WADC TECHNICAL REPORT 54-262

Cleared: December 27th, 1979
Clearing Authority: Air Force Avionics Laboratory

THE DESIGN OF DIGITS

Mason N. Crook
Frances Schulze Baxter

Tufts College

June 1954

Aero Medical Laboratory
Contract No. W33-038 ac-14559
RDO No. 694-51

Wright Air Development Center
Air Research and Development Command
United States Air Force
Wright-Patterson Air Force Base, Ohio

Carpenter Litho & Prtg. Co., Springfield, O.
200 - 16 November 1954

ABSTRACT

Design characteristics of transilluminated numerals in the Air Force-Navy Aeronautical style at very low brightness were investigated. The experimental results indicate narrow stroke widths and close spacing were unfavorable. With area held constant, the effect of height/width ratio varied from digit to digit. Configurational characteristics could not readily be improved, but a round-tipped "3" in an open style was found to be superior to the standard flat-topped design. Data on confusions among both intact digits and digits with small stroke defects showed reasonable degrees of consistency. A number of special observations for which the test situation was well suited threw light on problems of experimental technique.

From an analysis of the literature it was concluded that most of the contradictions on digits by different investigators can be resolved and the data fitted into a coherent pattern. This report is of value to the Air Force in that the principles given can be used as guide lines for charts and checklist designs to give better readability.

PUBLICATION REVIEW

This report has been reviewed and is approved.

FOR THE COMMANDER:

Anderson Colonel USAF (MSC)
JACK BOLLERUD
Colonel, USAF (MC)
Chief, Aero Medical Laboratory
Directorate of Research

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

SECTION I

INTRODUCTION

This report presents the results of a series of experiments on the design of dial type numerals, and reviews a number of related studies by other investigators.

Work on the design of numerals as a factor in recognizability has been stimulated by the proliferation of instrument display systems in modern technology, and by the development of human engineering as an area of research. This frame of reference has operated to establish certain broad patterns in the work in question.

The orientation has usually been operational rather than theoretical, with the result that the experimentation has been planned to establish empirical optima in design variables rather than to determine characteristics of the receptor mechanisms from which such optima might be predicted. Operational situations differ in many ways (e.g., illumination level) and it is not always recognized that the generality of findings in a particular situation may be limited by such differences.

While the volume of work is not large, much of it is quite recent, and results have often appeared in special reports of limited accessibility rather than in the regular periodicals. This has made it difficult for those not working directly in the field to get more than a fragmentary picture.

Suggestions for improvement of design have usually been conservative. To make this point more concrete, it might be helpful to think in terms of four levels of novelty:

<u>Level 1</u>. Changes of detail resulting in only minor departures from some current design.

<u>Level 2</u>. Substantial innovations, which might or might not show a regard for aesthetic considerations and printing conventions, but which retain the essential features of traditional patterns.

<u>Level 3</u>. Radical innovations which constitute departures from existing patterns, but which have enough familiar elements to be correctly recognized at first glance or with a minimum of learning.

Level 4. Entirely new patterns, which could probably be made more discriminable than those now in use, but which would create a problem of relearning and habituation.

Most of the work, including our own, has put the emphasis either on comparisons of preexisting designs or on modifications limited to levels one and two above.

In certain situations, such as code messages and some types of meaningful text, letters and numbers tend to be mixed. In other situations, exemplified by dials and counters, numerals appear uncomplicated by letters. Optimal design of the digits "0", "1", and "8" might depend on whether the alphabetic symbols "O", "l", and "B" are possible sources of confusion. Both types of situation are included in the studies reviewed below, but our interest in this paper is limited primarily to numerals unmixed with letters.

Usages with respect to terminology and quantification have been somewhat haphazard. Certain digit designs have come to be identified by the names of their originators, as Berger and Mackworth. Later investigators have sometimes modified these designs in such characteristics as digit-width but retained the original labels. In specifying dimensions, height has sometimes been measured to outer margins, sometimes to mid-points of outer strokes. The premium on space in some kinds of displays makes it necessary to take cognizance of two measures which have commonly been slighted, over-all digit area and inter-digit spacing. For convenience and consistency, we shall establish a few definitions and conventions of measurement to be used in this paper. Data from other authors will be converted to these terms when necessary.

Glossary and Conventions of Measurement

Recognizability

The property measured by time, error, or other scores in recognition tests. A distinction is current in the literature between "legibility" or "readability", a property of meaningful text, and "perceptibility" or "visibility", a property of characters presented singly or in small groups, often under adverse conditions. In the latter type of situation, with which we are here concerned, "recognizability" appears to be a somewhat more descriptive term.

Digit

One of the ten symbols, "0" through "9".

Set of Digits

The ten symbols.

Numeral

One or more digits in a functional unit.

Dimensional Characteristics

Stroke-width and digit-width, in relation to height, the features most commonly and conveniently specified by measurement.

Configurational Characteristics

All design features other than digit-width and stroke-width.

Contrast Relation

The black-on-white vs. white-on-black relation.

Design

The complete dimensional and configurational characteristics of a set of digits as specified by the original author. Over-all size and contrast relation are excluded; design is the same in any size and contrast relation unless the author has alternative designs; we might indicate, for example, Berger design, black on white. "Design" is similar to but not identical with "type face".

Style

The configurational characteristics of design. Thus "Berger style" might refer to the Berger design in a modified stroke-width.

Type Face

Same as design except that details might vary among point-sizes, as happens in printing type.

Pattern

The major characteristics of form which differentiate the ten symbols, letters of the alphabet, geometric forms, etc.

Height and Width of Digit

Measured between the horizontal and vertical tangents respectively at the outer limits of the digit contours.

"Width" of digits in a set usually requires further specification. The "1" is ordinarily in a class by itself. Sometimes the remaining nine are the same width; frequently eight digits are the same width and the "4" is wider; occasionally other variations occur. The term "digit-width" without qualification will refer to the nine digits in a set other than the "1" when the nine are alike. The same term with exclusion of the "4" explicitly indicated will be used when eight digits in the set are alike. "Mean digit-width" will designate the mean of nine digits when they are not alike.

Spacing of Digits

Measured between adjacent vertical tangents.

Area of Digit

Area of the rectangle formed by the horizontal and vertical tangents.

Stroke-Width, Digit-Width, and Digit Spacing

To be expressed in percentages of digit height.

Unit of Dimension

Inch

Unit of Illumination

Foot-candle

Unit of Brightness

Foot-Lambert

At best, the available data do not permit the formulation of simple rules for optimizing designs in all situations. They do, however, show some consistencies and trends. In this paper we shall undertake to do three things: (1) analyze briefly the more relevant studies to illustrate the variety of factors coming into play and the complexity of the results; (2) report our own work in the quite limited area of dial-type numerals under low illumination; (3) indicate such generalizations as the results to date seem to suggest.

SECTION II

LITERATURE

A. Verbal Material

The legibility of letters and words has been studied more extensively than that of numerals. Of the research on verbal material, the larger part has been concerned with measuring some aspect of ordinary reading (e.g., Paterson and Tinker, 34, Luckiesh and Moss, 30). The results of such studies are of limited value for the recognizability of numerals. Numerals are more often presented singly, and in any case the recognition of individual characters is critical, unlike meaningful text in which recognition is by words and phrases.

In a limited number of investigations of the recognizability of letters the criterion has been distance at which individual characters could be identified, or number of errors under limiting conditions of illumination, distance, or exposure time. Such testing of individual letters is more likely to provide useful leads for the recognizability of numerals than are studies of ordinary reading. Examples are work by Sanford (39), Roethlein (37), Paterson, Walsh, and Higgins (33), Tinker (42), Forbes and Holmes (20), Uhlaner (44), Mackworth (32), Brown and Lowery (9), Forbes, Moskowitz, and Morgan (21), and Brown (8). A few points suggested by the work on letters which are of interest to us in the present context can be mentioned.

Rank orders of recognizability, and tables of confusion data by individual letters, have been reported by several investigators. On such results there is fair, but by no means complete, agreement. The suggestion is strong that the findings may depend in part on the type of testing situation.

Such processes as selective forgetting have been found to operate in the experimental routine to an extent that could bias the results from small groups of subjects (Sanford, 39).

Dimensional aspects of letter designs have been found to be more important than configurational aspects, within the limits of certain standard type faces (Roethlein, 37).

B. Numerals

Surveys of work on numerals going somewhat beyond the limited reviews included in the briefer experimental articles can be found in Gleason (22), Chapanis, Garner, and Morgan (12, pp. 170-180), Kappauf (23, pp. 24-35, 24, pp. 78-81), Brown, Lowery and Willis (11), and Soar (41).

Under conditions of poor visibility, visual discriminative capacities become relatively more important. Therefore the classical literature on visual performance as a function of size, brightness, contrast, and exposure time constitutes valuable background material. Among the authors who have given more than passing attention to this literature are Gleason (22), who reviewed it in relation to numeral design in general, and Berger (4, 5, 6) and Brown (7), who discussed it in relation to their own experimental work.

Experimental work on the recognizability of numerals falls into two categories: (1) that concerned primarily with the effect of numeral size and of various secondary variables such as brightness and vibration; reports by Craik (14), Rock (36), and Crook and associates (16, 17) typify this category; (2) that concerned more specifically with numeral design; this is the category in which we are most directly interested.

Before we turn to individual studies of design it will be useful to distinguish three broad types of operational situations in which numerals are used: (1) in printed materials to be read at conventional reading distance; (2) in displays to be read at relatively large distances, as license plates, highway signs, and wall charts; and (3) on dials and consoles, for which a typical viewing distance is 28 inches. These situations are differentiated to some extent not only in average viewing distance, but in various other respects such as style of numerals commonly used and lower limit of illumination likely to be encountered. It is of course necessary to distinguish between viewing distance in regular use and in experimental testing; distance thresholds have been the criterion of recognizability in experiments related to all three of the above types of operational situation.

The emphasis in our own studies has been on dial type numerals at low brightness. The extent to which optimal design is a function of the situation remains to be determined from such data as can be brought to bear on it. In any case, it is reasonable to assume that investigations having different operational referents can throw light on certain kinds of common problems such as the type of variable likely to be important, the relative frequencies of particular inter-digit confusions, and possible pitfalls in experimental methodology. So far as practical, we shall group the studies reviewed below into the above three categories.

The literature covered in the following sections includes all of the experimental investigations of numeral design which have come to our attention except for one or two items which either have not been released for distribution or have appeared only in classified reports.

