Pergamon

0042-6989(95)00098-4

Vision Res. Vol. 36, No. 2, pp. 295–306, 1996

0042-6989/96 $9.50 + 0.00

Testing the Indeterminacy of Linear Color Mechanisms from Color Discrimination Data

KENNETH KNOBLAUCH,*† LAURENCE T. MALONEY‡

Received 8 June 1994; in revised form 13 February 1995

It has previously been reported that, for some choices of the fixed spatial and temporal characteristics of test stimuli, it was possible to estimate the spectral sensitivities of chromatic mechanisms from chromatic discrimination data alone. If mechanism sensitivities could be reliably estimated for *any* choice of test stimuli characteristics, the influence of spatial and temporal factors on chromatic discrimination performance could be directly measured. Previous studies, using test stimuli with other spatio-temporal characteristics, have found equi-discrimination contours whose ellipsoidal shapes seem to preclude estimation of mechanisms. Since there is no commonly-accepted method for testing the adequacy of ellipsoidal fits of chromatic equi-discrimination contours, it is possible that alternative psychophysical procedures combined with more powerful statistical tests could detect the pattern of deviations from ellipticality reported previously. In this paper, we describe psychophysical tests and statistical analyses that, taken together, provide a more powerful test of the indeterminacy of mechanisms than previous methods. We develop a method based on analysis of residuals for detecting the pattern of deviations from ellipticality. We apply these tests under fixed experimental conditions similar to those in which other researchers have found ellipsoidal equi-discrimination contours. For these conditions, for any of the tests performed, we do *not* reject the hypothesis that equi-discrimination surfaces are ellipsoidal.

Color Color discrimination Crozier's Law

INTRODUCTION

In a color discrimination experiment, the observer attempts to detect whether a test light has been added to a conditioning background. The spatial and temporal properties of the test light are fixed, and its spectral properties are systematically varied. The observer's performance is typically summarized by specifying those test lights that are equally detectable. When these lights are assigned coordinates in a color matching space, they form nested, equi-discrimination surfaces or, if the experiment is confined to a single plane in color space, nested, equi-discrimination contours. The shapes of the equi-discrimination surfaces may change with the spatial and temporal properties of test stimuli, and the spatial, temporal and spectral properties of any conditioning stimuli. For brevity's sake, we will refer to these factors jointly as the *fixed experimental conditions*.

Typical models of color discrimination assume that color information is initially encoded as the excitations of *linear color mechanisms* and that performance is a scalar function of a *non-linear combination* of these excitations. For any choice of fixed experimental conditions, the experimenter can attempt to select linear mechanisms and a particular rule of combination that account for ('fit') the observed performance. The fit between model and data can be used to test the model as a hypothesis concerning early vision. The spectral sensitivities of the fitted mechanisms and the form of the rule of combination summarize discrimination performance for the fixed experimental conditions considered.

A complete model of color discrimination performance would also predict how changes in the fixed experimental conditions affect the form of the rule and the number and identities of the mechanisms. One straightforward approach to formulating such a model is to estimate mechanism spectral sensitivities and rules of combination for each of several choices of fixed experimental conditions independently. Cole, Hine and McIlhagga (1993, 1994), for example, measured color discrimination performance for several choices of fixed experimental conditions and estimated the spectral sensitivites of plausible linear mechanisms for each choice.

This approach assumes that the spectral sensitivities of linear mechanisms *can* in fact be estimated from chromatic discrimination data. Common models of color

*The Lighthouse Research Institute, The Lighthouse Inc, 111 East 59th Street, New York, NY 10022, U.S.A.

†To whom all correspondence should be addressed at: INSERM Unité 371, Cerveau et Vision, 18 avenue du Doyen Lépine, 69675 Bron Cedex, France [*Email* ken.knoblauch@cismibm.univ-lyon1.fr].

‡Center for Neural Science, Department of Psychology, New York University, New York, NY 10003, U.S.A.

discrimination share a striking 'singularity' pointed out by Poirson, Wandell, Varner and Brainard (1990). When an equi-discrimination surface is ellipsoidal, there are infinitely many choices of linear mechanisms that account for the data equally well: without additional assumptions, the spectral sensitivities of the mechanisms are indeterminate.

As Cole and colleagues note, they succeeded in estimating mechanism sensitivities precisely because they found non-ellipsoidal equi-discrimination contours for all choices of fixed experimental conditions. In contrast, Poirson and Wandell (1990a) reviewed previous studies of color discrimination performance and concluded that the outcomes of these studies were consistent with the hypothesis that discrimination contours were ellipsoidal in shape.

The apparent inconsistency in outcome could be due to any of several differences between the studies reviewed by Poirson and Wandell (1990a), and the studies of Cole and colleagues. It may be that mechanisms can be reliably estimated for some choices of fixed experimental conditions and not for others. Alternatively, it is possible that mechanism spectral sensitivities can be estimated for any choice of fixed experimental conditions, but that different experimental methods or more powerful data analytic methods are needed.

The condition, *that equi-discrimination surfaces are nested ellipsoids, all centered on the base light, all of the same shape*, guarantees that the spectral sensitivities of the mechanisms remain indeterminate even though single threshold measurements are replaced by full psychometric functions.* We refer to this condition as the *Indeterminacy Condition* and the hypothesis that this condition holds (for a given observer and choice of fixed experimental conditions) as the *Indeterminacy Hypothesis*.

In this paper, we develop a suite of psychophysical tests and statistical analyses that, taken together, provide a more powerful test of the Indeterminacy Hypothesis than any previously employed. We apply these tests (described below) under fixed experimental conditions in which other researchers have tested the Indeterminacy Hypothesis and failed to reject it. If the pattern of deviations from an ellipsoid reported by Cole and colleagues were present in our data, we would detect it with high probability.

Our applications of the methods we develop are based on assumptions shared by most models of color discrimination. We summarize these assumptions and explain the tests in the following two sections. We postpone further review of previous experimental work relevant to the Indeterminacy Hypothesis until the Discussion.

*This condition on equi-discrimination surfaces implies that a single linear transformation can simultaneously transform all of the nested equi-discrimination surfaces into spheres. It also implies that this same linear transformation renders all of the psychometric functions in all directions identical. It implies the condition of Poirson and colleagues for a single equi-discrimination surface, but is not implied by it.

NOTATION AND MODELING ASSUMPTIONS

A test stimulus of magnitude I is denoted $Is(t,x,y)\delta$ where $\delta = (\delta_1,\delta_2,\delta_3)$ is a unit vector of mean excitations of long-, medium-, and short-wavelength receptors (L, M and S, respectively) specifying the *color direction* of the test stimulus in LMS space, I denotes the *magnitude* of the stimulus, and $s(t,x,y)$ denotes the *spatiotemporal form* of the stimulus. Here, t denotes time since start of trial, and (x,y) denote retinal coordinates. Let P[DETECT] denote the probability of correctly detecting the presence of the test stimulus. A plot of P[DETECT] versus I is a *psychometric function* in the direction δ. The experimenter can measure distinct psychometric functions for each direction δ in color space.

We use the Quick–Weibull psychometric function (Quick, 1974; Weibull, 1951), based on the Weibull cumulative distribution function:

$$W(I;\alpha;\beta) = 1 - \exp\left[-\left(\frac{I}{\alpha}\right)^{\beta}\right]. \quad (1)$$

The α or 'threshold' parameter determines the equi-discrimination contour corresponding to $1 - 1/e \approx 0.63$ probability of correct discrimination, independent of β. The parameter β is proportional to the slope of the psychometric function at the point $I = \alpha$ in both linear and semi-logarithmic plots. We will refer to it as the *slope parameter*.

Models of color discrimination

Wyszecki and Stiles (1982) describe two broad classes of models of color discrimination, the first based on thresholding of the sum of power transforms of the rectified outputs of linear mechanisms (*one-detector models*), the second based on probability summation among three independent detectors each of which compares the output of a single linear mechanism to threshold (*three-detector models*). If we assume that the decision variables in either model (the quantities compared to a threshold) are distributed as Weibull random variables, then both models predict that equi-discrimination surfaces will be of the form,

$$C = \sum_{i=1}^{3} |\mu_i|^{\rho_i}, \quad (2)$$

where C is any constant and μ_i is the output of the ith linear mechanism. For the three-detector models, the parameters ρ_i are the slope parameters of the Weibull–Quick psychometric functions corresponding to the three detectors. For the one-detector model, the ρ_i are the exponents of the power transform applied to the rectified outputs of each mechanism.

