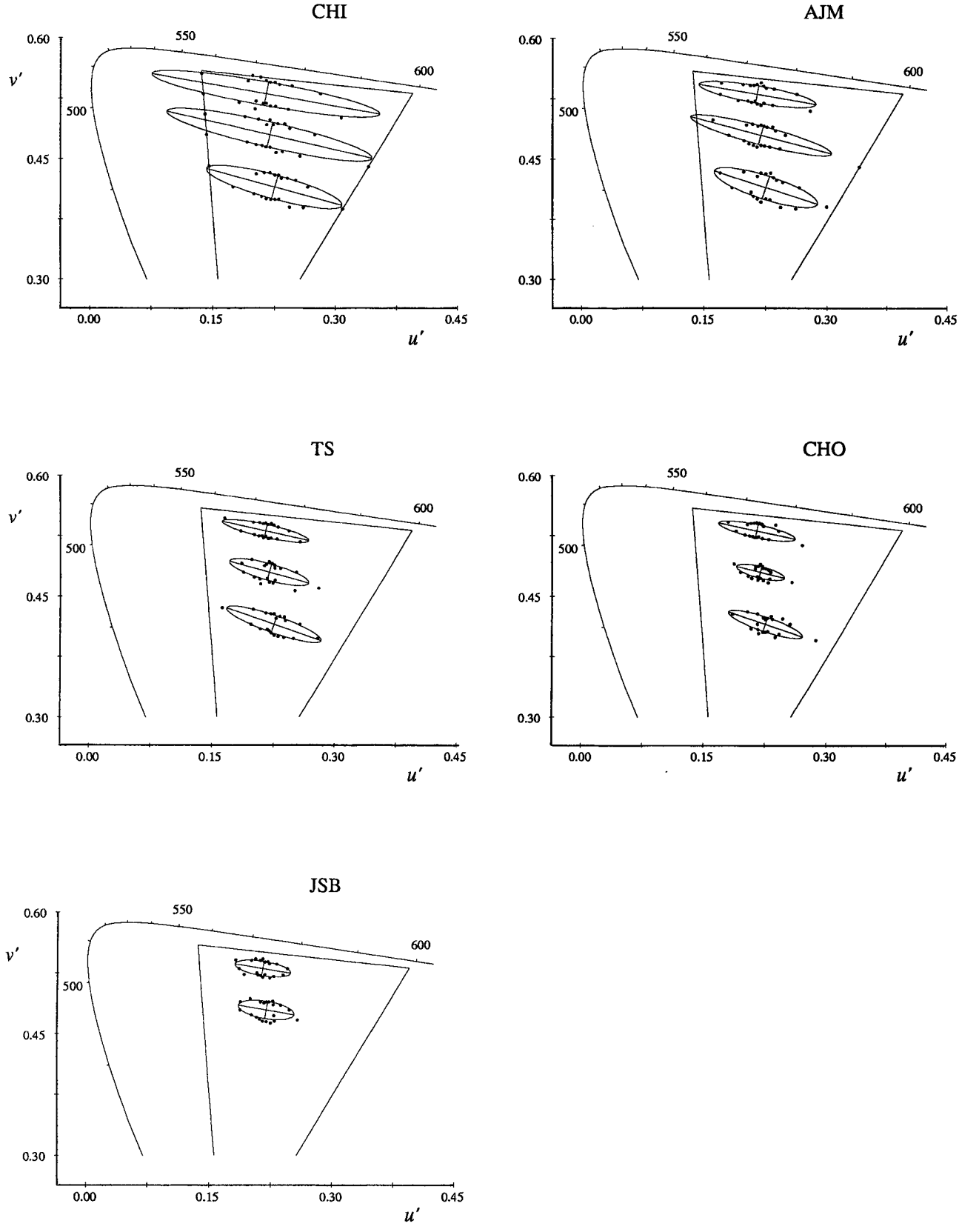

FIGURE 10. Discrimination ellipses obtained for five simple deuteranomalous subjects. All these subjcts' ellipses point in a deutan direction, although there is a wide variation in the length of the ellipses, corresponding to a wide variation in the severity of the colour deficiency.