1. Numerals in Print

Tinker (42) in 1928 made a study of printed letters, digits, and mathematical signs in black on white. The digits were in a Modern design of non-uniform stroke-width similar to that used in many textbooks. Type size was 8-point, giving a digit height of approximately 0.09 in. Viewing distance was 15.5 in. Constant artificial illumination was used. Digits were presented singly in tachistoscopic exposures, in one series mixed with the lower case letters and 12 mathematical symbols, in another series mixed with capital letters and 11 mathematical symbols. Criterion of recognizability was number of errors. Results were obtained from six subjects. Rank correlation between the recognizability scores for the digits in the two series was 0.27. The detailed results showed clearly that the lack of agreement between the rank orders was due in part to the difference between the upper- and lower-case letters in the two series with which the digits were confused. It was also found that digits with rounded contours (3, 6, 9, 0) tended to be confused with each other.

Tinker (43) subsequently studied Modern and Old Style printed numerals in black on white. Modern numerals are uniform in height and position on the line. In Old Style numerals the "0", "1", and "2" are relatively small, corresponding in height to the loops of lower case letters. The remaining digits correspond in height to Modern digits or to capital letters, but are positioned to extend above or below the line like lower case letters with ascending or descending stems. Both styles have non-uniform stroke-widths. Both are used in texts and tables. All numerals were presented in 10-point type, under artificial illumination. Criteria were: Method 1, distance at which digits could be read; Method 2, reading speed and errors when digits were presented in an ordinary reading situation. In Method 2, viewing distance was 20 inches. Digits were tested both singly and in groups by Method 1, in groups only by Method 2. Digit spacing for the group presentations, though not reported, was presumably in the range found in ordinary printing. Six subjects were used throughout, twenty additional subjects in Method 2. It was found that Old Style digits could be read at a slightly, but reliably, greater distance than Modern when presented in isolation, and at considerably greater distance in groups, but under conditions of ordinary reading no significant differences in speed or errors occurred. Confusion patterns were more consistent among the Old Style than the Modern digits, the most obvious confusion groups being the smaller digits (0, 1, 2), those which project above the line (6, 8), those which project below the line (3, 4, 5, 7, 9), and those with similar form (0, 2, 3, 5). The following rank difference correlations were found between the recognizability scores of the 10 digits as determined by Method 1 for the two styles and the single versus grouped presentation: Modern and Old Style presented singly, rho 0.90; Modern and Old Style presented in groups, rho 0.27; Modern presented singly and in groups, rho 0.65; Old Style presented singly and in groups, rho 0.31. We have in this experiment not only evidence of differences in recognizability related to digit style, but also evidence of the dependence of such differences on the testing criterion and the grouping arrangement.

McLaughlin, in an unpublished thesis (32), analyzed errors made in a previous study of numeral reading in the same laboratory. The numerals were in a Modern design, printed black on white in tabular form, under conditions of illumination and vibration sufficiently severe to produce a moderate percentage of errors when viewed at 14 in. Digit spacing was wide enough to eliminate any possible effect of crowding. It was found that confusions within any one of the three digit groups, completely curvilinear, completely rectilinear, and mixed, were about 30% greater than those among the groups. It was also found that the sequence of items read had an influence on the placement of errors.

McLaughlin proceeded to an experimental investigation of numeral design in general, without obvious reference to any of our three types of operational situation. The first step was a listing of possible ways in which digit elements could differ, and a selective narrowing down of the list to items which seemed suitable to the problem of improving design. Examples given were: size, slant, extensions above and below the type body limits of the other numerals, rectilinearity vs. curvilinearity, height/width ratio, and size ratio of top portion to bottom portion. Designs of individual digits were varied freely in these respects, through several steps of preliminary testing, until a set was arrived at which seemed to deserve a more formal test. In the final design, the digits were of uniform height with no projections above or below the line. Certain digits were moderately unconventional in configurational details but all were immediately recognizable. Stroke-widths were uniform within some digits while not in others, but somewhat wider on the average than in the AND design with which they were experimental compared. As a control for the stroke-width variable, a set of numerals in the AND style, but with strokes wide enough to approximate the average of the experimental digits, was included in the test. Digits were projected individually onto a viewing screen, in black on white, in exposures of 0.04 sec. Contrast with background was estimated at between 4% and 10%. Digit height was equivalent to that of 11-point type at 14 in. Criterion was number of errors. Results were based on groups of from three to six subjects.

The experimental digits were found to differ significantly from the standard AND digits at the 5% level of confidence, but not to differ significantly from the modified AND digits with which they were equated in average stroke-width. It therefore appears that the stroke-width variable introduced may have improved recognizability, but the configurational changes did not. The author pointed out the brief exposure technique may have favored stroke-width. The very low contrast could be expected to enhance this effect.

2. Numerals to Be Read at a Distance

Dunlap (19) is frequently cited for a study of which we have been unable to locate a copy and on which few details are available. He is reported to have analyzed 122 license plates and concluded that recognizability is favored by dark characters on a light ground rather than the reverse, by ample white spaces in and around the characters, and by relatively thin strokes.

Aldrich (1) tested four styles of license plate numerals, Round, Block, Open, and Vermont 1937, in black on white. Block differed from Round in that the curves were constructed of small straight line segments, the difference being hardly noticeable under near-threshold conditions. The Open style was designed to emphasize the distinctive features of individual digits. It differed from Round and Block in having a much more open loop on the "5", and in straightened out rather than recurved tails on the "6" and "9". Both a flat topped and a round topped "3", and a "1" both with and without a foot piece, were included in some of the sets. Vermont 1937 was an adaptation of Block. Width of Round, Block, and Open digits was 50% of height, that of Vermont 1937 42% of height. All styles but Vermont 1937 were tested in two stroke-widths, 12.5% and 8.3% of height, Vermont 1937 in 12.5% only. Round and Open, in the 12.5% stroke-width, are shown in Figure 1. Numerals were 3 in. high. Criterion was distance at which they could be recognized in outdoor daylight. Simulated license plate groupings were used. Data were obtained from 14 subjects. It was found that the Open style was definitely best and Vermont 1937 definitely worst, with little difference between Round and Block. The stroke-width 12.5% of height showed a small but consistent superiority over the 8.3%. The "1" with the foot piece was definitely superior to the "1" without in terms of recognition distance, and the round topped "3" was superior to the flat topped "3" in terms both of distance and number of errors. The "9", which was simply an inverted "6" in all styles, was among the digits most confused, while the "6" was among those least confused. In addition to the measurable differences among styles and patterns, this study demonstrates some advantage for the wider strokes under limiting conditions of seeing, and also impairment from reduced digit-width, which is the main feature differentiating the Vermont 1937 from the Block style.

In the same article Aldrich reports a companion study on the effectiveness of various ways of grouping the numerals and of combinations of numbers and letters, but the results are peripheral to the present context.

Berger (3), in a group of studies published in a two-part article, did what was in some respects the most analytic investigation of numeral design to date. He undertook to devise a set of license plate digits which were individually equated in recognizability and which made the most effective use of available space. Experimental designs were tested in black on white, white on black, and with transilluminated strokes. The designs produced as a result of the study were constructed essentially of straight line segments; the "0", a rectangle with rounded corners, can be considered a prototype for the set. (See Figure 1.) Digit-width and stroke-width were treated as experimental variables. Digit height was 3.15 in. Criterion was distance threshold for recognition, under both day and night illumination.

The first part of the study was done under outdoor daylight. The initial step was to determine optimum stroke-width for digits "2", "5", and "8" presented singly. Digit-width was 52% of height. Stroke-widths from 2.5% to 20.0% of height were tested. The optimum for white on black was found to be 7.7% and for black on white 12.5% of digit height. White on black at its optimum was better than black on white at its optimum stroke-width. The next step was to investigate configurational features of design, using white on black at the optimum stroke-width, and again testing by the distance criterion. Results led to the following adjustments in individual digits: for digit "2" the length of the hook in the upper left hand corner was set at between 12.5% and 16.3% of digit height and the diagonal stroke was made a straight line; for digit "3" straight lines were adopted, which gave a flat top, and the lower end of the diagonal stroke in the upper loop was positioned on the vertical mid-line; for digit "4", the optimum crossing points of the vertical and horizontal strokes were determined, and the third stroke was positioned to connect the ends of the first two strokes, thus forming a loop; for digit "5" the top line was set to extend the full width of the pattern; for digits "6" and "9" the rectangular loop was set at half the over-all area, and the tail was set at 45°; for digit "7" the stem was given maximum inclination; for digit "8" the diagonal cross lines were set at 90°. For digit "8" a further adjustment was arbitrarily made - the width was increased to 53.7% of digit height to equate the inner horizontal and vertical dimensions of the loops. A further step consisted in empirically adjusting the widths of all digits except the "1" to equate them in recognizability to the "8", which was taken as the standard. The widths so arrived at ranged from 49% of digit height for the "2" to 60% for the "4". The mean for the set (exclusive of the "1") was 52%. Finally, digit spacing was determined such that two adjacent digits could be resolved, but not necessarily recognized, at the distance at which the standard "8" could be recognized. For this determination each digit was paired with "0", separately on the right and left. The resultant spaces varied with the digits. It was found that wider spacing was needed for 5-digit than for 2-digit combinations, so 10% was added to the spacings as above determined. With this adjustment, the spacings ranged from 11.2% of digit height for two adjacent "7's" to 30.8% for two adjacent "0's", with a mean of 22.4% for all possible pairs.

In the second part of the study, Berger investigated stroke-width under night lighting, for light numerals on a dark ground, with both reflected and transmitted light. The reflected light was from a 15-watt bulb a few inches below and before the surface. If an intensity of six candlepower and a distance of not more than a foot can be assumed, the illumination would be at least six foot-candles. For transillumination, pieces of cardboard with the numerals cut out were placed in front of a light-tight box containing a 10-watt bulb. This probably gave higher brightness of digit strokes, certainly higher contrast, than the reflected light. Using digits "2" and "8" with reflected light, the optimum stroke-width was found to be the same as in daylight, 7.7% of digit height. With transillumination the results were quite different. Again with digits "2" and "8", legibility improved as strokes were made narrower until the strokes reached 2.5% of digit height.

In these experiments four subjects commonly constituted a group, but sometimes more were used. The results show a clear interaction between stroke-width on one hand and contrast relation and mode of illumination on the other. In evaluating the data on digit-width it should be borne in mind that the approximate proportions of Berger's standard "8", to which his other digits were equated, was determined by license plate conventions rather than by experimental test. The term "Berger design" as used in this paper will imply Berger's optimum stroke-width for the appropriate contrast relation and illumination condition.

Berger reported two subsequent studies (4, 5) which were concerned with the variables of width and height. The digits were in Berger's own style, black on white, presented singly. Subjects ranged from two to five in a group.

