COLOR CODING OF CONTROLS:

RT Improvement Found When Monochromatic Labels

Were Replaced by Polychromatic Labels

by

Warren D. Cheney

A Thesis Submitted to the Faculty of the

DEPARTMENT OF PHILOSOPHY AND PSYCHOLOGY

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF ARTS

In the Graduate College

UNIVERSITY OF ARIZONA

1 9 6 0

STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in their judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Warren D. Cheney

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

Neil R. Bartlett — Jan 22, 1960

NEIL R. BARTLETT, PhD. — Date

Professor of Philosophy and Psychology

Head of the Department

TABLE OF CONTENTS

TABLE OF CONTENTS (cont'd)

TABLE OF CONTENTS (cont'd)

LIST OF TABLES

LIST OF FIGURES

I. PRELIMINARY DISCUSSION

A. THE PROBLEM

Will the speed and accuracy of controls selection be improved by the addition of color coding to labels, or by substitution of color coding for labels?

Ten years ago, when the first textbook in the field of human factors engineering was published (12), authors Chapanis, Garner and Morgan appeared quite certain that color would indeed improve the distinguishability of controls. "Painting different controls different colors is another obvious way of making controls distinguishable", they wrote.

Inherent in that statement, however, was the fact that it constituted an hypothesis, not a scientifically proven principle of psychophysics. Aware of the hypothetical nature of their postulate, they added this comment. "Despite the fact that color coding would be an obvious advantage, there has been relatively little research on the problem, and relatively little use has been made of color coding of controls . . . Only systematic research on this problem, however, can tell us how to use color coding of controls most effectively."

Also writing of the problem in 1949, Stellar (41) postulated "Controls should be made as distinctive as possible". He cites the early work of Weitz (45) and comments:

> Although these results clearly show that shape and color coding or shape alone help reduce confusion errors when the positions of controls are interchanged, it should be pointed out that further work on this problem is necessary....Further work must be done on the use of size coding and the use of shape and color coding on controls other than levers... where the overall efficiency of the operator can be measured with and without the benefit of coding.

Ten years later it is still true that relatively little use has been made of color coding controls. The control panels of the electronics field, the motor vehicle field, the aircraft field and now the missile field remain generally monochromatic as to controls coding.

While color has been introduced for aesthetic effect into automobile dashboards, with matching steering wheels and seat covers, color has not been used to particularly improve the discriminability of controls.

Ten years later it also is true that relatively little research on the color coding problem has been accomplished. While important amounts of human factors research time has been invested in anthropometrics, in the legibility of dials and radar scopes, in symbolic and pictorial displays, in controls knobs, switches, handwheels, cranks, levers, footpedals, covers and cases, as well as in factors of ambient environment, hazards and safety, there are to date perhaps less than ten studies reported in the literature which deal directly with the basic problems of color coding controls and indicators. Subsequent sections of this paper will offer review and comment on the few studies found to have been concerned with those problems.

B. INDICES THAT COLOR COULD BE PARAMOUNT IN CONTROLS AND INDICATOR CODING

Studies of the psychophysics of vision may be said by now to have rather firmly established several fundamental human capabilities as to color perception. A review of the literature reveals considerable agreement on certain basic data pertinent to color coding problems.

For example, persons with color-normal vision can distinguish a very large number of different colors, recognizing quite subtle variations on three levels, namely hue, brightness and saturation. Judd (28) reports "about 10 million surface colors can be distinguished by the normal human eye under optimum observing conditions . . . perhaps half a million considered to be commercially different."

Boring, Langfeld and Weld (9) indicate some 300,000 different colors can be distinguished by the human eye. Maerz and Paul (33) list 7000 colors in their Dictionary of Color. Kelly and Judd (29), using the ISCC-NBS method of designating colors, have named 7500 colors in their Bureau of Standards Circular No. 553.

While recognizing the marked spread of these statistics, there is no doubt that color perception for the normal human eye has great ranges of discriminability. Furthermore, studies of memory for color, (25) and (11), have demonstrated that subjects with normal color vision can be trained to distinguish and identify colors from memory with high degrees of accuracy.

At least two studies, (19) and (43), offer evidence that color used as a cue for discrimination of controls produces significantly faster RT than cues of form, size, or brightness.

Research reported in this paper presents evidence that controls using polychromatic coding produce significantly faster selection RTs than controls using monochromatic labels only.

In summary, it seems clear that the capability of the normal human eye to recognize and distinguish differences in color, including the variations of hue, brightness and saturation, is a paramount capability which can be of prime importance to human factors engineering. It is also clear that the usefulness of this capability has yet to be fully recognized and thoroughly studied. In fact, the area of color coding is perhaps one of the poorest studied in human factors research.

It is likely that a good deal of the resistance to color coding visual displays has stemmed from the difficulties inherent in controlling front illumination of aircraft pilots' cockpits. At night, with his requirements to take-off, fly and land as efficiently and safely as during the day, the pilot needs control panel illumination of the lowest level consistent with legibility.

However, there is little reason to condemn all other controls and indicator panels to monochromatic designs merely because of the pilot's special problem. Indeed, aboard aircraft having multimembered crews, there is no need to treat all control stations as being similar in limitations to the pilot's cockpit. The navigator, the engineer, the radio operator, as well as the bombardier, perform their respective functions by use of instruments only, have no operational need for daylight or windows. Their stations can be artificially lighted and held under steady-state controlled illumination without impairing in any way their operational duties.

Once the designer's attention moves from aircraft to the hundreds of other fields wherein the visual display is a standard item, it should be obvious that the special limitations of the pilot's cockpit are not present. Thus, in any situation which permits panel illumination to be controlled, and controlled lighting is possible in the great majority of situations, color coding of controls and indicators can be undertaken without fear of color distortion due to lighting changes.

At this point in the development of human factors engineering, it is true that the body of validated data on color coding is small, incomplete, and beset with ancillary problems of vision and perception also in need of further research.

But there are now on record probably enough independently conducted studies of controls and indicator color coding to reveal more than a trend or inclination favoring the use of color as a coding method. The data established by the major studies reviewed in this paper are sufficiently firm to permit the following prognositication to be made:

Color coding may be the superior method for coding controls and indicators, both as to speed and accuracy of discriminability, for the majority of instrumentation in the majority of visual display applications.

C. SOME STUDIES RELATED TO THE PROBLEM

1. USAF Published Data

a. The Aero Medical Laboratory Reports

A human factors engineer reviewing the literature to evaluate the effectiveness of color as a coding device for visual displays will discover contradictory reports.

If he turns to the file of research published by Wright Air Development Center Aero Medical Laboratory he will note, for example, that WADC TR 55-472 (36) and WADC TR 56-226 (5) both report color coding was found inferior to other types of coding. Again in WADC TR 53-221 (26) and WADC TR 55-471 (15) he will find sizable lists of reasons why color coding is disadvantageous.

When he reviews publications of the Air Force Ballistic Missile Division, such as AFBM Exhibit 57-8A (1) and the weighty tomes of the Air Research and Development Command, ARDCM 80-5 (3) and 80-6 (4), he may not find much greater enlightenment. Missiles engineers concerned with human factors are advised to use AFBM Exhibit 57-8A virtually as a "Bible". Thus, the color coding of controls currently permitted in the huge USAF missile program is generally restricted to the requirements of a few sparse paragraphs laying down color coding parameters. An examination of those paragraphs, therefore, becomes particularly pertinent to the problem under investigation. For instance, AFBM Exhibit 57-8A, paragraph 3.1.2, Control Coding, reads:

> Common coding methods will be used when available. The applicability (advantages and disadvantages) of each coding method will be determined in accordance with Table III, pp. 32-33 WADC TR 56-172 (18).
>
> 3.1.2.2 The five most common methods of coding are: Location, Shape, Size, Labeling, Color.
>
> 3.1.2.2.5 Color Coding. Presently, no standard color code for missile controls exists. If color coding is selected as a coding scheme for controls, the integrating authority will insure uniform application of the code throughout the system. In general only four colors (in addition to the customary black or gray control color) will be used for control coding: red, green, amber (yellow), and white. Additional colors should be used only if necessary and should be selected from those listed in WADC TR 54-160 (6), p. 96.

Referring to WADC TR 56-172 (18), as required above, one will find:

> 2.3.6 Color Coding. Color coding is most effective when a specific meaning can be attached to the color (e.g. red for danger). The color of the control depends largely upon the color of the illuminant; controls will reflect their own colors only if illuminated by a white light. As intensity of illumination is reduced, the color of the control changes and is gradually lost. Hence, for color coding, at least a moderate amount of white light must be used. In general, color should not be used as the primary method for coding controls; it is effective when combined with other methods.
>
> In general, only five colors should be used: red, orange, yellow, green, blue.
>
> Even under ideal conditions, an operator has difficulty in using effectively more than 10 or 12 colors. He can recognize many more but is limited primarily by his ability to attach a name to each.*

The asterisk refers the reader of WADC TR 56-172 to the 1949 volume of Chapanis, Garner, and Morgan (12), as well as to WADC TR

53-221 (26) which also refers back to this same volume for authority for that statement.

An examination of the original text reveals, however, that WADC 56-172 has misquoted Chapanis, Garner, and Morgan who actually wrote:

> It is conceivable, for example, that four or five colors would be the optimal number of colors for color coding and that such a large number as 10 or 12 colors would be too many for most operators to keep in mind and use effectively. Only systematic research on this problem, however, can tell us how to use color coding of controls most effectively.

By omitting the phrase, "It is conceivable", the WADC document gives the impression that Chapanis, Garner and Morgan had conducted research to prove their color coding postulates. Rather, their position in 1949 was strictly theoretical, as the phraseology of their writing makes clear.

The effect of that omission, however, is to give the impression that four or five colors are indeed the "optimal" limit for color coding, backed by the authority of the most widely known testbook in the field. Thus we are presently getting in our new missile control panels, for instance, color coding that overlaps to an absurd extent. When the 65-2 ATLAS missile launching pad mobile roof is drawn to the closed position, an indicator light illuminates "green". When the mobile roof is moved to the fully open position, an indicator light is illuminated the same "green".

Of course, this may result in some confusion on the part of the launch control officer who cannot see the launcher from his blockhouse console. But the color coding requirements of Air Force Ballistic Missile Exhibit 57-8A have not been violated. Unfortunately, to correct this confusion-producing coding is not at all a simple matter inasmuch as there are no alternate "go" colors permitted.

If the authors of WADC TR 56-172 had wished to quote experimental evidence of the number of absolutely identifiable spectral hues, the research of Halsey and Chapanis conducted in 1951 (23) could have been cited. That study is reviewed later in this paper, as is a 1958 research of the same problem conducted by Conover and Kraft (15). The reports of both investigating teams indicate, however, that further study is required before the practical number of absolutely identifiable spectral hues can be established. It would be advisable, therefore, to word government dictates to designers on the subject of color coding in terms more consistent with experimental findings and avoid absolute statements which in effect prevent advances in the state of the art.

b. The ARDC Manuals

Our earnest searcher for human factors engineering advice on color coding will no doubt be referred, sooner or later, to another group of USAF publications considered highly authorative, the ARDC Manuals for designers. Served on the missiles industry as the top governing documents of ARDC and of its Ballistic Missile Division, AFBM, they are: HIAD, ARDC Manual 80-1, Vol. II. Handbook of

Instructions for Aircraft Designers (Guided Missiles) (2); HIGED, ARDC Manual 80-5, Handbook of Instructions for Group Equipment Designers (3); and HIAGSED, ARDC Manual 80-6, Handbook of Instructions for Aircraft Ground Equipment Designers (4).

Of the three, HIGED contains the most extensive and pertinent data regarding use of color coding for visual displays. Yet here again, the human factors engineer may be less than satisfied. For although HIGED carries a 37 page section headed Human Engineering, there are but three brief paragraphs dealing with color coding of controls:

> 7.3.8 Controls Identification.
>
> 7.3.8.3 Identification by Color. Identification of controls may be established through the assignment of different colors to each control or group of controls having different functions. However, colors seen by reflected light may lose their identity at low levels of illumination or under colored lighting conditions (such as photographic dark-room safe-lighting) and should not be employed for coding when such conditions are anticipated.
>
> 7.3.8.3.1 Distinguishable Colors. Approximately 11 or 12 colors are clearly distinguishable in the visible color spec trum, but four or five is the maximum number recommended for use at one time when immediate discrimination is desired. White, yellow, red, blue and green are colors most readily distinguished from each other. The difference between black and brown, in certain plastics such as phenolics, is negligible for purposes of identification.
>
> 7.3.8.3.2 Color Coding Method. No one method of color coding is recommended since each situation requires an application of color especially suited to it.

These paragraphs on color coding reveal the presently accepted philosophy of the USAF regarding coding by color, and amount to "law" as far as controls and indicator coding is concerned, both for the huge missiles industry and for the aircraft industry.

The influence of the previously discussed WADC Technical Reports shows up in the "Distinguishable Colors" paragraph, complete with postulate distorted into fact. But HIGED's color coding paragraphs show other signs of weakness as well, particularly in the matter of how colors can be varied to accomplish distinguishability and in the matter of color specification. Indeed, it must be obvious that neither color nor vision experts could have been consulted in the preparation of this "Controls Identification" section.

To anyone properly backgrounded in the science of color, it is obvious that many combinations of colors other than the five chosen by HIGED and its WADC antecedent reports can be so spaced in hue, brightness, and saturation as to be distinguished from each other as easily as "white, yellow, red, blue and green". Again it is obvious that perfectly distinguishable differences can be given the two colors, black and brown, whether phenolic plastic or any of several other materials is the medium.

To specify colors merely by naming generic hues is a practice which ought to be replaced, of course, with proper tristimulus specification. Whether the 1929 Munsell method or the 1931 CIE method is employed, authors of color coding sections could specify with scientific awareness which ranges of variation in hue, brightness and saturation are acceptable for the "white", "yellow", "red", "blue", and "green".

Also regrettable is the tendency of publications such as HIGED to give the air of finality to their philosophies of color coding.

This kind of criticism applies, of course, to many areas of engineering documentation other than color coding. One explanation, not a defense, of the practice lies in the fact that relatively little basic research has been conducted on color coding per se.

Authors of controlling documents really do not possess, as yet, enough experimentally proven data to establish a valid, comprehensive philosophy of color coding.

c. Two Laboratory of Aviation Psychology Investigations

Two significant studies of color coding not referenced in the previously discussed USAF are, however, products of USAF commissioned research. Contracted for by WADC's Aero Medical Laboratory, the two studies were both carried out by the Laboratory of Aviation Psychology at Ohio State University and published as WADC TR 55-375 (35) and WADC TR 55-471 (15).

The earlier report by Muller, Sidorsky, Slivinske, Alluisi & Fitts, (35), reports the results of a series of eleven studies of the feasibility of several different types of symbols for the coding of information on cathode ray tubes and similar displays for use in future air traffic control and related systems.

Experiment 9 was designed to measure the efficiency of check-reading as a function of the type of code and the number of targets to be checked. The contrasted codes were (1) inclination, consisting of 1/4-in. diameter outline circles each containing a single radius line at one of the 12 major clock positions, and (2) color, consisting of 12 1/8-in. diameter color circles obtained by transilluminated Wratten filters.

Each target was one of 12 symbols. The task was to report as conflicts any two or more targets that were similarly coded.

Procedures were such that information processing time could be measured by a Standard Electric Timer. "Results demonstrated that the color code was markedly superior to the inclination code." The Muller et al research thus constitutes another piece of evidence supporting the results obtained by Eriksen (19) wherein color was found superior to form in comparable tasks of visual display symbol selection.

The more recent Laboratory of Aviation Psychology study by Conover & Kraft, (15), October 1958, had a three-fold purpose: (a) to determine the maximum number of absolutely identifiable stimulus categories in the dimension of hue, (b) to construct a scale of equal discriminability for hue, and (c) to validate the scale on an independent population sample.

Ten color-normal subjects were tested with a set of 25 Munsell maximum saturation color patches. The colors were viewed through a 3° aperture of a neutral mask and illuminated by 21.4 ft.c. of 6800° K. light.

Test results showed that the number of absolutely discriminable hues ranged from 5 to 16, depending upon the individual tested. Under idealized viewing conditions, half of the subjects tested could discriminate "without appreciable error" nine maximally saturated colors.

In the interest of rendering the proposed color code usable by "somewhat more than half the population", the experimenters proposed reduction of the number of colors to eight. The set of 8 surface colors nominated for their scale of equal discriminability for hue is as follows: (notations are Munsell Company color standards) 1R 999, 9R 892, 1Y 946, 7GY 960, 9G 1099, 5B 1087, 1P 1135, 3RP 1003.

Recognizing that ideal illumination might not prevail under some viewing conditions, the experimenters offered three alternate sets of colors for coding use comprised of 7,6 and 5 colors respectively. The recommendation for colored phosphors, used to color code CRT displays, prescribes holding the number of colors to 4. Trans-illuminated, filter-controlled colored indicators should not number more than 6 colors to insure absolutely identifiable code symbols. Although the Conover & Kraft study was not concerned with testing color coding versus other coding for visual displays, it is significant as evidence of a continuing concern by the USAF with the problems of color coding. The senior investigator, D. W. Conover, had made this study the subject of his doctoral thesis, and because of the high value placed upon the work by the Aero Medical Laboratory, a second, more detailed report of the research is to be published. Listed as WADC Technical Note 58-262, the report is to comprise the complete content of the PhD thesis including the techniques and computations involved in the development of a scale for equal discrimination for hue, and in the estimation of the amount of information that can be transmitted per symbol by the use of surface colors.

2. U.S. Navy and Other Research Related to the Problem

The U. S. Navy through its Office of Naval Research also has been concerned with color as a discrimination cue in visual displays. One of the most pertinent studies commissioned by ONR was that of Eriksen (19) who conceived a means of testing the relative merits of form, hue, size and brightness as visual display cues. That study is reviewed in detail later in this section. But as preface to consideration of Eriksen's work, the results of an earlier experiment which produced evidence contrary to his deserves discussion. Indeed, it is possible that ONR could have been prompted to start Eriksen on his experiment partly because of that challenging early study entitled "The relative difficulty of the number, form and color concepts of a Weigl-type problem," by Grant, Jones, and Tallantis (20).

a. Grant, Jones and Tallantis

Using the Wisconsin Card Sorting Test, the experimenters undertook "to determine the relative difficulty of the three sorting categories, color, number, and form." Their procedure and discussion was as follows.

Procedure: Briefly stated, there were 64 response cards each carrying one to four figures of a single color. There were four colors, red, green, yellow and blue; four figures, stars, crosses, triangles and circles. Thus, each card could be sorted according to form, color, number.

Ss had 4 double-compartment sorting boxes, with stimulus cards placed in the respective upper halves of the compartments: one red triangle, two green stars, three yellow crosses and four blue circles. Ss were told to sort the 64 cards into four boxes according to Ss' concept of what the stimulus cards respective categories were.

Only instructions were 'right' or 'wrong' after each sort.

Initially, the category of color was the correct key, but after S. had sorted 10 correctly by color, E. shifted category to form (but without warning S.), and after 10 correct form sorts, category was changed to number.