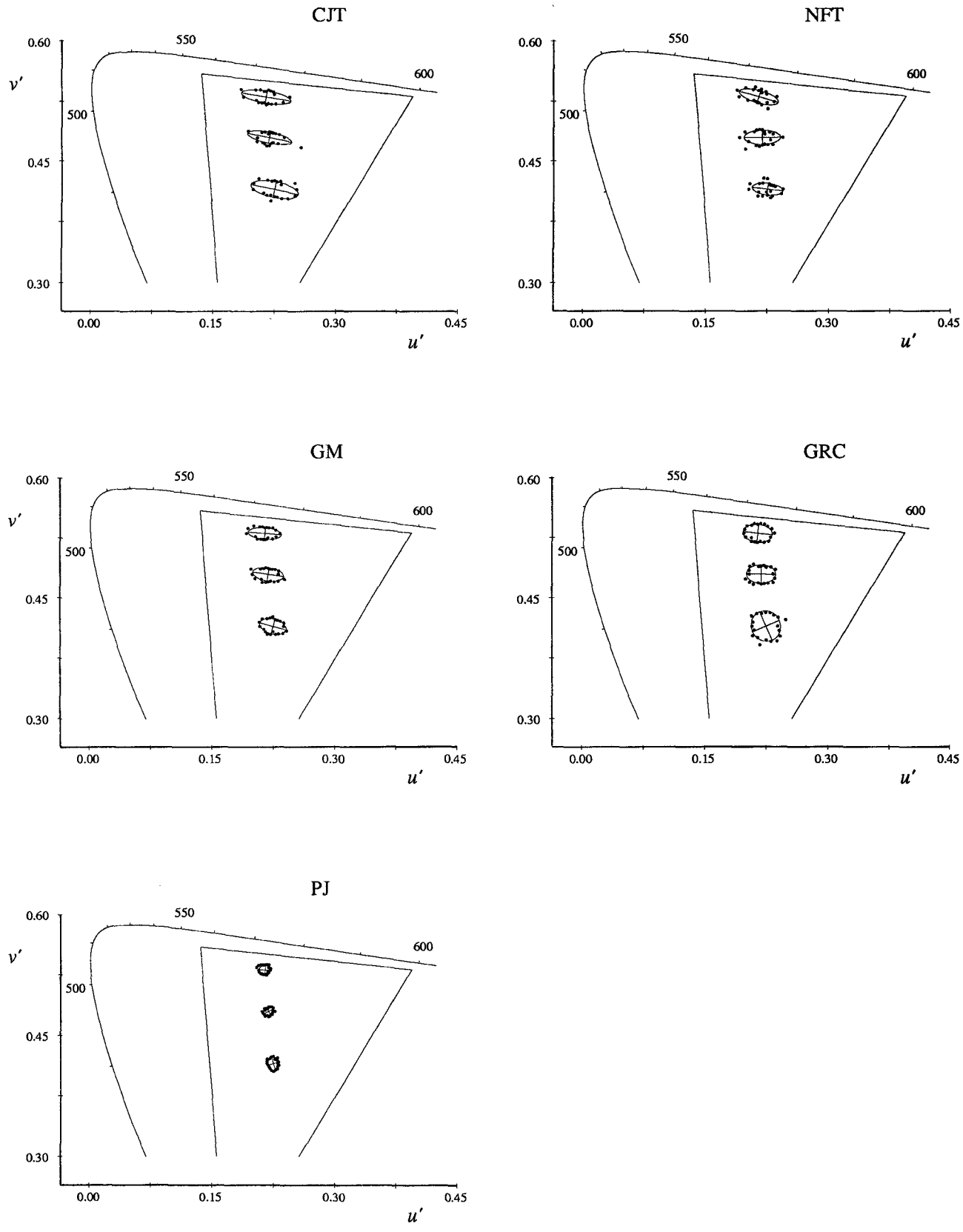

FIGURE 11. Discrimination ellipses obtained for five simple deuteranomalous subjects. Of particular interest are the ellipses of PJ and GRC: although these observers are simple deuteranomalous, their ellipses show no direction of worst discrimination, and furthermore, PJ produces ellipses that are smaller and rounder than those of the colour-normals in our sample.