In the first of these studies (4) width only was investigated. Digits "0" and "5" were used. Stroke-width approximated the optimum for black on white in the original design. Widths tested were 25.0%, 33.3%, 45.8%, 55.0%, and 69.2% of digit height. Height was 0.24 in. throughout. Three criteria were used, (1) distance thresholds under 58[a] foot-candles, (2) Luckiesh-Moss Visibility Meter thresholds, and (3) brightness thresholds. Results by distance and brightness thresholds were in good agreement, showing an increase in recognizability with increasing width throughout the range, but at a much faster rate for the "0" than for the "5". Results with the Visibility Meter were only partly in agreement, and were more variable. Berger concluded that the instrument was not suitable for this type of investigation.

In the later study (5) the primary variable was height, but a range of widths was included. Test symbols were "0", "5", and the letter "E". Stroke-width was considerably narrower than the optimum in the original design. Each symbol was presented in five heights, at three widths for each height, as follows: for height 0.10 in., widths 40%, 100%, and 160% of height; for height 0.16 in., widths 25%, 62%, and 100%; for height 0.22 in., widths 18%, 45%, and 73%; for height 0.28 in., widths 14%, 36%, and 57%; and for height 0.33 in., widths 12%, 29%, and 47%. Criterion was distance thresholds under 58 foot-candles. Recognizability was found to increase with height for the three widths and the three different symbols; the increase was somewhat more rapid for the larger widths, but very similar for all three symbols. In the previous study it had been found that recognizability increased with width more rapidly for the "0" than the "5", height being constant. From the data of the later experiment it was possible to construct curves to check this effect, and the earlier finding was confirmed. The curves for the two digits crossed at widths in the region 20% to 30% of digit height. The implication of this in terms of

---

[a] This figure was given in the original article (4) as 22 foot-candles, but was corrected in the subsequent publication (5) to 58.

absolute thresholds is that in the wider designs the "0" was more recognizable, in very narrow designs the "5" was more recognizable. General results for the "E" were very similar to those for the "5". The author offered a detailed interpretation in terms of acuity principles and spacing of elements in the retina.

The interaction between width and individual digit pattern as determiners of recognizability is shown convincingly in these two experiments. It should be noted that the test symbols were so designed that an increase in width or height produced a corresponding increase in area, so these results tell us nothing directly about the effect of height-width ratio with area held constant. In Berger's procedures the threshold was approached a certain number of times from above and a corresponding number from below. In at least the former case, and probably in the latter, the subject knew in advance what symbol he was judging. This is psychologically quite different from a test situation in which the subject does not know the symbol in advance. It can be expected to produce lower thresholds. But a more important question, which must remain unanswered for the present is, would knowledge have the same effect on all designs, or would the effect be related to relative difficulty and the types of possible confusion.

Mackworth (31) did a study during the war directed at improving the recognizability of letters and numbers to be viewed at varying distances in operations rooms. Some testing was done on the effect of size of characters and viewing angle, but our present concern is with the variable of design. The digits in use prior to Mackworth's study were similar in most details of style to the Leroy digits shown in Figure 1, but narrower, the width being about 58% of height. Stroke-width was about 14% of height.

In Mackworth's new design the diagonal stroke of the "2" was straightened, the top bar of the "3" widened, the loop of the "5" opened, and the loops of the "6" and "9" reduced and the tails straightened. Width was reduced to about 50% of height while stroke-width was kept at about 14%. This is the set commonly known as the Mackworth digits (Fig. 1). Contrast arrangement in the testing was black on orange for the old set, black on yellow, giving slightly higher contrast, for the new. Characters approximately 0.80 in. high were presented singly at viewing distances of 25 to 40 ft. Illumination was 10 foot-candles. Exposure time was 1.62 sec. Letters and numbers were mixed in the test series. Criterion was number of errors. Subjects were used in groups of 30 to 40. The new display of letters and numbers together was found to be reliably better than the old, and the results by individual characters show that on most of the new digits fewer errors were made than on the old. These data cannot, however, be taken as adequate evidence for the superiority of the new digit design. In the experimental test, the new characters had a small advantage in contrast. It is probable, also, that scores on the digits were determined in part by confusability with the letters presented in the same experimental series. In the new set the letters were modified, and probably improved, in design, and were also wider on the average than in the old set, giving a somewhat greater over-all size. The tests did not, therefore, provide a clear comparison between old and new digit designs per se.

3. Dial-Type Numerals

The Air Force-Navy standard letters and numerals for use on dials have grown out of the experience of manufacturers and service agencies rather than experimental tests. In use, these numerals ordinarily appear light on a dark ground, commonly illuminated by floodlighting, but sometimes transilluminated or fluorescent. The 1944 specification sheet (45) served as the starting point for our own research described below. This sheet specifies the width and stroke-width of all characters, but the reproductions shown on the sheet do not precisely agree with the specifications. The 1948 revision (46) shows the same drawings, omits the statement about width of characters, and specifies stroke-width as 12.5% of height unless characters are not over 0.125 in. high, in which case strokes shall be 16.7%. Mean width of the digits exclusive of the "1", based on measurements of the 1948 drawings, is 65% of digit height, and stroke-width is 12.9% of height. In the text to follow, these latter values will be taken as correct for the current AND design.

Craik (14), working in England in 1941, investigated white dial-type numerals (Fig. 1) generally similar in style (except for the "4") to the AND. The brightness levels involved were probably from 0.0001 to 0.01 foot-Lambert, which is the range covered by Craik's work on numeral size. The main recommendation for design, presumably based on experimentation the details of which were not reported, was that for low brightnesses the strokes be equal in width to the mean of the enclosed blank spaces. This is equivalent to about 20.0% of digit height, considerably heavier than the AND. Mean width of the Craik digits, exclusive of the "1", as determined from measurements of the drawings shown in his report, is approximately 61% of height.

Loucks (26, 27, 28, 29), in 1944, investigated the accuracy with which various instrument dials could be read. Of the four reports, the second and third contain material on numeral design. Numerals appearing on the dial faces were generally similar to the current AND but showed some variations in detail and were not uniform on all dials. The variations, of which somewhat less open loops were characteristic, presumably reflected earlier stages in the AND series. Among the large number of experimental comparisons which were made, some involved numeral dimensions combined with other variables, some involved numeral dimensions alone. Two illumination conditions were used, (1) reflected light from an incandescent source and (2) a relatively high intensity of ultraviolet light. Entire dials were exposed for intervals of 0.75 or 1.50 sec. Criterion was number of errors in reading the pointer position. Results were obtained from groups of about 20 subjects.

In Loucks's second report (27) five comparisons were described. The first of these involved numeral size rather than proportion or configuration, but the results illustrate a problem of interpretation. Of two dials otherwise alike, one had numerals 0.19 in. high with strokes 16.7% of height, the other had numerals approximately 50% larger in over-all dimensions and in stroke-width. Results showed the smaller numerals to be better under reflected light, contrary to the obvious expectation. As a possible explanation, the author pointed out that it was easier to relate the smaller numerals to the scale marks with which they belonged.

In comparisons two and three the 0.19 in. numerals were tested in two stroke-widths, 16.7% and 25.0% of digit height. The former were found to be better under reflected light and equally good under ultraviolet.

Comparisons four and five again illustrated the complication of variables. One dial had the 0.19 in. numerals with strokes 16.7% of digit height at every fifth scale mark, while the other dial had numerals 0.12 in. high with strokes 25.0%[a] of digit height at each scale mark. The arrangement of smaller numerals was found to be better under reflected light at 1.50 sec. exposure, there was no significant difference under reflected light at 0.75 sec. exposure nor under ultraviolet light at 1.50 sec. exposure, while the larger numerals were better under ultraviolet light at 0.75 sec. exposure. A definitive interpretation of these results is not possible because stroke-width was confounded with size, but it seems probable that the advantage of the smaller numerals under the most favorable seeing conditions (reflected light, 1.50 sec. exposure) resulted from the greater ease with which they could be related to their proper scale marks, while the loss of advantage as seeing conditions got worse is what would be expected on the basis of size.

In Loucks's third report (28) the several comparisons described included four which involved stroke-width, the results of which can be summarized together. Numerals were 0.25 in. high. Stroke-widths were 6.2%, 12.5%, and 18.9% of digit height. Under reflected light the stroke-width 12.5% of height was better than the 18.9% but not significantly different from the 6.2%, while under ultraviolet light the 12.5% was better than the 6.2% but not significantly different from the 18.9%. This indicates an increase in the optimum stroke-width with decreasing illumination.

Kuntz and Sleight (25) studied stroke-width in Leroy type numerals in both black on white and white on black. Mean width

[a] There appears to be a discrepancy between the stroke-width dimension for the smaller digits given in the text and the reproduction in Loucks's Figure 3. The 25.0% was estimated from Figure 3.

of Leroy digits, which are widely used in hand lettering, is about 72% of height. The style differs from the AND in a number of details, the most noticeable of which are closed loop on the "4", a less open loop on the "5", recurved tails on the "6" and "9", and a curved stem on the "7". Stroke-widths tested were 22.2%, 20.0%, 18.2%, 16.7%, 15.4%, 14.3%, and 13.3% of digit height. These figures are in terms of over-all digit height, and have been computed from those given by Kuntz and Sleight which were based on measurements to the mid-points of top and bottom strokes rather than the outer limits. Three illumination levels were used which gave white surface brightnesses of 3, 10, and 31 foot-Lamberts. Digits were presented 30 on a card with wide spacing. Distance thresholds were determined by starting the cards at a relatively large distance and moving closer by steps, the subject identifying as many digits as possible at each step. Results were obtained from 14 subjects. Variance analysis showed brightness, stroke-width, individual digit patterns, and subjects to be significant, as well as most of the first-order interactions. Contrast relation was not significant. For the three brightnesses and two contrast relations combined, the optimum stroke-width was 16.7% of digit height, with very little difference in the range from 14.3% to 20.0%. The stroke-width curves for black on white and white on black were very similar, with somewhat more irregularity in the latter. The stroke-width curves plotted for the three brightness levels separately suggested a tendency for the optimum to shift toward wider strokes as brightness decreased. The optimum stroke-widths found here tended to be considerably broader than those found by Berger. We shall attempt to account for this discrepancy in a later section.

On the relative recognizability of individual digits, the authors reported that there appeared to be "....a consistent tendency for those numerals having most straight lines with least configuration to be most legible".

Brown and associates, at the Aeronautical Medical Equipment Laboratory of the Philadelphia Naval Base, studied the requirements for letters, numbers, and markings to be used on transilluminated aircraft control panels. Results have appeared in a series of six reports (2, 8, 9, 10, 11, 18), of which two contain material on numerals. On such panels the characters appear white under reflected light in the daytime and red by transilluminated light at night. The numeral styles investigated were the AND, Berger, and a new design intended to incorporate the best features of the other two, designated the AMEL (Fig. 1). Numerals were 0.16 in. high. Viewing distance was 28 in. Sets of three or four digits, with wide spacing, were presented in brief exposures. Criterion was number of errors. Results were based on groups of 20 to 48 subjects.