Experimental design: categories were successively color, form, number, color, form, number. Results were tabulated for each category: total errors, perseverative errors, non-perseverative errors, and total correct answers.

Results of this experiment showed that Ss were able to sort selectively for number most easily, for form next most easily, while selecting for color was most difficult. However, the only scores on which the differences between sorting categories were statistically different were the total correct responses and the perseverative errors on the succeeding category. In other words, it required less reinforcement to acquire the number-sorting response and, once learned, the response of sorting number tended to perseverate more than from form or color sorting responses.

Configurational aspects are involved. The number <u>one</u> consisted of a single figure, centered; <u>two</u> consisted of two figures, one in upper left-hand corner, the other in lower right; <u>three</u> consisted of three figures forming an inverted equilateral triangle; and <u>four</u> consisted of four figures forming a square.

Thus the Ss could sort according to configuration rather than number, and verbal reports indicate that some of them did.

If the results reported by Grant et al were considered valid, evidence of the relative difficulty of sorting would seem to establish color as the most difficult. Furthermore, this conclusion could be taken as proof that color would be the least efficient method of coding a visual display as against coding by form or by number.

Careful evaluation of the apparatus and procedure will reveal, however, that the Grant methodology was not designed to cleanly isolate the color variable from those of form and number. The same card which was sorted for number was also sorted for form. Furthermore, there was no change of cards when the test for color sorting was made. Color was present at all times on each card, hence it cannot be assumed that color had absolutely no effect upon Ss sorting for number or form. Again, both form and number cues were present on cards sorted for color, hence it cannot be demonstrated that form and number had no effect on color sorting.

b. <u>Eriksen</u>

When Eriksen began his study some three years later (19), his announced purpose was to determine the speed with which objects could be located on a visual display under the following conditions:

a) when the various classes of objects on the display differed from one another on only one of the four visual dimensions, form, hue, size, and brightness.

b) when classes differed on two, and, on three of these dimensions.

Method: Display was a 3-ft. square, flat white in color, ruled into 81 4-inch squares with black lines; only the 49 central group of squares were used. Within each square an object-card was hung; each square mounted a $\frac{1}{2}$-in. electrode contact. Illumination was an overhead Macbeth Daylight lamp, mounted so as to provide even, shadow-free light.

Seven classes of objects were used: circles, hexagons, diamonds, triangles, crosses, stars, squares. These forms were cut from paper and made to look phenomenally the s ame size, all circumscribable by a 1-inch circle. Colors were Munsell standards: R, YR, Y, GY, G, BG, B, with brightness varied from N 8/ to n 1/ respectively, and with sizes varied from 4/8 in. to 10/8 in. respectively (note inversion of size and brightness). The ten patterns used were pre-tested and found to be of equal difficulty.

Procedure: A sample card was placed at left of the display. S. was told to look at sample then find six object-cards that matched. When S. located the matching card in the display he touched the contact under card with electrode probe in hand. This provided accurate timing of the S's discrimination operation. Speed was stressed, but Ss were cautioned against making errors.

Subjects: N=60, ages 19-35, found to have normal color vision by test.

Results: For single dimensions of discrimination, hue (color) is significantly faster than form, and both hue and form are significantly faster than brightness or size. Location is fastest when classes differ in hue-form, hue, and, hue-form-brightness.

The multidimensional location times obtained by compounding dimensions fail to show any consistent advantages over single dimension location times. Hue-form is the only case where a compound gives an even slightly faster location time than the best of the single dimensions of which it is composed.

Compounding differences in hue and form with differences in brightness or size results in slower location than is obtained for differences in either hue or form alone. Also, the compounding of brightness and size with hue and form does not yield uniform results. Brightness or size, when combined with hue yields much quicker location than do either brightness or size singly.

There is a marked degree of consistency in the location times for the compounds. The compounds of two, as well as those of three dimensions tend to rank in the same order, with respect to location time, as do the single dimensions of which they are composed.

It may be noted that in this study Ss could have located target objects for compounds on the basis of any one of the dimensions in the sample. For example, in the form-brightness-size compound, target objects could have been correctly located on the basis of form differences only, irrespective of brightness or size. The Ss were not aware of this, however. Their instructions were to locate 'six objects that matched the sample.'

This argument seems plausible, but there are two important points against it in the data. If the argument were valid, we would expect compounds of three dimensions to give slower location times than compounds of two. Also, the compound brightness-size would be more difficult to locate than either brightness or size singly. Both of these deductions are contradicted by the data.

The finding that color is significantly faster than form, brightness or size for single dimensions of discrimination reversed the color results reported by Grant and associates (20). It is also of pertinent interest to note that location times obtained by compounding dimensions (i.e. cues) failed to show any consistent advantages over single dimension location times. However, Eriksen adds that the compound of hue-form gave a slightly faster location time than either hue or form singly.

Finally, it is significant that Eriksen's findings in the use of compound cues did not result in significantly slower location times.

c. Other Related Studies

Another reversal of the Grant and associates findings was reported in 1955 by Wohwill (43) who used the Wisconsin Card Sorting Test and essentially the same procedures followed by Grant. Employing the same four kinds of stimulus cards, and the same double-compartment sorting boxes, Wohwill's only deviation was to narrow the number of cards to be sorted from 64 to 48. His results demonstrated that form was less readily abstracted than color or number. He also reported that speed of discrimination was fastest when color was the sorting cue.

Two Massachusetts Institute of Technology studies of color coding in a visual search task conducted at the Lincoln Laboratory, (21), and (22), also reported superiority of color coding. The earlier study, 1953, conducted by Green, McGill and Jenkins (22), employed a display of transilluminated symbols comprised of 3-digit numbers randomly arranged. Half the numbers were blue, half were yellow. Os were given a target number to locate and the time required for finding the target was recorded. Results showed that Os who were told the color of the target number found the target in about half the time required by Os who were not told the color.

In 1956, Green and Anderson (21) conducted a variation of the 1953 MIT experiment. Using a display of transilluminated symbols comprised of 2-digit numbers in the range 10-69, arranged in random order in a matrix of 10 rows and 6 columns, the symbols were coded either red or green. Results showed that when Os knew the target number color, search time was proportional to the number of symbols having the same color as the target. When the target color was not known, search time was proportional to total number of symbols on the matrix.

Another investigation can be noted here although its primary concern was with reaction time for a series of correct responses rather than with color coding per se. Reed (39) set up a submarine "Christmas Tree" indicator panel, comprised of 24 red and green lights randomly arranged, and recorded the respective speed of identification for two groups of Os, one group color-normal, the other group partially color-blind. He found that the color-blind Os were slower to identify target indicators than were the color-normal. He also reported, however, that when speed was rendered irrelevant, the color-blind were able to identify indicators as well as the color-normal.

While these results imply color coding is advantageous, it would be inaccurate to offer Reed's work as clearly supportive of the hypotheses tested by the present investigation.

More recently, in a dial reading experiment by Bartz (8), attention value as a function of illuminant color change was tested. Two panels were employed, 16 dials mounted on each panel in identical layouts. One panel was illumined by a steady red light; the other panel, also illumined red at the outset, permitted individual dials to illumined green when pointers deviated. Results showed that deviations were responded to faster and more accurately for dials given color change than for dials permitted no color change.

D. INDICATIONS OF NEED FOR THE PRESENT STUDY

As the preceding sections have indicated, the amount of basic research accomplished to date dealing with color coding has been rather spasmodic, scattered, and lacking in planned continuity. Without in any way disparaging the several studies carried out on color coding during the past 15 years, the human factors engineer in need of a comprehensive body of validated data for the subject cannot but admit many lacks and gaps exist, deficiencies which stand out in glaring relief when the entire literature of color coding is gathered into a single collection and reviewed as a whole. After due analysis of that body of data, it seems evident that one of the most fundamental questions in the field of color coding of controls and indicators still has need of experimental validation, namely, is color a positive aid when used either as an addition to labels or used alone for controls and indicator identification?

Intimately related with that question is another, also in need of experimental validation: How many variations of hue, brightness and saturation can be given to a controls or indicator coding system without presenting operators indiscriminable color differences?

Specific research areas relating to these questions are discussed at the conclusion of this paper together with some recommendations as to studies which need to be undertaken in order to provide experimental validation for answers pioneer research has so lightly established. Also noted are several other areas of study which are as yet almost untouched in the field of color coding.

At this juncture there is, however, a very simple query to answer: why was the present study undertaken?

Examination of the literature revealed many basic questions in need of study and validation; some of those questions have been stated previously. Viewing the spread of unproven postulates, this investigator finally selected the present study as one which would provide, perhaps as much as any other, some of the basic experimental data upon which several corollary postulates essential to the philosophy of color coding depend.

E. HYPOTHESES TESTED

Two hypotheses were tested. Given a control panel problem, it is postulated that:

1. Speed and accuracy of controls selection will be significantly improved if polychromatic rather than monochromatic color coding of labels is employed.
2. Speed and accuracy of controls selection will be significantly improved if polychromatic color coding only, without labels, is employed to identify controls.

II. DESCRIPTION OF APPARATUS AND PROCEDURES

A. APPARATUS AND TESTING MATERIALS

1. The Control Panel and Electronic Chronoscope

As shown in Figure 2, a group of five pushbutton microswitches were mounted in a single row 37.0 inches from the floor. Built by the Micro Switch Division of Minneapolis-Honneywell, the pushbutton switches were spaced 1.75 inches apart, employed black plastic caps 0.50 inches in diameter, required an operating force of 5 oz. max., permitted a total travel of 0.25 inches, had a minimum breaking distance of 0.010 inches.

Custom built label holders designed to carry standard Munsell 1.5x 3.0 in. color patches were fabricated to fit into metal pockets centered under each pushbutton. Adequate clearance permitted the label holder to be readily interchanged during administration of the tests, a technique which prevented subjects from using position as a selection cue.

As seen in Figure 1, the apparatus provided a screen which, in pre-trial position, masked the pushbutton controls and allowed the experimenter to change the label positions without their being viewed by the subject. The screen was of 0.25 in. Masonite, 24.0 x 20.0 inches, mounted in channels 8.0 in. in front of the control panel. Painted the same neutral gray as the control panel, the screen was held in the pre-trial position by pull-pins mounted on the channels.

Squeezing a hand-held intercom switch, visible in Figure 1, the subject was able to close the circuit activating a solenoid, mounted on the back of the control panel, which would withdraw the pull-pins and drop the screen to the post-trial position as shown in Figure 2.

As the screen dropped it activated a microswitch that started the electronic chronoscope counting selection time, cut off by depression of the target pushbutton. The chronoscope start switch was so placed that the screen would not activate the counter until the upper edge of the screen had fallen three inches below the bottom level of the pushbutton labels. This design feature meant that start of RT was not counted until the controls and their labels were fully in view.

Line voltage variation and time lag of the electronic chronoscope and associated circuitry were measured by depressing a control pushbutton then releasing the screen. During repeated tests of this kind on different days, the shortest system lag time recorded was 5.6 msec, longest was 9.1 msec. Maximum line voltage variation thus amounted to the difference, 3.5 msec, while the average system lag time amounted to the mean, 7.3 msec.

Had the prime objective been the recording of absolute selection RTs, with intent to precisely establish the speed of color perception and hand travel, the error factors represented by the line voltage

Figure 1. Subject at control panel, screen up.

Figure 2. Screen down, subject striking target button, experimenter watching hand travel. (Part of electronic chronoscope is seen in background.)

variation and the system lag time could have assumed importance. But here the focus of investigation was the relative amount of improvement in selection RT which color coding of controls would provide. Each subject was competing with his own scores, not with scores achieved by other subjects. Consequently, since line voltage variations reached a maximum of only 3.5 msec. this error factor was considered negligible. System time lag, being relatively constant for all subjects, was not fed into the raw data of Figures 7-14, but was subtracted from recorded RTs of hand-travel time in the compilation of Table 12.

2. Command Cards

A command card window, 2.0 x 4.0 in., was cut in the control panel to the subject's left, on line with the top edge of the screen. The window was so positioned as to provide unobstructed visibility to the seated subject.

Five command words in Ryko lettering were inked one each on 5 command cards respectively. Letters were 0.015 in. high, the same in size and style as the corresponding command words carried by the 5 pushbutton labels. Command words were: BOMB, FIRE, HOLD, DE-ICE, CHOP, vocabulary typically employed in defense aviation squadrons.

In selecting the command words, some danger conditions were included, lending dynamics to the commands employed. The icing threat was injected by the use of DE-ICE. Since several kinds of danger to an aircraft in flight call for cutting off motors, the word CHOP was introduced, short for aircrew slang "chop the gas". HOLD was chosen in view of the heavy use to which the word is put during flight, viz. hold bearing, hold altitude, hold airspeed, etc. FIRE and BOMB were "naturals" inasmuch as modern military planes typically carry controls labeled with these command words.

3. Cue Tabs

Three configurations of cue tabs were employed, 5 tabs per set:

(1) Monochromatic with labels (Mono-L), Figure 3.

(2) Polychromatic with labels (Poly-L), Figure 4.

(3) Polychromatic without labels (Poly no-L), Figure 5.

Labels read: BOMB, FIRE, HOLD, DE-ICE, CHOP. The words were inked on white paper the same size and with the same Ryko lettering as were the command card words. The white paper labels with black lettering were cemented to the top of the 1.5 x 3.0 in. cue tabs so that no tab background showed above the labels. This brought the label upper edge to approximately 0.4 in. below the pushbutton rim when the cue tab was in position ready for testing.

4. Determination of Color Cues

Tabs carrying labels for the monochromatic cues were of a 6/0 neutral gray, approximately the same in hue, value and chroma as the gray face of the control panel. Illumination was provided by distortion-free daylight, as detailed on pages 37-38, insuring normal color aspect of cue tabs.

Tabs carrying labels for the 5 polychromatic cues were colored papers which can be specified in Munsell notation as 10.0 RP 4/4, 7.5 GY 8/6, 7.5 R 6/10, 2.5 PB 3/6, and 10.0 R 3/6.

Polychromatic cue tabs without labels were comprised of a set of 5 of the same colored papers.

Determination of color spacing provided by the polychromatic cue tabs was made after running a pilot test of more than 500 trials using as tabs a conventional set of Munsell color patches matched for chroma and value, each hue being in the 5/6 bracket. The hues were 2.5 R, 5P, 5 BG, GY, and 7.5YR. When transposed and plotted in terms of CIE chromaticity coordinates, it was found that these five hues were approximately equally spaced.

During testing, however, subjects stated that the 5 BG seemed "more intense" than the other hues. One subject repeatedly made slightly faster scores when this hue was the target cue. In effect, the 5 BG became a "favorite color", seeming to have more dynamic visual appeal than the other polychromatic cue tabs.

At least two studies of the unequal luminance of equally bright colors, one by MacAdam (32), another by Chapanis and Halsey (13), provide explanation of this effect. Their findings demonstrated that equally luminous hues do not all appear to have equal saturation, and that equally saturated hues do not all appear to have equal brightness. In the experimental testing noted above, the 5BG, although of the same saturation and brightness as the other hues, appeared to have greater saturation, an effect which was heightened when the 5BG was placed between the 7.5YR and the 2.5 R.

Figure 3. Pushbutton controls with mono-L cue tabs. (Labels mounted on 6/0 gray.)

Figure 4. Pushbutton controls with poly-L cue tabs. (Left to right: 7.5 R 6/10, 10.0 R 3/6, 2.5 PB 3/6, 7.5 GY 8/6, 10.0 RP 4/4.)

Figure 5. Pushbutton controls with poly no-L cue tabs. (Left to right: 7.5 R 6/10, 10.0 R 3/6, 2.5 PB 3/6, 7.5 GY 8/6, 10.0 RP 4/4.)

To avoid this effect in the final testing, colored papers were chosen for distinguishability of all three characteristics, hue, value and chroma. At the same time, care was exerted to achieve psychological balance, i.e. a balance in which no single color tab appeared more luminous or more saturated than other tabs in the set.

Examination of the specifications of the color tabs finally chosen will reveal that while the same chroma or saturation /6 was characteristic of three hues, 2.5 PB, 7.5 GY, 10.0 R, the other two hues varied in value, /4 and /10. Viewing the five color tabs as a group, it is seen that contrast, basic ingredient of distinguishability, was heightened by using tabs of high, medium and low key or value.

5. Assignment of Hues to Color Cue Tabs

For the pilot trials, assignment of command words to color tabs was essentially arbitrary, the only deference to color-word association being that BOMB was given the hue 2.5 R. Since this was the only "red" among the 5 color tabs, the remaining 4 colors were assigned to the remaining 4 command words without special design, FIRE being coded GY.

Almost from the outset during polychromatic trials, subjects showed a tendency to reach for the red color tab when the command FIRE was the stimulus. Sometimes the subject would actually hit the red cued pushbutton despite its label, BOMB; sometimes he would almost hit BOMB then move across to the GY cued FIRE pushbutton. If he hit the red cued button he scored an error; if he started to hit it but swerved, his hand was delayed and he scored a hesitation with attendant longer RT.

This result exposed the factor of word-color association, obviously a problem affecting choice of color tabs for the control panel experiment. Accordingly, a separate study of command-word color association, detailed in another report (14), was undertaken. That study may be summarized as follows:

A group of 28 college Air Force ROTC upperclassmen, all of whom had been tested by the USAF and found to have color-normal vision, were used as subjects. Phase I of the experiment presented each subject with a list of 60 command words drawn from terminology commonly employed by military air-crews, the subject requested to write down names of colors stimulated by the command words. Phase II of the experiment presented the subjects with a set of 30 paper color tabs selected for their distinguishability, Ss being instructed to write down whatever command word came to mind when viewing the individual color tab.

While results for a majority of the command-word color associations were inconclusive, as was expected, there were a group of four which ranked in the top 5 for both Phase I and for Phase II tests.

Colors evoked by command words compared with command words evoked by colors revealed that virtual unanimity prevailed between the two tests for ranks 1 to 5. Phase I rank order, first through fifth, was: FIRE-RED, GO-GREEN, STOP-RED, HEAT-RED, SLOW-YELLOW.

Phase II rank order, first through fifth was: GREEN-GO, RED-FIRE, RED-STOP, BLUE-UP, YELLOW-SLOW.

Excepting for the 4th ranked item, the word-color associations are the same. Fact that FIRE and GO are 1st, 2nd and 2nd, 1st respectively, is not significant. It is significant that correlation between the above noted Phase I ranking 4 word-color associations and the Phase II ranking 4 color-word associations was 1.0.

Implication of these results is that designers of color coding for controls and/or indicators can expect errors or hesitation or both if operators are required to use code colors markedly dissimilar to red, green, yellow, red, for the commands fire, go, slow, stop, respectively.

It may be postulated, of course, that operators could be trained to ignore previously learned command word color associations when being taught a new code. Experimental evidence offers testimony, however, that such re-training may not always eradicate long standing, habitual command-word color association such as STOP-RED, GO-GREEN, FIRE-RED, and SLOW-YELLOW. The experience of this experimenter cited previously is a case in point wherein subjects repeatedly veered away from GY when searching for FIRE even though the green-yellow tab was plainly labeled FIRE, and the men had thoroughly mastered the color coding system.