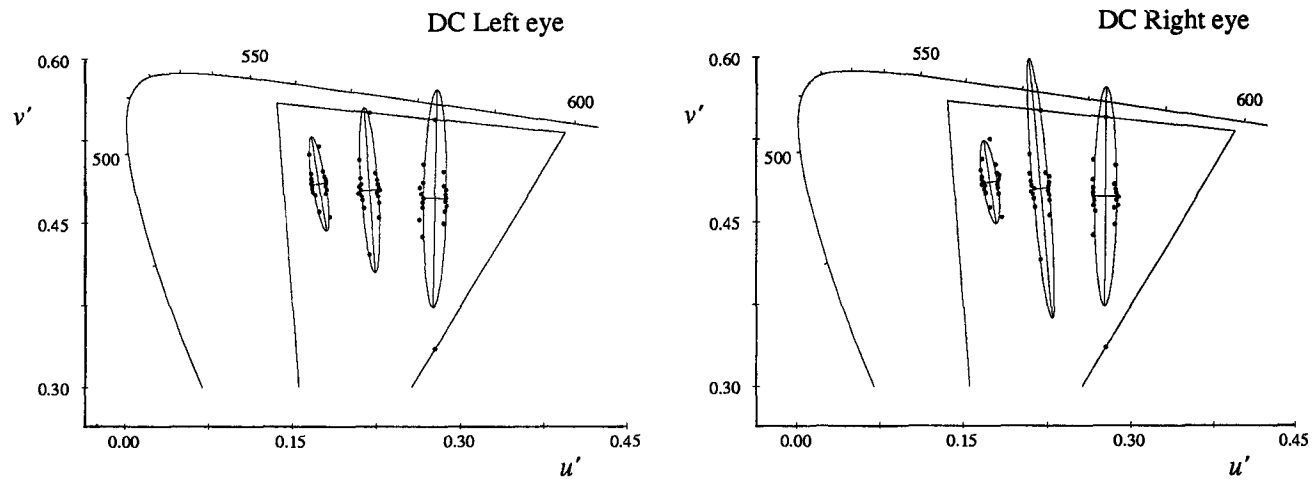

FIGURE 12. Discrimination ellipses obtained from a subject with acquired tritanopia. Ellipses for the left and right eyes are shown separately and show remarkable agreement.

DISCUSSION

Assessment of test

The combination of the principles of Stilling and Chibret has given a test that is rapid to administer and requires no preliminary equation of luminance: screening along three lines in colour space can be completed in 4 min, and one full discrimination ellipse can be obtained from a naïve subject in 20 min. The basic form of the test achieves a good separation of protan and deutan subjects and gives a quantitative measure of colour discrimination. A positive feature of the test is that it distinguishes anomalous observers of differing discriminative ability. A limitation—common to all tests that measure discrimination—is that there exists a small minority of anomalous observers who enjoy a chromatic discrimination within the normal range and reveal themselves only on the anomaloscope or other colour-matching test.

The principle underlying the test is that the spatial luminance noise in the array forces the subject to rely on chromatic signals in distinguishing target from ground: he or she is obliged to use the ratio rather than the sum of the signals from different classes of cone. To ensure even better isolation of the chromatic signal, a possible modification would be to introduce temporal luminance noise as well, a suggestion made by Mollon (1982) and incorporated in a version of our test by Birch, Barbur and Harlow (1992). But this modification has a potential disadvantage: the activity that flicker would induce in the large-axoned fibres of the optic nerve might, by ephaptic transmission (Shepherd, 1988, Chap. 7), introduce noise into the smaller fibres that carry chromatic information and so might elevate colour thresholds. This possibility would be of particular concern when testing patients with demyelinating pathology of the optic nerve. Moreover, a flickering array might act more centrally to distract elderly or brain-injured patients. Since spatial noise seems sufficient to achieve our purpose, we have therefore eschewed the use of temporal modulation.

Residual chromatic discrimination in dichromats

Our most curious finding is that dichromats, *as a population*, have elevated thresholds on the tritan axis of the test. An excursion along this line alters the ratio S/M for the protanope and the ratio S/L for the deuteranope, and it is commonly assumed that these ratios are extracted by a phylogenetically ancient subsystem that remains unimpaired in red–green colour blindness.*

Our data on this issue are preliminary ones, but if they prove reliable they need an explanation. Three possibilities can be mentioned briefly here:

(1) The first is that dichromacy may sometimes be associated with a (very) mild cone dystrophy. In genetic alterations that delete or disable both the long- and the middle-wave genes and so produce blue cone monochromacy, a slow macular degeneration is often seen (Nathans, Davenport, Maumenee, Lewis, Hejtmancik, Litt, Lovrien, Weleber, Bachynski & Zwas, 1989). It is possible that the presence of malfunctioning cones in some dichromats leads to a very mild form of cone dystrophy.