Brown, Lowery, and Willis (11), in the third report of the above series, tested the recognizability of the AND and the Berger styles, somewhat modified from the original designs to produce the same ratio of digit-width to height in the two styles. Mean width of the original AND digits exclusive of the "1" is 65% of height, as determined by measurement of drawings shown on the specification sheet. The corresponding figure for the original Berger design is 52% of height. In the Brown, Lowery and Willis experiment the width of all digits except the "1" was adjusted to 60% of height. This modified the appearance of the Berger digits more than that of the AND. In the new dimensions, the two styles appeared to differ mainly in the greater angularity of the modified Berger digits, which had the effect of filling out the corners of the circumscribed rectangles. A larger difference showed in the case of the "4", the AND "4" having an open loop and the Berger "4" a closed loop. Testing was done under both simulated daylight and low level transillumination. For the former, illumination levels of 40 and 80 foot-candles were used, exposure time was 0.007 sec., and stroke-widths were 6.7%, 11.6%, and 16.7% of digit height. With red transillumination, brightnesses were 0.33, 0.79, 1.63, 2.60, and 3.35 foot-Lamberts; exposure times were 0.04 and 0.20 sec., and stroke-widths were 6.7%, 11.6%, 12.5%, 14.3%, and 16.7% of digit height.

Results by individual digits were presented in tables, but the authors' analyses of the stroke-width effects were mainly in terms of all the digits pooled. With floodlighting, there was very little difference between stroke-widths of 11.6% and 16.7% of digit height, but errors were much larger for the narrowest stroke (6.7%). Illumination level showed little effect at the two heavier stroke-widths, but considerable effect for the narrowest stroke. It should be noted that the very brief exposure could be expected to interact with other conditions which might tend to impair recognizability. It is unlikely that such a marked stroke-width effect would show under these general conditions with longer exposure. With transillumination, more errors were made at the shorter exposure as might have been expected, and also at the lower brightnesses, the brightness effect being more marked at the shorter exposure. Stroke-width showed a somewhat complicated effect in relation to the other variables. A tendency for more errors to occur with stroke-width 16.7% than with 12.5% of digit height was manifest at the shorter exposure, and the effect became quite marked at the longer exposure. The narrowest stroke, 6.7% of digit height, was relatively unfavorable, more especially at the lower brightnesses and shorter exposures. Comparing the two numeral styles digit by digit, the Berger style "4" with the closed loop proved reliably better than the AND style "4". There was some evidence for superiority of the AND style "9" over the Berger style "9" under floodlighting. For other digits, no reliable differences were found. The authors concluded that, over the range of lighting conditions studied by them, a stroke-width 12.5% of digit height affords the best compromise.

On the basis of their own work, the literature, typographical conventions, aesthetic considerations, and common sense, the authors formulated five principles of digit design for the type of cockpit displays with which they were concerned:

(1) All digits except the "1" and "4" should have a width 60% of height; the "4" should be one stroke-width wider.

(2) Each digit should occupy all of its allotted area (i.e., strokes should be extended to the edges of the circumscribed rectangle).

(3) All vertically segmented digits except the "4" (the "3", "5", "6", "8", and "9") should be divided at the mid-point.

(4) Outlines should be curvilinear.

(5) Stroke-width should be 12.5% of digit height.

Following these principles they designed a set of digits, identified by them as the AMEL font, which combined some features of the AND and Berger styles. It can be noted that of the five principles, only that on stroke-width had been validated by direct experimental test.

Atkinson, Crumley, and Willis (2) did a follow-up study to the above, the report of which appeared as number five in the series on control panel markings. The problem was to experimentally compare the AND and Berger styles used in the previous study with the AMEL. Stroke-width was 12.5% of digit height in all parts of the experiment. For simulated daylight, illumination levels of 11, 24, and 34 foot-candles were used, with an exposure time of 0.005 sec. With red transillumination, brightnesses were 0.10, 0.30, 0.80, 1.60, 2.60, and 3.30 foot-Lamberts and exposure time was 0.20 sec.

Total scores for the 10 AMEL digits could be compared with total scores for each of the two other styles at six levels of transillumination and three levels of simulated daylight. The AMEL digits showed fewer errors in 16 of the 18 comparisons. With 10 digits and six transilluminations pooled, the superiority of the AMEL set over each of the other two sets was significant at the 1% level, and the same was true for the 10 digits and the three daylight illuminations pooled. Evaluating differences by individual digits in terms of a 5% significance level, the AMEL "5" was better than the "5" in each of the other styles under each of the two illumination conditions; under red transillumination the AMEL "1", "3", and "4" were better than their counterparts in the AND style, and the AMEL "2" and "9" were better than the Berger style "2" and "9"; under simulated daylight the AMEL "4" was better than the AND style "4" and the AMEL "2" was better than the Berger style "2". The AMEL "4" probably benefited from being one stroke-width wider than the "4's" in the other sets. In no instance was an AMEL digit significantly less recognizable than its counterpart in one of the other styles. Comparing the AND and Berger styles, in this as in the preceding experiment the Berger style as a whole showed a statistically insignificant superiority. A few significant differences between the two styles showed on individual digits, but these were not entirely consistent from one experiment to the other. It should be remembered that this was not a test of the original AND and Berger designs, but of those styles with digit-widths adjusted to 60% of height.

A secondary finding of some interest was a loss of recognizability, in the transillumination situation, as brightness increased from 2.60 to 3.30 foot-Lamberts. A similar effect had been found with letters in an earlier study in the same laboratory. The explanation suggested by the authors is a broadening of the strokes beyond the optimum width by irradiation.

Schapiro (40) compared four numeral styles, the AND, Craik, Berger, and Mackworth, at stroke-widths 10.0%, 12.5%, 16.0%, and 20.0% of digit height. Digit-widths varied considerably among the styles. In the original designs, the mean digit-widths for the set (exclusive of the "1") are: AND, 65% of height; Craik, 61%; Berger, 52%; Mackworth, 50%. Corresponding figures in the Schapiro sets, averaged over all stroke-widths, were approximately as follows: AND, 65%; Craik, 62%; Berger, 56%; Mackworth, 54%.[a] The spread was therefore somewhat reduced, but remained wide enough to cause substantial differences among the styles in over-all area. Digit height was constant at 0.26 in. Viewing distance was 28 in. Digits were presented singly, in exposures of 0.95 sec., and appeared as trans-illuminated images on a projection screen. Brightness levels used were 0.011, 0.044, 0.145, and 0.975 foot-Lambert, a lower range than in any of the above studies except that of Craik and possibly Loucks. The primary criterion was time for the subject to respond orally to the digit presented. Errors were also recorded. Results were based on 12 subjects.

Time scores were variance analyzed with respect to subjects, brightness, style, stroke-width, and digit. In main effects, only brightness and subjects showed significant variances, but the four stroke-width means showed a systematic trend, the broadest stroke being the most recognizable. In first and second order interactions, many of the experimental combinations were significant. The two variables of most interest to us in the present context are stroke-width and style, both of which showed a definite interaction with brightness. At the three higher brightnesses performance was not significantly related to either stroke-width or style, but at 0.011 foot-Lambert both variables had a definite effect. For the four stroke-widths, 10.0%, 12.5%, 16.0% and 20.0% of digit height, at this low brightness level, the mean response times were 0.978, 0.879, 0.812, and 0.772 sec. respectively, all differences being

---

[a] The present authors are indebted to Dr. Schapiro for making available to them his complete set of original drawings. The figures for digit-width are based on our measurements of these drawings.

significant except that between the two broader strokes. For the four digit styles, AND, Craik, Berger, and Mackworth, at the same brightness level, the scores were 0.832, 0.911, 0.854, and 0.838 sec. respectively, the differences between Craik and each of the others being significant. Error scores by digit styles at 0.011 foot-Lambert were also shown, the total errors for AND, Craik, Berger, and Mackworth being 309, 420, 358, and 239 respectively.

The stroke-width data of this experiment are consistent with a pattern which can be detected in the results of the studies reviewed above. The accumulated evidence suggests that the optimum stroke-width is narrower (and possibly more critical) for trans-illuminated digits than for those illuminated by reflection, but it also suggests that in either case the optimum gets wider as illumination decreases. Schapiro's lowest brightness, 0.011 foot-Lambert, is in the range where sheer amount of stimulation, as determined by brightness, exposure time, and size of bright area, can be assumed to be important. Therefore Schapiro's finding that, even with transillumination, the optimum stroke-width is at least 16.0% of digit height, and perhaps wider, seems quite reasonable. It will be recalled that Craik, presumably working in a brightness range from 0.01 foot-Lambert down, recommended strokes approximately 20.0% of digit height.

The results on digit style are more difficult to evaluate, partly because of a dearth of comparable data in the literature. At 0.011 foot-Lambert, Schapiro found the Mackworth type digits second best of the four styles in time scores, and the best by a considerable margin in error scores. It will be recalled that the digit sets used varied in width, the AND style at 65% of height being the widest, the Mackworth at 54% of height the narrowest. Berger (4) found an increase in recognizability with increasing width (resulting in increasing area) throughout this range. It might be expected that legibility would be particularly favored by greater width at the low brightness level of 0.011 foot-Lambert. The possibility should be considered that the narrow Mackworth style digits presented a more recognizable design on the basis of height/width ratio alone. Evidence on the effect of height/width ratio available from other sources presents some difficulties of interpretation, as will subsequently appear. In Schapiro's test situation, any special merit the Mackworth digit might possess would have to outweigh advantages in width (i.e., in area) for the other styles ranging up to 20%. Schapiro presents a full tabulation of errors by digits, stroke-widths, and styles, for the one brightness level 0.011 foot-Lambert. For all styles but the Mackworth, nearly half the errors occur at the narrowest stroke (10.0% of digit height). For this stroke-width, as for all stroke-widths pooled, the Craik style shows the largest number of errors. If we compare the Craik and Mackworth styles digit by digit at this narrow stroke-width, the difference in errors favoring the Mackworth style is seen to be distributed over all the digits, being relatively large on the "0", "1", and "8" for which the configurational differences appear to be negligible, and relatively small on the "6", for which the configurational difference is conspicuous. Other digit-by-digit comparisons over the Table fail to suggest meaningful generalizations. A satisfactory clarification of these results, therefore, must await further evidence.

