COLOR DISCRIMINATION AT THRESHOLD: 
THE APPROACH THROUGH INCREMENT 
THRESHOLD SENSITIVITY

D. B. KIRK

Vision Research Laboratory, 5044 Kresge II, Box 56, The University of Michigan, 
Ann Arbor, MI 48109 U.S.A.

(Received 24 April 1981; in revised form 30 November 1981)

Abstract—A close correlation between Stiles's increment threshold (π-mechanism) approach and color discrimination at threshold was found: a sharp transition from detection of two lights via two π-mechanisms (π1 and π2) to detection via only one π-mechanism (π1) was paralleled by a sharp decline from excellent discrimination between those two lights (even with the rods bleached) to no discrimination. This parallel between detection and discrimination suggests that Stiles's approach can isolate single visual pathways with distinct perceptual correlates.

INTRODUCTION

In his classic 1939 investigation, W. S. Stiles found a surprisingly systematic relationship between the wavelength and radianse of a large, steadily viewed background field, and the wavelength and threshold radianse of a 1', 63 msec test flash. In plotting foveal log test threshold vs log field radianse curves (t.v.r. curves), Stiles (1939) found three distinct component branches, each branch having a fixed shape regardless of the test or field wavelength. Each branch obeyed the invariance properties that changes in the test wavelength caused shifts of each branch strickly parallel to the log test radianse axis, and similarly changes in the field wavelength caused shifts parallel to the log field radianse axis. Stiles (1953) later termed these branches the π1, π4 and π5 mechanisms. Although one additional branch was later found (π2; Stiles, 1953) in some observers' data, and three additional branches (π3, π4 and π5; Stiles, 1953, 1959) had to be admitted when high field radianse were employed, all seven branches had approximately the same shape and all obeyed the above invariance properties.

A π mechanism is defined to be a branch in a t.v.r. curve obeying the above invariance properties, so that these mechanisms are strictly only elements in a mathematical analysis of increment threshold data. However, current evidence indicates that these mechanisms are closely related to many perceptual phenomena as well, e.g. metamerie matching (Pugh and Sigel, 1978), brightness matching (Whittle, 1973), hue cancellation and chromatic adaptation (Larimer, 1981; Pugh and Larimer, 1980; Pugh and Mollon, 1978; Thornton, 1981), and color naming (Krauskopf, 1978). Perceptual correlates for these mechanisms have also been described. For example, Stiles (1949, p. 152) observed that "...in passing from the low to the high intensity range of t.v.i. [or t.v.r.] curves showing the transition from the "green" [π4] to the "blue" [π1] mechanism...the appearance of the test stimulus at just above the threshold changes. The apparent colour becomes purplish instead of blue or blue-green and there is a loss in sharpness of outline." Similarly, Whittel (1974, p. 599) observed that as a violet flash is increased in radianse and crosses the π4 or π5 threshold, "...the flash changes hue and acquires the crisp appearance characteristic of vision with "green" or "red" mechanisms."

One particularly simple and appealing basis for such correlates and relations would be for each π-mechanism to represent the threshold response of a single unidimensional channel or visual pathway. Such a relationship would yield the following application to color discrimination between threshold flashes (i.e. discrimination between two equally detectable stimuli differing only in their spectral composition). Signals equated for the same channel (i.e. π-mechanism) would not be discriminable, whereas signals equated for different channels might well be easily discriminable. Such a relationship is supported by the preliminary experiments of Stiles (1949, cf. below) which suggest that lights detected only via π1 are not discriminable, but that lights detected separately via π1 and π4 are discriminable. Similarly, Rollman and Nachmias (1972, cf. below) found that two lights, one which should be detected via π4 and the other via π5, are discriminable.

A further test of this relationship can be made if, for a fixed field (i.e. steady background) wavelength, two test flash wavelengths can be chosen so that a change in the field radianse causes a fairly sharp shift from detection of each flash via different π-mechanisms to detection via the same π-mechanism. Assuming good discrimination between the flashes when detected via different π-mechanisms, the sharp transition in detection caused by the field radianse change should be matched by a similarly sharp decline in discrimination between the two test flashes.