When $\rho = \rho_1 = \rho_2 = \rho_3$, the measured slopes will be the same in all directions:

$$C = \sum_{i=1}^{3} |\mu_i|^{\rho}. \quad (3)$$

The case $\rho = 2$ in equation (3) is consistent with the Indeterminacy Condition: the equi-discrimination contours are then concentric spheres that, when plotted in any color space, will be nested concentric ellipsoids.

DESCRIPTION OF TESTS AND ANALYSES

Residuals analysis

The first test that we propose is an analysis of residuals and will be designed to test whether a *single* equi-discrimination surface (the α surface) described by equation (3) is an ellipsoid. Figure 1(A) is a plot of two equi-discrimination contours based on equation (3). The circular contour is obtained by setting $\rho = 2$ and the non-circular contour by setting $\rho = 4$. Cole *et al.* (1993, 1994) estimated values of ρ that ranged from 2.7 to 6.6. The non-circular contour in Fig. 1(A) is representative of the contours they observed. The radius of the circular contour was set to the mean distance of the non-circular contour from the origin. The plotted circle is the least-square fit of the non-circular contour by a circle and it accounts for 99.6% of the variance, an excellent fit. For each point on the non-circular contour, we record *the residual* of the fit, the distance along the radius from the circle to the non-circular contour (negative if toward the origin, positive if away from the origin). Figure 1(B) is a plot of the residual along each radius versus the direction of the radius θ measured in degrees. In our experiments, the stimuli are symmetric about the base light, and a stimulus in direction θ is the same stimulus as in direction $\theta + 180$. Hence, we only plot residuals for angles between 0 and 180 deg in this example and in the data analysis below. The residuals amount to only 0.4% of the variance, but they are obviously patterned. It is possible to account for an arbitrarily high proportion of the variance and still have patterned residuals.

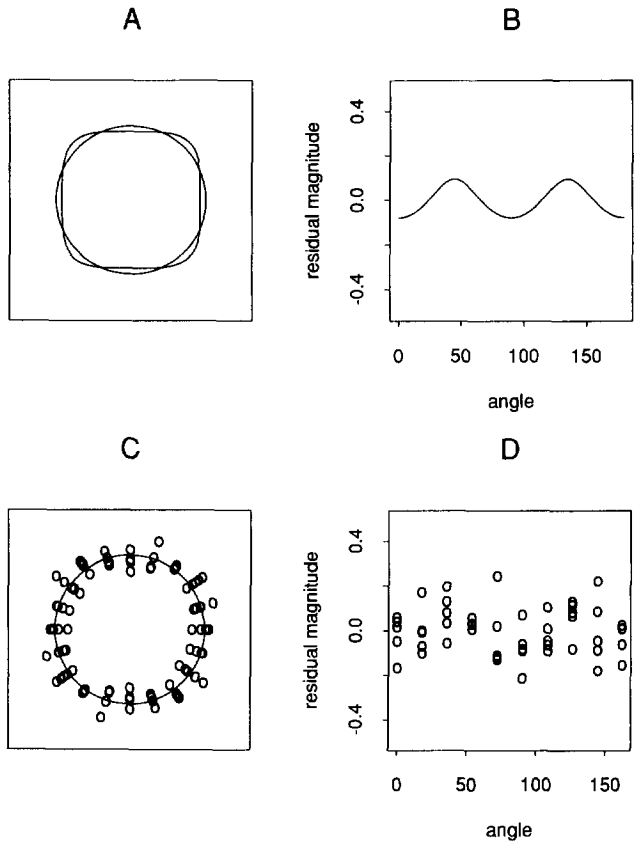

FIGURE 1. (A) Threshold contours for $\rho = 2$ (circle) and $\rho = 4$. (B) A plot of the radial difference between the contours in (A) versus angle from 0 to 180 (the upper half circle). (C) Simulated data. See text. (D) The plot of residuals versus angle for (C).

Suppose that we have measured an equi-discrimination contour in 10 directions with five replications of the measurement in each direction. We fit the resulting data to an ellipse and transform the data and fitted ellipse so that the ellipse is circular. The result might be the simulated data in Fig. 1(C).

The simulated data in Fig. 1(C) were prepared as follows: threshold values for $\rho = 3$ were computed from equation (3) in a plane of color space. The thresholds were normalized across directions to have average length 1. Five threshold settings in each of 10 directions were simulated by adding independent, identically-distributed Gaussian noise with mean 0, standard deviation 0.1 to the normalized thresholds. The directions and number of threshold settings are similar to those used in the experiments below. The magnitude of the Gaussian error is the magnitude estimated from the data in these experiments. The resulting simulated data are plotted in Fig. 1(C). Would the residuals test reject the Indeterminacy Hypothesis for these simulated data?

The magnitudes of the residual errors versus direction are plotted in Fig. 1(D). We would like to test whether the pattern of data in Fig. 1(C and D) is sufficiently inconsistent with the hypothesis $\rho = 2$ (the Indeterminacy Hypothesis) that we should reject that hypothesis.

We first correlated the patterns in Fig. 1(B and D) by multiplying each of the replications in Fig. 1(D) by the value of the curve for the same direction θ in Fig. 1(B), and summing the products. We denote the sum by R. If $\rho_{true} = 2$, then the distribution of positive and negative values of the residuals would be independent of direction. Since the mean of the residual 'template' in Fig. 1(B) is 0, the expected value of R is also 0. If $\rho_{true} > 2$, though, then the residuals in Fig. 1(D) would tend to be positive in the regions where the template residuals are positive, and negative where they are negative. Consequently, when $\rho_{true} > 2$, the products of corresponding residuals in Fig. 1(B and D) will tend to be positive, as will R. If, therefore, R is 'large' we will reject the hypothesis that $\rho_{true} = 2$ in favor of the alternative hypothesis $\rho_{true} > 2$. We chose the threshold for rejection R_{FA} so that the False Alarm probability (rejecting the hypothesis when $\rho_{true} = 2$) is set to some fixed value, typically 0.05 or 0.01.* Once we have

*In this example, for which the distribution of error is known, the values R_{FA} needed can be computed from a table of the z-statistic (Hays, 1988). $R_{FA} = z_{FA}S$ where S is the estimated standard deviation of the residuals and $z_{FA} = 1.65$ for FA = 0.05, $z_{FA} = 2.33$ for FA = 0.01. This formula assumes that the template residuals (see text) have been adjusted to have mean 0 and standard deviation 1.

precomputed the False Alarm rate, by choice of R_{FA}, we would like to find out how powerful our test is: what is $\pi(\rho_{true}) = P[\text{REJECT } \rho = 2 | \rho = \rho_{true}]$, the probability of rejecting the hypothesis as a function of the true value of ρ? This function $\pi(\rho_{true})$ is the power function of the statistical test.

Power of the residuals test

We estimated the function $\pi(\rho_{true}) = P[R > R_{FA} | \rho = \rho_{true}]$, the power function of the test, by simulation. We computed thresholds R_{FA} so that $P[R > R_{FA} | \rho = 2] = FA$ where the 'false alarm' rate FA was set to either 0.01 or 0.05. We considered the values of ρ_{true} shown in the first row of Table 1. For each combination of ρ_{true} and FA, we repeated the following simulation 100 times. We generated fifty residuals, five in each of ten equally-spaced directions, that were independent, identically-distributed Gaussians with mean 0, and standard deviations 0.1. (The number of residuals and the standard deviation were chosen to match those observed in the experiments reported below.) We added the residuals to a contour based on the specific value of ρ_{true} and fit the best circular contour to the resulting data, as above and as in the experiments below. We computed the correlation R and compared it to R_{FA} for each of the False Alarm rates. Note that in all cases we used the residual pattern of Fig. 1(B) as the template ($\rho = 4$).

Table 1 reports the estimated power of the statistic R for selected value of ρ_{true} for false alarm rates of 0.05 and 0.01. This test would almost certainly reject for $\rho > 3.5$ in the conditions of our experiment, and with high probability would reject for $\rho = 3$ or greater. For the particular simulated data set in Fig. 1(C and D), generated with the assumption that $\rho = 3$, the test rejects the hypothesis $\rho = 2$ at the 0.01 level. This test is very likely to detect deviations from $\rho = 2$ smaller than those reported by Cole and colleagues with the number and pattern of data trials that we employed, and with measurement error comparable to that estimated from our own data (see below).