Soar, in an unpublished Doctor's thesis (41), reported the first direct investigation of height/width ratio as a design variable, area of circumscribed rectangle being held constant. Stroke-width was also varied and the interaction evaluated. Numerals were of the AMEL style developed by Brown, Lowery, and Willis (11) and experimentally tested by Atkinson, Crumley, and Willis (2). All digits except the "1" were prepared in four height/width ratios, the two dimensions being adjusted to keep area constant. The resulting widths were 30%, 45%, 60%, and 75% of height. The "1" was varied in height only, to conform to the other digits in the sets. At each height/width ratio, three stroke-widths were used. The stroke-widths were constant in absolute terms over the four height/width ratios, and were therefore not constant when expressed as per cent of digit height. In the latter terms, the figures were as follows: for digit-width 30% of height, stroke-widths were 6.4%, 9.5%, and 12.7% of height; for digit-width 45%, stroke-widths 7.8%, 11.7%, and 15.6%; for digit-width 60%, stroke-widths 8.9%, 13.4%, and 17.9%; for digit-width 75%, stroke-widths 10.0%, 15.0%, and 20.0%. Digits appeared in black on white, the heights for the four ratios varying around that of 10-point type (about 0.1 in.). Viewing distance was 24 in. Illumination was 1 foot-candle. Digits were presented singly, in exposures of 0.04 sec. Criterion was number of correct responses. A total of 72 subjects was used, six being assigned to each of the 12 combinations of experimental variables.

Variance analyses were made of the results by individual digits, subjects' variance being isolated by means of a covariance analysis based on performance in practice trials in which the test material was alike for all subjects. Effects of the experimental variables differed markedly from digit to digit. Height/width variance was significant at the 5% level or better for the following six digits, but maximum recognizability occurred at different ratios: for digit "0", width 75% of height was most recognizable; for "3", width 75% of height; for "4", 45%; for "6", 75%; for "7", 60%; for "9", 45%. For the following three digits, height/width was not significant: digit "2", maximum recognizability at width 60% of height; for "5", maximum at 45%; for "8", maximum at 60%. For digit "1", height/width is not meaningful in the same terms as for the other digits. Considering nine digits, the modal width for maximum recognizability was 60% of height, but for those in the "significant" category, the mode was 75%. The "0" particularly added weight to this latter group, as it was the least recognizable of the ten digits under Soar's conditions, and showed one of the largest height/width effects, with maximum recognizability at width 75%.

Stroke-width variance was significant at 5% or better for digits "0", "6", and "8", all with maximum recognizability at the narrowest stroke. Of the digits for which stroke-width was not significant, maximum recognizability was at the narrowest stroke for the "9", at the widest stroke for all the others. Thus the "0", "6", "8", and "9" showed maximum recognizability at the narrowest stroke, the "1", "2", "3", "4", "5", and "7" at the widest stroke. This grouping is suggestive, as all the digits for which the narrowest stroke is favored have enclosed spaces, while all those for which the widest stroke is favored except the "4" do not. Soar did not present pooled data for the ten digits, but if we compute average stroke-width scores for the set from the data by individual digits given in his Table 24, the differences between the three stroke-widths are found to be quite small, with the narrowest stroke showing a slight advantage. When we consider that the narrowest strokes varied from 6.4% to 10.0% of digit height (depending on the height/width ratio) this result appears to be at variance with the trend toward wider optimum strokes at low illumination levels which characterized several previous studies. It would be difficult to account for this conflict of evidence in terms of a simple interaction between height/width ratio and stroke-width, because in Soar's study this interaction was not significant for any digit. The two effects demonstrated by a substantial mass of data in this investigation are the influence of height/width ratio and the interaction between individual digit pattern and other design variables. The latter effect was demonstrated on a smaller scale by Berger (4, 5). As will be indicated below, there is reason to believe that the rank order of the digits is in part a function of characteristics of the testing situation such as illumination level. It may be that some of the apparent conflicts between the results of different investigators can be accounted for only by assuming multiple interactions which would be very difficult to demonstrate statistically.

Reinwald (35) investigated the recognizability of the AND, Mackworth, and Berger white-on-black designs as a function of vertical and horizontal viewing angles from 0 to 75°. Stroke-widths in the AND and Mackworth sets were approximately 12.5% of height, that in the Berger 7.7%. Characters were presented singly, in white on black, illuminated by reflection. Brightness of white figure was about 1.28 and of black background 0.126 foot-Lamberts. Digits were interspersed in the test series with AND and Mackworth letters. Criterion was recognition distance. Scores were better at vertical than at corresponding horizontal angles of presentation, which is explainable in terms of the height/width ratio. The AND and Mackworth digits were consistently better than the Berger. The AND showed a small and unreliable superiority over the Mackworth. The relatively poor showing of the Berger was presumably related to the narrower stroke-width, which appears anomalous at first glance as the stroke-width was Berger's optimum for white on black. We shall revert to this problem in a later section.

4. The Ordnance Numerals

The Ordnance numerals, shown in Figure 2, are of interest because of their extensive use on graduated scales. Like the AND, they have been developed over a period of time on the basis of experience in production and use rather than through formal testing. The exhibit of Figure 2 was adapted from a Naval Ordnance specification sheet. It illustrates a style rather than a design, because the specifications vary in stroke-width and letter-width depending on the particular application. The most distinctive characteristic of this style is that some digits extend above or below the line. When the digits are used in groups under good illumination, this could be expected to afford the kind of advantage that Tinker found for Old Style over Modern numerals (43). Under poor illumination, it would be necessary to weigh this advantage against the gain that might be realized by increasing the height of all digits to the same outer limits.

SECTION III

EXPERIMENTATION

A. The General Program

The experimental work here reported was aimed at exploring a number of design characteristics of dial type numerals with respect to recognizability at low brightness levels. Brightnesses were in the neighborhood of 0.01 foot-Lambert, a region well below that covered by most of the work reviewed above but one of some operational importance when it is necessary to preserve dark adaptation.

Numerals in the test situation were transilluminated. In terms of visual effectiveness, the most characteristic difference between transillumination and reflected illumination is that the former normally provides greater contrast between figure and ground. Brightness level can be either high or low in either type of system, and it is sometimes necessary to distinguish between the effects of contrast and those of brightness. The results below pertain to conditions of very low brightness but relatively high contrast.

Certain limiting conditions were assumed because of current practice and technological considerations. Digits of the AND style as ordinarily used, especially when they appear in groups, are uniform in height, stroke-width, and horizontal alignment, and are not usually interspersed with letters. Our experimental digits and testing conditions conformed to these limitations unless otherwise specified.

Poor visibility can often be compensated for by increasing the size of the characters that have to be read. But on instrument dials space is limited, and its economical use becomes important. Therefore in introducing experimental design modifications we were careful to keep over-all area constant, except in the few instances where size itself was a variable.

Variations in design were limited for the most part to Levels 1 and 2 described in the Introduction, i.e., variations within the range of conventional patterns.

The equipment and procedure to be described called for preparing experimental materials on kodalith film slides for projection. It became apparent early in the program that small variations in size and stroke-width resulting from inadequate control of the photographic processes were affecting the test scores. Control of the photography was therefore made more rigid and the size of characters on the slides was increased to minimize the percentage effect of any variation that remained. Results from experiments reported below were obtained with accurate slides made by the improved technioue.

It also became clear in this period that, even with accurate slides, small differences in digit-width (height being constant) were more likely to be reflected in scores than were modifications of design which did not involve changes in width or stroke-width.

These developments directed our attention to the distinction between dimensional and configurational variables. This is a distinction of convenience. Any change can be stated in dimensional terms if a sufficiently elaborate specification system is worked out. But the two dimensions, stroke-width and height/width ratio, are easy to specify, and it is useful to separate these from such factors as shapes of loops and slopes of lines. Thus we have two kinds of design variable, (1) dimensional, consisting of stroke-width and height/width ratio, and (2) all others. Overall size, internal relations being constant, is not treated as a design variable.

In planning experiments, dimensional and configurational variables were treated separately so far as possible. This policy was subject to one limitation. When stroke-width or height/width ratio is varied, an impression is sometimes created of an accompanying configurational change. With relations reversed the effect is less equivocal; configurational modifications can be made independently of stroke-width and height/width ratio.

In most cases, experimental digits were compared with the AND as a reference standard. When a digit was experimentally modified in one respect, e.g., stroke-width, configuration, or height/width ratio, it was kept as nearly identical as possible with its AND counterpart in all other respects. In the AND design as illustrated on the specification sheet, the width differs somewhat from digit to digit. Our reproductions for experimental comparison did not always conform precisely to the widths shown on the specification sheet, but these discrepancies were small, and experimental digits were carefully matched in width to the particular standards with which they were compared.

This project was launched before some of the more recent literature reviewed above had become available, and its direction was determined in certain respects by the status of the problem at that time. The work was carried out in several overlapping phases, and the outline below represents a simplified organization rather than a chronological sequence. A certain amount of experimentation, done while techniques were in process of development, was superseded and is therefore not reported. Certain additional units of work for one reason or another failed to produce data of any value and are reported only in summary form. The results below are those which survived this selection process.

The sections on Apparatus and General Procedure apply, with minor exceptions, to the entire program. Conditions and results from individual experiments will be described separately, and a general discussion will follow.

B. Apparatus

The apparatus consisted of a special projector with accessories described in an earlier report (16). The optical system was designed to permit continuous variation in the size of a projected image over a range of four to one without loss of definition or change of brightness. Brightness could be set at the level desired by means of a system of filters.

Among the accessories was a throat microphone circuit by means of which the interval between exposure of a digit and subject's response could be recorded on a polygraph. This part of the equipment was different in detail from, but similar in function to, a circuit described by Roush and Hamburger (38).

A feature not previously described was a shutter control circuit with timer which could be used for single exposures of 1/2 second duration.

The subject faced a translucent screen on which the digits were projected from behind. Viewing distance was 28 inches. A series of hoods shielded the subject and the screen from extraneous light. The screen area was 12 x 12 in. When a digit image was projected in total darkness disturbing autokinetic movements tended to occur. This could be prevented by providing a visual frame of reference in either of two ways: (1) by flooding the entire screen area with very faint illumination, and (2) by flooding only strips about 1-1/2 in. wide by 12 in. high at the lateral edges of the screen. The brightness of this field light was estimated to be about 5×10^{-6} foot-Lamberts, and it could not be seen until after some dark adaptation.

Brightness of digit images was measured directly on the screen with a Macbeth illuminometer. Mean readings of a series were reproducible within plus or minus 2%.

The first step in making digit slides was to prepare cardboard cut-outs about 1.56 in. high. These were pasted on glass, and then reduced in a converted enlarger to a convenient final height, in most cases 0.19 in. Kodalith transparencies in this size constituted the slides. The projected image could be varied from one-half to twice the height of that on the slide.

The experimenter changed image size by rotating a crank. In the experimental routine the rotation was paced by a flashing light. Image size was indicated by a pointer on a continuous scale.

C. General Procedure

The equipment was used in three ways.