Using appropriate test and field wavelengths, such a sharp transition of detection via π4 and π1 to detect-
tion via \( n_1 \) only can be obtained. For example, using 600 nm background fields, Stiles obtained approximately the lower envelope of the solid curves in Fig. 1a and 1b as the thresholds for 435 and 480 nm flashes, respectively. The dashed lines show the predicted positions of other \( n \)-mechanisms based on the invariance properties given earlier. Under such conditions, the 435 nm flash was detected via \( n_2 \) for low field radiances and via \( n_1 \) for higher field radiances.

For the 480 nm flash, a sharp transition between the \( n_4 \) and \( n_1 \) mechanisms occurs, so that slightly above 435 nm flashes were detected via \( n_1 \) only.

Thus by measuring the degree of discrimination between the 480 and 435 nm flashes at various radiances of the 600 nm field it is possible to determine if changes in discrimination (if any) are correlated with transitions between detection via different \( n \)-mechanisms. In particular, one can test if equally detectable flashes, one flash detected via \( n_4 \) and the other via \( n_1 \) or \( n_3 \), are discriminable. Similarly, one can test if equally detectable flashes detected only via \( n_1 \) are discriminable. Such a test of discrimination between the 435 and 480 nm flashes detected via \( n_1 \) should be very nearly the most sensitive possible test of such discrimination, since wavelengths appreciably longer than 480 nm cannot be used if \( n_1 \) is to be isolated.

Discrimination between flashes detected via \( n_4 \) and \( n_1 \) or \( n_3 \) was found to be excellent, whereas discrimination between flashes detected only via \( n_1 \) was very poor or absent. The sharp t.v.r. curve transition between detection via \( n_4 \) and \( n_1 \) to detection via \( n_1 \) alone was indeed found to be well-matched with a similarly sharp decline in discrimination.
Color discrimination at threshold

(±10%) from 450 to >900 nm (for an equal energy spectrum), which was verified by calibrating the photodiode from 400 to 700 nm with a United Detector Technology UDT-500 photodiode, which in turn was calibrated by the manufacturer. The Wratten neutral density wedges were calibrated with the PIN-10DF for each wavelength of light they transmitted in the experiment, and for every density used, except for the wedges used in setting the 435 and 480 nm flash thresholds. Here the density was measured every 0.05 log units, linearly interpolating for density values not directly measured.

Photodiode readings were made at the end of each experimental session, and immediately after the absolute calibration session in which the MacBeth illuminometer was used, so the between session change in photodiode voltage, and thus the correct solute calibration session in which the MacBeth illuminometer was used, so the between session change in photodiode voltage, and thus the correct absolute radiance levels, could be calculated. Within session variability was monitored and found to be negligible.

**Procedure**

Two observers participated in this study; both observers used their right eye and had normal color vision as tested by the Farnsworth-Munsell 100 hue test and anomaloscope matching, and neither required corrective lenses. They were told about their performance during and after sessions requiring discrimination between test flashes, but were kept naive of the purpose and theoretical framework of the experiment. The first observer, M.B., was a 25-yr old male with extensive experience as a psychophysical observer, and the second observer, D.W., was a 20-yr old female who had not previously participated in a psychophysical experiment.

Observers were aligned in the apparatus by a combination of (1) determining if the tungsten filament image of the bleaching light fell on the observer's sclera when the observer looked to one side, (2) a rough determination of the Stiles-Crawford effect of the first kind, and (3) by centering the field in its "halo" of scattered light. No attempt was made to dark adapt the observer at the beginning of any session, either the observer's eye was bleached (see below), or the observer was initially adapted to the background field for at least 1 min, practice threshold measurements being made during this time.

The stimuli employed were circular and centrally fixated, consisting of a 3/4", 200 msec, 435 or 480 nm test flash concentric with a steady 5' 600 nm background field. Threshold vs radiance (t.v.r.) curves similar to those of Stiles (cf. Fig. 1a and 1b) were obtained for M.B. by first bleaching his eye for 5 min with a 5.65 log td tungsten "white" light subtending slightly more than 10' of visual angle (to eliminate rod contribution to threshold). Then following 5 min of dark adaptation, a full increment threshold curve was obtained, allowing about 1 min of adaptation time for the lowest 2 or 3 field radiances, increasing the adaptation time to 3-7 min for the higher field radiances, always allowing threshold to stabilize before taking measurements. Threshold settings were always made using the method of adjustment, one 200 msec flash being presented every 1.5 sec.