Of course, we do not really know what direction in this coordinate frame to count as 0 deg, since the fit is invariant under rotation. Hence we must look for the pattern of residuals shown in Fig. 1(B) rotated to some arbitrary angle and we must adjust the threshold R_{FA} to obtain the desired False Alarm rate. The necessary changes in the test and the method of setting R_{FA} are discussed below.

Suppose that, in equation (3), $1 < \rho_{true} < 2$. The residual patterns for values $1 < \rho < 2$ are similar to those for values of $\rho > 2$, but rotated 90 deg. Since we will test for the pattern at all angles, the test needs no change to test the null hypothesis $\rho = 2$ (the α surface is an ellipsoid) versus the alternative $\rho \neq 2$ and $\rho > 1$.

TABLE 1. The power of the residuals test

ρ_{true}	2.0	2.3	2.7	3.0	3.3	3.7	4.0
FA = 0.05	0.06	0.32	0.74	0.90	1.0	1.0	1.0
FA = 0.01	0.02	0.10	0.54	0.75	0.93	1.0	0.97

The table summarizes estimates of the probability of rejecting the hypothesis $\rho = 2$ as a function of ρ for two values of threshold R_{FA} chosen to set the False Alarm rate to 0.05 and 0.01, respectively. Each estimate is based on 100 replications of a simulation.

Slope of the psychometric function

The second test is an analysis of the slopes of psychometric functions. Even if the α discrimination contour is indiscriminable from an ellipsoid, we may still be able to reject the indeterminacy hypothesis by consideration of the slope parameters of psychometric functions in different directions in color space. For either model, if the ρ_i are unequal, then the slope parameters of psychometric functions in different directions will differ. The Indeterminacy Condition implies that all the equi-discrimination contours of equation (2) be concentric scaled copies of one another. This is equivalent to requiring that equation (2) can be written in the form

$$1 = \sum_{i=1}^{3} |a_C \mu_i|^{\rho_i}. \qquad (2')$$

for some choice of the scale factor a_C. This is only possible if $\rho_1 = \rho_2 = \rho_3 = \rho$. We may reject the Indeterminacy Hypothesis by showing that slopes in different directions in color space are reliably different. In particular, for either class of discrimination model, the directions in which maximum and minimum slopes are found correspond to the mechanisms with maximum and minimum values of ρ_i. Cole *et al.* (1993, 1994) reported such differences.

Crozier's Law

The third test is an analysis of Crozier's Law. Suppose that the experimenter establishes a direction δ in color space and requests the observer to set the stimulus $\pi + I\delta$ that is just discriminable from π. After N such settings $I_1, \ldots, I_N$. The experimenter computes the variance σ^2 of the settings:

$$\sigma^2 = \frac{\sum_{i=1}^{N} (I_i - \bar{I})^2}{N - 1}$$

where $\bar{I}$ is the mean of the settings. Crozier's Law is the assertion that σ is proportional to the threshold $I_\alpha = \alpha$ for which the test light is detected with fixed probability $1 - 1/e$ measured by the method of constant stimuli (Le Grand, 1949, 1957). There is considerable experimental support for Crozier's Law in brightness discrimination experiments and other sensory modalities (Holway, 1937; Holway & Hurvich, 1937; Crozier, 1950).

Le Grand (1949, 1957) presented a model of discrimination that incorporates Crozier's Law. Our interest here is not the theoretical mechanism underlying the law, but rather to note that the Indeterminacy Hypothesis predicts Crozier's Law whenever the linear mechanisms and rule of combination that control the threshold-setting procedure are the same mechanisms as control performance in a FORCED-CHOICE or YES–NO procedure. If the variability in the settings does not obey Crozier's Law, then, when the experimental results

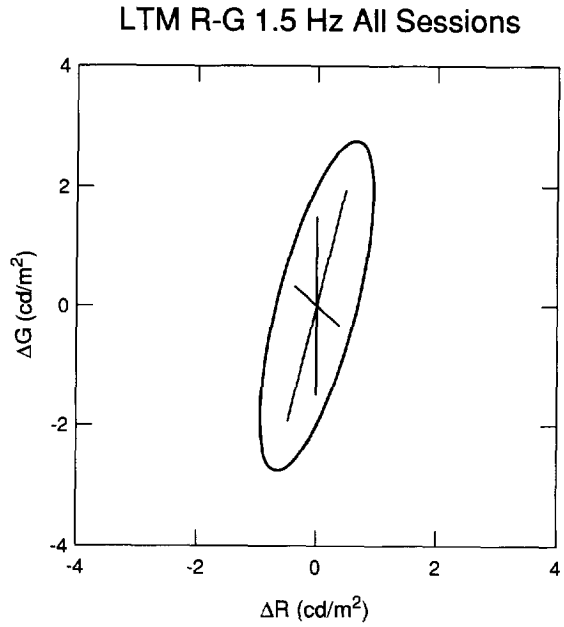

FIGURE 2. Directions investigated using the method of constant stimuli, two-alternative temporal forced choice. (Observer LTM.) The coordinates of the unit vectors in these directions are (0.237,0.972), (0.0,1.0), (−0.739,0.632).

are transformed to a coordinate frame in which all thresholds are equal the values of σ will vary with direction, establishing privileged directions in color space. This outcome would leave open the possibility that these variations in σ could be used to estimate linear mechanisms.

To summarize, we may test the Indeterminacy Hypothesis by testing (a) whether an equi-discrimination contour is elliptical (residuals analysis), (b) whether the slope parameter of the psychometric function is independent of direction, and (c) whether Crozier's Law holds.

In the experiments described next, we will test (a) and (c) for two choices of a plane in color space, and (b) for three directions. If all of the planar cross-sections of a surface are elliptical, then the surface is an ellipsoid. We choose to concentrate our test trials in planes in color space to increase our ability to reject the Indeterminacy Hypothesis by the residuals test described below. The residuals test will measure small, patterned deviations from the elliptical shape and, if the test is to be of any use, we must measure in many co-planar directions in color space. An analogous residuals test in three dimensions could detect patterned deviations from an ellipsoidal shape, but the number of color directions needed, if the test is to have any power, would be large.

METHODS

Experiments were performed using either two-alternative temporal forced choice [2AFC] method or the method of adjustment [MOA]. The results of the MOA experiments allow us to test the Indeterminacy Hypothesis (1) by testing whether the threshold contour is elliptical, and (2) by testing Crozier's Law. The 2AFC results are used to test equality of slope parameters in different directions in color space.

Stimuli

Stimuli were presented on an RGB display (Electrohome) under computer control. The display was calibrated with a combined spectroradiometer and a luminance meter (EG&G) and checked regularly with a hand-held luminance meter (Minolta, CS-100). The test stimulus was always a one degree disk presented on a dark background and viewed foveally. It consisted of a mixture of the three primaries of the television display, under computer control (PC-XT, Data Translation 12-bit D/A converters, linearized with software lookup tables) with CIE chromaticity coordinates of the guns: B (0.161, 0.072), G (0.255, 0.610), R (0.606, 0.356). In the MOA experiments, the temporal waveform was either a 1.5 Hz sine wave or a 12.5 Hz square wave. In the 2AFC experiments, the temporal waveform was a 1.5 Hz cosine wave presented in a Gaussian envelope with standard deviation 335 msec and duration 2 sec. The cosinusoid was symmetric with respect to the Gaussian envelope.

Directions tested in color space were selected by modulating the guns of the display in phase or antiphase in a series of preselected amplitude ratios. All test stimuli were modulated with mean chromaticity coordinates (0.3,0.3) and time-averaged luminance 100 cd/m^2. We considered two planes through this base light. For the first plane, the B gun was held constant and the R and G guns varied to select directions within the plane. The second plane comprised mixtures of the B gun and R + G, the R and G guns added in a fixed 1:1 ratio in lookup table units (in luminance units, the ratio was 1:2.7). Note that stimuli corresponding to opposite directions in color space differ only in temporal phase. Pilot experiments on subject KK showed no evidence of asymmetries in threshold under these conditions. While we measured only one of the two phases, we will plot each threshold point twice, once at its true location and once at the location symmetric through the base light. We plot all data with respect to the luminances of the three guns of the CRT.