Explanation of this result may lie in an analysis of color code training which young subjects normally undergo as they progress toward maturity in the typical framework of American culture. Most of our children begin learning the color coding of traffic lights as soon as they are old enough to cross the street alone. Thus they arrive at adulthood with some fifteen years of experience in associating go with green, stop with red. Likewise, all their training with fire signs, alarm boxes, fire-engines and fire-extinguishers has repeatedly taught red as the color for fire.

The coding of slow as yellow has usually come firmly into the training as teen-agers learn to drive automobiles, with most traffic signals using a yellow light as the warning signal between the go light's extinguishment and the stop light's illumination. Again, in most states of the U.S.A., motor vehicle warning signs used by highways typically employ yellow as backgrounds.

Further research investigating command-word color association is required before any final list of words and colors that have become generally associated can be established and validated. Results of the study noted here suggest there may be several more than 4 word-color associations established by the culture which can effect controls and indicator coding.

Applying command-word color association principles indicated by the above noted experiment, there was deliberate evaluation employed in assigning the command, FIRE, color 10.0 RP 4/4 for the control panel color coding study. Selection of a color suitable for the command BOMB was not as easily accomplished, however.

Results of the command-word color association test found color associations for BOMB to be varied, with black and gray as well as red listed. While red was first choice of 28.6% of the subjects for BOMB in the Phase I test, (i.e. colors evoked by command words), responses to the Phase II test, (i.e. command words evoked by colors), failed to find BOMB listed as a response to red by as many as 3 subjects. The latter result may not be taken as significant from one point of view, however. Subjects tested were as yet without any but academic knowledge of military action, none of them having had even a single day of combat training nor any first-hand contact with bombing missions, simulated or real.

When queried as to why colors such as black and gray were listed for BOMB, subjects admitted their knowledge of bombing derived essentially from motion pictures of bomb detonation, with most of said movies being viewed on television screens of the non-color variety.

It can be assumed that color association listed by veteran combat crews for the command BOMB would show a far higher unanimity of choice, with the colors probably nominated from the "hot" side of the spectrum, hues typical of fire-balls produced by the initial blasts of detonating bombs.

Since the control panel color coding tests were to be run with subjects who had spent several years in active duty with the USAF as fighter pilots and interceptor radio operators, it was no assumption that these men knew the detonation colors of live bombs and live rockets. Thus the "orange" hue of 7.5 R 6/10 chosen for coding BOMB was considered a color which was well related rather than at variance with the subjects' word-color association experience.

Reasoning behind selection of the other three cue tab colors included the following. It was the brownish stain of dried airplane fuel that suggested the use of "brown", 10.0 R 3/6, used to code CHOP; the coolness of the high key "green", 7.5 GY 8/6, seemed a logical reason to so code DE-ICE. The "blue" of 2.5 PB 3/6 used to code HOLD was selected because the hue possessed none of the stop, go, or warning associations characteristic of red, green, yellow, respectively. Since the airborne operator finds the word HOLD most frequently used in connection with neutral, non-emergency types of commands, the 2.5 PB 3/6 was selected as being relatively neutral and word-association free.

Use of "orange" 7.5 R 6/10 and "red" 10.0 RP 4/4 together in a set of controls color cues was deliberately done as a test of the possible confusion which might result from the two "hot" colors being seen in the same group of cue tabs. With subjects exerting maximal effort to search for and depress the target pushbutton in the shortest possible time, there was a condition presented by these "neighbor" hues which would test their distinguishability.

Experimental evidence of the degree of distinguishability afforded by use of the "red" and "orange" cue tabs is presented in the Results section of this paper, and analyzed in the section headed Discussion.

6. <u>Illumination of the Control Panel</u>

To insure distortion-free illumination of the colored cue tabs, the testing was conducted in a room with windows reaching to the ceiling and admitting north light only. No artificial light was employed; testing was conducted during hours which were clearly free of sunrise and sunset spectral limitations. Earliest testing began no earlier than 1.5 hours after sunrise; latest testing ended no later than 2.0 hours before sunset.

Measurement of incident daylight reaching the control panel gave readings of approximately 130 ft. candles illumination.

7. Establishing Random Presentation of Color Coding

In order to thoroughly eliminate any use of position as a cue for distinguishability, the location of the five cue tabs was changed for each trial, whether monochromatic or polychromatic. Systematic randomization of the relocation was determined by use of a 5x5 Latin Square randomly selected from Fisher and Yates Statistical Tables. Twenty permutations provided further random selection of the order in which the cue tabs were to be placed. In actual use during more than 2000 trials, the random order thus derived was never successfully anticipated by any subject.

The order of presentation of command cards was also randomized by use of the 5x5 Latin Square, this procedure supplying still another discouragement to "set" or advance knowledge of the command to be employed. Thus, although the variety of commands was limited to 5, during testing subjects never knew which of the commands would appear next.

B. PROCEDURE

Experimenter placed a command card in the control panel window. Subject was instructed not to hurry while reading the command since timing of his response was not begun until the screen dropped clear of the control pushbuttons and their cue tabs.

This procedure eliminated the addition of another variable, namely, RT required for viewing and perceiving the meaning of the command word.

Thus, the subject was able to transfer his gaze from the window and command card to a point on the screen approximately opposite the cluster of pushbuttons, concentrate upon the command word and the problem of finding the corresponding cue tab, then squeeze the intercom switch releasing the screen at the instant he himself elected.

This meant the tests were self-paced by the subject, rather than experimenter paced, a factor which undoubtedly contributed to the very fast RTs recorded. (See Tables 2 through 10).

With the pushbuttons and cue tabs in view, the subject sought out the target button and hit it in the shortest time he could. Electronic timing of his response cut off when the target button was depressed, opening the counter circuit. Since all pushbuttons were wired in parallel, depression of any button stopped the counter, an apparatus feature which permitted timing of wrong button hits as well as correct hits.

The order of testing was altered for each of three groups to permit analysis of order effect. Group I subjects were tested in the order: (1) Mono-L (2) Poly-L (3) Poly No-L. Group II subjects were tested in the order: (1) Poly-L (2) Poly No-L (3) Mono-L. Group III subjects were tested in the order: (1) Poly No-L (2) Mono-L (3) Poly-L.

When Mono-L cue tabs were the first to be tested, orientation trials consisted in running once each of the 5 command cards, with the subject instructed to release the screen when he felt ready to select the target pushbutton. Recording of RT began after these 5 orientation trials.

When Poly-L cue tabs were the first to be tested, orientation consisted of a five minute training period during which the subject learned which colors coded which command labels. During the first ten recorded trials, the subject was allowed to repeat aloud the color cue to be sought when he dropped the screen, a technique which proved adequate for firming-up color code learning.

When Poly No-L cue tabs were the first to be tested, orientation was also held to five minutes, and again the subject was allowed to repeat aloud the color to be sought when he dropped the screen. Comparison of test results for the first ten trials of Poly-L first subjects with the results of the Poly No-L first subjects showed that the latter group were slower to reach their mean RT, seemed to require longer warm-up. This suggests the orientation for Poly-No-L-first subjects should have been longer than for the Poly-L first.

With subjects exerting maximum effort for each trial in the attempt to reach peak performance, the first 30 trials usually consumed a full hour of testing time. At this point, subjects were given a five minute break before beginning the second block of 30 trials. Subjects typically approached their mean RT during the period of trials 20 to 40, hence warm-up effect was considered essentially negligible, if present at all, during the final 20 trials of the 60 given for each cue type.

Notation of the RT of each trial was made on raw data sheets after confirmation of the chronoscope reading had been verified by both subject and experimenter. Records included observations on hesitations as well as errors, and on how the search task was performed, i.e. left to right, right to left, or center start.

Subject was seated in a writing-arm chair whose seat was 18.0 in. from the floor. Direct-line distance from the front edge of the arm to the pushbuttons was 18.5 inches. At the start of a trial, the subject was required to keep the right hand resting on the arm so that palm and fingers were behind the arm's front edge, thereby establishing 18.5 in. as the minimal reach required.

With the pushbuttons mounted in line 37.0 in. from the floor, there was an easy lift for the hand rising from the 26.0 in. elevation of the writing arm to the target. According to anthropometry tables established by USAF records, (39), the reach required was well within the mean anterior arm reach of aviation cadets: 35.2 ± 1.6 inches.

C. SUBJECTS

Subjects were members of a USAF fighter-interceptor squadron on active duty flying two-place jet aircraft. Individuals selected comprised 4 pilots and 5 radio operators. These were divided into three groups; Group I and Group II each had one pilot and two ROs, Group III had two pilots and one RO.

Ages varied from 22 through 26, all subjects were married. Six Ss, including the 4 pilots, were USAF career men; the remaining three were non-career ROs.

Training within the USAF varied with length of service. The youngest member of the group had more than $2\frac{1}{2}$ years Air Force experience; the oldest had slightly more than 5 years.

Age and training did not seem to have a measureable effect on RT scores, however, as Table 11 shows. Ranked in terms of RTs, Pilot Mc. age 25, placed 1st; RO M. age 22 placed 2nd.

Biggest single factor affecting RT speed seemed to be motivation. In general, the career men appeared more eager to rack up fast scores than did the non-career men. Individual differences in personality also may have been a factor, although there was no attempt made to evaluate personality systematically.

Individual differences were very marked during tests of hand-travel RT, timing of which was accomplished by having Ss drop the screen and hit any button, no search for a given target being required. Table 12 lists the results of this phase of the testing, while the significance of hand-travel time for over-all results is discussed in Section IV.

III. RESULTS

A. TYPICALITY OF TRIALS ON WHICH RESULTS ARE BASED

Nine subjects went through exactly 180 trials each, making a grand total of 1620 trials. Each subject was tested with 60 trials for each of the three cue types, monochromatic with labels (mono-L), polychromatic with labels (poly-L), and polychromatic without labels (poly no-L).

Since the core objective of the experiment was to test whether the addition of color coding to controls labels would improve the speed and accuracy of selection rather than to test absolute speed and accuracy, it was considered irrelevant to indulge in an extended series of practice trials. Examination of Table 11 and the column of mean RTs scored for the mono-L cue type will show a result which was expected: the great majority of subjects (8 out of 9), were unable to bring their selection speed under 600 milliseconds.

This result was expected in recognition of the fact that with monochromatic cues, subjects were forced to search and find the target pushbutton by <u>reading</u> the labels. This operation might not have been too difficult if the spread of the 5 labels had not been 9.6 inches, which rendered impossible reading them as a single block by focusing on the center label. Rather, it was found that each label had to be read one at a time, and since the positions of the labels were altered after each trial, the test became a matter of developing maximal speed of search with reading, accompanied by recognition, understanding, or perception.

When the subjects tried to increase selection speed by reading faster, they often scanned past the target label, seeing but not perceiving, then had to scan back until they located it. The delay due to scan-past invariably caused a longer RT, often equal to a full standard deviation.

The problem of optimal scan speed for search requiring reading is discussed in some detail in the next section. This problem obviously had to be solved instinctively by each subject as he tried to avoid scan-past yet read with maximal speed.

Group I subjects, starting with mono-L cues, had to discover by trial and error their individual reading speed thresholds during the first trials undertaken. While struggling to reduce the errors and/or hesitations due to scan-past they may be considered to have been undergoing warm-up. Typically, warm-up extended through the first 20 to 30 trials at the end of which subjects were given a 5 minute break. Returning for the second hour of testing, Ss usually hit their maximal performance during trials 31 to 40.

Thus, performance during the final 20 trials of the mono-L cue type found subjects essentially free of warm-up. Orientation was well behind them, they seemed to have made friends with the apparatus. It was therefore considered best to use the final 20 trials as the most typical performance, and these scores comprise the raw data results reported.

Similarly with the poly-L and poly no-L tests, it was found that the final 20 trials represented the most typical performance.

Accordingly, derived data presented in the tables and reported in the results are based on raw data recorded for the final 20 trials of each cue type. Prepared in graphic form comprised of curves for performance with each of the cue types for each of the 9 subjects, the actual scores have been attached to this paper in the Appendix, Figures 6 through 14.

B. MAJOR RESULTS

Results obtained in this control color coding test program amount to unusually strong support for the postulate that polychromatic coding is superior to monochromatic coding. As Table 1 reveals, the combined grand mean percent of improvement over monochromatic scored by use of polychromatic cues was found to be 57.17% with poly-L mean improvement amounting to 54.66% and poly no-L mean improvement rising to 59.68%.

The mean RTs and SDs for each subject for each of the three cue types during the final 20 trials are presented in Tables 2 through 10.

Minimal mean percent improvement for poly no-L was 21.79% scored by Pilot S. of Group III, tested with poly-L cues first in the series, which meant least amount of benefit from practice. (This minimal figure was calculated by use of the mono-L mean RT minus one SD, a remainder considered representative of maximal or fastest mono-L mean RT speed, and use of the poly no-L mean RT plus one SD, considered representative of the minimal or slowest poly no-L mean RT speed.)

Combined minimal percent polychromatic improvement for all subjects amounted to 29.47%, a figure far stronger than borderline significance.

TABLE 1. COMBINED MEAN % IMPROVEMENT FOR ALL SUBJECTS

ALL GROUPS	Cue Type	Mean % Improvement Over Mono-L	Grand Mean % Improvement Poly Over Mono
Combined Results	Poly-L	54.66	57.17
	Poly no-L	59.68	

TABLE 2. GROUP I ORDER OF TESTING

	Cue Type	Order of Testing
	Mono-L	1
Group I	Poly-L	2
	Poly no-L	3

TABLE 3. GROUP I INDIVIDUAL MEAN RTs, SDs, % IMPROVEMENT

Subject	Cue Type	Mean RT (msec.)	Standard Deviation	Mean % Improvement Over Mono-L
	Mono-L	671	228	
Pilot Hi.	Poly-L	342	80	49.03
	Poly no-L	245	46	63.46
	Mono-L	808	212	
RO N.	Poly-L	538	171	33.42
	Poly no-L	310	107	61.63
	Mono-L	701	182	
RO T.	Poly-L	353	86	49.64
	Poly no-L	281	35	59.91

Mean RTs are those for final 20 trials for each cue type.

TABLE 4. GROUP I COMBINED MEAN % IMPROVEMENT

	Cue Type	% Improvement Over Mono-L	Grand Mean % Improvement Poly Over Mono
GROUP I Combined Results	Poly-L	44.03	52.85
	Poly no-L	61.67	

TABLE 5. GROUP II ORDER OF TESTING

	Cue Type	Order of Testing
Group II	Poly-L	1
	Poly no-L	2
	Mono-L	3

TABLE 6. GROUP II INDIVIDUAL MEAN RTs, SDs, % IMPROVEMENT

Subject	Cue Type	Mean RT (msec.)	Standard Deviation (msec.)	Mean % Improvement Over Mono-L
	Mono-L	664	171	
Pilot Ha.	Poly-L	348	29	47.59
	Poly no-L	295	38	55.57
	Mono-L	860	285	
RO K.	Poly-L	473	116	45.00
	Poly no-L	313	61	63.60
	Mono-L	644	201	
RO L.	Poly-L	251	47	61.02
	Poly no-L	231	37	64.13

Mean RTs are those for final 20 trials for each cue type.

TABLE 7. GROUP II COMBINED MEAN % IMPROVEMENT

	Cue Type	Mean % Improvement Over Mono-L	Grand Mean % Improvement Poly Over Mono
GROUP II Combined Results	Poly-L	51.20	56.15
	Poly no-L	61.10	

TABLE 8. GROUP III ORDER OF TESTING

	Cue Type	Order of Testing
Group III	Poly no-L	1
	Mono-L	2
	Poly-L	3

TABLE 9. GROUP III INDIVIDUAL MEAN RTs, SDs, % IMPROVEMENT

Subject	Cue Type	Mean RT (msec.)	Standard Deviation (msec.)	Mean % Improvement Over Mono-L
	Mono-L	465	202	
Pilot Mc.	Poly-L	113	32	75.70
	Poly no-L	233	95	49.89
	Mono-L	609	128	
RO M.	Poly-L	198	25	67.49
	Poly no-L	217	29	64.37
	Mono-L	623	146	
Pilot S.	Poly-L	230	136	63.08
	Poly no-L	281	92	54.92

Mean RTs are those for final 20 trials for each cue type.

TABLE 10. GROUP III COMBINED MEAN % IMPROVEMENT

	Cue Type	Mean % Improvement Over Mono-L	Grand Mean % Improvement Poly Over Mono
GROUP III Combined Results	Poly-L	68.76	62.57
	Poly no-L	56.39	

C. ORDER EFFECT AND INTERACTION ANALYSIS

Tabulation of mean test scores for poly-L and poly no-L, Table 14, provides visibility of whatever order effect and interactions may have been present during the testing.

At first glance, it would appear that since 6 of the 9 subjects demonstrated significant improvement when given polychromatic cues without labels, a conclusion could be drawn that selection speed is faster when controls are coded by color only than when coded by color plus label.

Examination of the scores recorded for Group III, however seems to show a reversal of that principle. Group III subjects scored significantly faster RTs for poly-L than for poly no-L.

Consideration of the variance in testing order suggests a different conclusion. The fastest mean scores were invariably made on the second polychromatic cues testing session by all three groups, whether the cue type was poly-L or poly no-L. Thus, the indications are that some degree of practice effect was at work.

Yet, if practice effect per se was of significant force throughout the testing, subjects should have made faster RTs for monochromatic label coding when this cue type was the third and last test of the series, as it was for Group II. Study of the mean RTs for mono-L in Table II reveals, however, that there was no significant difference between the mono-L scores of Group II and the mono-L scores of Group I which began with mono-L cues.