*Isolated antecedents of our finding can be found in the literature. As early as 1881, Bruno Kolbe remarked of one daltonian subject: "Auch beim Rotgrünblinden ist eine Herabsetzung des Blau-gelb-sinnes bemerkbar, doch ist dieselbe verhältnismässig geringer" (Kolbe, 1881). In the case of both normals and daltonians, wavelength discrimination near 500 nm and saturation thresholds near 570 nm are thought to depend only on the older subsystem of colour vision (Mollon, Estévez & Cavonius, 1990); and yet in several published studies the daltonians' performance is somewhat worse than that of normals. An example can be seen in the saturation thresholds of Chapanis (1944) which are conceptually similar to our own measurements: relative to normal observers, a severe deutan required more monochromatic yellow light to be added to white before he saw the field as coloured. Unfortunately, in this and similar studies, the "normals" were often more highly trained than the colour-deficient, and this must limit any conclusion drawn.

One example is already known, although the genetic rearrangement is an unusual one: Reichel *et al.* have described macular degeneration in protanopic male members of a family exhibiting a 6.5 kilobase deletion in the long-wave opsin gene (Reichel, Bruce, Sandberg & Berso, 1989).

(2) Secondly, owing to the absence of one class of cone, the opposed inputs to the residual colour channel of the dichromat might be less well balanced than in the normal case.

(3) A quite different category of explanation might be that colour in general is not a very salient cue for dichromat and that a lifetime of not attending to this feature of the world leads to poor performance when the task requires perceptual segregation on the basis of the residual colour signal. Explanations of this kind are *post hoc*, since it could equally be argued that the red–green dichromat would have learnt to pay particular attention to his residual colour signal.

The diversity of anomalous phenotypes

Anomalous trichromats show a very large variation in the extent to which their discrimination is impaired along the confusion axis of the corresponding dichromat, and indeed the discrimination ellipses of some deuteranomals do not exhibit a long axis that lies in the deutan direction when plotted in the CIE u', v' diagram (Fig. 11). Some observers classed as deuteranomalous reach the maximum excursion available on the colour monitor, whereas PJ exhibits a chromatic discrimination that is comparable to that of the normal observers in our sample. This heterogeneity of anomalous phenotypes is a classical observation. The primaries of the Nagel anomaloscope lie on a common protan and deutan confusion line, and it was early established that there is little relationship between the mid-point of an anomal's Rayleigh match and the extent of the matching range that he will accept (Köllner, 1915; Nelson, 1938; Willis & Farnsworth, 1952; Jameson & Hurvich, 1956; Pokorny *et al.*, 1979; Jameson, Hurvich & Varner, 1982). Nelson's wavelength-discrimination curves for seven deuteranomalous observers exhibit a range from those who resemble dichromats to those who approach a normal performance (Nelson, 1938), and a similar range is seen for protanomalous subjects (McKeon & Wright, 1940). Our own results can in particular be compared with those of Chapanis, who measured saturation thresholds by determining the proportion of monochromatic light that had to be added to a white of 5000 K in order for the subject to perceive the mixture as coloured (Chapanis, 1944): his procedure is equivalent to determining most of the central ellipse of our own measurements. For two of his deuteranomalous observers, the saturation function shows a pessimum close to the deuteranopic neutral point near 500 nm, i.e. the discrimination ellipse is elongated along a deutan confusion line. But one of his deuteranomals shows only a slight impairment in this spectral region (his Fig. 3).