In both the MOA and 2AFC sessions, the observer first dark-adapted for 5 min and subsequently adapted to the base light for 3 min before the first trial. In the 2AFC sessions the observer rested for 10 min in the middle of each session. In the MOA sessions, the observer adjusted the amplitude of the modulating stimulus to a value that was just discriminable from a steady light. Each setting was carried out in two stages. The observer initially adjusted the amplitude to near threshold in logarithmic steps. The controlling program then perturbed the observer's setting up or down in intensity at random. The step size was then reset to be equal with adjustments in linear units. The settings obtained with linear step sizes were taken to be the observer's threshold settings. Since we wish to test whether observer's errors are proportional to the

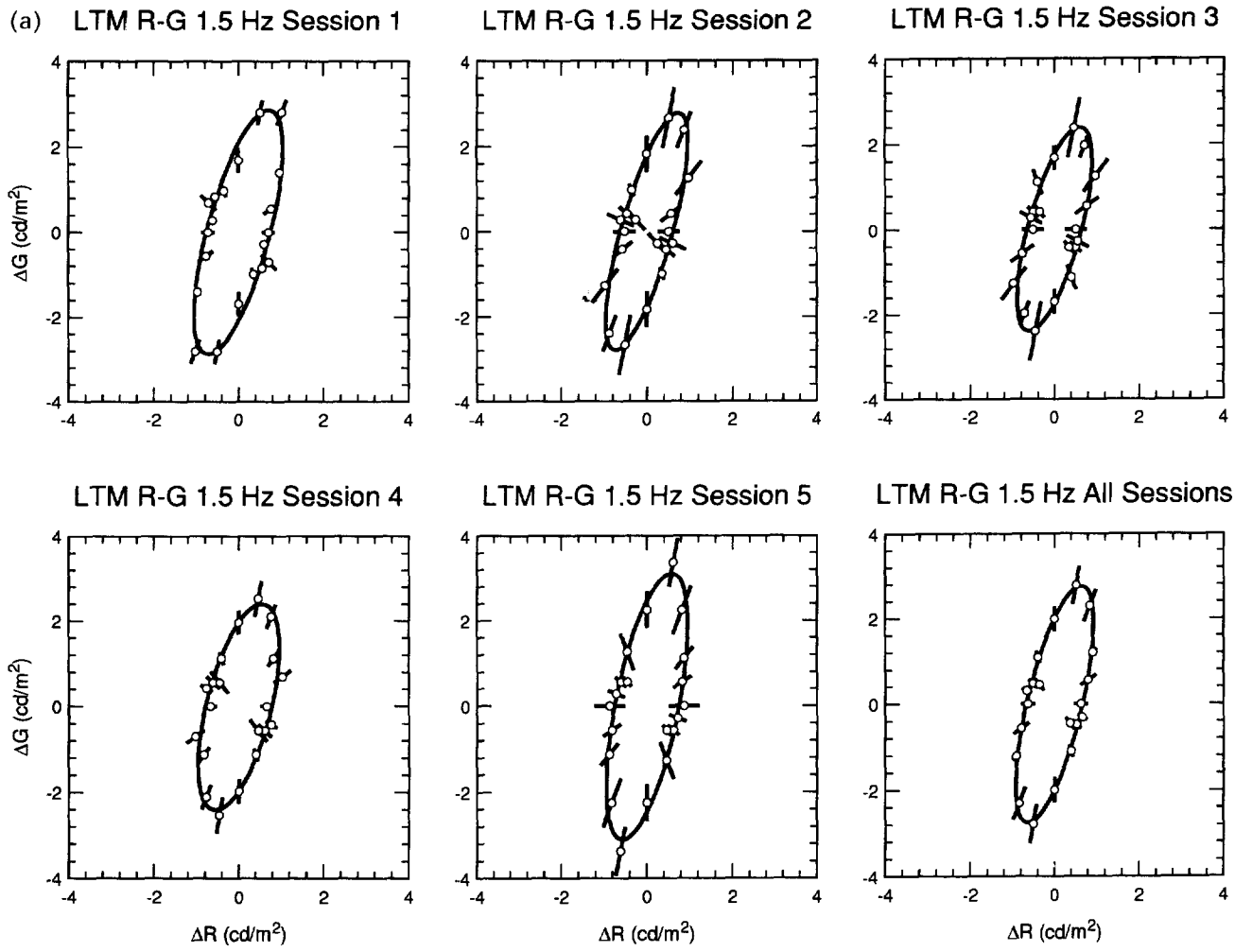

FIGURE 3(a). *Caption opposite.*

magnitude of threshold (Crozier's Law), we did not use step sizes that were proportional to threshold in the final stages of setting the threshold.

The procedure was carried out along each of 10 co-planar color directions tested in each of the two planes. Five sessions were run for each temporal frequency and color plane, with color directions presented in random order throughout a session. Six estimates of threshold were obtained for each direction within a session.

In the 2AFC session, 3 directions were tested, two approximating the longer and shorter directions, respectively, on the R–G discrimination ellipse defined by the MOA data and the third intermediate between those two. Figure 2 shows the directions tested using the 2AFC procedure. The method of constant stimuli was used. Within each of ten sessions, five modulation levels for each of the directions were presented in random order. The set of stimuli was presented 30 times for a total of 450 stimulus presentations per session. The psychometric functions for each color direction were fit with a Quick–Weibull psychometric function using a maximum likelihood procedure (Watson, 1979).

Subjects

Two normal trichromats and a dichromat (protanope) served as observers. One of the trichromatic observers and the dichromatic observer were authors. The third subject was naive to the purposes of the experiment. Analyses were not performed until all of the sessions were completed.

ANALYSIS AND RESULTS

Method of adjustment 1.5 Hz: fits of ellipses

For each observer and each condition, we estimated the best-fitting ellipse to each of the data sets collected in 5 sessions, and the best-fitting ellipse to the combined data from the 5 sessions. Let $(\hat{I}\delta_1^i, \hat{I}\delta_2^i, \hat{I}\delta_3^i)$ be a setting of threshold in direction δ^i, $i = 1, \ldots, N$. Using the method of Poirson *et al.* (1990), we minimize the term

