MEDICAL RESEARCH LABORATORY

U. S. Naval Submarine Base
New London

Volume 9 1950 pp. 54 - 89

THE DISCRIMINATION OF COLOR

I. An Experimental Evaluation of Four Methods for Measuring the Difference Limen of Chromaticity.

by
Harry G. Sperling
and
Dean Farnsworth, LCDR, MSC, USNR

Medical Research Laboratory Report No. 154

BuMed Project NM 003 041.19.01

APPROVED: T.L. Willmon, Captain, MC, USN, OinC 15 Sep 1950

THE DISCRIMINATION OF COLOR

I. An Experimental Evaluation of Four Methods for Measuring the Difference Limen of Chromaticity.

by

Harry G. Sperling

and

Dean Farnsworth, LCDR, MSC, USNR

Reference may be made to this report as follows: Sperling, H.G., Farnsworth, D., The Discrimination of Color, MRL No. 154, Vol. 9, pp. 54-89, 1950.

Color Vision Report No. 23

BuMed Project NM 003 041.19.01

INDEX

ABSTRACT

The difference limen in one dimension through a single point in the I.C.I. chromaticity plane was measured using four different psychophysical methods and four surround brightnesses with four highly trained, color normal observers. The psychophysical methods were: (a) Method of Constants - judged "same or different", (b) Method of Constants - judged "with direction", (c) Method of Paired Comparisons and (d) Method of Adjustments. The data were evaluated using a Latin square design and the analysis of variance.

The results showed: (a) differences in the magnitude of the D.L. over all methods were significant at the 1% level: the smallest D.L.s were obtained with Method of Adjustments and Constants "with direction", (b) inter-observer differences were significant at the 1% level, (c) there was no significant variation in the magnitude of the D.L. with variation of surround brightness, (d) no significant residual interactions were found, (e) comparisons of test-retest variability were made between the total within-subclass variance of each pair of methods resulting in the following order of increasing variability: Adjustments, Constants "with direction", Constants "same or different", Paired Comparisons.

A theoretical relation was derived relating the Constants D.L. to the Adjustments measure.

PREFACE

Visual experimentation depends heavily upon the use of the eye as a "null" instrument, that is as an instrument which detects whether two stimuli are different or not different. It is to be expected that of the many possible methods of setting up the psychophysical conditions of observation in experiments, certain methods would be more economical than others - would give more accurate and more stable results or would require less time to produce reliable answers to problems. The relative merits of the various procedures have been controversial for decades. This report presents the results of a direct comparison of four common standard psychophysical procedures applied to the judgment of color differences: two methods of "constant stimulus", a method of "paired comparisons" and a method of "adjustment".

The results can be applied to the design of laboratory and field experiments, to the design and use of optical equipment and to the comparison of data obtained by the several methods.

This experiment represents the first step in the experimental determination of scales of perceptual differences of chromaticity in the 1931 Standard Observer System.

THE DISCRIMINATION OF COLOR

I. An Experimental Evaluation of Four Methods for Measuring the Difference Limen of Chromaticity

INTRODUCTION

The I.C.I. Standard Observer and Coordinate System, as adopted in 1931[1], permits the specification of any color in terms of a hypothetical three color system based on specified photopic visibility values and the laws of additive color mixture. The chromaticity of a color may be identified by dominant wavelength and purity or by a position in the chromaticity plane in terms of the additive mixture of hypothetical red, green and blue that the "Standard Observer" would require to match it under photopic conditions.

Although the I.C.I. system locates the relative position (on a diagram) of a color it cannot indicate the perceptual distance between colors. Statements such as these can be made about the color of a substance; "it is redder", "it is bluer", "it is more saturated"; but, the present system does not include information which would permit us to say whether two color points representative of somewhat different colorimetric purities and dominant wavelengths will be perceived as the same or different under specified conditions of observation. Nor does it permit statements of how much redder, bluer, or more saturated one color is than another.

Several workers have undertaken the experimental study of perceptible color differences. Outstanding among these are Konig and Dieterici[2], Laurens and Hamilton[3] for spectral colors and Wright[4] and MacAdam[5] for the chromaticity plane. The latter two works offer measures which are intended to predict the perceptible difference of any two colors of the same apparent brightness*. Although the results of Wright and MacAdam's studies follow some general relationships, they fail to show close quantitative agreement, presumably because of the differences in procedures. Different psychophysical techniques and surround illuminations were employed.

There exist numerous results in the psychological literature of the effects of instructions, the serial and positional arrangement of stimuli and of surround conditions upon psychophysical estimation. The unit of perceptual color difference which will eventually enable us to construct a psychological color space must be selected with particular reference to these factors. It must be defined unambiguously in terms of the conditions and operations of measurement. In this study we will evaluate several psychophysical procedures and viewing conditions in order to find the way in which they differ in measuring color sensitivity and will attempt to specify the most stable measure of least noticeable color difference.

Four psychophysical techniques have been selected for evaluation. These are:

(a) Method of Constant Stimulus Differences with judgments of "Same" or "Different".

(b) Method of Constant Stimulus Differences with "Directional Judgments".

(c) Method of Paired Comparisons with judgments of "Same" or "Different".

(d) Method of Adjustment to equality.

---

*The work of the O.S.A. committee on the spacing of the Munsell Colors[6], while it deals with color discrimination was designed for equating supra-liminal distances and is not comparable to the studies of liminal differences.

THE PSYCHOPHYSICAL METHODS FOR SENSORY DISCRIMINATION

While it is impossible to give an extensive historical discussion of the four psychophysical methods in this paper, it is necessary that we treat some of their controversial aspects, especially where one of alternate procedures is chosen for use in this experiment. Extensive treatments may be found in Titchener[7], Kellogg[8], Irwin[9] and Guilford[10].

METHOD OF ADJUSTMENTS

In this method, the observer matches a variable stimulus to a fixed standard using a means of continuous adjustment. From the distribution of errors of match are calculated a measure of central tendency and of dispersion or variability around the central tendency. The difference in physical stimulus units between the standard and the average match yields a measure of Constant Error. The measure of dispersion, the Standard Deviation, Average Deviation or Probable Error serves as an index of the precision or sensitivity of match.

The following problems concerning Method of Adjustments have been raised:

1. Since Method of Adjustments requires matches or equality judgments of the observer, it might be expected to yield different results from those psychophysical methods which require judgments of difference.

2. The adjustments measure yields no separate estimates of discrimination and variability of judgments as do the other psychophysical methods, since the variability measure, S.D. or P.E., must be taken as the measure of discrimination.

It is therefore not possible to evaluate the precision of the discrimination measure as is done using the h or S.D. in Constants and Paired Comparisons.

3. It presents an observational task complicated by a motor adjustment task, thus confounding motor errors with perceptual errors[11, 12].

4. Method of Adjustments has also been criticized for the uncontrolled time of adjustment (Muller[11]). Other workers have attempted more vigorous control of Adjustments but the more precise control of conditions would seem to violate the spirit of the method (Guilford, 10, p. 67).

5. Another objection to Adjustments has been that the observer may obtain secondary visual, kinaesthetic, or tactual cues from the adjustment dial and its starting position.

In the present experiment every attempt was made to minimize the above difficulties. The task was perceptually simple, the linkage between adjustments dial and color mixers allowed delicate control of the stimulus, and secondary visual, tactual or kinaesthetic cues were minimized. The apparatus and procedures are described in the following sections.

METHOD OF CONSTANT STIMULI

The essential characteristics of the Constants Methods are the use of a fixed standard and the verbal estimation of comparison stimuli which are presented together with preceding or following a standard. The comparison stimuli are of fixed number and are of equal physical steps from the standard. The judgments in studies of intensity discrimination and weight lifting are typically "greater", "less" or "equal" (or "doubtful") to the standard, although many variations of instructions have been used. The stimuli are repeated a large number of times, usually in random order and yield a distribution of the proportions of "greater", "less" and "doubtful" estimates. The "greater" and "less" estimates when plotted usually yield ogival distributions. The medians of these two distributions are the limens above and below the standard. The stimulus interval between the two limens is labeled the "interval of uncertainty" and an average value of the stimulus distance from the point of equal occurrence of "greater" and "less" estimates to the two limens is taken as the Difference Limen or measure of discrimination. The stimulus distance from the physical standard to the point of equal occurrence of "greater" and "less" judgments (Point of Subjective Equality) is taken as a measure of Constant Error. The precision of estimation may be obtained from the Average Deviation, Standard Deviation or Probable Error of each of the "greater" and "less" functions.