Most of the discrimination experiments were done with M.B. as the observer; D.W.'s data were collected only to determine if the main features of M.B.'s data held for other observers, and no t.v.r. curves were obtained for D.W. For some of the discrimination sessions, M.B.'s (but not D.W.'s) eye was bleached in the same manner as when the t.v.r. curves were obtained (again to eliminate the rods), except that only one background field was used, introduced 2-3 min after the offset of the bleaching field. Measurements were taken from 6 to 15 min after offset of the bleaching field, a period in which rod threshold should be elevated more than about 1 log unit above their dark-adapted threshold.

The procedure for conducting the discrimination trials was the same regardless of whether or not the observer's eye had been bleached. The discrimination trials for a given field intensity were grouped into blocks, each block consisting of a practice session followed by the actual data gathering period in which thirty test flashes were presented, with 15 presentations of the threshold 480 nm test randomly intermixed with 15 presentations of the threshold 435 nm test. In addition, 5-10 blank trials were randomly intermixed in these 30 flash presentations in some sessions to obtain an estimate of the observer's false alarm rate (found to be between 5 and 10%). The practice session consisted of (1) 40 presentations of each threshold flash, alternating flash wavelengths in groups of 4, (2) 10 practice discrimination trials, and (3) about 15 presentations of each flash, as in (1). The purpose of the practice session was to enable the observer to learn the flash labels and apply them correctly, if possible, to each flash. During this practice session, the 480 nm flash was given the label A or B, the 435 nm flash was given the remaining label, and the observer was given the correct label (only A or B, and not any descriptive labels) as each light was presented. Labels were chosen randomly between each block of 30 trials, but were fixed within a block of trials.

Thirty discrimination trials were then given, the temporal sequence consisting of (1) the investigator raising the appropriate monochromator flap (if the trial was not a blank) and giving a verbal ready signal, (2) the observer actuating the shutter, thus opening it for 200 msec, (3) the observer pressing one of four buttons to signal if either (a) the flash was seen and was probably light A, or (b) the flash was seen and was probably light B, or (c) the flash was not seen, but was most probably A, or (d) the flash was not seen but was most probably light B, (4) the investigator both recording and repeating to the observer that response, and (5) similarly recording and informing the observer if the flash was actually light A, B, or a blank.
RESULTS

M.B.'s t.v.r. curves for \( \lambda = 480 \) and 435, \( \mu = 600 \) are plotted in Fig. 2, fitted by eye with Stiles's standard template (Wyszecki and Stiles, 1967, Table 7.5). The squares and triangles represent the thresholds for the 435 and 480 nm test flashes, respectively. The test and field sensitivities for \( \pi_1 \), \( \pi_2 \), and \( \pi_4 \) fit Stiles's average values (c.f. Fig. 1) with a difference of less than the \( \pi_4 \) and \( \pi_1 \) test and field sensitivity standard deviations found by Stiles (1946, p. 431) (applying Stiles's \( \pi_1 \), standard deviations to M.B.'s \( \pi_2 \) as well as his \( \pi_1 \)).

These agreements, as well as (1) the obvious flatting of M.B.'s t.v.r. curve obtained with a 480 nm test (due to the \( \pi_4 \) to \( \pi_1 \) transition), which Stiles (1946) found for all 20 observers, and (2) the strictly vertical shift of the probable \( \pi_1 \) branch as the test was changed from 435 to 480 nm, together ensure that the branches are properly labeled. T.v.r. curves for \( \lambda = 480 \) and 435 nm and \( \mu = 535 \) and 638 nm were also obtained for M.B.; each \( \pi \)-mechanism was displaced (at least approximately) only parallel to the field radiance axis and agreed with Stiles's average values to within Stiles's (1946) 1 standard deviation values.

Note that, for Fig. 2, the \( \pi_4 \) to \( \pi_1 \) transition for detection of the 480 nm test occurs at 8.5 log quanta sec\(^{-1}\) deg\(^{-2}\), as it did for Stiles's average curves in Fig. 1. Thus above 9.0 (understanding the units to be log quanta sec\(^{-1}\) deg\(^{-2}\) of the 600 nm field) \( \pi_1 \) should be isolated from \( \pi_4 \) by at least about 0.4 log units. Hence by selecting a field radiance between 9.0 and 11.0, both the 480 and 435 nm test flashes should be detected only via \( \pi_4 \), so this range has been labeled the \( \pi_1 \), \( \pi_4 \) range. Between 8.5 and 9.0, it is not clear if \( \pi_4 \) will be isolated, so this ambiguous range has been labeled with a question mark. But for more moderate field radiances between 8.0 and 8.5, the 435 and 480 nm flashes should be detected via \( \pi_1 \) and \( \pi_4 \) respectively, so this range has been labeled the \( \pi_1 \), \( \pi_4 \) range. For field radiances less than 8.0, the 435 and 480 nm lights should be detected via \( \pi_2 \) and \( \pi_4 \) respectively, so this range has been labeled the \( \pi_2 \), \( \pi_4 \) range.