$$\sum_{i=1}^{N} (\hat{I} \| A\delta^i \| - 1)^2 \qquad (4)$$

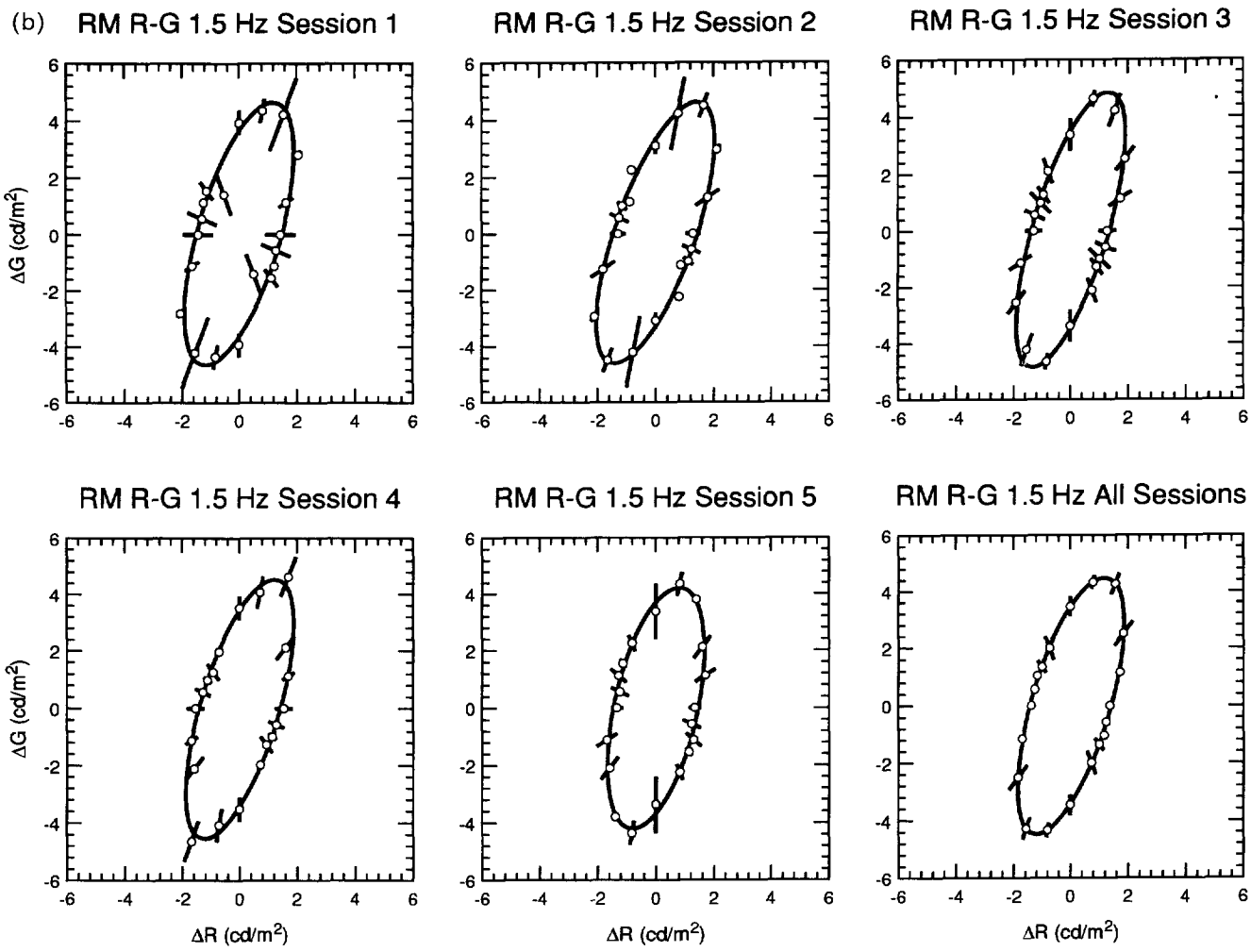

FIGURE 3. Results of the method of adjustment fits for two observers. R–G condition, 1.5 Hz stimuli. The best-fitting ellipses are shown. The error bars are ±1 standard deviation of the observer's settings of threshold.

by choice of the matrix A (where

$$\|\delta^i\| = \sqrt{\sum_{j=1}^{3} (\delta_j^i)^2},$$

the Euclidean length of the vector δ). As Poirson and colleagues demonstrated, there is no unique choice of A to minimize the above. We use their method to determine an A, recognizing that any rotation or reflection of the resulting transformed space is also a minimum of equation (4) above.

At 1.5 Hz, the data were well-described by a fitted ellipse oriented with the long axis in quadrants I and III and an aspect ratio of about 3:1 (in gun coordinates). Figure 3 shows the data for each of 5 sessions for the two trichromatic observers in the R–G plane with the best-fitting ellipses for those data. The mean data for the five sessions and the best-fitting ellipse to those data are also shown. The fitted ellipse for the mean data accounted for 98.0% of the variance of the data for Observer LTM and 99.4% for Observer RM.*

The error bars on each threshold setting are ±1 standard deviation of the observer's settings.

Figure 4 shows the data for each of five sessions for the Y–B plane for two observers, one a trichromat, the other a dichromat. The fitted ellipse for the mean data accounted for 99.6% of the variance of the data for Observer LTM and 98.7% of the variance of the data for Observer KK.

*Let us follow the terminology of Poirson *et al.* (1990) by calling the coordinate system to which the matrix A maps the data the *privileged coordinate frame*. We computed proportion of variance accounted for as one minus the sum of the squared residuals in the privileged coordinate frame divided by the sum of the squared magnitudes of the measured thresholds in the privileged coordinate frame. The resulting number ranges from 1 (the measured thresholds fall precisely on the unit cycle) to near 0 (many thresholds are well inside the unit circle, many well outside).

Method of adjustment 1.5 Hz: residual analyses

The residuals of the fitting procedure in equation (4) can be computed as $\hat{I}\hat{A}\delta^i - \hat{A}\delta^i/\|\hat{A}\delta^i\|$: the vector difference between the transformed data point $\hat{I}\hat{A}\delta^i$ (where $\hat{A}$ is the matrix estimated in the preceding section) and the

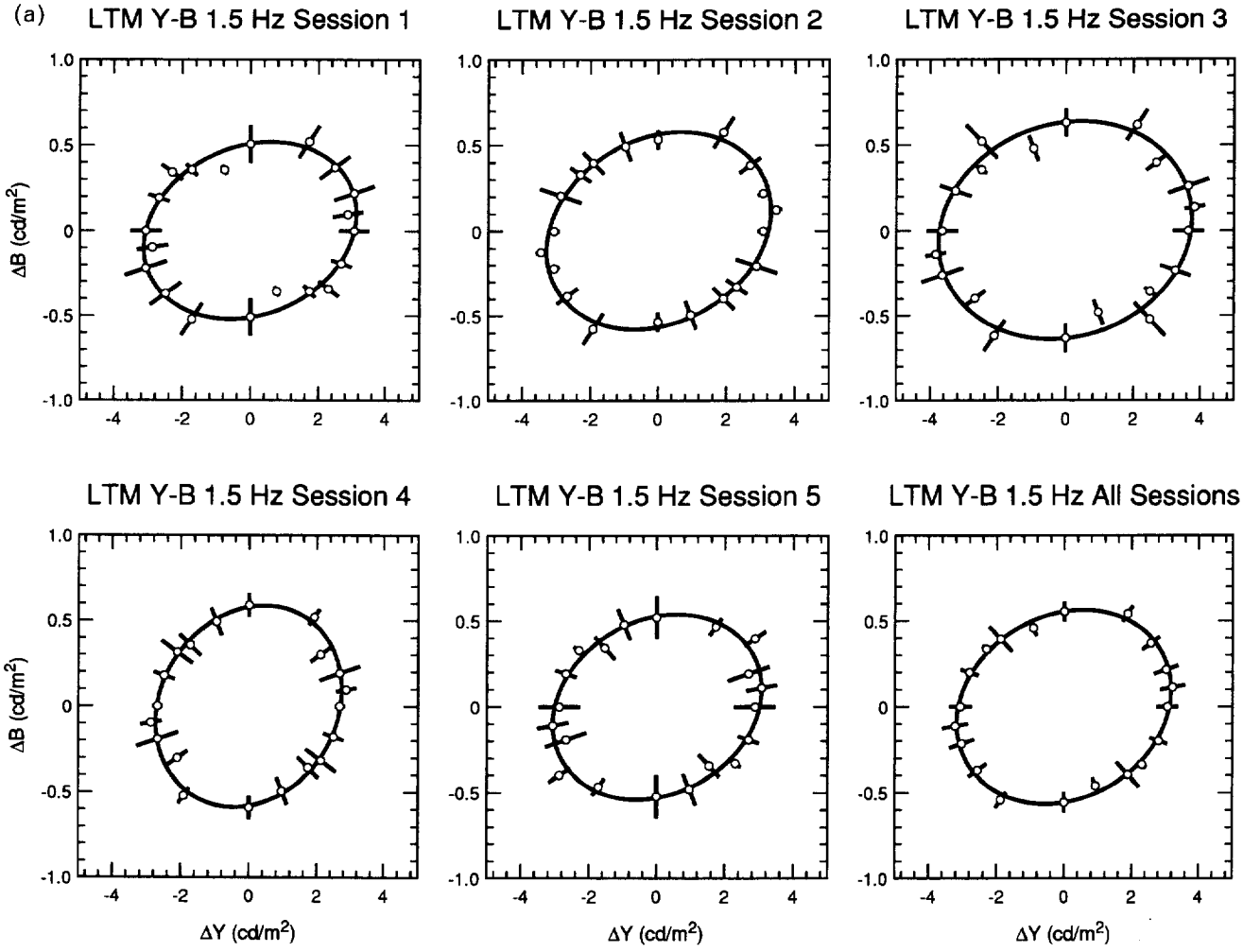

FIGURE 4(a). *Caption opposite.*

unit vector in the same direction $\hat{A}\delta^i/\|\hat{A}\delta^i\|$. The 'variance-accounted-for' measures reported above reflect the quality of the fit obtained to the data by an ellipse. They indicate that the residuals were small: 73.5% were <0.1 in absolute value and 84.0% were <0.2 in absolute value.

The residual data were transformed to polar coordinates and represented as θ_i, R_i where θ_i is the angle between the x-axis and the vector $\hat{A}\delta^i$ and R_i is the signed magnitude of the residual for that vector, positive for outside the circle, negative for inside.

Figure 5 plots the residuals as a function of angle for all of the 1.5 Hz data sets. The pattern of residual failure we seek to detect is the pattern shown in Fig. 1(B) and is defined as follows. Let $\hat{M}_\rho(\cdot)$ denote the inverse Minkowski function*

$$M_\rho(\theta) = 1/(|\cos\theta|^\rho + |\sin\theta|^\rho)^{\frac{1}{\rho}}. \tag{5}$$

and let

$$\bar{M}_\rho(\theta) = M_\rho(\theta) - \frac{1}{2\pi}\int_0^{2\pi} M_\rho(\theta)\mathrm{d}\theta, \tag{6}$$

the inverse Minkowski function adjusted to have mean 0 across the interval $[0,2\pi]$. The contour shown in Fig. 1(B) is a plot of θ versus $\bar{M}_4(\theta)$.