The application of Constants to chromaticity discrimination requires certain changes over the usual procedure. The judgment of chromaticity is markedly different from that of weights and of auditory and visual intensity stimuli. There exists no simple quantitative agreement between sensory impressions and physical stimulus magnitudes throughout the chromaticity plane. Stimuli spaced along a line in the chromaticity plane vary in both hue and in saturation in such a way that judgments equivalent to "greater" and "less" are difficult to formulate. Any series of stimuli through a point about which we wish to determine a D.L. may involve two or more hues. Thus, if we chose a series of stimuli that fall on a straight line in the I.C.I. diagram, an observer must judge in terms not of a graduated series of perceived magnitudes but of a gradation towards two different hue and saturation points. For example, if we select a series of stimuli which are equally spaced physical mixtures of given blue-green and yellow-green chromaticities, and select as standard a 50% mixture of those two colors, the observer must judge comparison stimuli which change from yellower than the standard to bluer than the standard. We could randomize the stimuli on both sides of the standard, and present them in the same series as is done where graduated magnitudes are dealt with. The judgments, then, would have to be "more yellow" or "more blue" than the Standard. By thus including in the same series stimuli in two directions from the standard, we would require of the observer two criteria to be used together, one for the "more yellow" and one for the "more blue". This might not be a more complicated task for the observer than to judge "heavier" or "lighter" in weight lifting if it were true that chromaticity involved the single perceived dimension of hue. However, since all points within the spectral locus involve changes in saturation as well as hue and since, also, hue alone shifts irregularly in some parts of the chromaticity plane, it appears unlikely that the observer could successfully maintain two criteria simultaneously.

Therefore, a Constants procedure which seemed better suited to chromaticity discrimination than the usual procedure was chosen for evaluation in this study. Stimuli in each direction from the standard were explored separately. Where the D.L. was to be determined on a line which graded from blue-green to yellow-green, a point mid-way was chosen and variable stimuli judged which were spaced in the yellow direction, yielding a psychometric function from which the Limen for the perceptible increment in that direction could be calculated. The same was then done in the blue direction from the standard. The stimulus distance between the two limens became the interval of uncertainty. The P.S.E., D.L., and a standard deviation could be calculated from the slope of each psychometric function. This technqiue has previously been employed by Dimmick[14] for color discrimination. It fulfills the original definition of the D.L. and allows the observer a single criterion on which to judge. At the same time the simplicity of the judgment may be expected to overcome some of the variability due to cautious observers who tend to make a disproportionately high number of "equal" or "doubtful" responses.

PAIRED COMPARISONS

The Method of Paired Comparisons differs from Method of Constants in that random pairs of stimuli from the series are judged together in such a way that each stimulus is judged with each other stimulus an equal number of times. This technique was proposed for psychophysics by Culler[15] who offered it as a modification of the Method of Constants. Culler combined the stimuli in difference steps of physical separation and accumulated proportions of plus and minus judgments which yielded an ogive distribution, just as in Method of Constants. Culler used the P.E. of this distribution as the D.L. It is also possible to use the median, which measures the stimulus difference that is perceived half the time. This method, according to Culler, has the following advantages:

"(1) A maximum range of differences can be obtained with minimum equipment.

(2) It guarantees some very high and some very low proportions.

(3) It permits more very small R (stimulus) differences and stimulates O's caution.

(4) It eliminates the error of 'absolute' impression, by which is meant a constant judgment tendency due to the repeated application of the same standard with every pair." (15, p. 206-207).

SURROUND-TARGET CONTRAST

Each of the psychophysical methods has been studied under four conditions of surround illumination in order that the effects of brightness contrast upon color discrimination may be evaluated. This, as well as being of theoretical interest, is of immediate importance since the several previous studies of color discrimination have either failed to control surround or have arbitrarily chosen a specific surround brightness.

Further, if we may believe the surround to have any effect upon the D.L., then the result cannot be meaningfully reported without relation to the surround.

APPARATUS

A projection colorimeter, constructed by the Bausch & Lomb Optical Company for this Laboratory, was the apparatus used. It provides the following advantages:

(a) comfortable observing conditions using large fields projected on a screen rather than observed through an eyepiece.

(b) bipartite comparison field and surround projected from a single calibrated lamp source.

(c) additive mixture of two filtered beams for both halves of the comparison field, utilizing glass or Wratten color filters.

(d) independent variation of the illuminosity of comparison mixtures and surround.

(e) variation of comparison and surround fields in size over a wide range.

(f) independent variation of the color of surround.

(g) ability to set to any mixture of two chromaticities for comparison of colors at fixed steps or to adjust one mixture field by remote control.

(h) the chromaticity components of both mixture fields equated for luminosity by appropriate filter selection.

The optical system of the instrument is illustrated in horizontal cross-section in Figure 1. In principle, there are three lantern slide projectors combined to project light from a single source utilizing a system of mirrors, condensors and projection lenses. Light from the source A, is condensed C_1, upon the center projection lens L_1, and projected directly upon the screen. Lenses L_2 and L_3 received light from A through condensors C_2 and C_3 reflected at right angles from front surface mirrors M_2 and M_3 through auxiliary lenses L'_2 and L'_3. The light from projection lenses L_2 and L_3 is projected upon the screen through color filters located at F_2 and F_3.

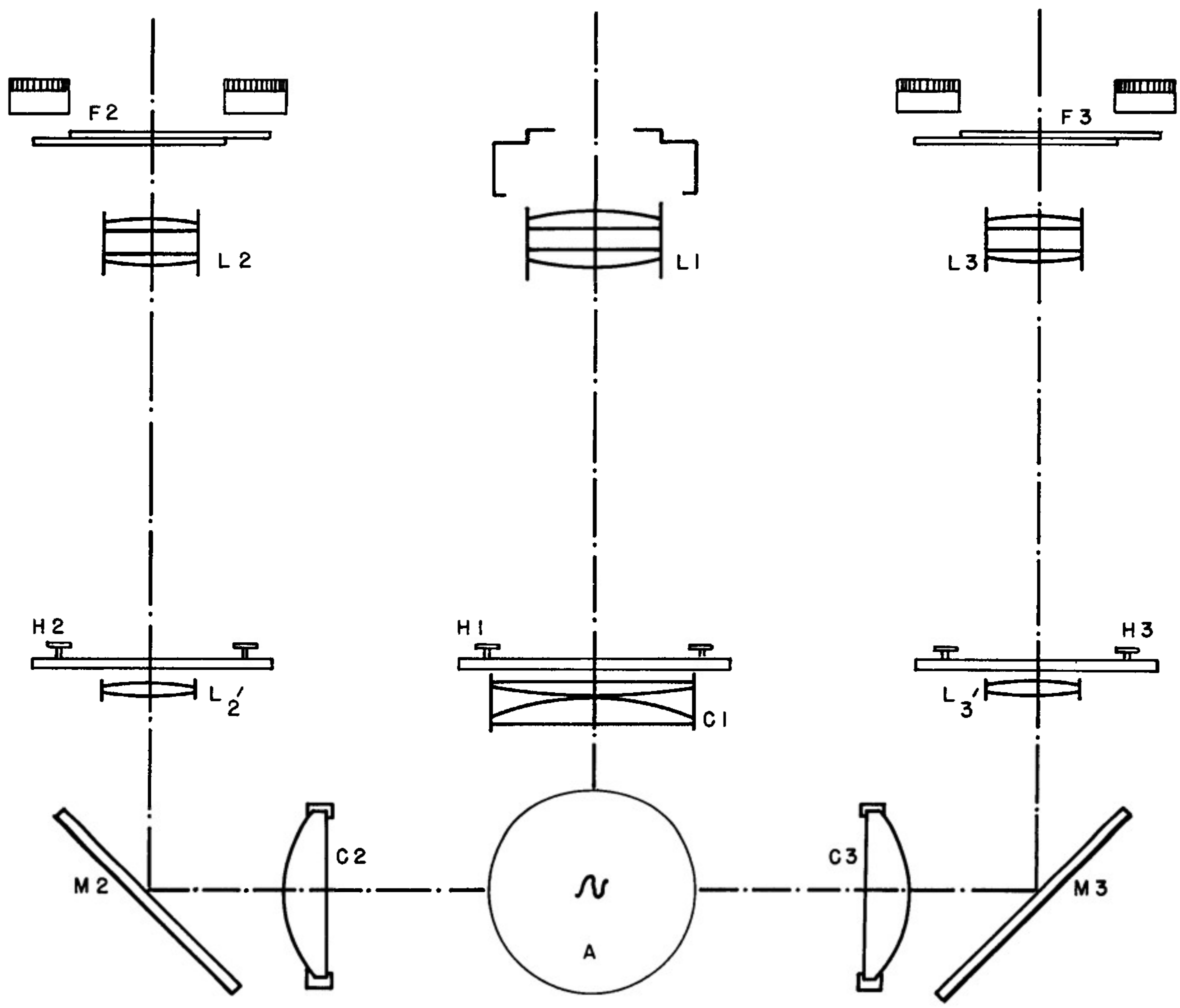

Figure 1

Optical System of Projection Colorimeter

The color filter holders, illustrated in Fig. 2, permit the additive mixture upon the screen of light from two filters of selected chromaticity and the regulation of the brightness of the mixture. This is accomplished by the adjacent positioning of the two filters in a slide carriage which moves in a straight line directly in front of the projection lens. The carriage is controlled by a drum through a rack and pinion connection. The drum is calibrated from 0 to 100 in equally spaced steps. Since the mixture is linear these are read directly as percent mixture of the two component chromaticities. Luminous transmittance is equated by varying the thickness of glass filters or by combining neutral Wrattens with color Wratten filters. The luminance of the mixture fields is controlled by means of variable rectangular apertures which operate from drums through rack and pinion connections. Since these drums are calibrated in equal steps from 0 to 100, the apertures are rectangular, and the utilized central area of the projected beam is homogeneous, the luminance of the mixture on the screen is a linear function of the aperture width.