How are we to explain the diversity of anomalous phenotypes? Neitz, Neitz and Jacobs (1991) have suggested that only three amino acids control the spectral sensitivity of the middle- and long-wave pigments, those at sites 180, 277 and 285 in the amino-acid sequence of the opsin (the protein component of the molecule). The larger part of the difference between the normal middle-wave and normal long-wave pigments is accounted for by residues 277 and 285, and since the codons for these two amino acids are close together in the same exon of the gene, they are unlikely to become unyoked by crossing-over events so as to yield an intermediate "anomalous" pigment. Site 180 is thought to be polymorphic in the case of both the middle- and the long-wave genes. Neitz and Neitz (1992) have therefore argued that there exist only two common forms of the long-wave pigment and two common forms of the middle-wave pigment; in each case the pigments differ in spectral sensitivity according to whether serine or alanine is found at site 180. Neitz and Neitz propose that the protanomalous observer has lost the normal long-wave pigment(s) and that his residual red–green discrimination depends on comparison of the quantum catches in the two alternative forms of the middle-wave pigment. Similarly, the deuteranomalous observer has lost the normal middle-wave pigment(s) and, depends on the two forms of long-wave pigment for red–green discrimination. If the anomal's discrimination depends in this way on two pigments with fixed spectral separation, then some separate assumption would be needed to explain why anomalous trichromats vary so vastly in their chromatic discrimination.*

A more traditional hypothesis would be that there *is* variation in the spectral separation of the two residual photopigments present in the long-/middle-wave region: as the two pigments more closely approach each other, the poorer the anomal's discrimination becomes. We may call this explanation the *spectral proximity hypothesis*. A version of this hypothesis can be seen in the account of anomalous trichromacy favoured by M. Alpern *et al.*, who have proposed that the two long-wave

*One such assumption might be that the anomal's two long-wave cones, or two middle-wave cones, are sometimes present in very disproportionate numbers, so in turn reducing the numbers or the sensitivities of post-receptoral chromatic units. *Numerical disproportion hypotheses* of this kind have a long history (Jameson *et al.*, 1982; Pokorny & Smith, 1987). Deeb *et al.* have advanced a general theory of the control of expression, which could be used to explain a variation in the relative numbers of different cone types. They propose that a promoter region lies just upstream of each gene in the opsin array and that a locus-control region further upstream (Nathans *et al.*, 1989) must bind to the proximal region in order for transcription to occur. Only one gene in the array can be thus selected, and the probablity of expression is lower the more distant is the gene from the promoter region (Winderickx, Sanocki, Lindsey, Teller, Motulsky & Deeb, 1992c; Winderickx, Battisti, Motulsky & Deeb, 1992a; Deeb, Lindsey, Hibiya, Sanocki, Winderickx, Teller & Motulsky, 1992). Suppose, following Neitz and Neitz, that a deuteranomalous opsin array contains several genes, all with the codons for long-wave residues at sites 277 and 285 but differing at the codon for 180. Chromatic discrimination will be best when the two genes at the 5′ end of the array carry opposite codons for site 180.

pigments of the deuteranomalous subject are both drawn from among several alternative forms of erythrolabe that can occur in the normal population, whereas the two middle-wave pigments of the protanomalous correspond to alternative forms of chlorolabe (Alpern, 1987). In fact, evidence from human and primate studies suggests that sites 180, 277 and 285 are not the only ones that control the spectral sensitivity of the pigment and that there are others that act, perhaps non-additively, to produce small shifts in the λ_{max} of the photopigment (Winderickx, Lindsay, Sanocki, Teller, Motulsky & Deeb, 1992b; Williams, Hunt, Bowmaker & Mollon, 1992). Merbs and Nathans (1992) have constructed hybrid genes consisting in part of exons drawn from the long-wave gene and in part of exons from the middle-wave gene, and have expressed them *in vitro*: they report that the spectral absorbance of the artificial pigments depends on exons 2 and 4 as well as on exon 3 (which includes the codon for site 180) and exon 5 (which codes for sites 277 and 285).

So it remains possible that there is a family of long-wave pigments with a range of slightly different λ_{max} values, and similarly a family of middle-wave pigments. The spectral proximity hypothesis thus remains a plausible explanation of the diversity of anomalous phenotypes. However, we do offer two alternative proposals.

The first possibility is that the normal long-wave and middle-wave pigments are distinguished by more than just those amino-acid sites (such as 180, 277 and 285) that control spectral tuning. The molecule may also carry a signal, perhaps in one of its extramembrane loops, that indicates its identity and, directly or indirectly, provides a label for post-receptoral cells. If anomalous vision depends on the presence of two types of erythrolabe or two types of chlorolabe, then the precision of colour vision may depend on the extent to which recombination events have left the two pigments with different labels and thus on the extent to which post-receptoral processes are suitably wired. A hypothesis of this kind must assume, with Reid and Shapley (1992) and against Lennie, Haake and Williams (1991), that midget ganglion cells normally draw inputs of opposite sign from distinct classes of cone.*

The second possibility is that recombination events sometimes produce a cone opsin that is not only altered in its spectral sensitivity but also is functionally abnormal (Williams *et al.*, 1992): for example, the molecule may be less readily transported to (and from) the cell membrane, it may be less stable, its quantum efficiency may be altered, its signalling may be impaired. The result may be to decrease, or even increase, the effective optical density of the cone,† or to increase the noisiness of its signal.