1. The Size Threshold Procedure

In this procedure the image size was set well below the threshold of recognition, the shutter was opened, and the size was steadily increased until the subject named the digit. At this point the image size was read from the scale.

2. The Recognition Time Procedure

In this method the size was set at a predetermined value, the experimenter pressed a key which simultaneously opened the shutter and made a mark on a polygraph strip, the subject named the digit as soon as he recognized it, and the response circuit activated by the throat microphone put a second mark on the polygraph strip. To keep time scores within a reasonable range, size of the digit image was adjusted for individual subjects on the basis of performance in preliminary trials. The goal was to produce time scores a little longer than the minimum obtainable under favorable conditions, but shorter than 10 seconds. The practice was adopted of terminating the exposure in 10 seconds if no response had been made, and recording 10 seconds as the score. In most experiments the incidence of such items was quite small. In a few experiments it became obvious early in the routine that a larger number of longer delays was to be expected, and in these cases the practice of arbitrarily terminating exposures was abandoned.

3. The Error Scoring Procedure

In this procedure the 1/2 second exposure was used and size of test image was adjusted to produce a useful percentage of errors.

Choice of procedure in various phases of the program depended on availability of components of the equipment and relative efficiency in gathering data. The recognition time procedure, for example, was capable of producing more scores per experimental period than the size threshold, but became available later in the series and required an additional time investment in analyzing the polygraph record. The error scoring procedure was relatively efficient both in data produced per unit of experimental time and in time required to analyze results.

The size threshold and recognition time procedures were not intended to produce errors. Nevertheless a digit was occasionally miscalled in each of them, and these instances were recorded. In such cases the miscalled digit was repeated later in the series to obtain the size or time score for correct recognition. Errors produced in this way, averaging less than 10% of responses, were too few per digit to provide anything more than a sidelight on the basic data. In the third procedure, on the other hand, results were in terms of errors only, and the difficulty level was set accordingly.

Individual experiments were always planned with balanced designs, particularly, all test items were balanced with respect to mean serial position in the experimental sequence. The repetitions of miscalled digits introduced occasional irregularities which were believed to be of no practical consequence.

In certain experiments the test digits consisted of only three out of the set of 10. In all such cases enough dummy items, made up of the remaining digits in the set, were included to constitute at least 20% of all trials. Digits were presented singly unless otherwise indicated.

All subjects had a minimum of four, usually more, practice trials before the first regular test trial. Some subjects were used for one period only, some for several periods. No systematic relation between the amount of practice and the experimental comparisons was observed.

Except as otherwise indicated subjects were given an opportunity to examine in advance the designs on which they would be tested.

The subject was instructed to watch for the appearance of the digit at the center of the dark area. In the size threshold and recognition time procedures he was told to respond as soon as he could be reasonably sure that he could identify the digit. In the error scoring procedure he was told that some of the items would be difficult, and was encouraged, but not required, to make a judgment on each one. An experimental period was preceded by 20 minutes of dark adaptation.

Subjects were male college students with visual acuity 20/25 or better on the Lebensohn Test at 14 inches. Twenty/twenty-five was accepted as the cut-off point because, though we made no systematic test of it, the Lebensohn Test seemed to be rejecting a larger percentage of cases at the 20/20 level than analogous tests, and also because the exact level was not critical for the comparative evaluations planned. Number of subjects in a group varied from four to 16.

D. Detailed Procedure and Results

1. Digit-Width

A change in digit-width, other dimensions being held constant, means a change in over-all size. Under marginal conditions of visibility size can be critical. There is a corresponding change in height/width ratio which might or might not be of consequence. As a first step in the evaluation of these two factors at our low illumination level, small changes in digit-width were tested.

The four digits, "2", "4", "5", and "8", were prepared in three widths, standard, 10% wider than standard, and 10% narrower than standard. Stroke-width was constant in absolute dimensions. Testing was by the size threshold procedure. The screen was floodlighted to a brightness of about 5×10^{-6} foot-Lamberts. Brightness of digit images projected on this background was 0.017 foot-Lamberts, the one instance among the experiments to be reported in which brightness exceeded 0.011 foot-Lamberts. Eight subjects were given two experimental periods each. In each period, 12 test trials were interspersed with four dummy trials on non-experimental digits. Thus the eight subjects in two periods each provided a total of 16 responses on each of the 12 test items.

Results in terms of size thresholds in inches for the four digits in three widths appear in Table I. As the smaller sizes mean lower thresholds, the trend in Table I clearly favors the wider digits. But the means show a relatively larger difference between the wide and the standard digits than between the standard and the narrow. The former of these differences is significant at the 1% level, the latter is not significant.

The fact that a significant difference was obtained as a consequence of only a 10% increase in width throws some light on the importance of dimensional variables under these conditions. The fact that the narrower digits were not reliably worse than the standard suggests the possibility that the corresponding change in height/width ratio may have compensated for the reduction in size. Experiments were therefore planned to test the height/width ratio as a separate variable.

2. Height/Width Ratio

The plan in these experiments was to vary height/width ratio while keeping area of circumscribed rectangle constant. The first step was the preparation of digits "2", "5", "7", and "8" in five ratios, the widths in percentage of height being 50%, 57%, 67%, 80%, and 100%. The 67% approximated the standard AND design. Stroke-width was constant in absolute dimensions. For digit-widths in the above order, stroke-widths in percentage of height were 10.8%, 11.5%, 12.5%, 13.7%, and 15.3%.

The four digits in five height/width ratios were tried on four preliminary subjects. Results pointed toward a continuous improvement in recognizability as the height/width ratio increased.

More formal experimentation was then planned. The width 100% of height was dropped as the most unpromising of the series. Slides were prepared in four ratios for all of the remaining digits except the "1". (See Figure 3.) Three experiments were done, (1) by the size threshold method, digits presented singly, (2) by the recognition time method, digits presented singly, and (3) by the recognition time method, digits presented in groups.

The lateral strips were faintly illuminated but the 9 by 12 in. area in the center of the screen was kept dark. Brightness of the digit image seen on this dark ground was 0.011 foot-Lambert.

a. Experiment 1, size thresholds, single digits

This experiment was done in three parts, the digits being divided into three sets (258, 067, 349) and each set tested with a different group of subjects. In each of the three parts, 12 subjects had one experimental period each. Two trials on each of the three digits in the four ratios, for a total of 24 test trials, were mixed with eight trials on dummy digits. The 12 subjects thus produced 24 responses on each test item.

Results in terms of size thresholds are shown in Table II. For convenience of comparison with other data the sizes were converted to percentages of those for the AND ratio (width 67% of height) and in this form they appear in Table II and Figure 6.

b. Experiment 2, recognition time, single digits

In the recognition time procedure more trials could be done in an experimental period, therefore all nine digits in their four ratios were shown to each subject. This made it possible to achieve a further economy by eliminating the dummy trials. Twelve subjects had two experimental periods each. Each period included three trials on each of the nine digits in the four ratios, for a total of 108. The 12 subjects produced 72 responses on each test item.

Results in terms of recognition time in seconds are shown in Table III. The data have also been converted to percentages of the scores for the AND ratio (width 67% of height), and in this form they appear in Table III and Figure 6.

c. Experiment 3, recognition time, grouped digits

In this experiment, each test digit was flanked by two other digits, the "0" and the "5". Both the OD5 and the 5DO arrangements were included. Only two ratios were tested, widths 57% and 67% of height. The digit "1" was added to the set, but was varied in height only. Spacing between digits was approximately 5.4% of height, i.e., slightly wider in absolute terms for the taller digits. This is very close spacing, but similar to that found on some dials. Subjects were instructed to identify the middle digit.

Four subjects had two experimental periods each. In each period, four trials were given on each of the ten digits in two grouping arrangements at two ratios, for a total of 160 test trials. The four subjects therefore provided 32 responses on each test item, or pooling the two grouping arrangements for each digit at each ratio, 64 responses per item.

Results in terms of recognition time in seconds are shown in Table IV, and in percentages of values for the AND ratio (width 67%), in Table IV and Figure 6.

In Figure 6 three sets of results are plotted, from Experiments 1, 2, and 3 just described, and also averages of data from Experiments 1 and 2. The latter are arithmetic means of the percentage scores. The lower scores, it will be remembered, indicate better recognizability. These results show that (1) recognizability varies as a function of height/width ratio in quite different ways for different digits, and (2) the optimum for the set of 9 digits is a ratio which gives a width 57% of height. The three sets of data agree reasonably well in supporting these generalizations. The shapes of individual curves, and the occasional contradictory trends shown in different sets of data, cannot be explained in terms of any convenient generalizations. We shall return to certain aspects of this problem in the discussion. At this point, one relevant fact about the more detailed results should be indicated. The grossest discrepancy shown by these data is that between the single and grouped digits in the case of the "4". The mean for the former drops slightly between widths 67% and 57% of height, but the latter rises at width 57% to a much higher value than any other point on any curve. Examination of the original scores by trials shows that two subjects had special trouble with this one test item. One subject in both experimental periods, and the other in one period, failed on all four trials to recognize within 10 seconds the "4" in width 57% when it was flanked on the left by a "5" and on the right by a "0". No similar tendency appeared on any other item. The fact that this happened to two subjects suggests that this particular combination of digits may have presented special difficulties for proper perceptual structuring.

3. Stroke-Width

A fair amount of work on the effect of stroke-width has appeared in the literature, but little of it at the low brightness level of the present study. Our investigation provided a limited amount of reportable data to help fill this gap. The size threshold procedure was used throughout. In a preliminary experiment with four subjects, at a brightness of 0.011 foot-Lambert, the AND "5" and "8", and somewhat modified designs of the "2" and "6" of about the same over-all dimensions, were tested in stroke-widths 10.0%, 12.5%, 15.0%, and 17.5% of digit height. Results showed the 10.0% to be clearly the worst for all four subjects, with no systematic differences among the other three. More formal experiments were then planned.

a. Experiment 1, AND style digits

For this experiment the AND digits "4", "5", "7", and "8" were prepared in the four stroke-widths, 11.2%, 12.5%, 13.7%, and 16.1% of height, a narrower range than in the preliminary test. Over-all height and width dimensions were held constant.

The lateral strips were faintly illuminated, and brightness of the digit images projected against the dark center of the screen was 0.011 foot-Lambert.

Eight subjects were given two experimental periods each. Each period included one trial on each digit at each stroke-width, for a total of 16 test trials, interspersed with four dummy trials. In this routine the eight subjects produced 16 responses on each test item.

Results are shown in Table V. The largest difference between two stroke-widths for a single digit, and the largest difference between means of the four digits, are quite unreliable. These data therefore show no evidence of a stroke-width effect in the range tested.

b. Experiment 2, digit-width 50% of height

In view of the possibility that a taller and thinner design might be advantageous, a stroke-width test was made with digits having a width 50% of height. This approximated Berger's height/width ratio.