To increase the sensitivities of the statistical tests employed here, M.B.'s discrimination data were pooled over 1/2 log unit bins (discrimination data for}
Color discrimination at threshold

Two such log odds ratios were computed for each bin, one for the data in which M.B. reported detecting the flash (the "detected" data), and one for the data in which M.B. reported not detecting the flash (the "not detected" data). Pooling over 1/2 log unit bins allowed the sample sizes to be sufficiently large so that the sampling distribution of log \( \hat{\alpha} \) was normally distributed with mean \( \log \alpha \), the population log odds ratio of which \( \log \hat{\alpha} \) is the maximum likelihood estimator (Everitt, 1977; Feinburg, 1977; Gart, 1971). The sampling variance is then approximated by

\[
s^2 = \frac{1}{x(1,1)} + \frac{1}{x(1,2)} + \frac{1}{x(2,1)} + \frac{1}{x(2,2)}.
\]

To test for independence (no discrimination), \( \log(\hat{\alpha})/s \) can be compared to a normal (0, 1) distribution, as can \((\log \hat{\alpha} - \log \hat{\alpha}_1)/\sqrt{(s^2 + s^2)} \) to test for significant differences between log odds ratios \( \log \alpha_1 \) and \( \log \alpha_2 \). Since the alternative to the null hypothesis of no discrimination is one of discrimination or positive log \( \hat{\alpha} \) values, the test for no discrimination is one-tailed. Since differences between different log odds ratios may be positive or negative, a test for such differences should be two-tailed.

Within either the \( \pi_1 \), \( \pi_4 \) or ambiguous ranges, no significant differences \( (P > 0.05) \) were found between the log odds ratios when M.B.'s eye had or had not been bleached: for the "detected" data in these ranges, the log odds ratios were nearly equal, differing by less than 0.3. The data were thus pooled over the bleached and non-bleached conditions within each of these ranges. Such pooling was not necessary in the other ranges: all of the data in the \( \pi_2 \), \( \pi_4 \) range were obtained in the bleached condition, and the rods should be saturated in the \( \pi_1 \), \( \pi_1 \) range (at 600 nm, 9.0 log quanta sec\(^{-1}\) deg\(^{-2}\) = 3.8 log scot. td; Aguilar and Stiles, 1954).

In the lower panel of Fig. 3 are plotted the log odds ratios for M.B.'s "detected" data, along with 95\% confidence intervals. The numbers directly below that plot give the obtained one-tailed significance level of departures from the null hypothesis of random responding (no discrimination). The middle panel in Fig. 3 gives t.v.r. curves from Fig. 2 for comparison. Discrimination is quite clearly excellent in the \( \pi_4 \), \( \pi_1 \) and ambiguous ranges, particularly in the \( \pi_4 \), \( \pi_1 \) range where there is less than a 1 in 1 billion chance that the null hypothesis of no discrimination is correct. But the data in the \( \pi_1 \), \( \pi_1 \) range indicate that discrimination is poor or absent here; all of the confidence intervals, even though fairly narrow, include 0, the value expected if there is no discrimination. Similarly, for the \( \pi_1 \), \( \pi_1 \) range there are no significant \( (P > 0.05) \) departures from the null hypothesis of no discrimination.

Again using a two-tailed test for differences between the log odds ratio (as was used to test for differences in bleaching conditions), no significant differences \( (P > 0.05) \) were found between the four \( \pi_1 \), \( \pi_4 \) 1/2 log unit bins. The \( \pi_2 \), \( \pi_4 \) log odds ratios, however, is significantly different at the \( P = 0.025 \) level from all of the other 6 log odds ratios. The ratios for the \( \pi_1 \), \( \pi_4 \) and ambiguous ranges are not significantly different from one another. However, both are significantly different (at the 0.0001 and 0.025 levels, respectively) from the ratio for the 9.0-9.5 range, indicating a rather large change in discrimination for a relatively small change in field radiance of about 1 and 0.5 log units respectively.