At positions H_1, H_2 and H_3 in Figure 1 are illustrated 3 1/4 by 4 inch slide holders which carry masking slides. These are focused by lenses L_1, L_2 and L_3 upon the screen. Depending upon the pattern on the slides, different shape, size and separation of the match field and surround is accomplished. In the present work a $2^{o}4'$ diameter circular comparison field was used, divided semicircularly and centered in a square surround field 15^{o} 10' on a side, illustrated in Fig. 3. The observer sat 15 feet from the screen and viewed it binocularly at an angle of approximately 5^{o} from normal. The center of the screen was at eye level. The screen was surfaced with a sphere white paint of high, nonselective, diffuse reflectance.

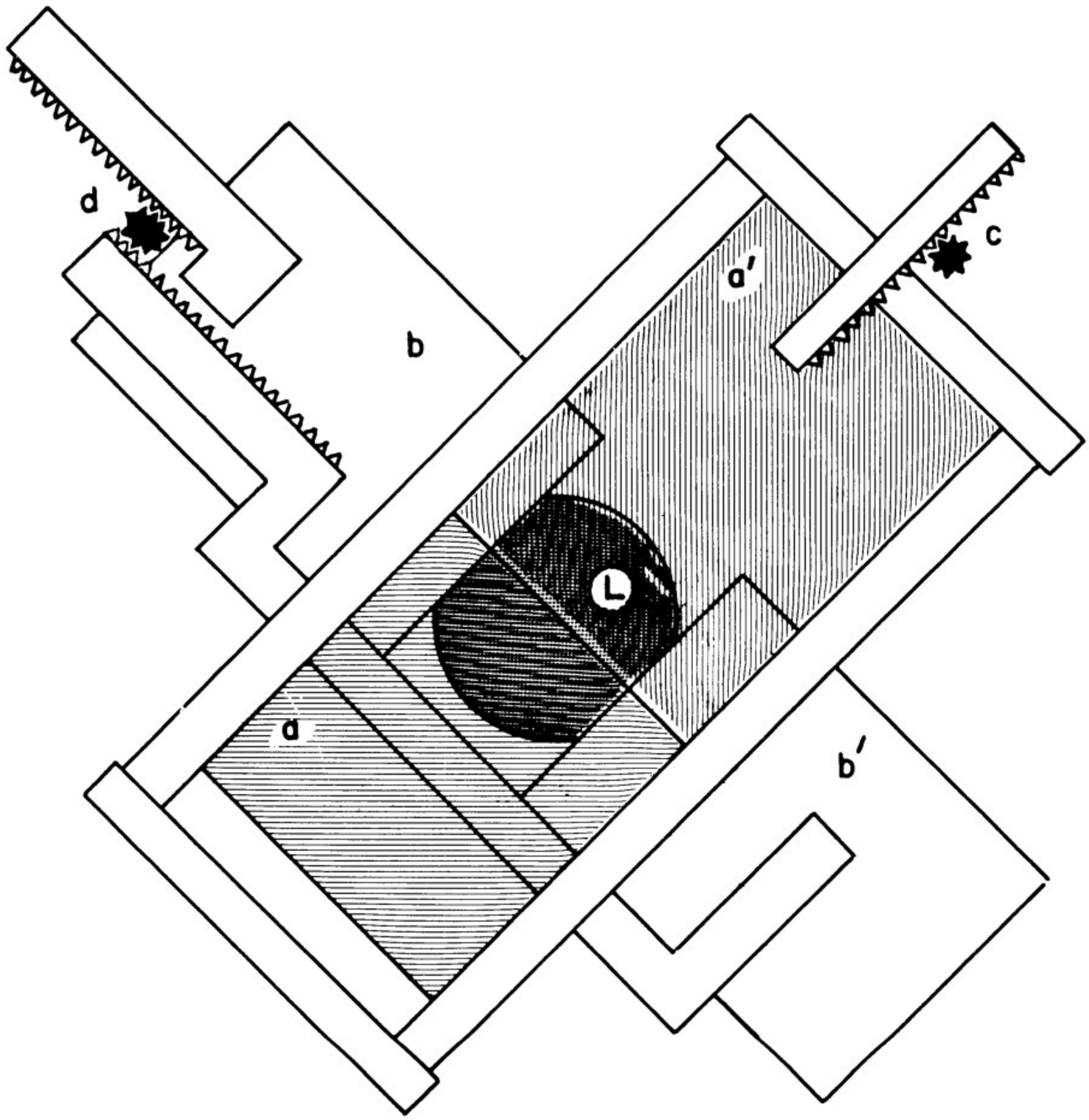

Fig. 2

FILTER HOLDERS OF PROJECTION COLORIMETER

a,a - filters b,b - aperture jaws
c - rack and pinion adjustment of filter slide
d - rack and pinion adjustment of aperture
L - projection lens

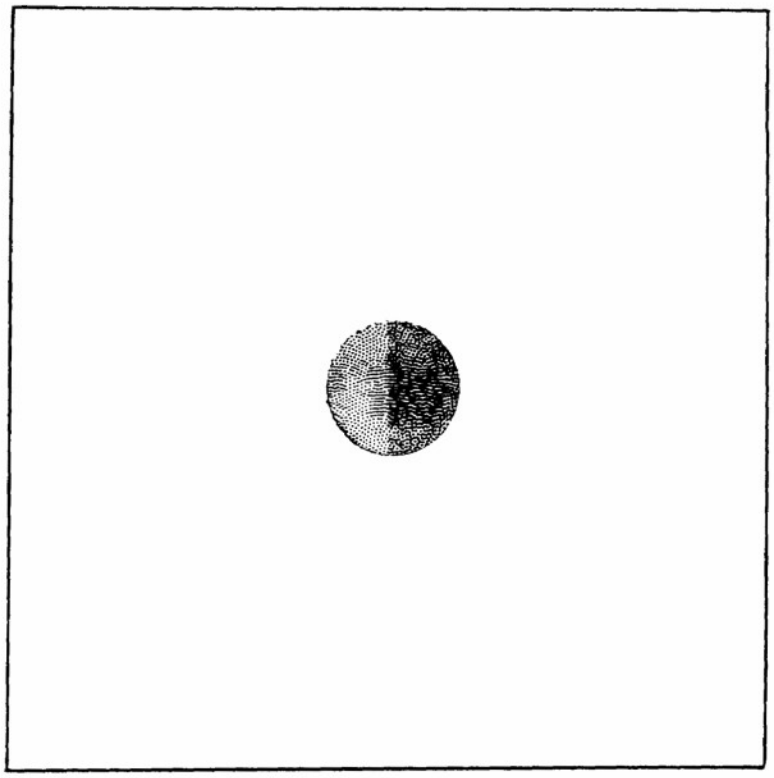

Figure 3

Diagram of Bipartite Comparison Field and Surround Showing Relative Size and Shapes

The light source used was a Westinghouse No. 4 Photoflood maintained throughout the experiment at a color temperature of 3300°K. It was matched visually by voltage regulation to an NBS standard bulb each 1 to $1\frac{1}{2}$ hours of burning and burned on stabilized voltage.

Two Corning glass color filters were selected, one in the yellow-green region of the chromaticity plane, the other in the blue-green region. They were equated for total visual transmission. Their color characteristics weighted for the illuminant which was used, are as follows in the I.C.I. notation:

Filter	x	y	$T_{3300°K}$
Yellow-Green	.386	.559	.384
Blue-Green	.230	.381	.370

Only a small portion of the distance from the standard to the endpoint chromaticities was found necessary to go from 0% to 100% different judgments, in each direction from the standard. The endpoints and range used are illustrated in Figure 4.

One of each pair of filters was held in the filter holders in front of lenses L_2 and L_3. Thus any additive mixture of the two chromaticities could be projected on each of the two halves of the comparison field.