*Jameson and Hurvich (1956) proposed that "... the color systems of anomalous observers differ systematically from the normal because of (1) progressive degrees of deviation in the photosensitive substances, and (2) independent reductions of increasing degree in the responsiveness of the paired red–green process ...". But it is uneconomical to suppose that many anomals just happen to inherit two independent alterations. And we today believe that the rearrangement of the opsin gene array is the primary genetic alteration in anomalous trichromacy (Nathans, Piantanida, Eddy, Shows & Hogness, 1986). The present hypothesis offers a route by which rearrangement of the opsin gene array could lead to changes in postreceptoral wiring. Numerical disproportion hypotheses offer another.

†Pokorny and Smith (1987) made the suggestion that abnormal alleles may differ in the amount of pigment produced. It is interesting that Merbs and Nathans (1992) record a tenfold variation in the optical densities of the hybrid pigments they have prepared *in vitro* (their Table 1), although this could merely reflect experimental variation.

REFERENCES

Alpern, M. (1987). Variation in the visual pigments of human dichromats and normal human trichromats. In Committee on Vision (Eds), *Frontiers of visual science*. Washington, D.C.: National Academy Press.

Arden, G., Gündüz, K., & Perry, S. (1988). Color vision testing with a computer graphics system: Preliminary results. *Documenta Ophthalmologica, 69*, 167–174.

Birch, J., Barbur, J. L. & Harlow, A. J. (1992). New method based on random luminance masking for measuring isochromatic zones using high resolution colour displays. *Ophthalmic and Physiological Optics, 12*, 133–136.

Cavonius, C. R., Müller, M. & Mollon, J. D. (1990). Difficulties faced by color-anomalous observers in interpreting color displays. *Society of Photo Optical Instrumentation Engineers Proceedings, 1250*, 190–195.

Chapanis, A. (1944). Spectral saturation and its relation to color-vision defects. *Journal of Experimental Psychology, 34*, 24–44.

Chibret, J. B. P. L. (1887). Contribution à l'étude du sens chromatique au moyen du chromatophotomètre de Colardeau, Izarn et Dr. Chibret. *Revue Générale d'Ophtalmologie*, 49–59.

Cornsweet, T. N. (1962). The staircase method in psychophysics. *American Journal of Psychology, 75*, 485–491.

Cuvinot, P. (1992) *Mesures et développement de tests ophtalmologiques sur le GEPCO*. Mémoire d'ingénieur, Conservatoire National des Arts et Métiers, Paris.

Deeb, S. S., Lindsey, D. T., Hibiya, Y., Sanocki, E., Winderickx, J., Teller, D. Y. & Motulsky, A. G. (1992). Genotype-phenotype relationships in human red/green color-vision defects: Molecular and psychophysical studies. *American Journal of Human Genetics, 51*, 687–700.

Estévez, O., Spekreijse, H., van Dalen, J. T. W. & Verduyn Lunel, H. F. E. (1983). The Oscar color vision test: Theory and evaluation (objective screening of color anomalies and reductions). *American Journal of Optometry and Physiological Optics, 60*, 892–901.

Fallowfield, L. & Krauskopf, J. (1984). Selective loss of chromatic sensitivity in demyelinating disease. *Investigative Ophthalmology and Visual Science, 25*, 771–773.

Galezowski, X. (1868). *Du diagnostic des maladies des yeux par la chromatoscopie rétinienne précédé d'une étude sur les lois physiques et physiologiques des couleurs*. Paris: J.-B. Baillière et fils.

Hart, W. M., Hartz, R. K., Hagen, R. W. & Clark, K. W. (1984). Color contrast perimetry. *Investigative Ophthalmology and Visual Science, 25*, 400–413.

Heard, P. F., Stone, C. J., Gregory, R. L. & Marmion, V. J. (1987). A new computer graphics test for red/green colour anomaly. In Verriest G. (Eds), *Colour vision deficiencies VIII* (pp. 181–194). Dordrecht: Junk.