Of course, the direction counted as 0 deg is arbitrary, and we may add an arbitrary shift φ to all the angles (modulus 180 deg) without changing the empirical data. We therefore developed a measure of patternedness of the data which could be maximized by choice of angular shift φ. The measure of patternedness is correlation with $\bar{M}_4(\theta)$:

$$\max_{\varphi} \sum_{i=1}^{50} \bar{M}_4(\theta_i + \varphi)R_i. \tag{7}$$

All maximizations were performed using the STEPT package (Chandler, 1975).

Table 2 reports the values of equation (7) found for each of the four conditions for which residuals are plotted in Fig. 5. For subject LTM in the R–G condition, the measure of patternedness in equation (7) was maximized by shifting the data by 99.9 deg. The patternedness score was 0.148 at this shift.

*The inverse Minkowski function $M_\rho(\theta)$ defines the contours of equation (3) (restricted to a plane in color space) for a given value of ρ. The points $[\theta, M_\rho(\theta)]$ specify a contour in polar coordinates. All of the other contours are scaled copies of it. Fig. 1(A) shows such contours for $\rho = 2$ and $\rho = 4$.

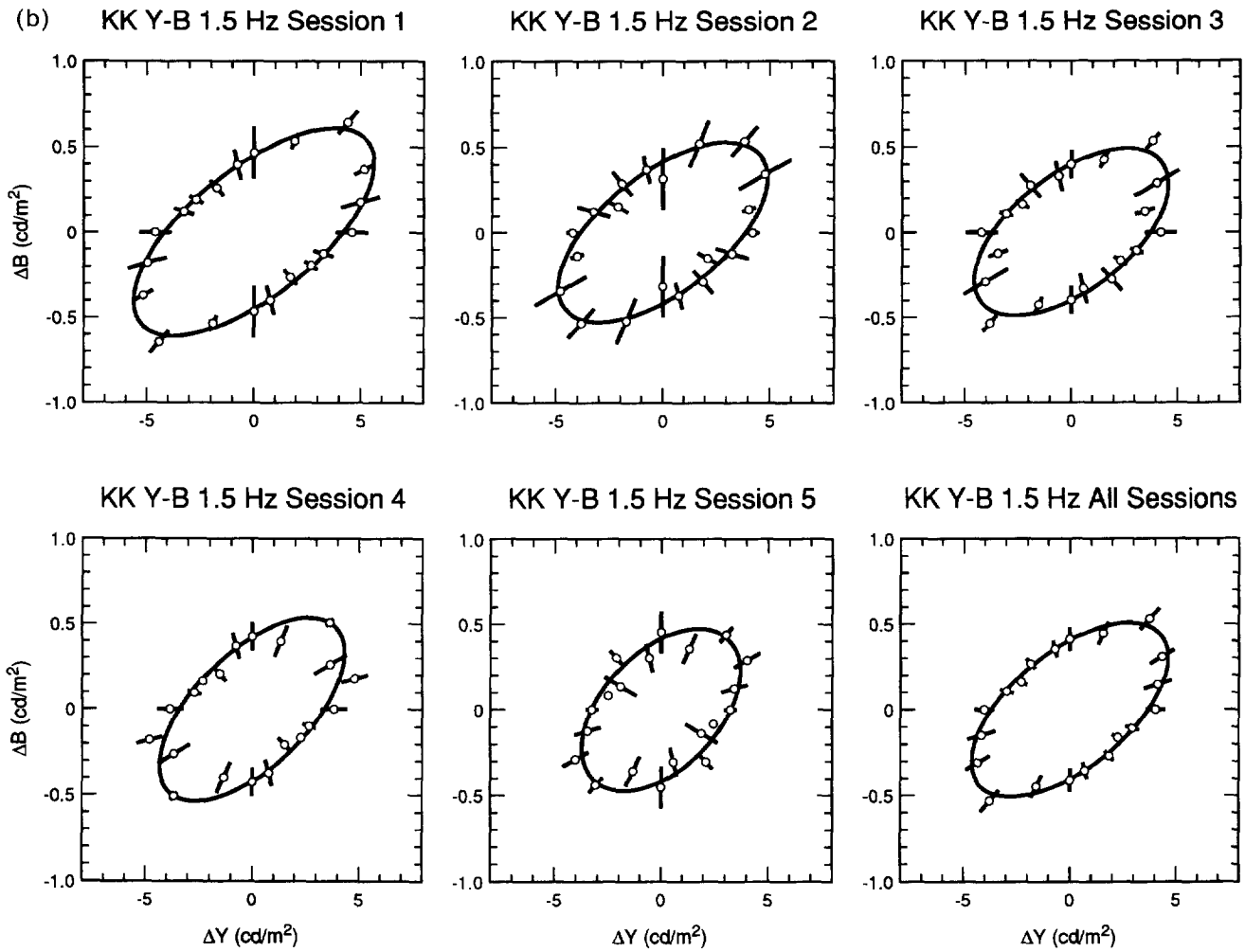

FIGURE 4. Results of the method of adjustment fits for two observers. Y–B condition, 1.5 Hz stimuli. The best-fitting ellipses are shown. The error bars are ±1 standard deviation of the observer's settings of threshold.

We next must decide if any of the patternedness scores are so large that we should reject the Indeterminancy Hypothesis. As in the example earlier in the paper, we must select a threshold value for the patternedness score so that the probability of a False Alarm (the level of the test) is 0.05. We used Efron's bootstrap method (Efron, 1979, 1982; Efron & Tibirishani, 1993) to estimate the 50th, 95th, and 99th percentiles of the patternedness measure in equation (7) for the computed values in Table 2. The Bootstrap method is a general method for estimating the variances and biases of complicated measures such as our patternedness measure. The details of the computation of the confidence intervals reported in Table 2 are given in Knoblauch and Maloney (1994).

Examining Table 2, we see that we do not reject the hypothesis for any of the four observer/conditions. We see that the patternedness value for observer/condition LTM/Y–B exceeds the corresponding 95th percentile. However, with four independent tests,* the probability that one or more will be significant at the 0.05 level is more than 0.12. We conclude that we could not reject the Indeterminacy Hypothesis for any of the four observers when residuals from all sessions are combined.†

Method of adjustment 12.5 Hz

We also fitted ellipses and performed the same statistical tests for the 12.5 Hz data. Similar results were found at a temporal frequency of 12.5 Hz with the major axis now pointed into quandrants II and IV and the aspect ratio increased to 7:1. We could not reject the Indeterminacy Hypothesis.

*Since we are testing four independent hypotheses, one for each of the observer/conditions, we must also correct for multiple tests. We employ the Bonferroni correction procedure (Hays, 1988). If we test each of the data sets at the 0.01 level then the overall probability of a False Alarm is 0.05 level or less.

†We repeated all of the above maximizations and bootstrap computations for two other patterns of residuals in addition to the Minkowski $\rho = 4$ residuals of Fig. 1(B). We used Minkowski residuals corresponding to $\rho = 8$ and also used a sinusoid with period $\pi/2$. The patternedness measure for the sinusoid is, of course, the power spectrum of the data evaluated at frequency 4 cycles per 2π radians. We also repeated the computations with a different patternedness function that allowed the phase shift φ to be different for each of the five sessions. If, for some reason, the observer's residuals were patterned but shifted phase between sessions, this latter measure would detect the shifting pattern. None of these tests reached the appropriate significance level for the test. We could not reject the Indeterminacy Hypothesis by any of these tests.