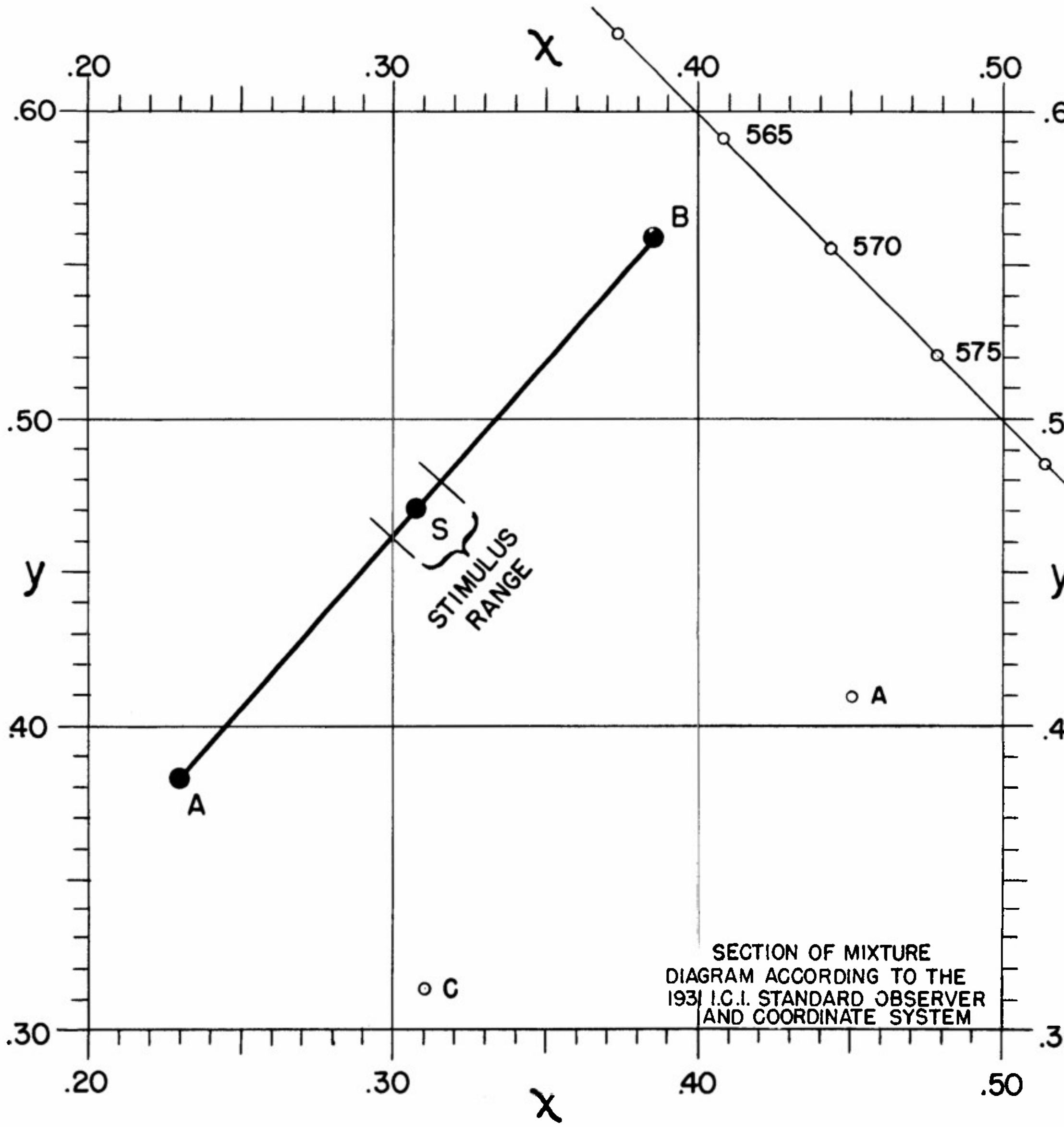

Figure 4

I.C.I. Plot Showing End-point Chromaticities A and B, Standard S and Stimulus Range

PROCEDURE

Each of four observers judged under four surround brightness conditions and by each of the four psychophysical methods. A four by four Latin square design resulted in a different combination of method and brightness to each observer, thus balancing possible remaining practice effects and possible extraneous influences from session to session, while permitting the separate treatment of methods effects, brightness effects and intersubject differences, using Analysis of Variance.

The Latin square design is shown in Table I. Observers are represented by arabic numerals, surround brightnesses by Roman numerals, and psychophysical methods by letters. Each box represents the combination of observer, method and surround brightness under which data for the D.L. were collected.

TABLE I.

Four by four Latin square design employed in studying effects of psychophysical method, surround contrast and observer differences upon the difference limen of chromaticity.

		Surround to Target Contrast			
		I 3:1	II 2:1	III 1:1	IV .05:1
Observers	1	A	B	C	D
	2	B	A	D	C
	3	C	D	B	A
	4	D	C	A	B

A - Method of Constants - "Same or Different"

B - Method of Constants - "With Direction"

C - Method of Paired Comparisons

D - Method of Adjustments

The order of presentation for subjects 1 to 4 is read from left to right in each row. The entire square was repeated once to afford an estimate of test-retest reliability.

The four psychophysical methods were applied as follows:

METHOD A

The Method of Constants Stimuli with Judgments of "same" or "different"

The right half of the comparison field was set at a 50% mixture of the two chromaticities and used as a standard. The variable mixtures were projected on the left half of the comparison field. Ten equal mixture steps along the line connecting the two initial chromaticities were used, five on each side of the standard. Each step was presented 25 times in predetermined random order, such that all five were presented before being repeated. The five on the blue side of the standard were not run together with those on the yellow side. Rather, 25 estimates for the five stimuli on one side of the standard were completed before exploring the other direction. From subject to subject and condition to condition, the blue and yellow series were alternately presented first. The instructions to the observer were as follows:

"On the screen in front of you, there appears a circular colored area, which is split in half. The right half will always remain the same. The left half may appear the same as the right or different. On each trial judge whether the left half appears the same as or different from the right half; say 'same' or 'different'. Between trials fixate a corner of the screen."

The observed proportions were treated and the D.L. calculated using Urban's Constants Process.

METHOD B

The Method of Constants Stimulus Differences

"with direction"

The procedure here was the same as for Method A, except that the observer was given a specific direction in which to judge. Instructions were:

"On the screen in front of you there appears a circular colored area, which is split in half. The right half will always remain the same."

The observer was then shown the 100% mixture in whichever direction, yellow or blue, was to be explored first and he identified the direction of change from the standard. He was then instructed, "Your task is to estimate on each trial whether the left side is 'more blue' or 'not more blue' than the right. Say 'yes' for it is more blue, 'no' for it is not. Between trials fixate a corner of the screen." The same instructions were used substituting "more yellow" for the other side of the standard.

The data were treated in the same way as were those of Method A.

METHOD C

The Method of Paired Comparisons with Judgments of "Same" or "Different"

This method differs from that of Constants in that each stimulus is compared with every other stimulus rather than with a fixed standard. Thus, it yields a D.L. over the interval between the endpoint chromaticities rather than the D.L. around a 50% mixture of the endpoints. The same stimulus series was used as for the constants methods. The instructions were:

"On the screen in front of you there appears a circular colored area which is split in half. On each trial judge whether the two halves appear the same or different. Say 'same' or 'different'. Between trials fixate a corner of the screen."

The proportion of "different" estimates for each difference step between the comparison chromaticities was fitted to the normal ogive using the Constants Process.

The D.L. was computed as the median chromaticity difference.

METHOD D

Method of Adjustment

Here, the observer makes equality matches to a fixed standard and the Probable Error about the mean setting is taken as the D.L. One hundred matches were made, from starting points randomly distributed on either side of the standard well outside the range of uncertainty.

The instructions were:

"On the screen in front of you appears a circular colored area which is split in half. The right half will always remain the same. The left half will appear different from the right on each trial. Your task is to turn the dial on the chair arm in front of you until the two halves appear the same. Then, hold the dial in that position until the experimenter says 'ready' for the next trial. Between trials fixate a corner of the screen."

The four brightness conditions used are summarized in Table II.

TABLE II

Brightness of Screen in foot lamberts measured with bulb at 3300°K after one hour burning.

Condition	Comparison Field	Surround	Contrast Ratio
I	3.3 f.l.	10 f.l.	3:1
II	3.3 f.l.	6.6 f.l.	2:1
III	3.3 f.l.	3.3 f.l.	1:1
IV	3.3 f.l.	.17 f.l.	1:05

OBSERVERS

The four observers used in this experiment had previous training in psychophysical judgment. Two had received this experience as undergraduate psychology majors as well as in subsequent laboratory work. The other two, in addition, received training in graduate school. All observers had normal color vision as demonstrated on the American Optical Company Pseudo-Isochromatic plates, the Double Wedge Anomaloscope, and the Farnsworth-Munsell 100 Hue test.