Huddart, J. (1777). An account of persons who could not distinguish colours. *Philosophical Transactions of the Royal Society of London, 67*, 260–265.

Jameson, D. & Hurvich, L. M. (1956). Theoretical analysis of anomalous trichromatic color vision. *Journal of the Optical Society of America, 46*, 1075–1089.

Jameson, D., Hurvich, L. M. & Varner, D. (1982). Discrimination mechanisms in color deficient systems. In Verriest G. (Eds), *Colour vision deficiences VI* (pp. 295–301). The Hague: Junk.

Jordan, G., Sarkies, N. & Mollon, J. D. (1990). Clinical evidence for the independence of two sub-systems of colour vision. *Perception, 19*, 360.

King-Smith, P. E., Vingrys, A. J. & Benes, S. C. (1987). Visual thresholds measured with color video monitors. *Vision Research, 12*, 73–80.

King-Smith, P. E., Chioran, G. M., Sellers, K. L. & Alvarez, S. L. (1983). Normal and deficient colour discrimination analysed by colour television. In Mollon, J. D. & Sharpe, L. T. (Eds), *Colour vision: Physiology and psychophysics* (pp. 167–172). London: Academic Press.

Kolbe, B. (1881). *Geometrische Darstellung der Farbenblindheit*. St Petersburg: Buchdruckerei der Kaiserlichen Akademie der Wissenschaften.

Köllner, H. (1915). Über Übergänge zwischen normalen Farbensinn und angeborene Rotgrünblindheit und über die Möglichkeit ihrer quantitativen Bestimmung. *Archiv für Augenheilkunde, 78*, 302–335.

Lakowski, R. (1966). A critical evaluation of colour vision tests. *British Journal of Physiological Optics, 23*, 186–209.

Lennie, P., Haake, P. W. & Williams, D. R. (1991). The design of chromatically opponent receptive fields. In Landy, M. S. & Movshon, J. A. (Eds), *Computational models of visual processing* (pp. 71–82). Cambridge, Mass: MIT Press.

MacAdam, D. L. (1942). Visual sensitivities to color differences in daylight. *Journal of the Optical Society of America, 32*, 247–281.

McKeon, K. M. & Wright, W. D. (1940). The characteristics of protanomalous vision. *Proceedings of the Physical Society, 52*, 464–479.

Merbs, S. L. & Nathans, J. (1992). Absorption spectra of the hybrid pigments responsible for anomalous color vision. *Science, 258*, 464–466.

Mollon, J. D. (1982). Color vision. *Annual Review of Psychology, 33*, 41–85.

Mollon, J. D. (1987). On the origins of polymorphisms. In Committee on Vision (Eds), *Frontiers of visual science: Proceedings of the 1985 symposium* (pp. 160–168). Washington, D.C.: National Academy Press.

Mollon, J. D. (1989). "Tho' she kneel'd in that Place where they grew ...". *Journal of Experimental Biology, 146*, 21–38.

Mollon, J. D. & Jordan, G. (1988). Eine evolutionäre Interpretation des menschlichen Farbensehens. *Die Farbe, 35/36*, 139–170.

Mollon, J. D. & Reffin, J. P. (1989). A computer-controlled colour vision test that combines the principles of Chibret and of Stilling. *Journal of Physiology, 414*, 5P.

Mollon, J. D., Estévez, O. & Cavonius, C. R. (1990). The two subsystems of colour vision and their roles in wavelength discrimination. In Blakemore C. (Ed.), *Vision: Coding and efficiency* (pp. 119–131). Cambridge: Cambridge University Press.

Nathans, J., Piantanida, T. P., Eddy, R. L., Shows, T. B. & Hogness, D. S. (1986). Molecular genetics of inherited variation in human color vision. *Science, 232*, 203–210.

Nathans, J., Davenport, I. H., Maumenee, R. A., Lewis, R. A., Hejtmancik, M., Litt, E., Lovrien, R., Weleber, B., Bachynski, B. & Zwas, F. (1989). Molecular genetics of human blue cone monochromacy. *Science, 245*, 831–838.

Neitz, J. & Neitz, M. (1992). Do people with anomalous color vision have anomalous cones? *Investigative Ophthalmology and Visual Science, 33*, 754.