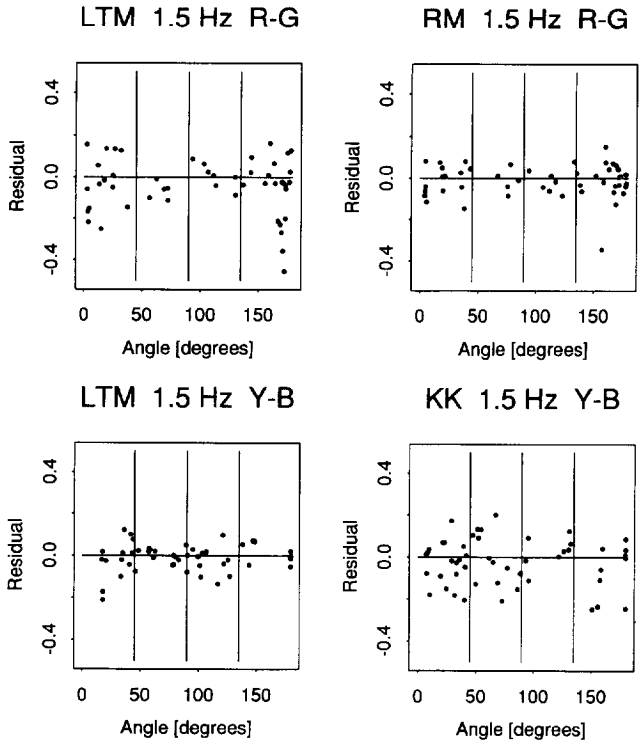

FIGURE 5. Plots of the residuals versus angle for four observer/conditions.

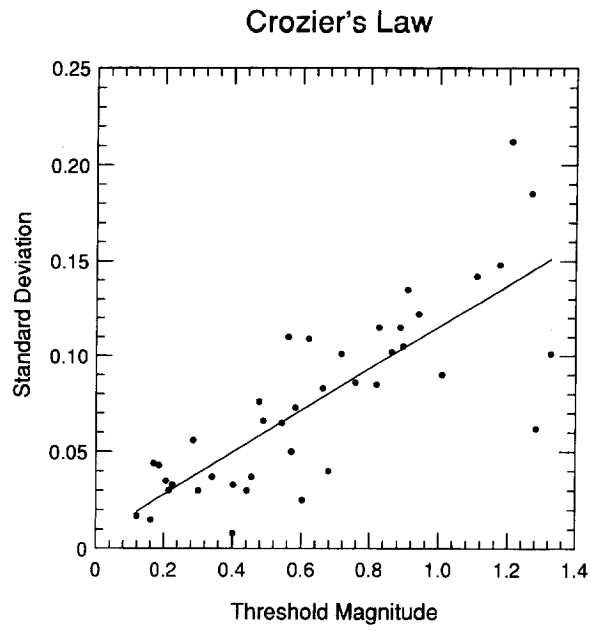

FIGURE 6. Plots of the standard deviations of settings versus mean setting for the method of adjustment data with the best linear fits. 1.5 Hz data. The line shown is the least-squares fit to the data: $y = 0.109x + 0.006$, $R^2 = 0.64$.

Method of adjustment 1.5 Hz and 12.5 Hz: Crozier's Law

Significant linear associations were found between the mean-threshold amplitude and its standard deviation for all observers (Fig. 6) We tested the linearity of the association by means of a Wald–Wolfowitz Runs Test, a standard method for assessing the patternedness of residuals (Hays, 1988). We did not reject the hypothesis that a linear fit is appropriate ($P = 0.122$). We could not reject Crozier's Law or reject the Indeterminancy Hypothesis by rejecting Crozier's Law.

We have not rejected the hypothesis that the single threshold contour measured by Method of Adjustment is an ellipse, consistent with the Indeterminacy Hypothesis. It is still possible that it could be rejected by the data collected using Method of Constant Stimuli with two-alternative forced choice trials.

Fitting the 2AFC data

We fit the data of each session using a generalization of the maximum likelihood method of Watson (1979). We generalized the method to permit maximization of the likelihood of several data sets simultaneously, sometimes with independent parameter settings, sometimes with parameter settings constrained as described below. The maximization was performed using the STEPT Package (Chandler, 1975). Recall that the data included the results of 10 sessions of 150 trials in each of three separate directions (30 sessions total). We will index the sessions by i and the directions by j in the following discussion.

We first fit the data with separate threshold α_{ij} and slope parameters β_{ij} for each session (60 free parameters). The log of the maximum likelihood was -2405.62. We next fit the data with separate threshold α_{ij} for each session and direction but a common slope parameter β_j for each direction (33 free parameters: log maximum likelihood $= -2416.78$). The resulting slope estimates were: Direction (0.237,0.972): $\beta = 2.003$, Direction (0.0,1.0): $\beta = 2.075$, and Direction (-0.739,0.632): $\beta = 2.024$. Last, we fit the data with separate threshold α_{ij} for each session and direction but a common slope parameter β for all sessions and directions (31 free parameters: log maximum likelihood $= -2416.82$). The estimated value of the third fit (one common value across sessions and directions) was $\beta = 2.04$ based on approximately 4500 2AFC trials. Again, in all fits, a separate threshold parameter was

TABLE 2. Analysis of patternedness of shifted residuals by condition and subject with Bootstrap estimates of quantiles of the patternedness measure (see text)

Condition	Subject	Shift (deg)	Patternedness score	50%-tile (Bootstrap)	95%-tile (Bootstrap)	99%-tile (Bootstrap)
R–G	LTM	99.9	0.148	0.089	0.179	0.220
R–G	RM	91.9	0.021	0.044	0.094	0.116
Y–B	LTM	93.7	0.085*	0.036	0.076	0.095
Y–B	KK	77.3	0.074	0.062	0.128	0.160

estimated for each session for each direction so that any session-to-session variability in threshold would not lead to a biased estimate of β (Maloney, 1990).

The three maximum likelihood fits are fits of nested models. We may use a *nested hypothesis test* (Mood, Graybill & Boes, 1974) to test whether the fit is significantly improved by (a) allowing a separate β for each direction, or (b) allowing a separate β for each direction and sessions. The test is computed by taking the difference in the log likelihood ratio λ of two nested models, multiplying it by 2, and treating the result as a χ^2-statistic with degrees of freedom the difference in the number of parameters between the two nested models. The test of the least-restrictive model against the model with separate β for each direction is $2\lambda = 22.28$ with 27 *df*. We could not reject the constrained model in favor of the less constrained at the 0.01 level.

We also could not reject the most nested model (one β for each direction and session) in favor of the second model: $2\lambda = 0.08$ with 2 *df*.* The hypothesis of a single psychometric function with slope = 2 (in log coordinates) could not be rejected by a nested-hypothesis test.

DISCUSSION

We could not reject the Indeterminacy Hypothesis (a) by testing the goodness-of-fit of elliptical contours to the method of adjustment data, (b) by testing the patternedness of the residuals to the elliptical fits of the method of adjustment data, or (c) by testing Crozier's Law. Nor could we (d) find differences in the slopes of the psychometric function for method of constant stimuli measures in three directions. Such differences would also serve to reject the Indeterminacy Hypothesis. We could not estimate the linear mechanisms underlying discrimination performance under these stimulus conditions.

Cole *et al.* (1993) measured color discrimination performance using 2 deg stimuli with Gaussian intensity profiles. The stimuli were flashed and of 200 msec duration, presented on an adapting background chosen to ensure that detection was in the Weber region. They employed a two-alternative forced choice task and fitted the resulting color discrimination data by means of a model of color discrimination based on probability summation among independent color mechanisms. The free parameters of this model included the spectral sensitivities of three color mechanisms and one additional parameter specifying the rule of combination. They reported success in estimating the spectral sensitivities of plausible linear color mechanisms accounting for the data.

In a subsequent paper (Cole *et al.*, 1994) they repeated these measurements with different fixed experimental conditions: spatial Gabor profiles (a Gaussian modulated by a sinusoid) and Craik–Cornsweet profiles flashed for 200 msec. The stimuli were, as before, presented on an adapting background chosen to ensure that detection was in the Weber region. They employed a different fitting procedure, not so closely tied to a particular model of color discrimination, and, again, they reported success in estimating the spectral sensitivities of plausible linear mechanisms. They rejected what we term the Indeterminacy Hypothesis for each of their choices of fixed experimental conditions.

Poirson *et al.* (1990) tested the Indeterminacy Hypothesis with 2 deg disks with a Gaussian timecourse. They estimated threshold in each of several color directions via a two-alternative forced choice staircase method. They fitted the resulting discrimination contour to three classes of surfaces one of which was ellipsoids. They could not reject the ellipsoidal fit in favor of either of the other classes, and concluded that discrimination performance does not permit estimation of 'distinguished directions' in color space. That is, they did not reject the Indeterminacy Hypothesis.

They evaluated the ellipsoidal fit, however, by considering only the proportion-of-variance-accounted-for. This procedure ignores potentially useful information present in the residuals of the fit (the differences between fitted and measured threshold).† Tests based on the patterning of residuals are commonly used in statistics, to test the fit of a model to data: "An appropriate plot of the residuals will often expose gross model violations when they are present" (Chatterjee & Price, 1991). We have described such a residuals test for testing the Indeterminacy Hypothesis and have illustrated its power.