The log odds ratios and the significance levels obtained when M.B. reported not detecting the flash...
are significantly different at the 0.05 level for either observer. The 95% confidence intervals for those data are given in Table 1. The two ratios for the "detected" data are not significantly different from each other at the P = 0.05 level.

**Table 1. Log odds ratios and 95% confidence intervals (observer D.W.)**

<table>
<thead>
<tr>
<th>Field radiance range (log quanta (600 nm) sec(^{-1}) deg(^{-2}))</th>
<th>&quot;Detected&quot; data</th>
<th>&quot;Not detected&quot; data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Log odds ratio</td>
<td>95% confidence interval</td>
</tr>
<tr>
<td>8.2-8.6</td>
<td>2.8†</td>
<td>2.8 ± 1.1</td>
</tr>
<tr>
<td>9.1-10.5</td>
<td>0.21*</td>
<td>0.21 ± 1.3</td>
</tr>
</tbody>
</table>

*Denotes P > 0.05; †denotes P < 10\(^{-6}\).
equated at threshold: Ikeda et al. (1970), Boynton et al. (1964), Guth (1965, 1967), Stiles (1967) and Krauskopf (1974) have consistently found a rather complex inhibitory and excitatory relation between cone classes, so that any test additivity between cone classes would be exceptional.

This strictly short-wavelength-sensitive cone input to $\pi_1$ is not contradicted by the finding of Pugh (1976) that $\pi_1$ does not obey the property of field additivity. Those nonadditivities involved fields of wavelength $\geq 550$ nm, but $\pi_1$ test sensitivity can only be measured between roughly 400 and 500 nm (Stiles, 1953), so those nonadditivities do not enter into the argument here. Indeed, Pugh (1976) showed that $\pi_1$ is field additive for fields of wavelength $\leq 500$ nm, a finding consistent with the hypothesis that both the $\pi_i$ field and test sensitivities for wavelengths $\leq 500$ nm are short-wavelength-sensitive cone action spectra.

**Related work**

Stiles (1949, pp. 150–152) proposed a similar experiment to the present one, using 480 and 435 nm stimuli placed side by side and flashed simultaneously on a 600 nm background field. Assuming both flashes are detected via $\pi_1$ at threshold, Stiles proposed that those flashes would be indiscriminable at threshold but would become indiscriminable if the radiance of both were increased to the point where the 480 nm flash is detected via $\pi_4$ as well as $\pi_1$ (cf. Whittle's (1974, p. 599) observations above). Stiles (1949, p. 151) found that "preliminary experiments on these lines indicate that the two stimuli do in fact appear similar over a certain range above the threshold, and begin to differ in apparent colour at about the expected intensity." The present study is, in part, a test of Stiles's proposition that the two lights detected via $\pi_1$ are, indeed, not discriminable, a proposition that appears incorrect. Stiles's finding supports the close correlation found between $\pi$-mechanisms and color discrimination found in this study.

Rollman and Nachmas (1972) employed an experimental paradigm very similar to the present one of presenting one of two equally detectable flashes (or a blank) and having the observer respond (1) whether or not the flash was detected and (2) if the flash was color A or B. The 43 msec flashes used were either green (Wratten 65) or red (Wratten 29) and 10° in diameter. No background field was used (except for one observer so that the green and red flashes should have been detected via $\pi_4$ and $\pi_5$, respectively. Thus since discrimination was quite good, equally detectable flashes, one detected via $\pi_4$ and the other via $\pi_5$, should be discriminable.

Rollman and Nachmas (1972) also found (in agreement with the present experiment) that, if the flashes are discriminable when they are reported detected, some discrimination is often evident even when the observer reports not detecting those flashes. Unless one supposes that the observer gives some "no detect" responses while in a detect state, this is evidence against a low (and high) threshold, two-state theory of detection.

**Acknowledgements**—This investigation has benefited greatly from the support and criticisms of Dr Mathew Alpern and Dr David H. Krantz. Dr Krantz provided the original concept for this investigation, which was generously supported by his grant NSF GB-36642X, as well as NIE training grant EY07022 for vision research to the University of Michigan.

**REFERENCES**