TABLE 11. MEAN RT RANK ORDER FOR ALL SUBJECTS

Group	Subject	Mean RT (msecs.) Mono-L	Poly-L	Poly no-L	Rank Order	Mean Fastest Rank	Group Rank Order
	Pilot Hi	671	342	245	6,5,4	5	
I	RO N.	808	538	310	8,9,8	8	3
	RO T.	701	353	281	7,7,6	7	
	Pilot Ha	664	348	295	5,6,7	6	
II	RO K.	860	473	313	9,8,9	9	2
	RO L.	644	251	231	4,4,2	4	
	RO M.	609	198	217	2,2,1	2	
III	Pilot Mc	465	113	233	1,1,3	1	1
	Pilot S.	623	230	281	3,3,6	3	

Order of Testing	1	2	3
Group I	Mono-L	Poly-L	Poly no-L
Group II	Poly-L	Poly no-L	Mono-L
Group III	Poly no-L	Mono L	Poly-L

TABLE 12. MEAN RT FOR HAND TRAVEL, ALL SUBJECTS

Group	Subject	Mean RT for Hand Travel (msec.)			Grand Mean Hand Travel (msec.)	RT Rank	Group Rank
		Mono-L	Poly-L	Poly no-L			
	Pilot Hi.	—	73	82	77	8	
I	RO N.	—	109	20	64	4	2
	RO T.	—	39	39	39	2	
	Pilot Ha.	37	93	33	54	3	
II	RO K.	82	127	60	90	9	3
	RO L.	88	61	61	70	7	
	RO M.	69	77	54	67	5	
III	Pilot Mc.	—	03	19	11	1	1
	Pilot S.	62	62	82	69	6	

Order of Testing	1	2	3
Group I	Mono-L	Poly-L	Poly no-L
Group II	Poly-L	Poly no-L	Mono-L
Group III	Poly no-L	Mono-L	Poly-L

TABLE 13. ERRORS AND HESITATIONS RECORD, ALL SUBJECTS

Group	Subject	Errors Mono-L	Errors Poly-L	Errors Poly no-L	Hesitations Mono-L	Hesitations Poly-L	Hesitations Poly no-L
	Pilot Hi.	2 (s) 2 (h)	1 (c)	0	1 (h) 1 (s) 1 (u)	2 (s) 3 (u)	1 (c) 3 (u)
I	RO N.	0	0	3 (c)	3 (s) 2 (u)	2 (h) 3 (c) 1 (s)	1 (c) 2 (u)
	RO T.	2 (h)	1 (s)	1 (c) 1 (s)	1 (s) 2 (u)	1 (u) 2 (s)	1 (c) 1 (u)
	Pilot Ha.	1 (s)	2 (c)	0	5 (u)	3 (u)	1 (c) 1 (u)
II	RO K.	0	0	0	1 (s) 2 (h) 3 (u)	2 (c) 2 (u)	2 (u)
	RO L.	2 (s) 1 (h)	0	0	1 (s) 2 (u)	1 (c) 3 (u)	1 (c) 1 (u)
	RO M.	3 (s)	1 (c)	1 (c)	3 (s)	0	2 (c) 1 (u)
III*	Pilot Mc.	3 (h)	0	0	1 (s) 1 (u)	0	1 (s) 1 (h)
	Pilot S.	4 (h)	2 (c)	1 (c)	1 (s) 2 (u)	1 (c) 1 (u)	1 (u)
	Totals	8 (s) 12 (h) 20	6 (c) 1 (s) 7	6 (c) 1 (s) 7	12 (s) 18 (u) 3 (h) 33	5 (s) 7 (c) 2 (h) 13 (u) 27	7 (c) 1 (s) 1 (h) 12 (u) 21

*Note: Group III was tested with poly-L cue tabs last, whereas the other groups were tested with poly no-L cue tabs last.

Symbols: (s)= scan or reading (c)= color confusion
(h)= hand travel (u)= undetermined

Mean hesitation improvement= 27.27%
Error improvement= 286%

TABLE 14. ORDER EFFECT & INTERACTION ANALYSIS

Group	Subject	Mean RT (msec.) Poly-L	Mean RT (msec.) Poly no-L	Msecs. Gain Over 1st Poly	% Gain Over 1st Poly	Group % Gain Over 1st Poly
	Pilot Hi.	342	245	97	28.36	
I	RO N.	538	310	188	34.94	27.89
	RO T.	353	281	72	20.39	
	Pilot Ha.	348	295	53	15.23	
II	RO K.	473	313	160	33.80	18.99
	RO L.	251	231	20	7.96	
	RO M.	198	217	19	51.50	
III*	Pilot Mc.	113	233	120	8.75	29.13
	Pilot S.	230	281	51	18.14	

*Note: Group III was tested with Poly-L as the 2nd polychromatic cue type; other groups had Poly-L as the 1st polychromatic cue type.

Order of Testing	1	2	3
Group I	Mono-L	Poly-L	Poly no-L
Group II	Poly-L	Poly no-L	Mono-L
Group III	Poly no-L	Mono-L	Poly-L

Viewing the scores of Group III in Table 11 for practice effects calls for recognition of still another factor affecting group performances, namely, individual differences. Group III happened to be comprised of the three subjects whose superior scores caused them to rank 1st, 2nd and 3rd in the rank order for fastest mean RTs. Since the assignment of subjects to the respective groups was made by an Operations Officer whose only concern was the convenience of his squadron's regular duties rather than suitability of individuals for the control panel experiment, there was no reason to assume that the men of Group III were not randomly selected.

While it could be argued that pilots are typically faster than ROs and Group III had two pilots while the other groups had but one each, examination of Table 11 and the rank order of pilots versus ROs will reveal that ROs placed 2nd and 4th with two pilots below them, 5th and 6th. Fastest mean score for poly no-L was 217 msecs., racked up by an RO, not by a pilot.

In view of these facts, it would be fallacious to pursue the assumption that pilots are typically faster and accept the assumption as an explanation of Group III's superior performance. At best it can only be safely concluded that Group III happened to have the three fastest individuals among the nine men tested.

Table 12, Mean RT for Hand Travel, All Subjects, presents another group of results bearing on the matter of group interactions and the effect of individual capabilities. A comparison of Table 12 with Table 11 shows that correlation between hand travel speed and performance speed is nil for all but one subject, Pilot Mc., who ranked 1st in both. Subjects ranking 2nd and 3rd in performance ranked 5th and 6th in hand travel; subjects ranking 2nd and 3rd in hand travel ranked 7th and 6th in performance.

Correlations between group rank order also fail to appear, with the exception of Group III which ranked first in both hand travel and performance. But Group I, which ranked 3rd in performance, ranked 2nd in hand travel; Group II, which ranked 2nd in performance, ranked 3rd in hand travel.

In summary, evidence points to three conclusions regarding order effect:

1. Order of presentation had no significant effect on RT for controls selection when coded by monochromatic labels. Reading speed did not improve significantly with practice.

2. Order of presentation had a significant effect on RT for controls selection when coded by polychromatic cues, practice effects appearing to have affected the mean percent of improvement, as Table 14 reveals. As subjects gained in practice with the color coding, they improved in selection speed.

3. Inconclusive data are provided by present results as to the relative superiority of polychromatic cues without labels over polychromatic cues with labels. While there was reason to believe that labels became of no use to subjects of Group III in scoring the fastest mean RTs when using the poly-L cue tabs, further testing is indicated to validate this conclusion.

IV. DISCUSSION

A. MOTOR RESPONSE VS. PERCEPTUAL RESPONSE

Comparison of Table 11, Mean RT for Hand Travel, with Table 12, Mean RT Rank Order for All Subjects, permits a view of motor response versus perceptual response and of how one effects the other in RT tests of the kind investigated.

It might have been expected that subjects showing the faster performance RTs would have proved to be those with the faster motor response or hand travel RTs. However, testing results proved otherwise, with subjects ranking 2nd and 3rd in performance speed scoring 5th and 6th in hand travel speed, for example. Again, subjects ranking 4th and 5th in performance ranked 7th and 8th in hand travel.

It is also noteworthy that motor response mean RT varied from cue type to cue type for the same subject. This result suggests that maximal speed of motor response is not always present, particularly if the problem of controls selection is less than perfectly encompassed by the perceptual processes. For example, most subjects decreased their hand travel time as they progressed from one cue type to the next, but some subjects did not. Pilot Hi.'s last group of trials found his mean hand travel speed slowed to 82 msecs. from the 72 msecs. mean RT scored in the 2nd group of trials although his performance mean RT for the final trials was faster by 97 msecs.

A similar slowing of hand travel speed in the final trials was recorded for RO M. while at the same time speeding up his performance RT.

Evaluation of the proportion of performance time consumed by hand travel brings forth the fact that motor response time varied from 9.83% (Pilot Mc.) to 33.83% (RO M.). If it can be assumed that the balance of time consumed was purely perceptual, the conclusion may be drawn that individual differences in perception speed constitute the most critical conditioning factors in controls selection performance scores.

While the actual mean RTs for hand travel varied from 11 to 90 msecs. among the nine subjects, the actual mean RTs for fastest performance varied from 113 to 313 msecs. Other experimental designs than that employed in this investigation no doubt would be better suited to a study of perception time. However, this area is obviously of vital concern to controls selection training as well as to the design of coding controls and indicators.

In the following section a more detailed discussion of the perceptual problems involved in controls selection is presented.

B. ERRORS AND HESITATIONS ANALYSIS

Table 13, Errors and Hesitations Record, All Subjects, presents results of the errors and hesitations analysis. Most dramatic difference appears in the errors totals, with monochromatic cue tabs causing 286% more errors than the polychromatic cue tabs.

Study of the errors columns reveals that polychromatic cue tabs caused no errors whatsoever of the hand travel type, whereas monochromatic cue tabs errors were 60% comprised of hand travel mistakes. While scanning errors comprised 40% of the monochromatic cue tab mistakes, polychromatic cue tab scanning errors comprised only 14.3% of the mistakes.

As noted previously, the potential of color confusion between the "red" 10.0 RP 4/4 and the "orange" 7.5 R 6/10 was deliberately introduced into the polychromatic cue tabs. Results show that all but 14.3% of the errors caused by this cue type were due to color confusion. Hence the conclusion may be drawn that color coding design should avoid use of colors as closely related as the "red" and "orange" employed in this experiment.

One explanation of the color confusion found in this testing program may be derived from authoritative studies of color perception, including that of the OSA Committee on Colorimetry (38), which report that fall-off of red-orange receptor sensitivity has been found outside a relatively small foveal area of the retina. Color confusion errors and hesitations were more often recorded where the "red" and "orange" cue tabs were at opposite ends of the cue tab line. This meant their images reached the perpheral areas of the retina rather than the fovea when S. used center start scanning as most Ss. did with polychromatic cue tabs. It is therefore indicated that "red" vs. "orange" distinctions of the kind required by the cue tabs employed present psychophysical perceptual limitations which are basic and fundamental. Implications of these color perception limitations are many, and obviously, each implication requires individual investigation, and with the research design dictated by the end needs of the color coding designer.

A further conclusion is also apparent. If the "orange" had been replaced by another color which would not arouse color confusion with the "red", there would very likely have been markedly fewer errors recorded for the polychromatic cue types. In that event, the degree of accuracy afforded by color coding as against monochromatic coding could have mounted from the present 286% improvement to perhaps 400% or better. Further testing should reveal how valid this conclusion may be.

Examination of the hesitations columns of Table 13 reveals that color confusion was also a prominent factor in the production of delayed RTs. With poly-L cue tabs, color confusion hesitations amounted to 25.9%, and with poly no-L cue tabs 33.0%.

Elimination or sharp reduction of this kind of RT delay could mean that hesitations improvement resulting from color coding might reach 60%, an important advance particularly significant for the control panels of hypersonic aircraft and manned missiles wherein a 60% delay in search time could spell the difference between survival and disaster.

Practice effect could be ascribed to the apparent errors improvement shown by Group II subjects who were tested with mono-L cue tabs last and scored 33% fewer errors than Group I subjects who began with mono-L. Affecting this situation, however, is the factor of individual differences, tabulated in Table 11. The 6 errors made by Group I subjects were recorded for men ranking 5th and 7th in mean RT speed hence it cannot be assumed they were equally capable as the two men of Group II who ranked 4th and 6th and scored only 4 errors.

Another query maybe aroused when examining the performance of Group III subjects for mono-L errors. Their testing with the monochromatic cue tabs came 2nd, rather than 1st, yet their errors total was 10, highest of all the three groups. Since no labels were used during Group III's first tests, poly no-L cue tabs being the cue type, a re-learning task was imposed on this group by the mono-L cue type not required of Group I which began with mono-L. Further consideration of the intergroup testing sequences shows, however, that Group II also faced a re-learning task when given their mono-L tests following poly no-L testing. Thus, both Group III and Group II had the same re-learning conditions imposed upon them when tested with mono-L cue tabs, yet Group II scored only 4 errors against Group III's 10.

These results are sufficiently contradictory that no clean-cut conclusion may be drawn as to the presence or lack of practice effect, or, as to what may have caused the marked differences in error totals recorded by the respective groups. Further testing with perhaps several hundred subjects involved may be required before definitive conclusions could be validated.

Looking ahead to further testing and analysis of hesitation delays in controls and indicator selection, it is the opinion of this investigator that only gross deviations can be satisfactorily noted by the primitive method of naked eye observation. With hand travel consuming elapsed time spans as small as 0.2 seconds, $\pm$ 0.1, there is far too short a period in which the human eye can function efficiently as a recording instrument. The proper instrument to employ here is the high-speed motion picture camera or the television camera using electronic tape recording, these equipments permitting projection and analysis of recorded hand travel action at importantly slower speed.

Since hesitations have now become almost as critical for efficient operation of certain types of control panels as have outright errors, careful studies of hesitations are very much in order, with expensive recording equipment not merely justified but actually essential to solving the questions of how hesitation originates. Knowledge of hesitations causes are required, of course, not only by control panel designers but also by designers of training programs. At present it would seem that too few fundamentals of learning limitations on the perceptual and visual levels, on scanning speed thresholds and on acuity of color perception have been made available to training commands, and educators in general. Why this is so may very well constitute the subject of a separate and extensive investigation. In the following section of this paper, some additional comment is offered regarding these interrelated factors of perception, learning, and operational efficiency.

Viewing the hesitation record of Table 13 at the totals level, it may be noted that hesitations were most numerous for the monochromatic cue tabs, with significantly less hesitations appearing when polychromatic cue tabs were employed. Practice effect, while unclear for mono-L tests, is rather positive for polychromatic tests. As subjects progressed from their first testing with polychromatic cue tabs to their second polychromatic testing, their RTs invariably improved.

Indications of the testing progress recorded points to the likelihood that with additional polychromatic cue tab training, subjects would continue to reduce the number of hesitations, widening further the significant gap between hesitations committed during monochromatic trials and those committed during polychromatic trials.

In summary, the following conclusions may be drawn:

1. Polychromatic cue tabs reduced the number of errors by an extraordinarily high degree, averaging 286% improvement over monochromatic cue tabs.

2. Polychromatic cue tabs reduced the number of hesitations by a very significant degree, averaging 27.3% improvement over monochromatic cue tabs.

3. If color confusion, the cause of 85.7% of the color selection errors, were reduced or eliminated, it is quite likely that improvement of selection accuracy due to color coding controls would rise from the 286% found in this investigation to perhaps 400% or better.

C. COMMENT ON THE STATISTICAL METHOD

At first glance, the experimental design of this investigation would seem to lend itself to parametric statistics, with classic analysis of variance methodology neatly applicable to the study of order effect and interactions.

There are three cue types, three groups, and three orders of testing. The balanced number of data recorded seems ready-made for blocks of cells into which derived results could be dropped, producing the kind of mathematical setting that would permit full scale use of formulae and nicely refined calculations.

However, little more than a casual glance at the actual scores and at the mean RTs recorded for each subject will reveal that the data do not display evidence of normal distribution. Rather, the results are severely skewed and therefore unsuited to statistical analysis by parametric techniques.

For example, Table 11 reveals that the mean RTs scored for mono-L cue tabs are distributed over a range of 465 msecs., (Pilot Mc.), to 860 msecs., (RO K.), a difference amounting to 84.9%.

Again, distribution of the mean RTs for the poly-L configuration soars from 113 msecs., (Pilot Mc.) to 538 msecs., (RO N.), a difference of 476%.

Also, the mean RTs for poly no-L vary from 217 msecs., (RO M.), to 313 msecs., (RO K.), a difference of 69.3%.

A review of Table 12, Mean RT for Hand Travel, All Subjects, provides evidence of another seriously skewed distribution: mean hand travel speed ranged from 11 msecs., (Pilot Mc.), to 90 msecs., (RO K.).

In view of these several evidences of skewness, it was felt unwise to attempt manipulation of the data into form which would permit parametric handling of the statistical analysis.

Thus, statistical calculations requiring normal distribution were rendered inapplicable. Instead, the data were studied on the basis of nonparametric statistics, with each subject's performance measured against his own record. This technique provided percent of improvement factors and allowed comparisons of the percentages, both on the individual level and on the group level.

Since each subject was competing against his own scores rather than against the scores of his squadron mates, and since each man was tested singly, interaction effects existed only in the matter of practice or training, not between subjects but for each S. individually. (The analysis of these effects as presented in the previous section of this paper will not be reviewed here.)

D. SIGNIFICANCE OF SUPPORT FOR THE HYPOTHESES

To recapitulate, the two hypotheses tested were as follows. Given a control panel problem, it is postulated that:

1. Speed and accuracy of controls selection will be significantly improved if polychromatic rather than monochromatic color coding of labels is employed.

2. Speed and accuracy of controls selection will be significantly improved if polychromatic color coding only, without labels, is employed to identify controls.

As Table 1 shows, the mean improvement in speed of controls selection when polychromatic color coding of labels was employed amounted to 57.17 percent.

With poly-L results showing 54.66% improvement, and poly no-L results reading 59.68% improvement, there seems to be nothing but highly significant support for both hypotheses as to speed.

As Table 13 shows, use of polychromatic color coding of labels reduced the errors by 286 percent, a figure several times more positive than required of experimental results to qualify as significant. Color coding also significantly reduced hesitations by 27.3 percent.

In summary, therefore, it can be stated without qualification that both hypotheses were supported for speed and accuracy to a very significant degree by the experimental results.

V. CONCLUSIONS

A. IMPLICATIONS IF RESULTS ARE VALID

If the results demonstrated by this investigation are valid, and if they prove to be confirmed by further experimental testing, then several areas of human factors engineering will find need of the advantages demonstrated by use of color coding controls and indicators.

Whenever speed and accuracy of controls selection is of vital importance, coding of controls with color will become a matter of first consideration for the control panel designer. As soon as it is determined that the panel can be operated under steady-state controlled illumination, the question will not be, is color coding advisable, but, how can color coding principles be used most effectively.

With present evidence demonstrating controls selection speed to be better than twice as fast when coding is polychromatic rather than monochromatic, and with selection errors reduced by more than 250 percent due to color coding, control panels are certain to provide greater efficiency and greater safety if they have been given color coded controls and indicators.

In applications where the control panels have critical functions to perform, when, for example, the selection of a wrong control through operator error can cause serious damage to equipment, buildings, or personnel, the specific advantages of color coding will become a first concern of the panel designers.

Certain branches of the chemicals and petroleum industries have already made use of color coding controls and indicators, of course. Many process engineering designers have employed color coding not only for control panels but for identification of pipes, tubes, and receptacles, as well as for process components. The production of explosive liquids, gases, and of solid explosives has called for utmost use of all safety methods conceivable. Among these, color coding has become a primary safety device.

Without making a special study of the origins and basic principles of color coding employed by process engineering, however, it can be only conjecture as to how much the color coding customs now followed in that engineering field have been founded on the fundamental data of psychophysics and of color perception and how much founded on color practice which had no reason for its design other than tradition or habit. In some process plants studied by this investigator, for instance, the color coding seemed to have been determined by little more than a simple need to render one chemical line or control distinguishable from another. There had been built up within a single industry a rather elaborate color coding system which the personnel of that industry came to know and recognize, but there were arbitrary uses made of green and yellow, for example, which violated the color-word association findings noted previously (14). Such practices could have negative effects, on occasion defeating the very safety conditions they were set up to protect.