The digits "2", "5", and "8" in this ratio were prepared in stroke-widths 8.3%, 10.6%, and 12.7% of height. The 8.3% was not far from Berger's optimum of 7.7% for white on black under reflected light. The same digits in the standard AND design, width 67% of height, were included for comparison. Thus each digit appeared in four dimensional arrangements.

Eight subjects were given one experimental period each. Two trials on each digit in each dimensional arrangement, for a total of 24 test trials, were interspersed with eight dummy trials. The eight subjects produced 16 responses on each test item.

Results appear in Table VI. The difference between the means for the stroke-widths 8.3% and 10.6% was found to be significant at the 1% level. Difference between stroke-widths 10.6% and 12.7% was not significant, nor were any of the differences between the AND design and the tall thin design. On the basis of the relatively poor showing of the stroke-width 8.3%, it seems clear that Berger's optimum for white on black is not favorable for transilluminated digits at very low brightness, even though transillumination gives high contrast. As a by-product of this experiment, it can be noted that the inferiority of the standard AND design to the taller and thinner design in similar stroke-width, while not statistically significant, is consistent with the data of Figure 4 on all three test digits.

4. Spacing

The very close spacing between digits sometimes found on dials may be an unfavorable condition. One experimental test of this was made with the AND design. The recognition time procedure was used with the faintly luminous lateral strips bounding the screen. Brightness of digit image was 0.011 foot-Lambert.

All ten digits were included. Each was presented in the middle position of a three-digit group as in the last height/width experiment. The "0" and "5" were again the flanking digits, and both grouping arrangements, OD5 and 5DO, were used. All groups were prepared with two inter-digit spacings, 5.4% and 16.2% of digit height. Subjects were instructed to identify the middle digit.

Four subjects were given two experimental periods each. Each period contained four trials on each digit in each spacing in each grouping arrangement. Combining the two periods and the two grouping arrangements, the four subjects gave 64 responses on each digit in each spacing.

Results appear in Table VII. Difference between the means for the 10 digits pooled, computed in the usual way with the four subjects treated as a population sample, is not significant. It can be seen, however, that differences by individual digits favor the wider spacing in seven cases out of 10 with one difference of zero. Another breakdown, by subjects, periods, and grouping arrangements, but not by digits, provides 16 pairs of means; of these, 15 favor the wider spacing. It seems safe to conclude, therefore, that at least for these four subjects the effect of spacing was demonstrated.

5. Configuration

It will be recalled that configurational modifications involved no change in stroke-width or height/width ratio, and that they generally fell within the limits of familiar designs for the particular digits. A considerable number of modifications was tested at one stage or another of the program. The general goal in devising new configurations was to minimize common elements in digits likely to be confused. Some of this testing was done before such variables as digit-width were controlled with sufficient precision. Controls improved as the program developed, and results from the early experiments, while of no value for the permanent record, provided leads for further modifications to be tested. The attempt to improve on the AND configurations met with very limited success. The AND "8", for example, ranks low in the set under our testing conditions, so several variations were tried. In nearly all cases scores on the modified "8's" were worse than on the original.

One new design proved reliably better than its AND counterpart, a round-topped "3" (Fig. 4). In a well-balanced experiment eight subjects produced a total of 96 responses on both the AND "3" and the round-topped "3". Scores favored the latter for seven of the eight subjects, and the mean difference was statistically significant. In several less well-controlled preliminary experiments also the trend of the differences was in this direction.

In no other case did the data conclusively favor a new design when all non-configurational factors were known to be equated.

6. Defective Strokes

Interference with strokes by overlying instrument pointers, dirt, and localized glare spots can give the effect of changing one digit into another. A measurement of the effect of defects in strokes systematically varied in position, therefore, might throw light on critical details of design. For the experimental test nine AND digits were used, the "1" being omitted. Each digit was prepared with a segment one stroke-width in thickness cut out in each of eight positions around the digit outline. Positions of the missing segments were adjusted within some limits on the basis of tentative assumptions about the relative importance of stroke elements. Figure 5 shows the positions used. Each test slide carried only one missing segment; eight slides were therefore required for each digit.

This experiment was done by the error scoring procedure, the difficulty level being set to produce a mean incidence of approximately 25% errors. The faint lateral strips were used on the screen, and brightness of digit image was 0.011 foot-Lambert.

Sixteen subjects had one experimental period. Two trials were given on each digit with the missing segment in each of the eight positions, for a total of 144 test trials. This provided a total of 32 responses on each test item by the 16 subjects.

Instructions were formulated to encourage, but not compel, a judgment on each item.

Results appear in Table VIII. Each block in the table represents a test digit. Within a block, rows correspond to positions of the missing segments. Cell frequencies and row totals represent errors made in 32 trials. Column totals show frequencies with which digits heading the columns were erroneously reported.

Several things can be pointed out about these data.

(1) The results are partly intelligible from the test conditions; in most cases, the largest cell entry in a block falls in a column that might have been predicted from the position of the stroke defect.

(2) In several cases this reflects a very small change, as clipping a corner in 4-D and 5-B.

(3) Even the errors which are less predictable from the position of the stroke defect show a tendency to pile up in the same columns as those which are more predictable.

(4) The general pattern of confusions is not very different from that in unmutilated forms. (See the following section.)

(5) Six is the digit most often reported by mistake, the column total for the "6" comprising 25% of all errors. There may be a number preference factor in this, but it is also true that the "6" is at or near the top of the list in terms of the number of other digits with which its contours have ostensible similarities.

In summary, these results suggest both that specific details of design can be important, and that random impairment of the forms will tend to accentuate pre-existing confusion tendencies.

7. Inter-Digit Confusions

The distribution of mistaken identifications among the digits should provide a helpful guide for improvements in design. Three sets of data in this form are shown in Table IX. The first set (IX-A) was compiled from several size threshold experiments which differed somewhat in experimental details. The second set (IX-B) is from a well-balanced recognition time experiment, but one in which the digit "0" was not used. The third set (IX-C) is the Table VIII data from the experiment on defective strokes, converted to percentage terms. This was a well-balanced experiment but one in which the digit "1" was not used. The entry in a cell shows the per cent of trials in which the test digit in the row was mistakenly reported as the digit at the head of the column.

The frequencies are lower in IX-A and IX-B than in IX-C because they are based on procedures not designed primarily to produce errors. Tables IX-A and IX-B include data for the AND design only. The Table IX-C data are for AND digits with defective segments, but all defects are pooled. The several differences among the experimental conditions could be expected to produce some differences among the tables. In spite of this certain consistencies are apparent.

(1) The "3"-"5", "4"-"6", and "0"-"6" appear to be the most troublesome combinations (allowing for the fact that "0" is not a test digit in Table IX-B). The "5"-"6" and "0"-"9" come next in order.

(2) The confusions within a pair are not symmetrical. E.g., "4" is more often mistaken for "6" than the reverse.

(3) The column totals, showing frequencies with which the digits are incorrectly substituted for digits presented, have generally similar distributions. "6" clearly leads the list, with "5", "4", "0", and "9" standing relatively high.

Among the departures from consistency, the most conspicuous shows in the row totals. These totals give errors by test digits. By this criterion, digit "8" has a relatively favorable rank position in the A and B sets of data, but the worst rank position in the C set. This is probably not a sampling accident, and we shall revert to it in the following section.

8. Rank Order of Recognizability

It happens that no single experiment of those here reported provides data for all ten intact digits of the AND design gathered under homogeneous conditions. The rank order of recognizability under our low brightness condition, therefore, can be established more satisfactorily by data from another experiment previously reported (16) in which the recognition time procedure was used. In that experiment, a well-controlled test of all 10 digits was made, at four brightnesses ranging from 0.003 to 0.1 foot-Lambert. With the four brightnesses pooled, the order of decreasing recognizability was found to be "7", "9", "2", "5", "3", "1", "6", "4", "8", and "0".

The "9" and "6" are of special interest because in most digit sets they are the same figure in different orientations. The much better showing of the "9" than the "6" in the above ranking reflects a recurrent tendency in our data and therefore appears to be related to the testing conditions. In such rank orders as have been reported in similar digit styles at higher brightness levels, the "6" usually places a little above the "9". Aldrich, in his study of license plate numerals (1), found the "6" to be among the digits confused least often, and the "9" the one confused most often.

The digit "8" is near the bottom. This in itself is consistent with data from other studies, and with our own data from other tests employing similar criteria. One example is the recognition time experiment from which the error scores in Table IX-B were obtained; rank order of the time scores in that experiment put the "8" in last place. The point of interest is that in terms of errors, in both the size threshold and recognition time procedures, the "8" makes a very good showing (Table IX-A and -B). The explanation for this paradox probably lies in a characteristic of those procedures. In either technique, the subject has some freedom to adjust speed in relation to accuracy. The adjustment can be expected to vary with the subject and the difficulty level, it is not necessarily conscious, and cannot be completely controlled by instructions. As difficulty increases, within a certain range, a tendency to slow down is likely to appear; this tendency is reflected in higher size threshold or recognition time scores, and may be accompanied by a decrease in errors. This probably accounts for the favorable position of the "8" in the hierarchy of error scores. It will be noted that in the error scoring procedure, where no compensatory adjustment by the subject is possible, the "8" is again in last place (Table IX-C).

9. Incidental Observations

Certain qualitative observations made in the experimental situation may throw light on interpretations of the data.

a. The subject's view of the digits

The size threshold procedure is especially suitable for observing details of the perceptual phenomena that determine recognition, because the viewing period extends over several seconds and the perception undergoes dynamic changes. The digit first appears as a small light spot. As the size increases it begins to take on a fluctuating structure. It is not uncommon at this stage for the form to shift rapidly from one digit to another in the test series. Finally one form becomes stabilized. The subject makes his report when the stability is sufficient to make him reasonably confident of the judgment. In this process, the "8" is the most persistently amorphous of the digits, and is likely to produce the longest delay before a clear perception is achieved. This affords another way of looking at the discrepancy between the rank position of the "8" in size or time scores and in error scores. Since the "8" is slow to take on any structure the report is delayed, and it is less likely to be miscalled.

The high confusion between the "4" and the "6" is puzzling at first impression. The distinctive detail about the "6" is the tail sloping up to the right. Under near threshold conditions this tends to be the salient feature of the open-loop "4" also, which probably accounts for mistaking the "4" for a "6". It is unlikely, however, that our conception of the distinctive characteristics of a "4" would put much emphasis on this sloping tail, which probably accounts for the less frequent mistaking of a "6" for a "4".