TRAINING AND COLLECTION OF DATA

The following schedule of observer training was maintained. The results to be reported were the measures obtained in the first and second experimental sessions.

(a) Preliminary Training:

Each observer underwent preliminary training with different chromaticities in the blue-yellow-green region. This served to familiarize him with the general task of color discrimination and with the four psychophysical procedures.

(b) Training Sessions:

The Latin square was run through once in entirety, with the four observers, maintaining the arranged order and the chromaticities and brightnesses to be used in data collection.

Step sizes were varied until the data fell approximately symmetrically about the 50% point and first order reversals were eliminated. Where repititions were necessary they were made before going on to the next prescribed session.

(c) First Experimental Sessions:

The Latin square was followed with one D.L. determination per day, falling in the morning, with ten minute rest periods between each block of 50 observations.

(d) Retraining:

After completing the first experimental sessions each observer was run through the square collecting one half the previous number of estimates per stimulus value.

(e) Second Experimental Session:

With a time elapse of four weeks and following the retraining, data was gathered from each observer as described for the first experimental session.

RESULTS

The difference limens are tabulated in the Latin square form in Table III. This table presents the summary of the difference limens for the inspection of the reader.

The D.L.s are reported in percent mixture of the endpoint lights. Thus, a limen of 5 would mean that on the average a mixture of 5 percent more of one endpoint light and 5 percent less of the other than is included in the standard would be required to reach the difference threshold in one direction. The D.L. is an average of two values, one in each direction from the standard, as computed by the Urban Process.

The D.L.s reported are those of the Urban Process for the Constants Methods and Paired Comparisons. The probable Error is reported as the Adjustments measure.

The Urban Process was chosen because it is the most reliable conventional way of handling Constants data. The P.E. was also chosen, at this stage, since it is the conventional measure in psychophysics.

Table IV summarizes the analysis of variance. There it may be seen that the error term used in all F tests was independent of the residual error term (subclass discrepance) and permitted an evaluation of interactions. The error term used was the within-subclass variance resulting from variability between first and second experimental sessions, with all other effects removed. The F test formed between subclass discrepance (1.13) and within-subclass variance (.56), given in row 5, Table IV, tests the significance of interactions over the entire Latin square. The significance is reported in the P column. The within-subclass variance is of further value since it permits an estimate of the relative reproducibility of the D.L.

TABLE III

COMPLETED LATIN SQUARE

Difference limens for four Psychophysical Methods. The Urban Process is used for Methods A, B and C, and P.E. for Method D.

		Background Contrast				
		I	II	III	IV	Sum X
Os.		3:1	2:1	1:1	.05:1	1 - 4
1	D.L. 1st Exp.	A 5.37	B 3.04	C 5.12	D 1.15	
	D.L. 2nd Exp.	3.22	2.26	2.10	1.37	23.63
	Subclass Total	8.59	5.30	7.22	2.52	
2	D.L. 1st Exp.	B 2.08	A 3.83	D 1.08	C 2.03	
	D.L. 2nd Exp.	1.72	2.18	1.14	1.71	15.77
	Subclass Total	3.80	6.01	2.22	3.74	
3	D.L. 1st Exp.	C 2.12	D 1.73	B 2.37	A 5.16	
	D.L. 2nd Exp.	2.60	1.69	2.65	4.98	23.30
	Subclass Total	4.72	3.42	5.02	10.14	
4	D.L. 1st Exp.	D 1.47	C 2.28	A 2.03	B 1.04	
	D.L. 2nd Exp.	1.44	2.12	2.55	1.20	14.13
	Subclass Total	2.91	4.40	4.58	2.24	
	Sum x I-IV	20.02	19.13	19.04	18.64	

	A	B	C	D
Sum x A -D	29.32	16.36	20.08	11.07

A - Method of Constants - "Same or Different"

B - Method of Constants - "With Direction"

C - Method of Paired Comparisons

D - Method of Adjustments

Table IV - Completed Analysis of Variance

Urban D.L.s and 1 F.E. of Adjustments

Source	SS	DF	$V (M^2)$	F	P
1 Between Observers	9.24	3	3.08	$\frac{V_{obs}}{V_{ind.Error}} = \frac{3.08}{.56} = 5.50$	$P < .01$
2 Between Background Contrasts	.13	3	.04	$\frac{V_{illum.}}{V_{ind.Error}} = \frac{.04}{.56} = .07$	$P > .05$
3 Between Methods	22.17	3	7.39	$\frac{V_{meth.}}{V_{ind.Error}} = \frac{7.39}{.56} = 13.20$	$P < .01$
4 Within-subclass variance (Independent Error)	9.02	16	.56		
5 Subclass-Discrepancy (interactions)	6.82	6	1.13	$\frac{V_{subcl.Discrep.}}{V_{ind.Error}} = \frac{1.13}{.56} = 2.02$	$P > .05$
6 Total	47.38	31			

Since it represents the variability from one experimental session to the next for each observer under the same conditions, it may be broken down and an F test performed between the average within-subclass variance of each of the four methods. Where a level of confidence is reached, we may state that one method is significantly more variable from one testing to the next, than another. This has been done and is reported in Table V.

Table Va

Within-subclass variance of the D.L. over 4 Psychophysical Methods.

Observer	A	B	C	D
1	2.36	.31	4.56	.03
2	1.36	.06	.06	.01
3	.02	.04	.11	.00
4	.13	.01	.01	.00
Total	3.87	.42	4.74	.04
Mean	.97	.10	1.19	.01

Table Vb

Comparison between methods of within-subclass variance.

F =	A	B	C	D
A		9.10*	1.24	95.50**
B			11.28	10.50*
C				118.50**

**P< .01 *P< .05

Table VI

Significance of Differences Between Mean D.L.s#

METHODS

METHODS	A	B	C	D
A		4.50**	2.62*	4.85**
B		t =	1.43	2.55*
C	N=8 (D.L.s)			2.87*

**P < .01 *P < .05

No assumption of homogeneity of variance has been made in evaluating the t ratio (Snedecor 15, p. 83).

The analysis of variance yields the following results:

1. There are significant differences at better than the .01 confidence level between the four psychophysical methods.

2. There is no significant variation in the D.L. as a function of surround brightness over the brightness range explored.

3. Observers differ significantly and consistently in the magnitude of their D.L.s over all methods.

4. There exist no significant interactions between methods, illumination levels and observers. That is, by no significant degree did one or more observers benefit from one or more light levels, nor did one or more methods show an advantage when combined with certain light levels and/or observers.

5. The comparison of within-subclass variance (Table V) by F tests between methods yielded the following:

Method of paired comparisons (C) showed the greatest test-retest variability, but not significantly greater than Method of Constants with "same" or "different" estimates (A). Both Method of Adjustments (D) and Method of Constants "with directions" (B) showed significantly lower test-retest variability than either of the former. Method of Adjustments showed significantly less test-retest variability than did Constants "with directions".

A breakdown of the contribution of each method to the difference between methods may be seen in Figure 6 where mean D.L.s for each method are plotted, together with standard deviations. The significance of differences between individual methods are shown in Table VI where the t test between mean D.L.s has been employed. The methods arrange themselves in the following order with respect to magnitude of the D.L.: A > C > B > D with mean differences significant except between methods B and C.

In addition to the Urban D.L.s and 1 P.E. Adjustments measure, the standard deviations are reported and analyzed in the same fashion in Tables VII, VIII, IX and X. For the Constants data, these are the combined standard deviations of the yellow and blue ogives obtained from the Urban h. The Paired Comparisons standard deviations were calculated in the same way. The Adjustments S.D. is calculated directly from the data by the usual formula.

In Table VII, the S.D.s are tabulated in the Latin square form. Table 8 shows the completed analysis of S.D.s according to the same form used in Table 4 for the Urban D.L.s.