Viewing the several major fields in which visual displays of controls and indicators are functional, essential components, it is easily demonstrated that color coding is as yet largely not practiced. Yet whenever and wherever control panels are required, color coding is due to become a major design consideration from the outset. Likewise, indicator displays which are to function with maximal efficiency, providing operators with the most readily recognized and interpreted kind of coding, are certain to require color coding. Such are the general implications of results found dominant in the experimental testing reported herein.

Specific applications of color coding are indicated for virtually all kinds of vehicles, whether they carry personnel and cargo via air, water, land, or combinations thereof. Aircraft, missiles, satellites and space vehicles, including the now elaborate ground support equipment necessary for their control, all have speed and accuracy requirements among their several kinds of control and indicator panels. Perhaps the majority of these have need of color coding, for the majority of panels used by the vehicles in our air-borne and space-traversing categories demand split-second decisions of the crews who operate them.

Although the trend is toward increasing use of automatic controls, with electronic computers and servomechanisms now performing guidance and navigational computing formerly supplied by the human brain, designers of the hypersonic aircraft, and of manned missiles, satellites and space vehicles, still find need for the monitoring and decision making capabilities of the live pilot and his fellow crewman. As long as man is to be placed in control of a vehicle just so long will the need remain to provide him with the most superior type of controls and indicator coding.

Watercraft, like aircraft, also have need for clearly coded controls and indicators, and, as the craft becomes larger, ranging from boat to ship, small ship to large ship to aircraft carrier and supercargo ship, the control problems multiply in both number and complexity. For the vessels which qualify as ships, with seven seas navigational requirements to be met under allweather conditions, controls may become extended from a few panels in number to racks of panels. Aboard the larger naval vessels and passenger liners, when the control room becomes a control center with several operators manning several control stations, the coding problems are extensive. With such control centers typically placed amidships to afford the greatest possible protection for the ship if under attack, there is no problem whatsoever as to controlled illumination. Thus, color coding can be employed without limit, the fully enclosed control center being under a steady-state illumination the spectral composition of which can be specified to suit the needs of the color coding techniques employed.

Were a comprehensive list prepared of control panel fields wherein color coding is indicated, perhaps several thousand words would be required before the listings were all-embracive. While it is outside the purpose of this paper to attempt any such list, there is reason, at least, to note that the list would be large rather than small, its size being one measure of the importance which must be given the matter of controls and indicator color coding.

B. OTHER AREAS OF RESEARCH AFFECTING COLOR CODING

At least seven other areas of psychological research directly affect the field of color coding, and no doubt more than seven could be named were a separate study of the wide scope of visual and perceptual problems involved undertaken. At this early stage in the research of human factors governing optimal use of color coding, an investigator cannot but feel the enormity of this field whose unsolved problems are still so numerous he hesitates to presume a capability for envisioning them all.

The following seven related research areas are discussed herewith not because they are necessarily the seven most important to the color coding field, but because they happen to have had reports published in the literature. This is not at all meant to imply, on the other hand, that these reports deal with factors unimportant to color coding. Rather, the comment is intended to make the reader aware that to date there has not been established a fully comprehensive research program for the study of color coding, and until such a long-range program is carried out, no single investigator can assume that the scattered studies of the field completed thus far represent a balanced and thoroughly validated body of experimental data upon which major applications of color coding can be founded.

1. Unidimensional vs. Multidimensional Absolute Judgments

For example, while the literature reports several studies of the number of absolutely identifiable colors, such as the study by Halsey and Chapanis in 1951 (23) and the study by Conover and Kraft in 1958 (15), it may be noted that only unidimensional absolute judgments were tested. Multidimensional absolute judgments generally remain, even today, largely untested and hence unknown as to the limits which human perception will place upon them. Clearly this is a matter of direct concern to color coding, and to the designers of visual displays who typically begin their coding with labels composed of digits or letters or combinations thereof and who now, with color coding becoming an added feature for distinguishability, are actually at this very moment dealing with multidimensional rather than unidimensional coding systems.

Perhaps one of the most provocative reports published in recent years touching on this area was the paper by G. A. Miller, "The Magical Number Seven, Plus or Minus Two: some limits on our capacity for processing information," published in 1956 (34). After tracing the studies of both unidimensional and of multidimensional absolute judgments he notes: "It seems by adding more dimensions, and requiring crude, binary, yes-no judgments on each attribute we can extend the span of absolute judgment from 7 to at least 150."

This position is quite different from that of Chapanis, Garner and Morgan in 1949 who postulated that "10 to 12 colors would be too many for most operators to keep in mind and use effectively" (12). Two years later Chapanis and Halsey seemed to have confirmed that postulate with a study "On the Number of Absolutely Identifiable Spectral Hues" (23).

Technically, there can be no complaint to the effect that Halsey and Chapanis were testing any but unidimensional absolute judgments. But the application of their research, clearly reflected in the government human engineering publications reviewed in the Preliminary Discussion section of this paper, has assumed that color coding must be designed as if it were unidimensional at all times. Yet, the moment a label is added to a color cue there is no longer a single, unidimensional factor for distinguishability. Controls and indicators identified by polychromatic labels are possessed of multidimensional cues, not unidimensional cues. Thus, the limitations of unidimensional absolute judgments no longer apply. Instead, the laws of multidimensional absolute judgments become the proper laws to apply.

At this writing it is doubtful if the human factors engineer can gather together enough validated experimental data to support a full set of laws for multidimensional controls and indicator identifying cues. He has studies which establish trends and indications of those laws, but not enough follow-up, experimentally validated conclusions upon which to proceed with authority. Miller reviews (34) two studies by Eriksen, "Absolute Judgments as a Function of Stimulus Range and Number of Stimulus and Response Categories", 1955, and "Multidimensional Stimulus Differences and Accuracy of Discrimination", 1955, noting:

> Eriksen found that when size, brightness and hue were all varied together in perfect correlation, the transmitted information is 4.1 bits as compared with an average of about 2.7 bits when those variables were varied one at a time. By compounding three attributes, Eriksen increased the dimensionality of the input without increasing the amount of input information. The result was a marked increase in the channel capacity for discrimination.

In a further penetration of the discrimination problems of coding displays, Miller points to the significant part played by the memory. (Use of the term "bit" derives from his likening the human observer to a communication system with certain channel capacities comparable to those of the electronic circuit. The "bit" is considered "the amount of information we need to make a decision between two equally likely alternatives".) He states:

> The span of immediate memory and the span of absolute judgment are quite different kinds of limitations that are imposed upon our ability to process information. Absolute judgment is limited by the amount of information. Immediate memory is limited by the number of items.
>
> In order to capture this distinction in somewhat picturesque terms I have fallen into the custom of distinguishing

> between <u>bits</u> of information and <u>chunks</u> of information. Then I can say that the number of bits of information is constant for absolute judgment, and the number of chunks of information for immediate memory is constant. The span of immediate memory seems to be almost independent of the number of bits per chunk, at least over the range that has been examined to date.

For the above to have full meaning, it is necessary to note that the formation of a "chunk" of information involves a recoding process, with "bits" of information first coded then renamed and memorized as a single "chunk". This permits storing in the memory single units of information so coded as to include quite large numbers of individual details, obviating the need to memorize the original inputs separately. Recoding thus constitutes an extremely powerful weapon for increasing the amount of information that we can deal with. In one form or another, we use recoding constantly in our daily behavior.

While Miller would seem to deprecate, perhaps, his treating the subject in electronic and computer terminology when he terms it "picturesque", the aptness of that terminology is perhaps more descriptive than most. Physiological psychology seems more certain than ever before that mental activity is essentially an intricate electronic process involving the principles of feed-back and go no-go impulses common to the modern-day electronic "brain". It is therefore a highly provocative and penetrating approach to pursue, but, as Miller points out, there is yet a good deal more to be discovered about recoding and memory. The factors dealt with by Miller in the preceding paragraphs are the same factors which were at work conditioning the errors and hesitations of the aircrew subjects tested for the controls coding techniques reported here. Ten years hence, when the profession has come to understand much more about multidimensional absolute judgments, it is likely that the comment offered now in explanation of the controls selection errors and hesitations performance will seem rudimentary, perhaps even primitive.

2. Photoreceptor Sensitivity

Although recent studies of the psychophysics of color perception have generally supported the tristimulus principles laid down by Grassman in 1853, several investigators have noted that the photosensitive receptors of the eye respond with more than tristimulus activity. The microphysiology of the retina has now become an area of study in which the photoreceptors are not nearly so well understood as in the days of Helmholtz and Grassman. One hundred years ago, early investigators in the psychophysics of vision were quite positive that the human retina was composed of only three major groups of receptors, one group sensitive to blue, one to green, and one to orange-red. But by 1950, when Morgan and Stellar brought out the second edition of their textbook, Physiological Psychology, there was a sufficiently large amount of research data as to confuse the issue and cause them to write:

> It is hard to say just how many different receptors might function in significantly different ways in color vision....Let us say, then, that there are at least four distinct kinds of

> receptors: the blue of 440 to 460 mu, the green of 520 to 540 mu, the yellow of 580 mu, and the red of 600 mu. Perhaps future research will give us grounds for further subdividing these groups, but for the present, these four appear to be the 'best bets'.

Five years later, some of the expected future research was under way and new information became available. One of the most provocative reports came from MacAdam (32) of the Research Laboratories of Eastman Kodak. After an extended series of experiments with chromatic adaptation, he offered this conclusion: "Apparently, the visual nervous system provides only three channels to the perceptual centers capable of handling only three independent responses for color perception. But each of these responses appears to be stimulated by a combination of two or more different photosensitive processes in the eye".

Again in 1956, a study conducted by Boynton (10) on the sensitivity functions of human color vision provided experimental results indicating support for a 4-component theory of color vision.

While recognizing the very severe research problem posed by the microscopic photoreceptors, with some 125 million rods and 6.8 million cones per human retina, present knowledge has grown considerably beyond the scope envisioned by Grassman and Helmholtz. The relatively recent techniques of perimetry and studies of the retinal color zones have both increased our comprehension of color vision and complicated our color theories as we add to the record temporal patterning and subjective hue responses. The trireceptor law "does not adequately predict corresponding colors for different chromatic adaptations. The discrepancies between the predictions of that law and the observed colors seems to be systematic", points out MacAdam (32).

Without carrying this commentary deeper into the photoreceptor question, it can be seen that there are important new fundamentals to be validated as to the manner in which color perception takes place, and these fundamentals are of immediate concern to the designer of color coding.

3. Luminance of Equally Bright Colors

A matter of simple practicality devolves from the problem of color choice when designing color coding: how should the balance of appearance or equality of color aspect be determined?

Until relatively recent investigations were made known, it was assumed that the CIE method of plotting equally bright colors with a polar coordinate system which permitted spacing equal hue differences by equal angles and equal saturation differences by equal radial distances would allow ready selection of balanced colors. If the colors chosen were equally spaced in the CIE chromaticity diagram, they were assumed to have equal steps of color differences on either side and therefore were possessed of maximal distinguishability.

In 1955, Chapanis and Halsey published the results of their study (13) which investigated the fact that equally spaced colors of mathematically equal luminance do not always appear to have equal brightness. They cited work of MacAdam which demonstrated that as color mixtures increase in saturation, less luminance is required for the mixtures to appear equally bright.

To test this phenomenal fact in an effort to determine if there were systematic differences between perceived balance and the CIE chromaticity diagram derived balance of colors, Chapanis and Halsey employed a set of 50 color samples produced by transilluminating standard Corning and Wratten color filters so chosen as to match the balanced CIE figures for equal spacing and equal luminance. To achieve true psychological balance wherein all 50 colors appeared equally bright, the experimenters added masks of neutral film having graded, measured degrees of density.

Their conclusions: "With the exception of a reversal in the area of the yellow region, it appears that for colors of equal brightness, saturated colors require less luminance than desaturated ones."

When Hurvich and Jameson published the results of the fourth phase of an extended research conducted at Eastman Kodak under the title, "Some quantitative aspects of an opponent-colors theory" in 1956 (27), they also included findings to the effect that "equal distances in the CIE chromaticity diagram do not represent equal psychological differences." To insure their meaning would not be misconstrued, they defined their use of the term "psychological" as denoting "perceptual... thus psychological color differences are changes perceived in terms of brightness, hue and saturation."

Findings of this nature at once pose the question, if the CIE chromaticity diagrams cannot be used to formulate choices of colors equally spaced as to hue, brightness, and saturation, what sources can be employed with confidence by the color coding designer when specifying controls and indicator polychromatic cues?

At this writing it would appear that as yet, no new standard has been established, suggesting that additional research is required before the most suitable corrections can be agreed upon. Since the matter of color perception is inherently characterized by a subjective element, with the attendant variations due to individual differences to be reconciled after proper testing of large population samples, it cannot be assumed the task of providing the necessary new color standards will be either short-termed, inexpensive, or easily accomplished.

Meanwhile, the human factors engineer is obliged to achieve psychological color balance as best he can using his own devices for testing and controlling the differences between hue, brightness and saturation.

4. <u>Absolutely Identifiable Spectral Hues</u>

While it would be reassuring to accept the statement that "recent experimental evidence has provided an estimate of the number of colors the average person can discriminate absolutely" written by Conover and Kraft (15) and published in WADC TR 55-471, October 1958, an examination of the three reports cited therein as evidence proves the matter is less than definite, with some important variables untested and some only partially investigated.

Two of the three studies listed by Conover and Kraft for support of that statement are actually concerned with the same experimental data, namely, results obtained by Halsey and Chapanis in 1951, first published under the title, "On the number of absolutely identifiable spectral hues", (23). Five years later, the same experimental data was used in a short report entitled "Absolute judgments of spectrum colors", with Chapanis taking his turn for being listed first, hence the entry appeared as Chapanis and Halsey. Since the data of the 1956 report did not change or augment the 1951 findings, there are no grounds for listing the later report as offering further evidence on the question.

The third study reported is thus actually only the second study of the subject listed, itself the work of co-author Conover. Briefly described, Conover's procedure was as follows.

Ten color-normal observers were used, the experimenters employing a set of 25 Munsell maximum saturation color patches. Subjects viewed the color patches through a 3° aperture of a neutral mask, with 21.4 ft. c. of 6800° illumination reaching the color under test. Presentation of the 25 colors was made in random order to total 75 trials, the task being to identify the color with the number learned as the code during practice trials.

Results showed that the number of absolutely discriminable hues may range from 5 to 16, depending upon the individual tested. Under idealized viewing conditions, half the observers tested could discriminate, without appreciable error, 9 maximally saturated surface colors. Therefore, if operators are required to make absolute identifications of colored code symbols with an arbitrarily small number of errors, the code must be usable by somewhat more than half the population. If colors are to be recognized on the basis of hue alone under a limited range of viewing conditions, then 8 surface colors would appear to be the maximum number of color symbols that should be used; a smaller number (5,6, or 7) will probably be necessary for most practical purposes.

While the Conover and Kraft report does not specifically so state, it is likely that the Munsell maximum saturation color patches were not only matched for saturation but also for value, hence varied as to hue alone.

If this were the condition, then Conover's subjects were making unidimensional judgments, not multidimensional ones. In the light of Miller's penetration of the problems of human perception and the capabilities thereof in making discriminations when more than one dimension of difference is present (34), it seems clear that Conover's experimentation should be considered only one phase of several required to determine the number of absolutely identifiable colors. His work, like that of Halsey and Chapanis (23) seems to have tested only the number of absolutely identifiable colors distinguishable when differences were unidimensional.

Thus, it would appear that before human factors engineers can be advised on the true limits of absolutely identifiable colors, further investigations are required wherein the colors vary not only for hue, but for brightness and for saturation as well.

The work of Halsey and Chapanis reported in 1951, "On the number of absolutely identifiable spectral hues" (23) involved use of four sets of colors, matched for value and saturation (or intensity), comprised of 17, 15, 12 and 10 colors respectively. Their results were:

17 hue series: 72.4% identified correctly

15 hue series: 94.8% identified correctly

12 hue series: 96.0% identified correctly

10 hue series: 97.5% identified correctly

A significant comment included in that report contains indication of the benefit which multidimensional differences might add to the discriminability of colors. As long as 1951 Chapanis and Halsey stated: "It is possible that if intensity (saturation) were allowed to vary throughout the series, more hues would be discriminable according to our criteria."

In 1954 the same investigators published the results of another study dealing with color discriminability under the title, "Chromaticity confusion contours in a complex viewing situation" (24).

Again holding the differences to unidimensional variations of hue only, they employed 58 standard colors distributed throughout the CIE constant-luminance chromaticity diagram. Colored lights were the media, the test display carrying a total of 171 lights surrounding a central light which was the standard to be matched. There were 80 variations of colored light test displays, and every standard was compared with every other standard twice. A complete set of observations for any subject required 12 to 24 hours of testing. (This suggests there was no requirement that the subjects complete their matching in the shortest possible time).

Results revealed that the maximum number of color standards found at the 10% level of discriminability was 15, but the authors note the theoretical maximum may be 18. At the 20% level, 19 standards were selected.

At the 5% level, 11 nonoverlapping color standards were discriminated. Again, the investigators believe that if more complete testing were conducted, 14 colors would prove to be discriminable.

In summary it can be stated:

1. Experimental testing of the number of absolutely identifiable colors seems to have been limited to unidimensional differences of hue while holding brightness and saturation constant. Further testing of colors with differences of hue, brightness and saturation is required before the full scope of color discriminability can be established.

2. Human factors engineers must keep in mind that the limitations on the number of colors permitted color coding controls and indicators by most government handbooks and specifications are limitations reflecting research of unidimensional color testing. If colors used for coding vary in hue only, then the limitations found by Halsey and Chapanis and by Conover are probably valid.

3. Since equally bright colors often appear unequal in saturation, although they are scientifically matched for brightness and saturation according to the CIE constant-luminance chromaticity diagrams, adjustment of both brightness and saturation may be required to accomplish psychologically balanced hues. Thus, in selecting color for coding, empirical testing for balance is required.

5. Reaction-Time Considerations

Since the response to a command or situation which requires immediate action by the operator is a matter of reaction-time speed, certain fundamental precepts concerning the speed with which human responses can be expected to function become precepts of concern to the color coding designer.

Previous comment has dealt with the precepts and limitations of color perception, but an almost separate set of limitations and capabilities become important when speed of perception and speed of motor response are factors of prime importance. Under the general heading of reaction-time, human factors engineers may find a sizable body of experimental data bearing directly on the operation of controls and response to indicators, data that support certain laws of human reaction essential to human factors engineering.

One of the most comprehensive studies of reaction-time problems is that by Teichner (42), published in 1954, who surveyed a wide field of RT reports and provided his paper with a bibliography of 163 references. Limiting the study to "simple reaction time", his work includes the following findings.

Most RT studies have found reaction time faster in response to sound stimuli than to other stimuli. Most studies have also found faster RTs with tactual than with visual stimuli. A table by Robinson shows median RTs compared thusly: audition RT 0.142 msecs., touch RT 0.155 msecs., vision RT 0.194 msecs.

Combined stimuli, such as light, sound and electric shock, if presented simultaneously were found to elicit faster RTs than if the stimuli were presented singly. Successive stimulation of different sense modalities produced slower RTs than when the stimuli were present singly.