In addition, the methods used by Poirson and colleagues are based on a single equi-discrimination surface, and do not provide a complete test of the Indeterminacy Hypothesis. It is possible for a single equi-discrimination surface to be a perfect ellipsoid and yet find evidence in the slopes of the psychometric function that is sufficient to reject the Indeterminacy Hypothesis.

The variability of observer's settings in color matching tasks potentially provide information concerning linear mechanisms.‡ MacAdam and colleagues

*The very small difference between the log likelihood for the most nested model and the next most constrained model is surprising. The nested hypothesis test is based on the asymptotic convergence of 2λ to a χ^2 variable. The value of λ obtained is improbably small, suggesting that 2λ is not approximately χ^2 for this amount of data. The outcome of the test, however, is not in doubt.

†We note that Poirson and colleagues were primarily interested in determining whether ellipsoidal fits to equi-discrimination contours provided an effective summary of discrimination performance, just as the parameters of a psychometric function describe detection/discrimination performance. Even if the true equi-discrimination contour were not precisely ellipsoidal, the parameters of the fitted ellipse could still provide a useful summary of discrimination performance (Poirson & Wandell, 1990a, b).

‡The connection between variability in setting and discrimination threshold is empirical, not theoretical. Le Grand (1957) presents a model linking variability in setting and discrimination performance, intended to explain Crozier's Law. Poirson and Wandell (1990b) found that discrimination performance varied with the discrimination task that the observer was asked to perform. Comparing performance across various discrimination and discrimination-related tasks may lead to a better understanding of performance in any single task.

examined whether the distribution of settings around a match point in color space was trivariate Gaussian (Brown & MacAdam, 1949; Silberstein & MacAdam, 1945). Their analyses, based on a very large number of settings, did not lead to rejection of the Gaussian hypothesis. The 'equi-variability' contours of a trivariate Gaussian are nested ellipsoids, all the same shape. MacAdam's conclusion and Crozier's Law together imply the Indeterminacy Hypothesis.

When mechanisms cannot be estimated from color discrimination data alone, it may still be possible to constrain the possible choices of mechanisms by a second criterion. Thornton and Pugh (1983) and Eskew and Kortick (1994) combined color discrimination and hue equilibrium judgments to identify mechanisms. Chaparro, Stromeyer, Kronauer and Eskew (1994) and Knoblauch (1995) combined color discrimination with hue identification judgments to similar effect. In an elegant series of experiments, Poirson and Wandell (Poirson, 1991; Poirson & Wandell, 1993, 1996) systematically varied the spatio-temporal as well as the spectral characteristics of stimuli and were thus able to estimate candidate linear color mechanisms. Our results indicate that there are choices of fixed experimental conditions under which it is very difficult to estimate the spectral sensitivities of mechanisms given only color discrimination data. For these choices of experimental conditions, these alternative approaches may prove to be especially valuable.

REFERENCES

Becker, R. A., Chambers, J. M. & Wilks, A. R. (1988). *The new S language: a programming environment for data analysis and graphics* (pp. 87–117). Pacific Grove, CA: Wadsworth.

Brown, W. R. J. & MacAdam, D. L. (1949). Visual sensitivities to combined chromaticity and luminance differences. *Journal of the Optical Society, 39*, 808–834.

Chaparro, A., Stromeyer, III, C. F., Kronauer, R. E. & Eskew, Jr, R. T. (1994). Separable red–green and luminance detectors for small flashes. *Vision Research, 34*, 751–762.

Chandler, J. P. (1975). STEPT—Direct search optimization; solution of least squares problems. Quantum Chemistry Program Exchange, QCEP Program No. 307, Chemistry Department, Indiana University.

Chatterjee S. & Price, B. (1991). *Regression analysis by example*, 2nd edn. New York: Wiley.

Cole, G. R., Hine, T. & McIlhagga, W. (1993). Detection mechanisms in L-, M-, and S-cone contrast space. *Journal of the Optical Society of America A, 10*, 38–51.

Cole, G. R., Hine, T. & McIlhagga, W. (1994). Estimation of linear detection mechanisms for stimuli of medium spatial frequency. *Vision Research, 34*, 1267–1278.

Crozier, W. J. (1950). On the visibility of radiation at the human fovea. *Journal of General Physiology, 34*, 87–136.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics, 7*, 1–26.

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Philadelphia: Society for Industrial and Applied Mathematics.

Efron, B. & Tbirishani, R. (1993). *An introduction to the bootstrap*. London: Chapman & Hall.

Eskew, Jr, R. T. & Kortick, P. M. (1994). Hue equilibria compared with chromatic detection in 3D cone contrast space. *Investigative Ophthalmology and Visual Science, Suppl., 34*, 1555.

Hays, W. L. (1988). *Statistics*. New York: Holt, Rinehart & Winston.

Holway, A. H. (1937). On the precision of photometric observations. *Journal of the Optical Society of America, 27*, 120–123.

Holway, A. H. & Hurvich, L. M. (1937). On the discrimination of minimal differences in weight: I. A theory of differential sensitivity. *The Journal of Psychology, 4*, 309–332.

Knoblauch, K. (1995). Dual bases in dichromatic color space. In Drum, B. (Ed.), *Colour vision deficiencies XII*. Dordrecht: Kluwer.

Knoblauch, K. & Maloney, L. T. (1994). Tests of the indeterminacy of chromatic mechanisms from chromatic discrimination data. *Mathematical studies in perception and cognition 94-1*. New York: Department of Psychology, New York University.

Le Grand, Y. (1949). Les seuils différentiels de couleurs dans la théorie de Young. *Revue d'Optique, 28*, 261–278.

Le Grand, Y. (1957). *Light, colour and vision*. London: Chapman & Hall.

Maloney, L. T. (1990). The slope of the psychometric function at different wavelengths. *Vision Research, 30*, 129–136.

Mood, A. M., Graybill, F. A. & Boes, D. C. (1974). *Introduction to the theory of statistics: Third edition*. New York: McGraw-Hill.

Poirson, A. B. (1991). Appearance and detection of colored patterns. Unpublished Doctoral Thesis, Stanford University, CA.

Poirson, A. B. & Wandell, B. A. (1990a). The ellipsoidal representation of spectral sensitivity. *Vision Research, 30*, 647–652.

Poirson, A. B. & Wandell, B. A. (1990b). Task-dependent color discrimination. *Journal of the Optical Society of America A, 7*, 776–782.

Poirson, A. B. & Wandell, B. A. (1993). Appearance of colored patterns: pattern-color separability. *Journal of the Optical Society of America A, 10*, 2458–2470.

Poirson, A. B. & Wandell, B. A. (1996). Pattern–color separable pathways predict sensitivity to simple colored patterns. *Vision Research*. In press.

Poirson, A. B., Wandell, B. A., Varner, D. C. & Brainard, D. H. (1990). Surface characterization of color thresholds. *Journal of the Optical Society of America A, 7*, 783–789.

Quick, R. F. (1974). A vector magnitude model of contrast detection. *Kybernetik, 16*, 65–67.

Silberstein, L. & MacAdam, D. L. (1949). The distribution of color matching around a color center. *Journal of the Optical Society of America, 35*, 32–39.

Thornton, J. & Pugh, Jr, E. N. (1983). Red/green opponency at detection threshold. *Science, 219*, 191–193.

Watson, A. B. (1979). Probability summation over time. *Vision Research, 19*, 515–522.

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics, 18*, 292–297.

Wyszecki, G. & Stiles, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulae, 2nd edn* (pp. 654–677, 680–689). New York: Wiley.

Acknowledgements—This work was supported by Grant EY07747 from the National Institute of Health to Kenneth Knoblauch, by Grant F49620-92-J-0187 from the Air Force Office of Scientific Research to Laurence T. Maloney, and by Grant EY08266 from the National Eye Institute. We thank Michael Landy, Alex Chaparro, and Charles Stromeyer III for carefully reading and commenting on an earlier draft of this paper. We thank Robert Picardi for technical assistance. Portions of this work have been presented at the Special Topical Meeting of the Optical Society of America, "Advances in color vision." Irvine, California, January–February 1992 and at the European Conference on Visual Perception, Eindhoven, The Netherlands, October, 1994. Most of the analysis was done using the New S system (Becker, Chambers & Wilks, 1988).