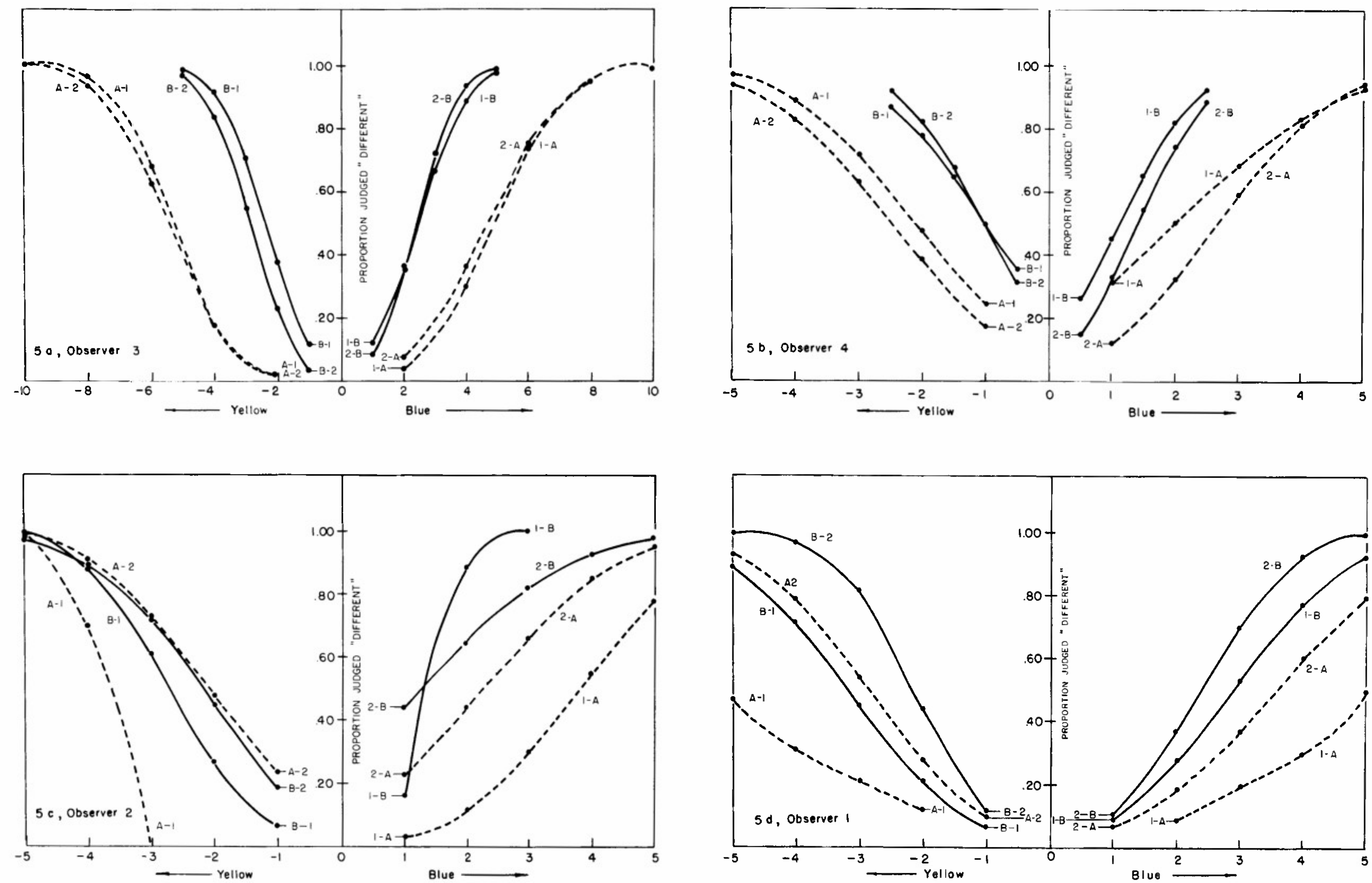

Figure 5

Normal Ogives Of Data From Method Of Constants, "Same Or Different" A And Method Of Constants, "With Direction" B, Repeated Once.

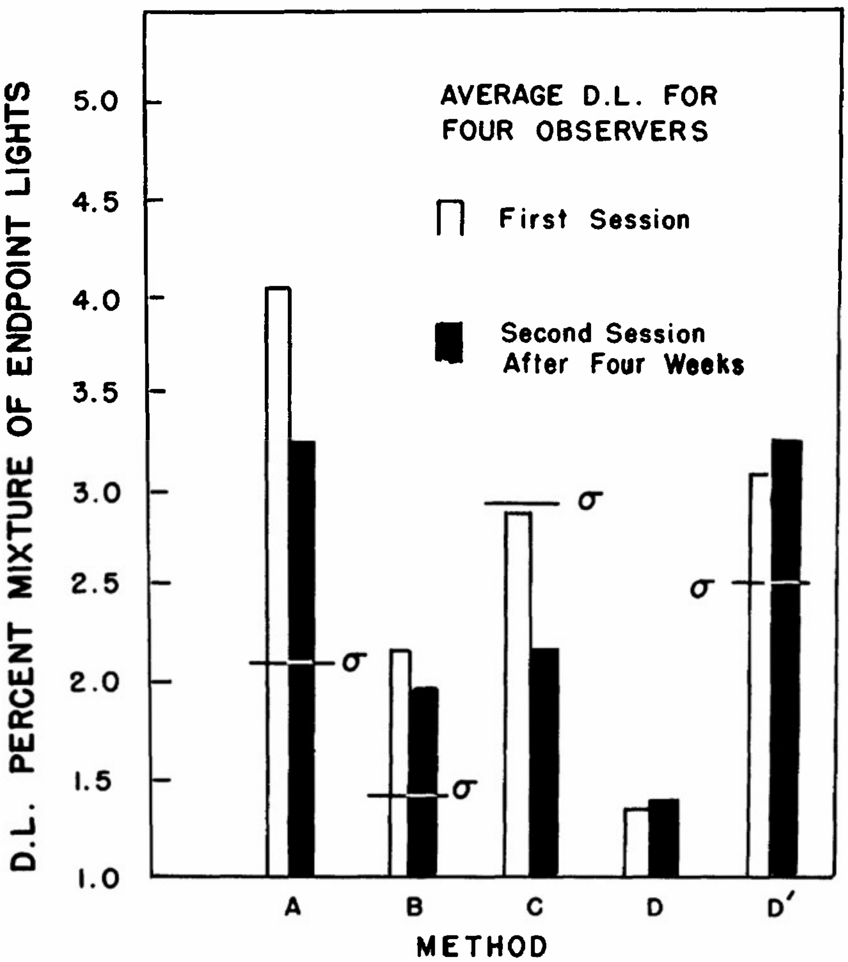

Figure 6

Mean Difference Limens for Four Psychophysical Methods.

(D represents the P.E. of Method of Adjustments; D′ represents the 1.5 standard deviations of Adjustments; σ 's represent the standard deviations of the distribution of judgments underlying each D.L.)

TABLE VII

Completed Latin square of standard deviations over four Psychophysical Methods, Background contrasts on four observers.

Os.		Surround to Target Contrast I 3:1	II 2:1	III 1:1	IV .05:1	Sum x 1 - 4
1	S.D. 1st Exp.	A 3.03	B 1.65	C 2.34	D 1.70	
	S.D. 2nd Exp.	1.80	1.18	1.51	2.03	15.24
	Total	4.83	2.83	3.85	3.73	
2	S.D. 1st Exp.	B .96	A 1.89	D 1.60	C 2.07	
	S.D. 2nd Exp.	1.82	1.83	1.69	1.96	13.82
	Total	2.78	3.72	3.29	4.03	
3	S.D. 1st Exp.	C 2.69	D 2.57	B 1.42	A 1.86	
	S.D. 2nd Exp.	4.18	2.51	1.37	1.89	18.49
	Total	6.87	5.08	2.79	3.75	
4	S.D. 1st Exp.	D 2.18	C 2.43	A 2.02	B 1.30	
	S.D. 2nd Exp.	2.13	2.61	1.71	1.11	15.49
	Total	4.31	5.04	3.73	2.41	
	Sum x I-IV	18.79	16.67	13.66	13.92	

Sums

Σx_A 16.03

Σx_B 10.81

Σx_C 19.79

Σx_D 16.41

Table VIII

Completed Analysis of Variance of Standard Deviations

Source	SS	DF	M^2(var.)	F	P
1 Between Observers	1.44	3	.48	$\frac{V_{obs.}}{V_{ind.Err.}} = \frac{.48}{.18} = 2.70$	.05= 3.24
2 Between background Contrasts	2.22	3	.74	$\frac{V_{backgr.}}{V_{ind.Err.}} = \frac{.74}{.18} = 4.11$	.05= 3.24
3 Between Methods	5.15	3	1.72	$\frac{V_{methods}}{V_{ind.Err.}} = \frac{1.72}{.18} = 9.56$	.01= 5.29
4 Within Subclass (independent Error)	2.84	16	.18		
5 Subclass Discrepance	.63	6	.10		
6 Total	12.28	31			

The following relations result:

1. The standard deviations obtained over all methods are significantly different at $P < .01$.

2. Variations in surround to target contrast produce significant differences in the magnitude of the S.D. at $P < .05$.

3. Observers do not differ significantly in the magnitude of the S.D.

4. The pattern of magnitudes of the S.D. show no significant interactions between the three main variables of Method, Contrast and Observers.

5. The breakdown of intra-subclass variance (Table IX) reveals no significant differences in the variability of the S.D. from one testing to the next between the Constants and Paired Comparisons Methods. Method of Adjustments shows significantly less variability of the S.D. from one session to the next than do each of the other psychophysical methods.