Among right handed subjects, it was found that RT for the left hand was slower than for the right hand. Again, the sum of RTs for each hand was larger when the S. tried to respond with both hands at once than when he moved them successively.

Teichner comments that "the cerebral factors involved have not been studied in a way to make them applicable to this discussion. ...The condition of the receptor and of the responding member are still among the conditions which have received insufficient attention."

Another phase of RT research has been concerned with the psycho-physical relationship between stimulus intensity and reaction time. For instance, a study by Bartlett and MacLeod (7) undertook to establish relationships between RT and signal and field luminance, for both foveal and peripheral stimulation over a wide stimulus range under precisely controlled conditions.

The investigators tested two major experimental conditions: field luminance level held constant while flash luminance was varied, and, flash luminance fixed while field luminance was varied. Results of testing variable flash at three field parameters confirmed Lemmon and Geisinger's finding that peripheral stimulation yields shorter RTs for a dim flash and also an agreement with the data of Galifret and Pieron who found, in the case of a relatively bright flash, that foveal latencies are shorter than peripheral.

Results of testing variable field at three flash parameters revealed that the latency-log field functions are characteristically flat, exhibiting little or no increases in RT over log luminance ranges where the field is clearly discriminable from the flash. Eventually, however, RT increases precipitously, becoming very large, as the field luminance approaches a level where the flash becomes indiscriminable. The study also states:

> Two trends may be noted in any one family of functions, as the flash parameter is decreased from 1500 to 0.15 mL: latency values become increasingly elevated over the flat portion of

> the curve; and the threshold upturn of each function occurs sooner, i.e., at a smaller field luminance value. As the flash is moved from the fovea to the periphery, the same two trends are in evidence, i.e., an elevation of latency, and threshold upturn at a dimmer field level.

This brief note on the problem of stimulus intensity provides indication of a key factor affecting the efficiency of color coding for controls and indicators. Intensity of front illumination for surface colors combined with the factors of saturation and brightness of the colors themselves may be recognized as inseparably interrelated considerations for the color coding designer. Still another kind of stimulus intensity factor appears with use of transilluminated indicators. Where the operator is given a wide console to monitor, with indicators to be watched not only by the foveal areas of the eye but by the peripheral as well, findings such as those reported by Bartlett and MacLeod become directly applicable. Indeed, careless planning of the illumination of control panels may result in sharply increased numbers of operator errors particularly with color coded controls.

6. <u>Search and Scanning Thresholds</u>

While observing the performance of pilots and ROs during the color coding tests reported herewith, this investigator was brought up against the problems posed by the requirement to scan and search with maximal speed for the target control and contact it without making an error or scoring a slow RT due to hesitation.

From the outset, beginning with the practice trials, it became apparent that an operator must discover his scan and search speed threshold before he can settle down to an error-free and hesitation reduced performance. In every case, when the subject exceeded his scan speed threshold his eyes would sweep past the target push-button, its presence seen but not perceived. The result of the sweep-past was either an error or a hesitation.

For example, the performance of RO K. as shown in Table 13 reveals him to be the only subject who scored no errors for any of the three cue types during the final 20 trials. Indeed, his full-length record contains only one error during 180 trials. Accordingly it might be concluded that he was the most reliable of all operators tested. Study of his overall performance, however, shows that his RTs were the slowest, his hesitations score was next to greatest, and his mean hand travel RT was slowest.

In the operation of hypersonic aircraft, manned missiles and satellites, a hesitation could be tantamount to an error. Thus, if crews were being selected for such vehicles, RO K. might be a far poorer choice than several others of the fighter squadron tested. Preferable would be the men of Group III who scored some errors but whose hesitation scores were lowest, ranking them 1st, 2nd and 2nd, while their performance mean RTs were fastest, ranking them 1st, 2nd, and 3rd.

Ideally, the best operators are those who learn early what their individual scan speed thresholds are and never exceed them. What seems to be needed, however, is a better method for teaching operators how to determine their scan speed thresholds. Under the conditions employed in this experiment, it was obvious that some subjects learned to sense their scan speed limitations faster and more effectively than others, but the awareness was much more a matter of feeling, almost instinct perhaps, rather than of scientific timing.

Looking at the problem from the human factors engineering viewpoint, it is evident that the planning of hypersonic and faster vehicular control requires accurate knowledge of human scan speed norms so as not to build control instrumentation that requires responses faster than operators can be expected to deliver with accuracy. Again, it is also a training command concern, with the need to avoid training programs that demand several hundred hours in the attempt to develop maximal response speeds at control stations.

The need, it would seem, is to investigate with use of precision timing devices how fast scan and search with perception can be. At least two methods are reportedly being used in this direction: the Miles technique of oculomyography, and, the high-speed motion picture cameras recently developed for missile firing documentation capable of precise, calibrated images recorded at speeds up to 2000 frames per second. It is hoped that the results of these studies now in progress by the USAF and the Navy will be published soon.

7. Integration of Human Engineering and Training Research

In a Journal of Aviation Medicine article, Klein and Gall (30) point out that integration of human engineering research with training research is necessary, even mandatory, if the operational efficiency of control and indicator panels is to be insured. While a special study probably would be required to determine how much or how little integration of the kind Klein and Gall call for actually exists, the fact that they, in their positions, were impelled to make the plea would indicate the need is real.

While some of our more advanced engineering teams may not have ignored training command problems in their human factors engineering, it is probably true that the majority of control and indicators panels are put into production without benefit of advice from training specialists. (Indeed, it is doubtful if 50 percent even have benefit of human engineering advice.)

Looking ahead to the time when color coding has become a standard feature of visual displays, it seems certain that training research should be regularly part of the color coding designer's concern as he tries to avoid introduction of coding systems too difficult or too complex to permit reasonably short training programs being accomplished.

Perhaps the ideal human factors engineering team will prove to be one having a training specialist as a regular, permanent member, part of his assignment being to insure training research data was promptly made available to human factors engineering and to insure that human factors research pertinent to training problems was made known to training commands.

One suspects that training is still considered by many top planning bodies as a service operation whose only mission is to take whatever equipment is bought, then teach personnel to operate it as best they can. Human factors engineers also may be prone, sometimes, to ignore training difficulties which control and indicator panels present, for example.

Considering human factors engineering in its fullest range of functions and responsibilities, is not a basic part of the job aimed at anticipation of training as well as of operational problems? If a control panel is given so great a complexity of design that it threatens to require a type of monitoring and motor response bordering on the upper limits of operator capabilities, is not that panel also apt to require a training program overlong and overstrenuous?

Analysis of the visual display while still on the drafting board, scrutinized both by human factors engineers and training specialists, would appear to be the preferable procedure. In this event, training difficulties as well as operational difficulties would be discussed, tested, and redesign instituted on the advice of specialists before the display went into production.

In short, it can be considered only good practice improving the state of the visual display art if integration of human factors engineering and training research is made a routine requirement. Since effective training techniques can be easily negated by ineffectively designed equipment, poor learning or slow learning is not necessarily a fault of the training program. Indeed, if follow-through by engineers were to include observation of their equipment effectiveness during training, it is likely that many items giving trouble during operational use could be caught in the training period and redesigned before production had been completed and delivery made on the entire allotment.

Color coding, like many other visual display features, will prove to be well designed or poorly designed only when put to the operational test. Follow-through observation of color coding effectiveness during training could well prove an important part of the overall checkout given the new display. Testing in the engineering lab may produce quite acceptable results, yet the same display prove difficult or slow when actually in used by operator personnel during training.

As a general rule it could be said that if a visual display proves difficult to master by operator personnel during training, the display needs redesign by human engineering and training specialists. In view of the now severely reduced lengths of time permitted the operator during flight to make command decisions and adjust the correct series of controls without error or hesitation, there is more need than ever before to so design the controls and indicator panels as to avoid any layout, coding, illumination, or anthropometric features which can cause operator confusion or hesitation however slight.

In 2 seconds, the average time required to observe an impending auto crash and push the brake that will avoid the accident, the 2000 mph aircraft travels more than one mile, while the manned satellite moving at orbital speed of 18,470 mph, travels more than 10 miles in the same 2 seconds.

This means that the training commands must so perfectly instruct the personnel who will man these hypersonic aircraft and satellites they will become aware of the significant danger which hesitations as small as 50 milliseconds could make. Accordingly, it will be essential that control and indicator panels be so carefully engineered as to layout, coding, dials, knobs, switches, anthropometrics and illumination that no operator hesitations or errors can be traced to poor design. It would seem clear, therefore, that designers of all kinds involved, including human factors engineers, would need the benefit of training specialists' consultation during design of the visual displays of these fastest of manned vehicles.

C. FURTHER STUDIES INDICATED

1. The Big Picture

Before proceeding with the several necessary further studies indicated, it seems to this investigator that a first essential would be the completion of a big picture, comprehensive survey of the entire field of color coding, with careful evaluation made of whatever research has been done that bears either directly or indirectly on this field.

Systematic analysis is needed of all experimental work undertaken to date, the end purpose being to evaluate each study as to the logic of procedures and basic assumptions, with consideration of possible flaws in its experimental design, and finally to reassess the validity of its conclusions. The report resulting from such a survey would provide the profession with probably the first comprehensive overview of the color coding field. It would allow new assessment of the presently disconnected and intermittent group of studies which comprise the presently known scope of research on color coding. It would provide visibility of the gaps, overlap, and fringe areas which require additional experimental work in order to insure a full spread of integrated research in color coding to be completed.

So far as this writer has been able to determine, there has not been instituted as yet any long range, comprehensively designed program of experimental research on color coding of controls and indicators which would thoroughly explore all the problems of psychophysics, color perception, reaction time, scan threshold, multidimensional absolute judgments, photoreceptor sensitivity, luminance of equally bright colors, the number of absolutely identifiable colors, and other ancilliary problems which must be encompassed by the color coding designer. If the paragraphs on color coding published in the ARDC Publications (1), (2), (3) and (4), are an index to what the profession now officially recognizes on the subject, then the field is in what amounts to desperate need of the big picture survey outlined.

Lacking that comprehensive survey there is no scientific way to determine which further studies of color coding should be undertaken first, no way to be certain how large the gaps visible at this writing actually are, no way to insure that the work going forward in neighboring fields such as vision and perception, reaction time et al includes study of the particular facets of those fields bearing directly on the problems of color coding visual displays.

2. Immediate Follow-up Indicated

Since the broad, big picture type of survey is one that may not be commissioned until a good deal of preparatory work has been spent in conferences with perhaps several high level agencies, involving time, travel and money, there are certain follow-up studies indicated by the testing results obtained in this experiment which could be placed on the immediate list without much risk of robbing the big picture program that may one day be instituted.

Since the objectives of these immediate follow-up studies have already been discussed in some detail, the notations which are offered here may be brief, amounting to little more than tabulations which would require detailed amplification before actual research would logically begin:

a. Repeat the presently reported color coding of controls experiment using a large population sample, testing the results obtained with the initial small sample. Experimental apparatus for the large sample tests should employ mechanized changing of cue tabs and electronic recording components that would allow both immediate read-out of RT and permanent recording of the data without need for RT notations by the experimenter.

b. Repeat the command word-color association experiment with large sample populations to determine if other word-color associations should be added to the list of the four noted herein, (Section II, A5).

c. Initiate a series of studies which would explore the number of absolutely identifiable colors resulting from varying hue, brightness and saturation.

d. Initiate a study comparing the discriminability range of transilluminated colored lights and indicators with the discriminability range of surface colors.

e. Initiate a study of control and indicator panel designs employing color coding for systems as well as for individual knobs, controls dials, switches, and indicators.

f. Initiate a study of the optimal size of color tabs used for coding controls and indicators.

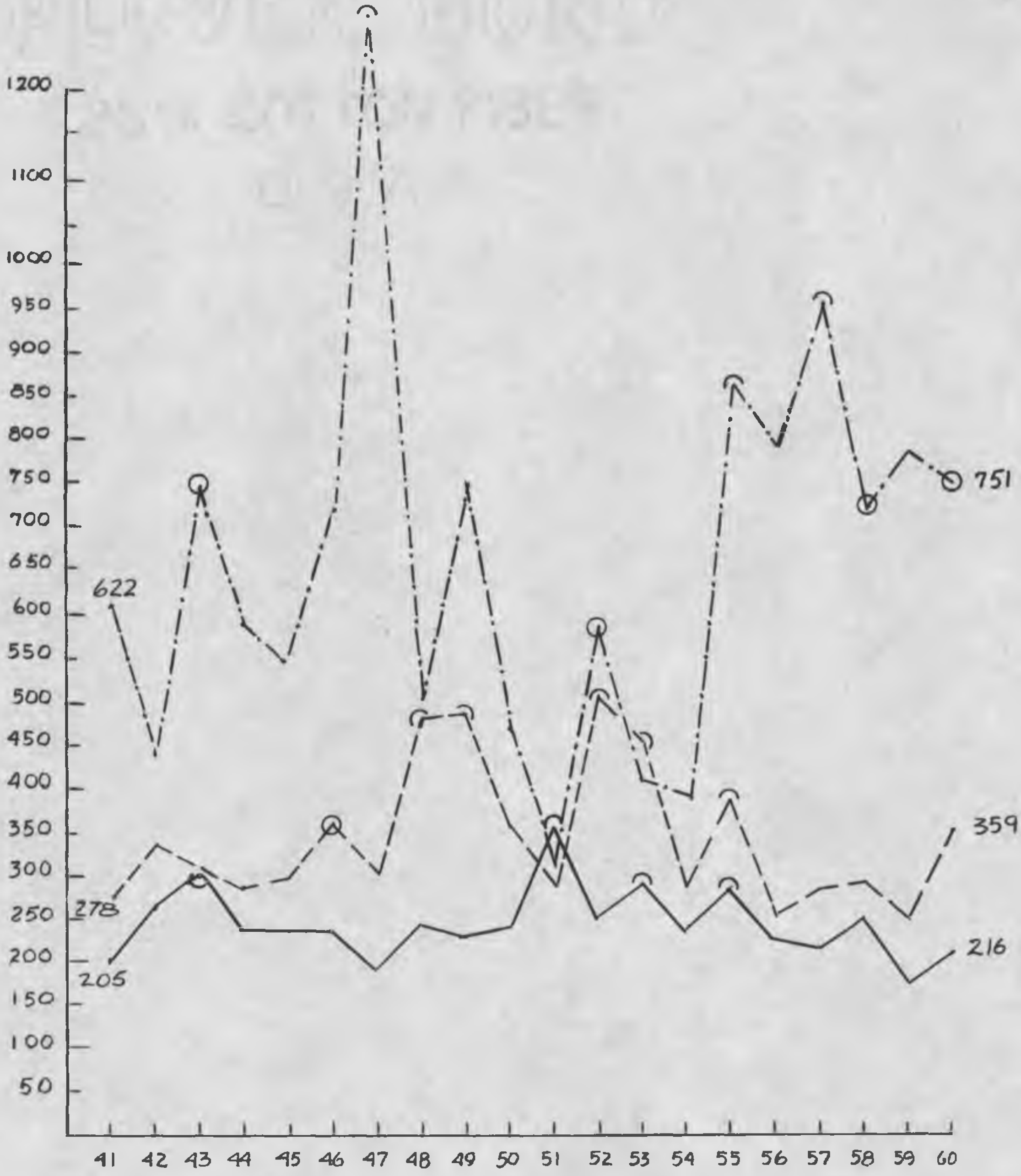
g. Initiate a study of errors and hesitations, using motion-picture or television cameras to record testing of subjects during the repeat experiment of the color coding of controls noted in (a) above.

h. Initiate a study of specific areas and levels on which training research and human factors research affecting the operational use of control and indicator panels, can or should be integrated.

VI. SUMMARY

Subjects using control pushbuttons coded with monochromatic, labeled cue tabs were timed electronically for search RTs, trials being numerous enough to demonstrate maximal speed for search, perception and motor response in hitting target buttons. Apparatus design prevented subjects learning target button location by position, the cue tab position being randomly varied for each trial. Two additional series of tests were run using polychromatic cue tabs, one series having labels, the other series having no labels. Results showed selection mean RT was 54.6% faster with polychromatic, labeled cue tabs than with monochromatic labeled cue tabs, and 59.7% faster with polychromatic unlabeled cue tabs. Mean improvement in speed when polychromatic cue tabs were used amounted to 57.2%. Hesitations and errors analysis showed that use of polychromatic cue tabs reduced the mean number of hesitations by 27.3%, while the polychromatic coded cue tabs reduced the mean number of errors by 286%.