Neitz, M., Neitz, J. & Jacobs, G. H. (1991). Spectral tuning of pigments underlying red–green color vision. *Science, 252*, 971–974.

Nelson, J. H. (1938). Anomalous trichromatism and its relation to normal trichromatism. *Proceedings of the Physical Society, 50*, 661–697.

Pokorny, J. & Smith, V. C. (1987). The functional nature of polymorphism of human color vision. In Committee on Vision (Eds), *Frontiers of visual science: Proceedings of the 1985 symposium* (pp. 150–159). Washington, D.C.: Academy of Sciences.

Pokorny, J., Smith, V. C., Verriest, G. & Pinckers, A. J. L. G. (1979). *Congenital and acquired color vision defects*. New York: Grune & Stratton.

Reffin, J. P., Astell, S. & Mollon, J. D. (1991). Trials of a computer-controlled colour vision test that preserves the advantages of pseudoisochromatic plates. In Drum, B., Moreland, J. D. & Serra, A. (Eds), *Colour vision deficiencies X* (pp. 69–76). Dordrecht: Kluwer.

Reichel, E., Bruce, A. M., Sandberg, M. A. & Berso, E. L. (1989). An electroretinographic and molecular genetic study of X-linked cone degeneration. *American Journal of Ophthalmology, 108*, 540–547.

Reid, R. C. & Shapley, R. M. (1992). Spatial structure of cone inputs to receptive fields in primate lateral geniculate nucleus. *Nature, 356*, 716–718.

Schaaff (1925). L'examen du sens chromatique dans le service des chemins de fer. *Bulletin de L'Académie de Médicine, 94*, 952–955.

Sellers, K. L., Chioran, G. M., Dain, S. J., Benes, S. C., Lubow, M., Rammohan, K. & King-Smith, P. E. (1986). Red–green mixture thresholds in congenital and acquired color defects. *Vision Research, 26*, 1083–1097.

Shepherd, G. M. (1988) *Neurobiology* (2nd edn). Oxford: Oxford University Press.

Steward, J. M. & Cole, L. (1989). What do color vision defectives say about everyday tasks? *Optometry and Visual Science, 66*, 288–295.

Stilling, J. (1877). *Die Prüfung des Farbensinnes beim Eisenbahn- und Marine-personal*. Cassel: Theodor Fischer.

Vierling, O. (1935). Die Farbensinnprüfung bei der deutschen Reichsbahn. Melsungen: Bernecker.

Vingrys, A. J. & King-Smith, P. E. (1986). Factors in using color video monitors for assessment of visual thresholds. *Color Research and Application, 11*, S57–S62.

Williams, A. J., Hunt, D. M., Bowmaker, J. K. & Mollon, J. D. (1992). The polymorphic photopigments of the marmoset: Spectral tuning and genetic basis. *EMBO Journal, 11*, 2039–2045.

Willis, M. P. & Farnsworth, D. (1952). *Comparative evaluation of anomaloscopes*, Rep. No. 190, Medical Research Laboratory, Bureau of Medicine and Surgery, US Navy.

Winderickx, J., Battisti, L., Motulsky, A. G. & Deeb, S. S. (1992a). Selective expression of human X chromosome-linked green opsin genes. *Proceedings of the National Academy of Science, U.S.A., 89*, 9710–9714.

Winderickx, J., Lindsey, D. T., Sanocki, E., Teller, D. Y., Motulsky, A. G. & Deeb, S. S. (1992b). Polymorphism in red photopigment underlies variation in colour matching. *Nature, 356*, 431–433.

Winderickx, J., Sanocki, E., Lindsey, D. T., Teller, D. Y., Motulsky, A. G. & Deeb, S. S. (1992c). Defective colour vision associated with a missense mutation in the human green visual pigment gene. *Nature Genetics, 1*, 251–255.

Wright, W. D. (1941). The sensitivity of the eye to small colour differences. *Proceedings of the Physical Society, 53*, 93–112.

Wyszecki, G. & Stiles, W. S. (1982). *Color science* (2nd edn). New York: Wiley.

Acknowledgements—Supported by SERC grant F24689 and MRC grant G9113630N. During part of the work, B. C. Regan was supported by a Wellcome Trust Vacation Research Scholarship. We thank Professor C. R. Cavonius, Dr G. Jordan, Dr J. Pokorny and Dr M. Webster for advice and discussion.