6. A breakdown of the contributions of each method to the differences in magnitude of the S.D. Table X, reveals no significant differences between the S.D.s of Constants "Same or Different" and of the other methods. Constants "with direction" shows significantly smaller S.D.s than do Paired Comparisons and Adjustments.

7. The order of magnitudes of the standard deviation over psychophysical methods are:

Paired Comparisons > Adjustments > Constants - "Same or Different" > Constants - "with direction". (Figure 6)

Since methods A and B were variations of the Method of Constants, yielded similar distributions and were computed identically, it is possible to directly compare their psychometric functions. These are plotted in Figure 5 for each observer. These are of the judgments "different".

Table IXa

Within-subclass variance of the standard deviation (as a measure) over 4 psychophysical methods.

Method

	A	B	C	D
Total (4 O's)	.81	.50	1.48	.06

Table IXb

Comparison of within-subclass variance of standard deviations between methods.

	A	B	C	D
A		1.615	1.831	13.117*
B			2.956	8.122*
C	F=			24.011**

$* < .05$ $** < .01$

Table X

F Tests between combined standard deviations* for each of four psychophysical methods.

	A	B	C	D
A		1.49	1.42	1.14
B		F=	2.11**	1.69**
C	$N_1=8$, $N_2=8$			1.25

**P< .01

*The unbiased estimate of the combined standard deviation was found as follows:

$$\hat{\sigma}^2 = \frac{N_1 S_1^2 + N_2 S_2^2}{N_1 + N_2 - 2}$$

see J. F. Kenney, (17, p. 126).

COMPARISON OF METHODS

Before evaluating the differences in magnitude, precision and reproducibility of the D.L. as determined the four different ways, the mathematical basis for comparison should be ascertained. If it were said that the four methods measured the same underlying function and did not differ in their sensitivity to it, then would the different ways employed to treat the data and compute the D.L. produce identical numerical results? Are there mathematical factors present in the methods which alone would be expected to produce different size D.L.s?

The two versions of Constants (Methods A and B) would be expected to yield comparable measures since they yielded similar distributions which were computed identically.

The psychometric function of Paired Comparisons would be expected to yield the same D.L. as that of Constants provided one assumption is made - that the function being measured does not change over the range of stimuli employed. Paired Comparisons differ from Constants only in that each stimulus in the series is used as a standard in turn rather than the use of a fixed standard midway in the series. Thus, if the function were the same over the stimuli employed, one of each pair of stimuli could be equated to the standard of constants and the resulting ogive would be expected, from the mathematical standpoint, to be identical with the Constants function. Its median would be equivalent to the Constants limen in one direction from the standard. Further, since stimulus pairs were presented randomly on either side of the 50-50 mixture, the D.L. of Paired Comparisons would be equivalent to the D.L. of Constants.

The Method of Adjustments does not yield a D.L., but a measure of dispersion. Figure 7 shows that the 1 P.E. dispersion measure includes a much smaller stimulus interval than do the medians above and below the standard of Method of Constants. In Figure 7, a unit normal curve is drawn to represent the distribution of matches in Method of Adjustments, with standard units on the abscissa and proportions 0 to 1.00 on the ordinate. On the same diagram, the relationship to the version of Method of Constants used in this experiment is shown. Starting at 0 standard units, the mean of the normal curve which is assumed to fall at the standard S, symmetrical normal ogives are drawn to represent the expected distribution of "different" or "more blue", "more yellow" estimates. Since this discussion assumes that the two methods measure the same underlying function and do not differ in their sensitivity to it, it is reasonable to assume that the ogive will approach 100% "different" judgments at the same stimulus difference at which the adjustments curve approaches 0 frequency of matches. We have chosen the plus and minus 3 standard unit points as cut-offs for the Adjustments distribution since both distributions sharply approach limits at 0 and 1.00 at these points; only .135% of the area of the normal curve is thus excluded. We have extended the normal ogives symmetrically between 0 (the standard) and $\pm$ 3 S.D. of the Adjustments curve. A comparison of the 1 P.E. and 1 D.L. shows that where 1 P.E. falls at .6745 S.D. of the Adjustments standard scale, the D.L. of Constants falls at 1.5 S.D. Thus, if we assume a normal curve for Adjustments, symmetrical normal ogives (Phi-Gamma hypotheses) for Constants and make the further assumption of 0 and 1.00 cutoffs at $\pm$ 3 S.D., 1.5 S.D. for Adjustments is mathematically equivalent to the D. L. of our version of Constants.

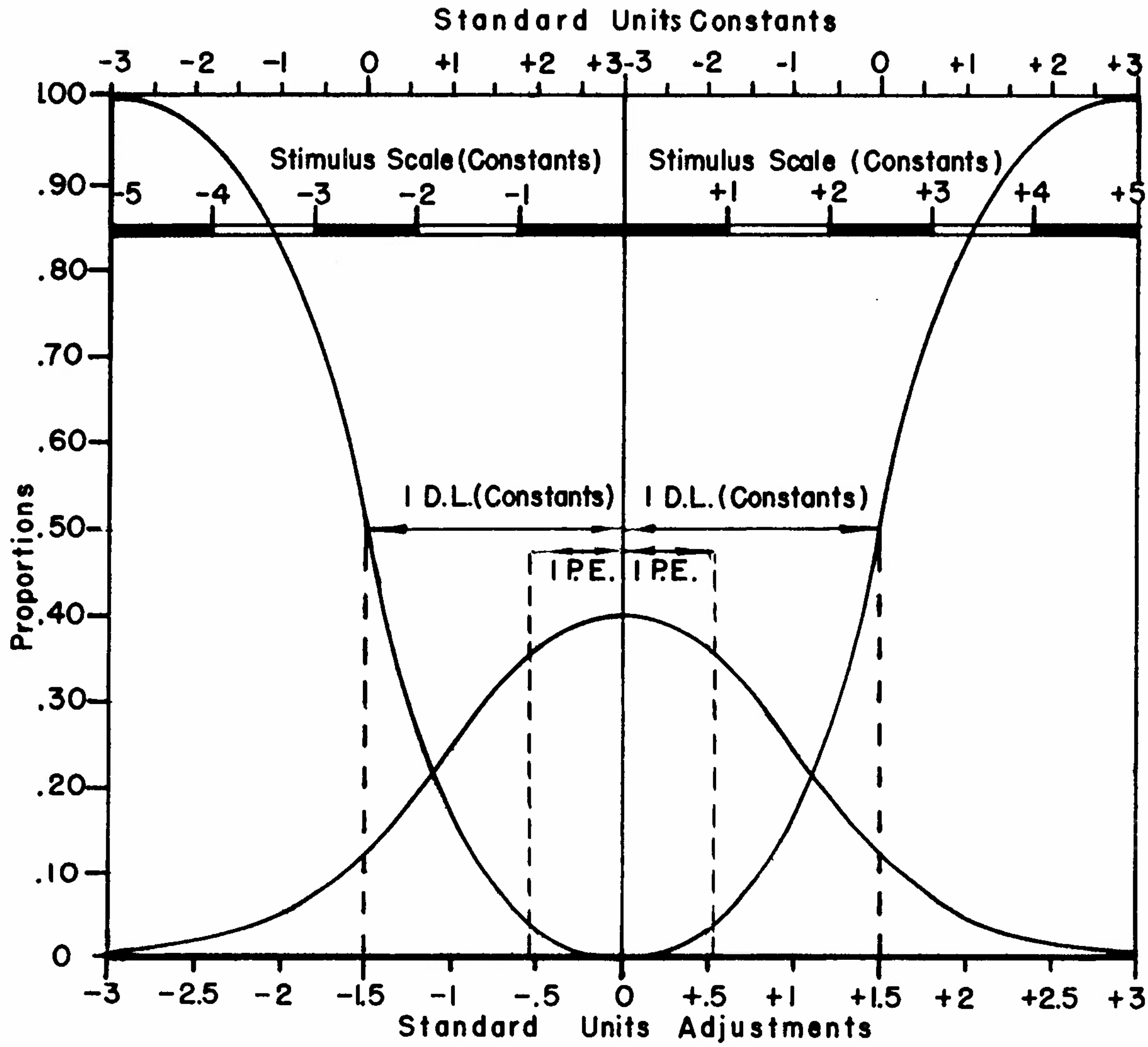

Figure 7

Theoretical Distributions for Method of Constants and Method of Adjustments

We find, making the above assumptions, that the 1.5 S.D. or adjustments is equivalent mathematically to the Paired Comparisons and Constants D.L.s which are equivalent each to the other. In the results we have quoted the P.E. (.6745 S.D.) that is generally used for Adjustments. If we revise this now to 1.5 S.D. and test the significance of the difference between combined measures by the four methods, we find that the 1.5 S.D. of Adjustments D' is not significantly different from the method A D.L. and is significantly larger than those of methods B and C, Table XI. If we substitute the 1.5 S.D. values for the D values in the analysis of variance, the sums of squares for methods differences reduces from 22.17 to 11.81 yielding an F ratio with the within-subclass variance of 6.91 which remains significant at $P < .01$. Thus, the remaining differences are significant and may be ascribed to other than mathematical differences between methods. Figure 6 shows the average D.L. for each of the methods with D' representing the 1.5 S.D. values.