APPENDIX



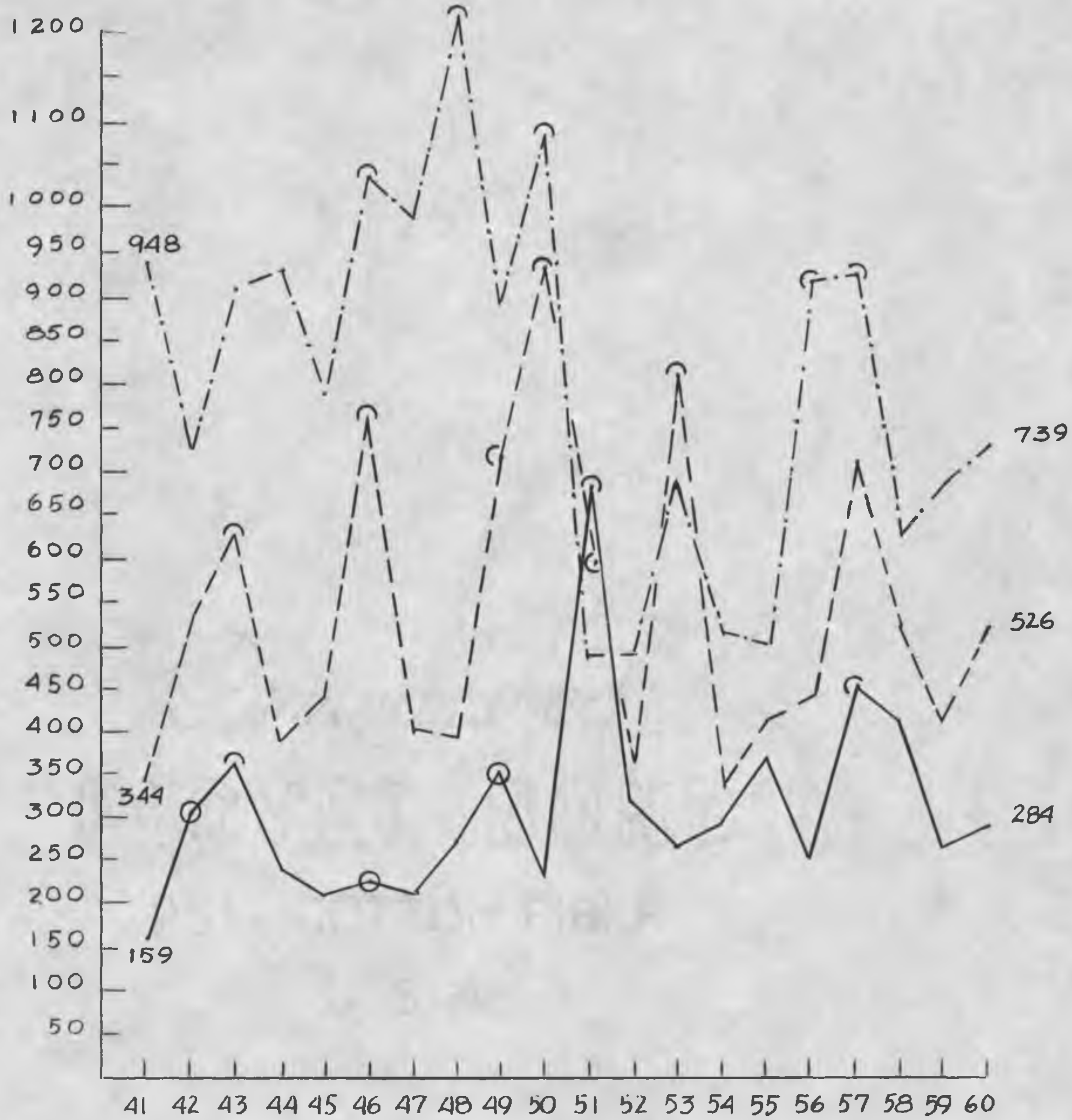
<u>Group I Subject:</u> 6' 1", 25 yrs, career, married

<u>Order of Testing</u>

<u>and Symbols:</u> Mono-L (-·-), Poly-L (---), Poly no-L (——)

<u>Error:</u> ⊙ <u>Hesitation:</u> ◠ ◡

Figure 6. PILOT Hi. ACTUAL SCORES: Final 20 for Each Cue Type



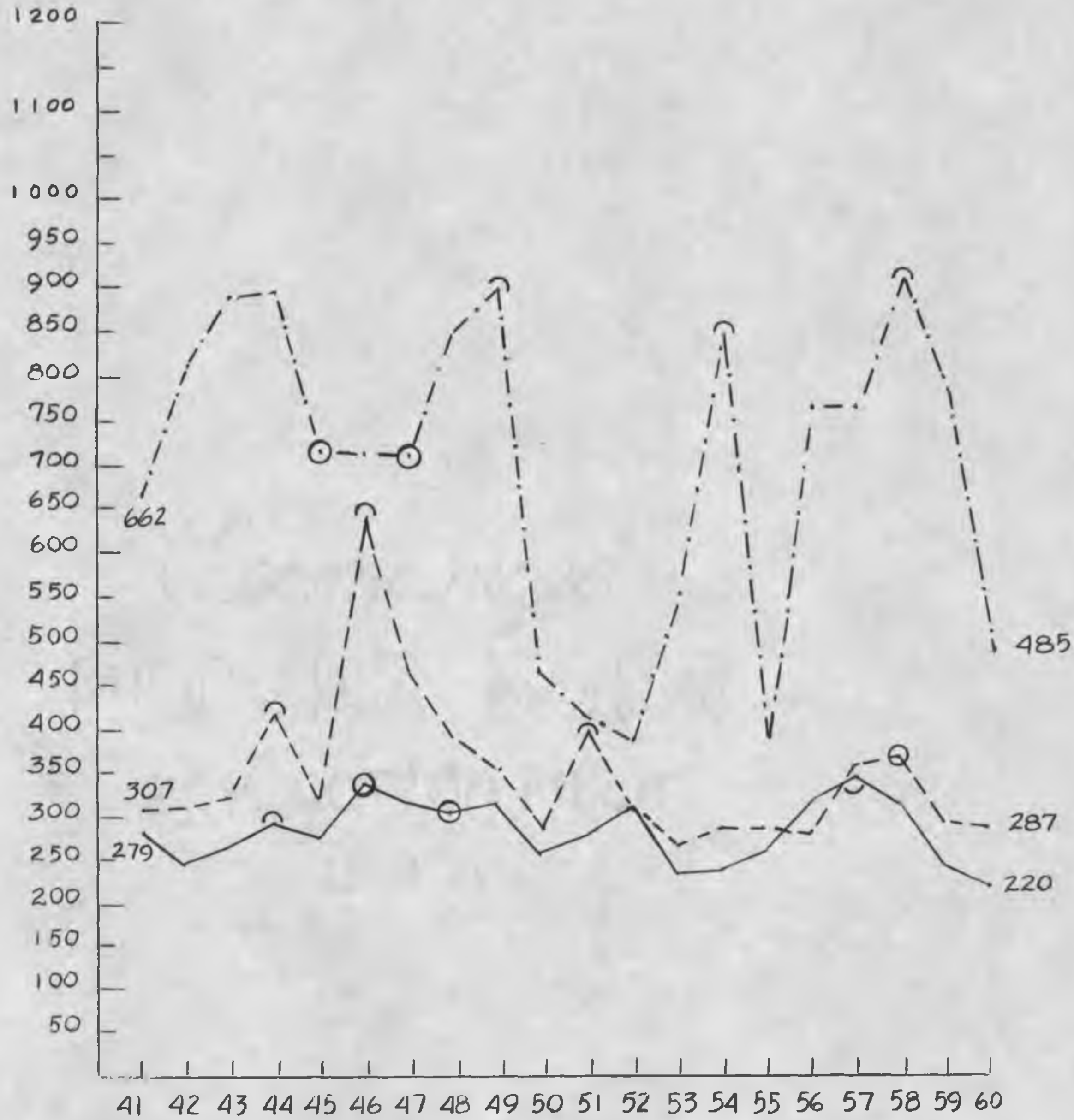
Group I Subject: 5' 11", 25 yrs, non-career, married

Order of Testing

and Symbols: Mono-L (–·–), Poly-L (– – –), Poly no-L (——)

Error: ⊙ Hesitation: ⌒ ◡

Figure 7. RO N. ACTUAL SCORES: Final 20 for Each Cue Type



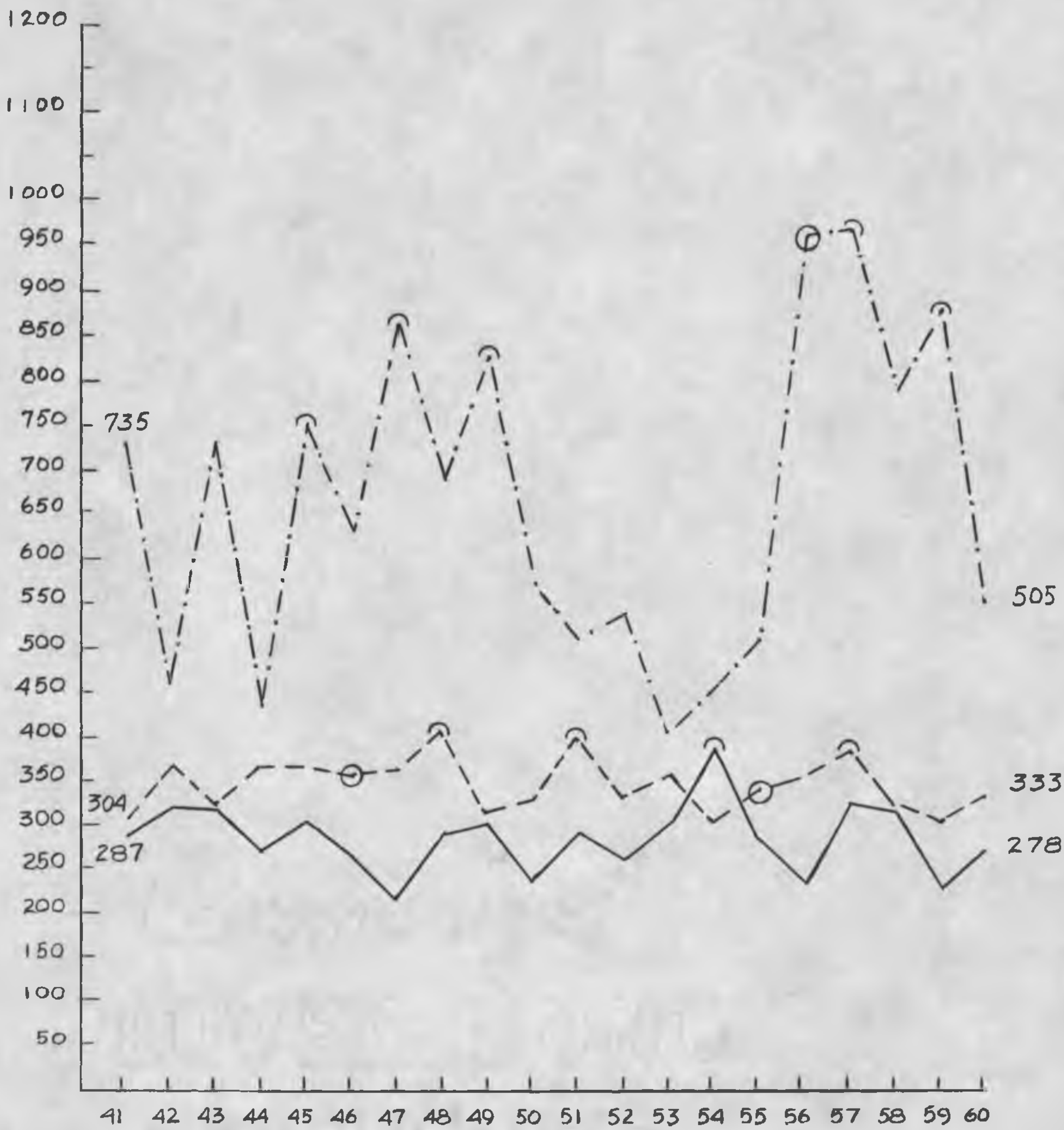
<u>Group I Subject:</u> 5' 8", 25 yrs, non-career, married

<u>Order of Testing</u>

<u>and Symbols:</u> Mono-L (-·-), Poly-L (---), Poly no-L (——)

<u>Error:</u> ⊙ <u>Hesitation:</u> ◠ ◡

Figure 8. RO T. ACTUAL SCORES: Final 20 for Each Cue Type



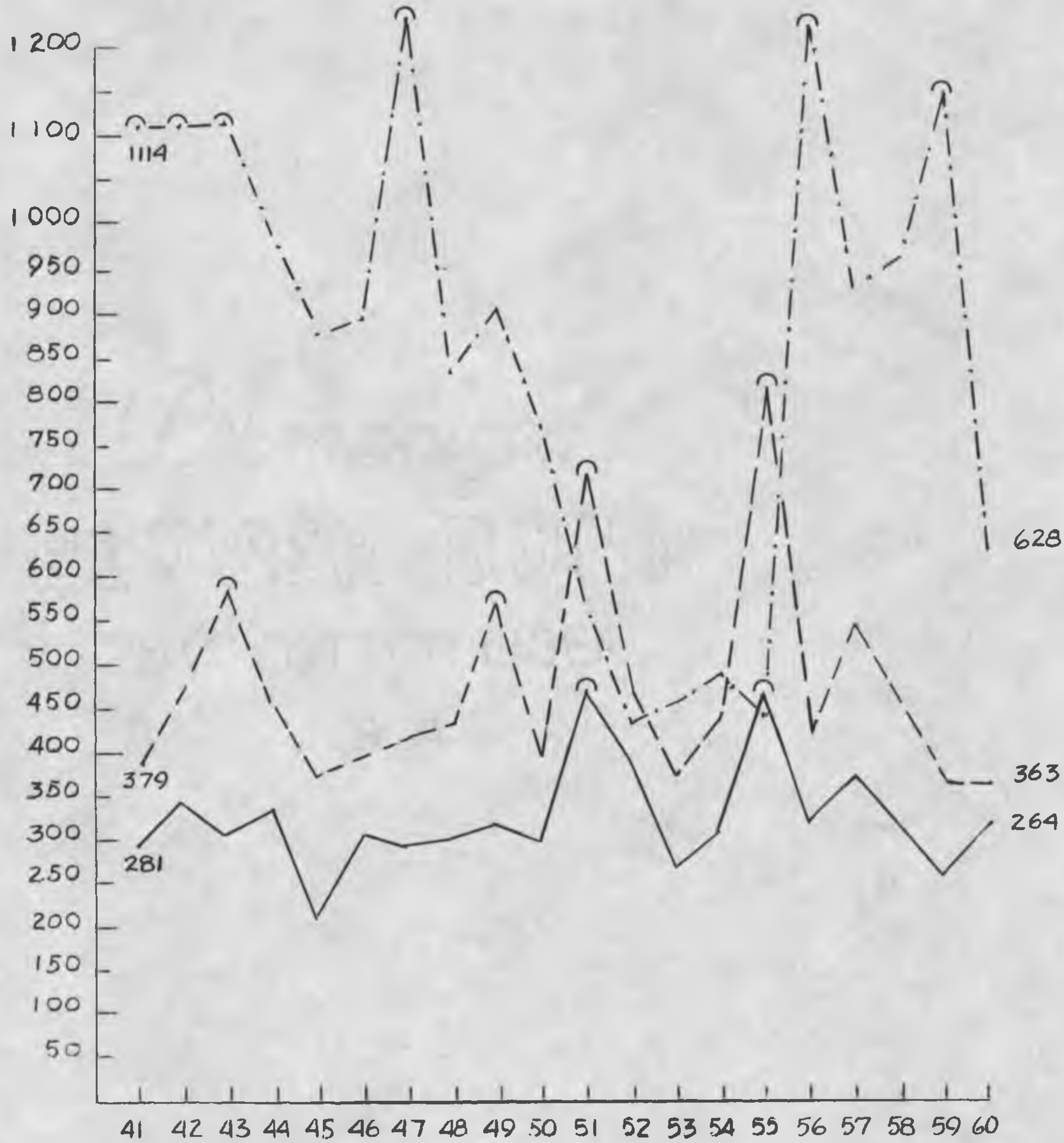
<u>Group II Subject:</u> 5' 10", 26 yrs, career, married

<u>Order of Testing</u>

<u>and Symbols:</u> Poly-L (- - -), Poly no-L (——), Mono-L (-·-)

<u>Error:</u> ⊙ <u>Hesitation:</u> ◠ ◡

Figure 9. PILOT Ha. ACTUAL SCORES: Final 20 for Each Cue Type



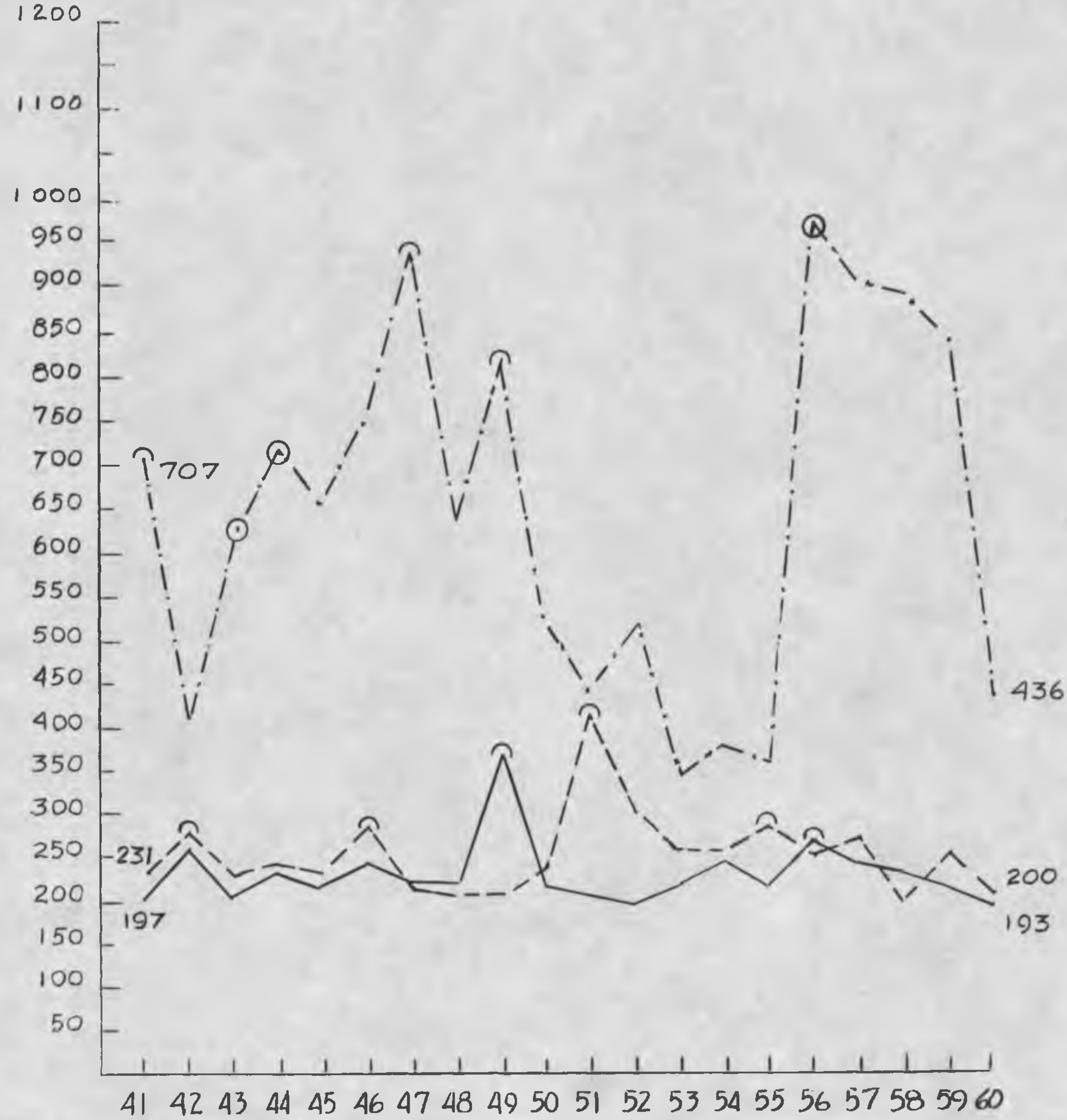
Group II Subject: 6' 0, 25 yrs, non-career, married

Order of Testing

and Symbols: Poly-L (---), Poly no-L (——), Mono-L (-·-)

Error: ⊙ Hesitation: ◠ ◡

Figure 10. RO K. ACTUAL SCORES: Final 20 for Each Cue Type



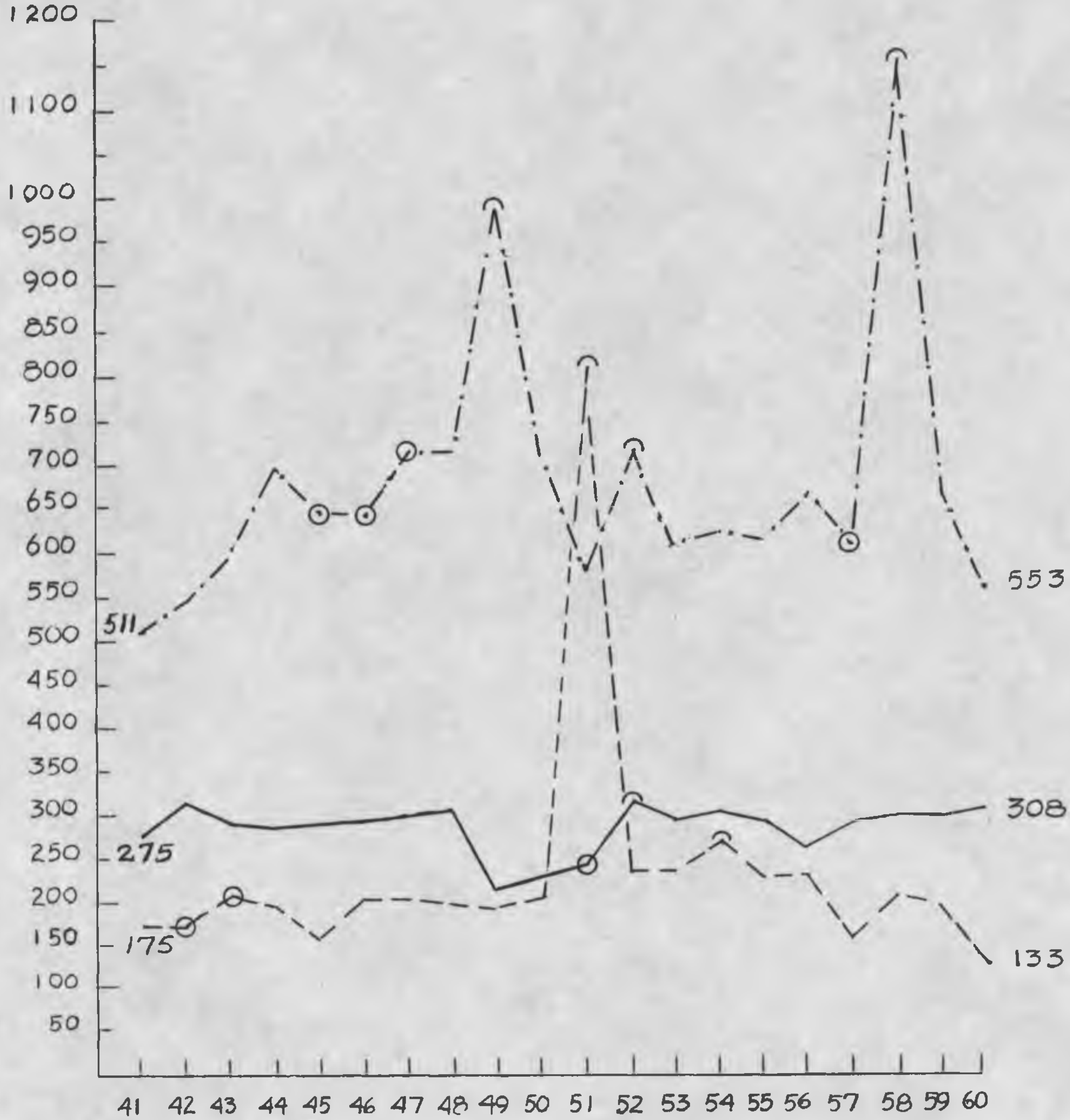
Group II Subject: 6' 0, 22 yrs, career, married

Order of Testing

and Symbols: Poly-L (---), Poly no-L (——), Mono-L (-·-)

Error: ⊙ Hesitation: ◠ ◡

Figure 11. RO L. ACTUAL SCORES: Final 20 for Each Cue Type



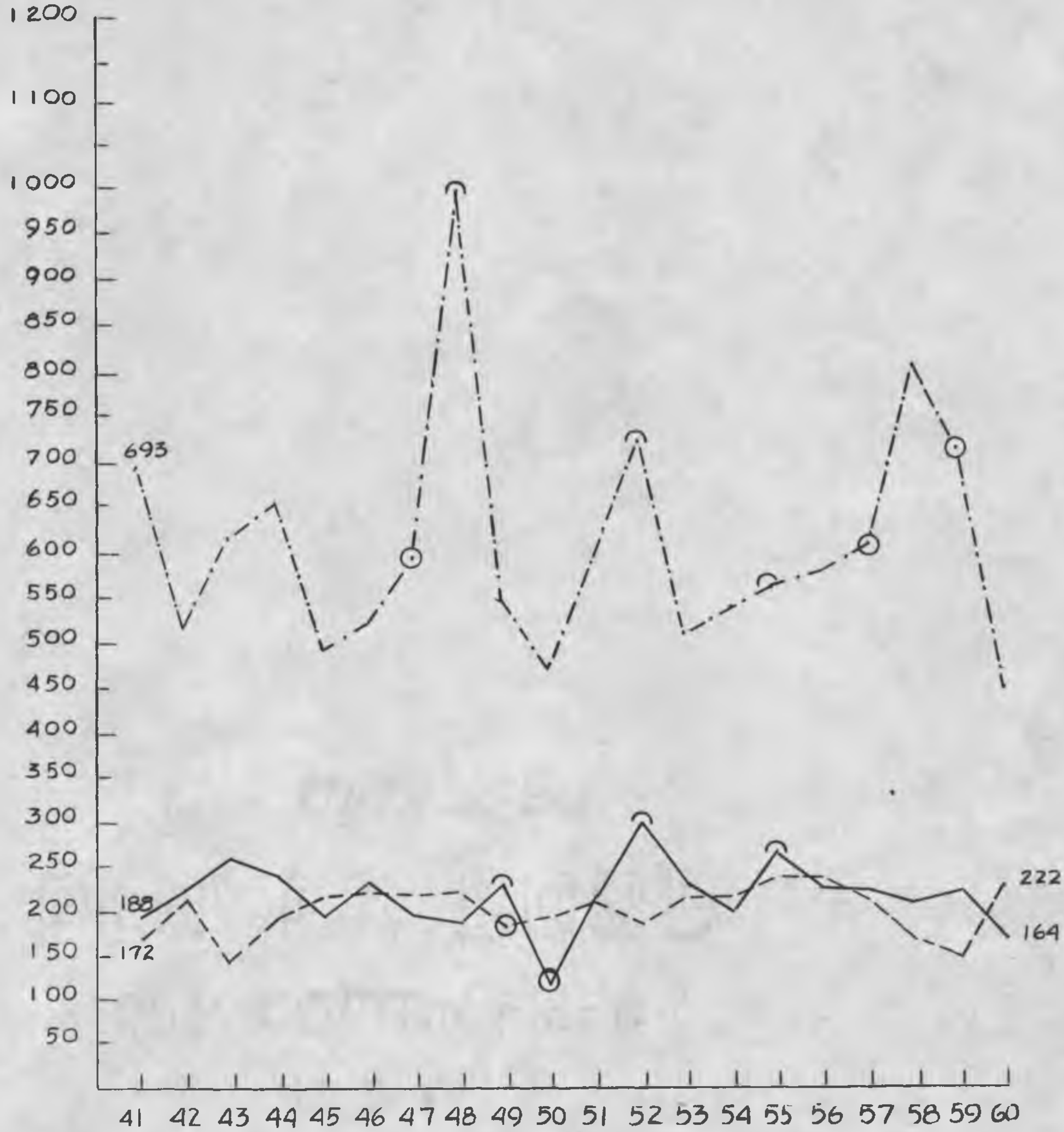
Group III Subject: 6' 0", 26 yrs, career, married

Order of Testing

and Symbols: Poly no-L (——), Mono-L (-·-), Poly-L (- - -)

Error: ⊙ Hesitation: ⌒ ◡

Figure 12. PILOT S. ACTUAL SCORES: Final 20 for Each Cue Type



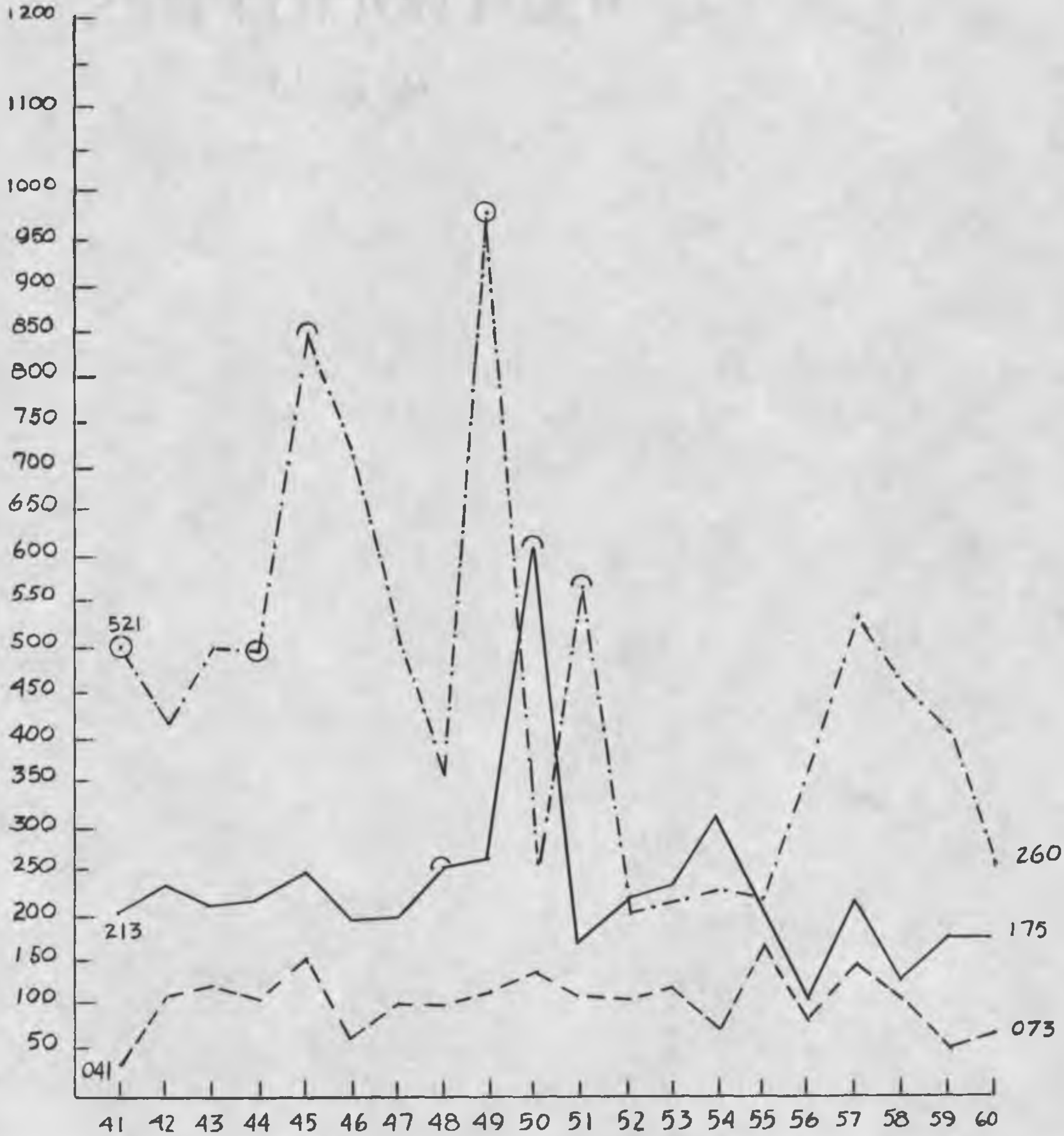
Group III Subject: 5' 10", 22 yrs, career, married

Order of Testing

and Symbols: Poly no-L (——), Mono-L (-·-), Poly-L, (---)

Error: ⊙ Hesitation: ⌒ ◡

Figure 13. RO M. ACTUAL SCORES: Final 20 for Each Cue Type



<u>Group III Subject:</u> 6' 1", 25 yrs, career, married

<u>Order of Testing</u>

<u>and Symbols:</u> Poly no-L (——), Mono-L (—·—), Poly-L (---)

<u>Error:</u> ⊙ <u>Hesitation:</u> ⌒ ◡

Figure 14. PILOT Mc. ACTUAL SCORES: Final 20 for Each Cue Type

BIBLIOGRAPHY

1. AFBMD, Human Engineering Design Standards for Missile System Equipment, AFBM Exhibit 57-8A. Air Force Ballistic Missile Division, ARDC. 1 Nov 1958.

2. AFBMD, HIAD, Handbook of Instructions for Aircraft Designers (Guided Missiles), ARDC Manual 80-1, Vol. II. Air Force Ballistic Missile Division, ARDC. 1 March 1958.

3. AFBMD, HIGED, Handbook of Instructions for Ground Equipment Designers, ARDC Manual 80-5. Air Force Ballistic Missile Division, ARDC, 1 August 1957.

4. AFBMD, HIAGSED, Handbook of Instructions for Aircraft Ground Support Equipment Designers, ARDC Manual 80-6. Air Force Ballistic Missile Division, ARDC, 15 August 1958.

5. ALLUISI, E.A., & MULLER, P.F. Jr. Rate of Information Transfer with Seven Symbolic Visual Codes: Motor and Verbal Responses, WADC TR 56-226, May 1956.

6. BAKER, C.A. & GRETHER, W.F. Visual Presentation of Information, WADC TR 54-160, August 1954.

7. BARTLETT, N.R. & MACLEOD, S. Effect of flash and field luminance upon human reaction time. J. Op. Soc. Amer. 1954, 44, 306-311

8. BARTZ, A.E. Attention value as a function of illuminant color change. J. Exp. Psychol. 1957, 41, 82-84

9. BORING, E.G., LANGFEID, H.S. & WELD, H.P. Foundations of Psychology, New York, John Wiley & Sons, 1948.

Bibliography (Cont'd)

10. BOYNTON, R.M. Rapid chromatic adaptation and the sensitivity functions of human vision. J. Op. Soc. Amer. 1956, 46, 172-179.

11. BURNHAM, R.W. & CLARK, J.R. A test of hue memory. J. App. Psychol. 1955, 59, 164-172.

12. CHAPANIS, A., GARNER, W.R., & MORGAN, C.T. Applied Experimental Psychology New York, John Wiley & Sons, 1949.

13. CHAPANIS, A. & HALSEY, R.M. Luminance of equally bright colors. J. Op. Soc. Amer. 1955, 45, 1-6.

14. CHENEY, W.D. Color associations evoked by command words. Unpublished study conducted at the University of Arizona, 1959.

15. CONOVER, D.W. & KRAFT, C.L. The Use of Color in Coding Displays WADC TR 55-471, October 1958.

16. CRAIN, K.J. The Specification and Description of Surface Colors RADC-TN-58-333, December 1958.

17. DIMMICK, F.L. & HUBBARD, M.R. The spectral location of psychologically unique yellow, green, and blue. Am. J. Psychol. 1939, 52, 242-249.

18. ELY, J.H., THOMSON, R.M. & ORLANSKY, J. Design of Controls WADC TR 56-172, November 1956.

19. ERIKSEN, C.W. Location of objects in a visual display as a function of the number of dimensions on which the subjects differ. J. Exp. Psychol. 1952, 44, 56-60.

Bibliography (Cont'd)

20. GRANT, D.A., JONES, O.R. & TALLANTIS, B. The relative difficulty of the number, form, and color concepts of a Weigl-type problem. J. Exp. Psychol. 1949, 39, 552-557.

21. GREEN, B.F. & ANDERSON, L.K. Color coding in a visual search task. J. Exp. Psychol. 1956, 51, 19-24.

22. GREEN, B.F., MCGILL, W.J. & JENKINS, H.M. The time required to search for numbers on large visual displays. Massachusetts Institute of Technology, Lincoln Laboratory, Technical Report No. 36, 18 Aug 1953.

23. HALSEY, R.M. & CHAPANIS, A. On the number of absolutely identifiable spectral hues. J. Op. Soc. Amer. 1951, 41, 1057-1058.

24. __________ Chromaticity-confusion contours in a complex viewing situation. J. Op. Soc. Amer. 1954, 44, 442-452.

25. HAMWI, V. & LANDIS, C. Memory for color. J. Psychology 1955, 39, 183-194.

26. HUNT, D.P. The Coding of Aircraft Controls. WADC TR 53-221, August 1953.

27. HURVICH, L.M. & JAMESON, D. Some quantitative aspects of an opponent-colors theory. IV. A psychological color specification system. J. Op. Soc. Amer. 1956, 46, 416-421.

28. JUDD, D.B. Color in Business, Science and Industry, New York, John Wiley & Sons, 1952.

Bibliography (Cont'd)

29. KELLY, L.L. & JUDD, D.B. The ISCC-NBS Method of Designating Colors and a Dictionary of Color Names. Circular U.S. Bur. Standards, 1955, No. 553, v. 158 p.

30. KLEIN, S.J. & GALL, C.F. Aviation human engineering is a scientific specialty. J. Aviation Med. March 1958.

31. LE GRAND, YVES Light, Colour and Vision, New York, J. Wiley & Sons, 1957 (Trans. from original French, published 1956).

32. MACADAM, D.L. Chromatic adaptation. J. Op. Soc. Amer. 1956, 46, 500-513.

33. MAERZ, A. & PAUL, M.R. A Dictionary of Color, 2nd Ed. New York, McGraw-Hill, 1951.

34. MILLER, G.A. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. 1956, 63, 81-97.

35. MULLER, P.F. Jr., SIDORSKY, R.C., SLIVINSKE, A.M., ALLUSISI, E.A., & FITTS, P.M. The Symbolic Coding of Information on Cathode Ray Tubes and Similar Displays. WADC TR 55-375, October 1955.

36. MULLER, P.F. Jr. Efficiency of Verbal versus Motor Responses in Handling Information Encoded by Means of Colors and Light Patterns, WADC TR 55-472, December 1955.

37. PANEL ON PSYCHOLOGY AND PHYSIOLOGY, COMMITTEE ON UNDERSEA WARFARE A Survey Report on Human Factors in Undersea Warfare, Washington, D.C. National Research Council, 1949.

Bibliography (Cont'd)

38. OSA COMMITTEE ON COLORIMETRY The Science of Color, New York, Thos. Y. Crowell, 1953.

39. REED, J.D. A note on reaction time as a test of color discrimination. J. Exp. Psychol. 1949, 39, 118-121.

40. ROEBUCK, J.A. Jr. Anthropometry in aircraft engineering design. J. Aviation Med. 1957, 28, 41-56.

41. STELLAR, E. Human factors in panel design. In Human Factors in Undersea Warfare, Washington, D.C. National Research Council, 1949.

42. TEICHNER, W.H. Recent studies in simple reaction time. Psychol. Bull. 1954, 51, 128-149.

43. WOHLWILL, J.F. The abstraction and conceptualization of form, color, and number. J. Exp. Psychol. 1957, 53, 304-309.

44. WOODWORTH, R.S. & SCHLOSBERG, H. Experimental Psychology, New York, Henry Holt & Co. Revised Edition, 1954.

45. WEITZ, J. The coding of airplane control knobs. In Fitts, P.M. (Ed.) Psychological Research on Equipment Design, Washington, D.C. U.S. Govt. Printing Office, 1947, 187-198.

46. WRIGHT, W.D. The Measurement of Color, New York, The MacMillan Co., Revised Edition, 1958.

47. __________ Researches on Normal and Defective Color Vision, St. Louis, C.V. Mosby Co., 1947.