Table XI

Significance of Differences Between Mean D.L.s# 1.5 S.D. (D')
Values for Method of Adjustments.

Methods

	A	B	C	D
A		4.50**	2.62*	1.25
B	t=		1.43	3.98**
C	N=8			2.87*

**P < .01 *P < .05

#No assumption of homogeneity of variance was made in evaluating the t ratio. (16, p. 83)

The above presents a way of relating Constants, Paired Comparisons and Adjustments in a mathematically valid fashion, using the Muller-Urban D.L. We may also define differential sensitivity in terms of a measure of dispersion as proposed by Fechner[13], Culler[15], and Boring[18]. A comparison of the standard deviations of the four methods has been presented in Tables VII, VIII, IX and X for this purpose.

EVALUATION OF RESULTS

Three criteria may be chosen upon which to evaluate the techniques for measuring color discrimination. These are (1) the magnitude of the measure - the smaller the value is, it may be argued, the closer it approaches the true discrimination of the eye. (2) The precision of the judgments from which the limen is calculated - considering a single set of data, the more widely dispersed the judgments or matches are the less precisely can the D.L. be specified. (3) Reproducibility from one session to the next; the more precisely a measure can be repeated on subsequent occasions, the more reliable it is. Information on each of the above three points has been obtained from the data.

The differences between Method of Adjustment and Method of Constants were analyzed into two parts. First, it was found that the difference in magnitude between the D.L. of Constants and the dispersion of Adjustments could be related by a constant of 1.5 S.D. applied to the Adjustments data. Provided that the usually accepted assumptions of the shape of the distribution yielded by the two methods were valid, it was possible to say that the differences in magnitude due to the mathematics alone were reconciled.

It was then possible to set up the hypothesis that the difference in instructions between methods would produce no difference between the magnitude of the D.L. and of the Adjustments dispersion measure. This hypothesis was tested by applying the 1.5 constant to the Adjustments S.D. and testing the significance of the difference between that measure and the two Constants D.L.s. It was found that the 1.5 S.D. adjustments measure was not significantly larger or smaller than the Constants "same or different" D.L. We, thus, could not reject the hypothesis that the instructions produced no differences in magnitude between the two methods. The Adjustments 1.5 S.D. was found significantly larger than the Constants "with direction" D.L. We, therefore, reject the hypothesis that the "with direction" instructions produced no difference in magnitude from Adjustments instructions and state that such instructions produce smaller measures.

The Adjustments measure, independently of magnitude shows significantly smaller variability, thus greater reproducibility from one session to the next than the other methods and a small standard deviation as compared with those of the other methods.

Paired Comparisons shows significantly smaller D.L.s than Constants "same or different" and the 1.5 S.D. Adjustments D.L. It shows significantly larger standard deviations than each of the other methods. Paired Comparisons also shows significantly greater variability from one testing to the next than Constants "with direction" and Adjustments. It shows higher but not significantly higher variability from one session to the next than Constants "same or different".

Constants "same or different" shows the largest D.L.s of the four methods. These are significantly larger than those of Constants "with direction", "Paired Comparisons" and the 1 P.E., Adjustments D.L., but not significantly different from the 1.5 S.D. Adjustments D.L. The magnitude of the Constants "same or different" standard deviations is not significantly different from those of the other three methods. This method shows significantly larger variability from one session to the next that Constants "with direction" and Adjustments, not significantly larger than Paired Comparisons.

Constants "with direction" shows significantly smaller D.L.s than Constants "same or different" and the 1.5 S.D. Adjustment but larger than the 1 P.E. Adjustment and not significantly smaller than the Paired Comparisons D.L. This method shows significantly smaller standard deviations than Paired Comparisons and Adjustments and smaller standard deviations but not significantly smaller than Constants "same or different". Constants "with direction" show variability from one session to the next which was larger than that of Adjustments but significantly smaller than the other methods.

CONCLUSIONS

1. Method of Constants "with direction" yields a smaller, more reproducible and more precise D.L. than does Constants "same or different". This difference may be ascribed to the difference in instructions.

2. Method of Adjustment may be analyzed into two sources of difference from the other methods in the magnitude of the discrimination measure. The first lies in the mathematical difference between the P.E. and D.L. which is reconciled by using 1.5 S.D. of the Adjustments data as the measure. The second source of difference lies in the procedure and instructions.

3. Adjustments (1.5 S.D.) yielded a measure not significantly different in magnitude from that of Constants "same or different" but significantly larger than Constants "with direction".

4. Paired Comparisons yielded small D.L.s not significantly different in magnitude from those of Constants "with direction", but of the lowest precision and reproducibility.

5. Variation of surround brightness produced no significant change in the magnitude of the D.L. over all methods. It does produce a significant variation in the standard deviation of the judgments.

6. Method of Adjustments yields significantly more reproducible measures of color discrimination than all other methods as well as high precision. Constants "with direction" shows considerably better results in these respects than do Constants "same or different" and Paired Comparisons.

ACKNOWLEDGMENTS

The authors wish to thank Miss Doris Newman, Head of the Statistics Facility of this Laboratory, for her review of the statistical procedures used here and for her collaboration in deriving the theoretical relationship between Method of Adjustments and Method of Constants. They are grateful to Dr. Forrest L. Dimmick, Head of the General Vision Facility, for his valuable suggestions on psychophysical procedures. They are especially grateful to Miss Ailene Morris, Miss Ruth Derby and Mr. Marshall Jones for the many hours spent in observations; to Mr. Jones for his assistance in computational work, Mr. John Verges for preparing the illustrations and Miss Natalie Nagle for her painstaking stenographic work.

BIBLIOGRAPHY

1. D.B. Judd, "The 1931 I.C.I. standard observer and coordinate system for colorimetry", J.Opt.Soc.Am. 23, 359 (1933).

2. A. Koenig and C. Dieterici, "Uber die Empfindlichkeit des normalen Auges fur Wellenlangenunterscheide des Lichtes", Wied.Ann.d.Physik und Chemie 22, 579, (1884)' Graefes Archiv. 30, (2), 158, (1884).

3. H. Laurens and W.F. Hamilton, "The sensibility of the eye to differences in wavelength", Am.J.Physiol. 65, 547, (1923).

4. W.D.Wright, "The sensitivity of the eye to small colour differences", Prac.Phys.Soc. 53, 93, (1941).

5. D.L. MacAdam, "Visual sensitivities to color differences in daylight", J.Opt.Soc.Am. 32, 247, (1942).

6. S.M. Newhall, D. Nickerson and D.B. Judd, "Final Report of the O.S.A. Subcommittee on the Spacing of the MunsellColors", J.Opt.Soc.Am. 33, 385, (1943).

7. E.B. Titchener, "Experimental Psychology, A Manual of Laboratory Practice", Vol. II, (The MacMillan Company, New York, 1915).

8. W.N. Kellogg, "An experimental comparison of psychophysical methods", Archiv. Psychol. 17, #126, (1929).

9. F. Irwin, "Psychophysical measurement methods", Psychol. Bull., 32, 140, (1935).

10. J.P. Guilford, "Psychometric Methods", (McGraw-Hill Book Company, New York, 1936).

11. G.E. Muller, "Zur Grundlegung der Psychophysik", (T. Griefen, Berlin, 1878).

12. G.S. Fullerton and J.M. Cattell, "On the perception of small differences", Pub. Univ. Penna., Phil. Series, No. 2, (1892).

13. G.T. Fechner, "Revision der Hauptpunkte der Psychophysik", (Breitkopf und Hartel, Leipzig, 1882).

14. F.L. Dimmick and M.R. Hubbard, "The spectral location of psychologically unique yellow, green and blue", Am.J. Psychol., 52, 242, (1939).

15. E. Culler, "Studies in psychometric theory", Psychol. Monog., 35, 56, (1926).

16. G.Y. Snedecor, "Statistical Methods", (Iowa State College Press, Ames, Iowa, 1948).

17. J.F. Kenny, "Mathematics of Statistics, Part 2", (D. Van Nostrand Co., New York, 1947).

18. E.G. Boring, "The Psychophysics of Color Tolerance", in the Inter-Society Color Council Symposium on Color Tolerance, Am. J. Psychol. 52, 384, (1939).