The digit "1" characteristically gives the appearance of starting farther below the threshold than any other digit, i.e., it looks dimmer. This is seemingly related to the fact that it has less mass, and consequently produces less light, than any other digit. Consequently it is slower to take on structure. If the subject's report is on the basis of a clear perception it would follow that the report should be delayed relative to other digits. It is probable, however, that some subjects make the judgment on the basis of faintness and lack of structure. In the rank orders reported from various investigations the position of the "1" jumps around in a way to suggest a subject variable of this or some similar type.

The same general effects characterize the recognition time as the size threshold procedure, though usually in a less obvious degree.

b. Mental sets

Sanford (39) pointed out that, in a test using letters of the alphabet, subjects would forget one letter or another, which tended to interfere with recognition. Even with digits, a much smaller class than letters, we have experienced this phenomenon. The subjective evidence of it, in the size threshold procedure, is a sudden realization, when a digit becomes clear, that it is not among those which had been held in mind as possibilities.

A complementary effect was sometimes encountered in the perseveration of an error. Occasionally a digit had to be repeated half a dozen times for a particular subject, evoking the same error each time, before correct recognition was obtained.

The extremely long delays by two subjects out of four in the third height/width experiment on the digit group "540" present a problem of interpretation. Either this is an especially difficult clustering of digits for the population in general, or we have another example of fixations by individual subjects. The sampling of subjects was too small to permit a choice between these alternatives.

The kinds of errors will depend in part on the subject's expectations as to whether he is going to be tested on the class of ten digits, a sub-class of three or four digits, or digits mixed with letters. In certain experiments containing only three test digits varied in one dimension or another, the series was padded out by the inclusion of dummy digits. Some sets of results suggested strongly that the dummies had not been adequate to set up the expectation that all ten digits were equally likely to appear. The expectation, of course, will be determined by the developing inferences the subject makes as he progresses through the test period, and this can be expected to vary from subject to subject.

SECTION IV

DISCUSSION

A. Evaluation and Summary of This Study

The exploratory nature of this investigation led to the coverage of a considerable number of variables and types of design in relatively small experiments. This limited the amount of data that could be accumulated per experiment.

Control of visual conditions presented certain problems. The plan was to establish limiting conditions in terms of both photopic and scotopic sensitivity, and of contrast between figure and immediate surrounds. Data obtained under such conditions could provide a baseline for future comparisons. In line with this policy, a minimum of accessory stimulation, and correspondingly a minimum of structure, was provided in the visual field by means of the faint lateral strips on the projection screen. These conditions resulted in some relaxation of control of fixation and accommodation, which could be expected to increase the variability of the results.

For such reasons, it can be inferred that trends in the data showed clearly only for variables having fairly marked effects.

With such considerations in mind, the findings can be summarized as follows.

(1) Under low brightness conditions, mere size was found to be critical. A small increment of digit-width is such a size variable. It seems indicated that the possibility of a size effect should be recognized in testing situations which are near threshold in other respects as well (exposure time, contrast, etc.).

(2) Though it would be difficult to demonstrate by pointing to particular sets of data, it seemed easier to measure the effect of dimensional changes than of configurational changes. There may be complex reasons for this, residing in characteristics of the experimental situation and in the nature of the cues that determine recognition. We doubt if it means that the effect of configurational differences within the range of familiar patterns is of no practical consequence, or cannot be measured.

(3) A definite and complex effect of height/width ratio (area held constant), within the range of ratios commonly used, was demonstrated. If it should be necessary to establish an optimum for the set of digits on the basis of our data, a mean digit-width in the range of 55% to 60% of digit height would be indicated. It is possible that something could be gained by adjusting width for the optimum by individual digits.

(4) Stroke-width effects were not entirely clear cut, but the range of stroke-widths tested was small. Results suggested that under our conditions a reduction in stroke-width below a value somewhere in the neighborhood of 10% of digit height would produce a substantial impairment.

(5) Some evidence was obtained showing an adverse effect of very close spacing.

(6) A round-topped "3" was found to be better than a flat-topped "3", on the basis of one well-balanced comparison and a body of less substantial supporting data.

(7) Small defects in digit contours were found to contribute both to types of confusion predictable from the defects, and also to types of confusion not readily predictable. A record of such effects should be helpful in improving digit design.

(8) Inter-digit confusions showed reasonable consistency from one low-brightness testing situation to another, and to some extent could be explained in terms of details of design.

(9) A number of observations on both the perceptual phenomena associated with digit recognition and on the performance of individual subjects suggested the probability of a large interaction of subjects with experimental variables and other aspects of the testing situation.

These several sets of data were developed in partly overlapping phases, and did not all become available for analysis and cross comparison until late in the program. An obvious next step would be the formulation of hypotheses for further testing on the basis of such an analysis.

Another step would be the testing of some designs on the third level of novelty listed in the Introduction---designs which transcend familiar patterns, but which are not too radical to be recognized at a glance or learned easily in one trial. The digit "8" in particular invites this treatment, because of its extraordinary refractoriness in conventional patterns. A start was made along these lines in the final stages of the present study, but the schedule did not permit an adequate follow-up.

A further investigation for further investigation might be measurement of the effect of reduced contrast between figure and ground at low brightness. In typical operational situations contrast is less than maximal, but varies considerably from one situation to another. Some of the stroke-width data in the literature suggest a complex interaction between brightness and contrast.

B. The Weight of Evidence on Principles of Design

Does the literature permit any useful generalizations? In answering this question we shall concern ourselves primarily with numeral styles which are uniform in height, stroke-width, and horizontal alignment. This includes dial-type numerals (AND, AMEL, Craik, Leroy), and numerals intended for distance viewing (Mackworth, Berger, and Aldrich's license plate numerals). The latter are characteristically narrower than the former, but the two types have sometimes been combined in experimental tests.

Among the studies more frequently cited, a number have been done with reference to particular applications. Commonly in these cases the authors have warned against extrapolating to other situations. Such warnings are well advised, because the records suggest a considerable degree of specificity in the effects.

A possible example of such specificity is the differences in rank order of the ten digits as reported by different investigators. The "7" most frequently stands at the top of the rank order, and the "8" at or near the bottom, but positions of the other digits are less predictable. A certain amount of random shifting is to be expected, but some shifts appear to be related to testing conditions. The better showing of the "9" than the "6" in our experiments, in contrast to the findings in other studies, was mentioned above. Two possible explanations for this discrepancy suggest themselves; (1) a dependence of perceptual similarities on brightness level, and (2) a dependence of confusions for a particular digit on the design of other digits in the set. The "4" with the open loop used in our tests was especially liable to confusion with the "6", and perhaps to some extent vice versa. In the Aldrich study (1), where the "6" showed to good advantage, a "4" with a closed loop was used.

Reports of interaction effects provide one line of evidence pointing toward specificity. Kuntz and Sleight (25) found all of the first order interactions among subjects, brightness, contrast relation, stroke-width, and digit, except that between stroke-width and contrast relation, to be significant at the 1% level. Second order interactions were not computed. Schapiro (40) carried out alternative computations on his data after certain transformations and adjustments. The first order interactions which survived these treatments best were illumination by subject, illumination by style, and illumination by stroke-width. Less conclusive indications of interaction effects were found for style by digit and subject by digit. All of the second order interactions among subject, illumination, style, stroke-width, and digit were significant at a high confidence level. Berger (4, 5) demonstrated an interaction between digit and width. Loucks (27, 28) and Brown, Lowery, and Willis (11) found interactions between brightness and stroke-width. The Atkinson, Crumley, and Willis data (2) suggested an interaction between brightness and style. In Soar's investigation (41) the variables were height/width ratio and stroke-width. Subjects' variance was eliminated by a covariance technique, and variance analyses were then done by individual digits. Results pointed to interactions between digit and both height/width and stroke-width, though a statistical measure of these effects was not provided. Interaction between height/width and stroke-width was computed for each digit, and found not to be significant in any instance. This is contrary to what might have been expected from the multiplicity of other interactions reported, and the literature provides no other data on this effect with which Soar's finding can be compared. Regardless of this one instance, the mass of evidence tells us that interactions are rather the rule than the exception.

With this warning in mind, we shall return to the question what consistencies can be discovered.

1. Digit-Width

In view of the fact that styles of different width are sometimes experimentally compared, it is worth while to emphasize that when height is held constant differences in width produce differences in size, and this in itself can determine recognition under near threshold conditions. In addition to the limited data of the present study, Aldrich (1) and Berger (4, 5) have reported evidence for this effect.

2. Stroke-Width

A cursory examination shows that it is important to relate stroke-width data to brightness level, and perhaps also to contrast relation and degree of contrast. The interaction at least with brightness level appears to be larger than that with digit-width or style. Therefore we shall tentatively lump together the several digit styles on which stroke-width has been investigated.

It can be provisionally assumed that if there is any difference in optimum stroke-width between black digits on white and white on black, the latter would have the narrower optimum. The literature on digit strokes supports this with one set of data (3) and shows no definite reversals. Background data on visual discriminations suggest that a difference is more likely to be found at high than at low brightnesses.

Transillumination typically produces higher contrast than illumination by reflection. It is a reasonable inference that high contrast favors narrower strokes, especially at high brightness. Therefore if any stroke-width difference exists, the optimum should be narrower for transillumination than for reflected illumination.

Fluorescent strokes are visually more similar to transilluminated strokes than to strokes illuminated by reflection, and unless the contrary is indicated, can be presumed to have low brightness.

If there is an effect of exposure time, a priori considerations suggest that very brief exposures would favor broader strokes.

These propositions will constitute tentative assumptions for our attempt to systematize the stroke-width data.

At illumination or brightness levels from daylight down to about 5 foot-Lamberts the types of numerals that concern us have been tested under reflected light. From 5 down to the neighborhood of 1 foot-Lambert both reflected illumination and transillumination have been used. At very low brightnesses we have either transillumination or fluorescence, the Craik study being an exception.

Most sets of data indicate fairly broad regions of negligible stroke-width effect. In the following analysis, therefore, an indicated optimum should be regarded as the mid-point of a region perhaps five percentage points wide.

a. Black-on-white, reflected light

Considering first black digits on a white ground under daylight, Berger (3) found a stroke-width optimum at about 12.5% of digit height, and Aldrich's data (1) are not inconsistent with this.

It is necessary to interpret the results of the Kuntz and Sleight study (25) in light of their method of measuring digit height to the mid-points of outer strokes. As a consequence of this technique an increase in stroke-width would produce an increase in over-all digit height and width. It could be expected that, with the distance criterion used, digits in wider strokes would produce better scores because of size, and the indicated stroke-width optimum would creep up accordingly. In comparison with other data, the Kuntz and Sleight optimum strokes do in fact seem rather broad, and it is probably unwise to accept them without some correction.

It is possible to get an estimate of the effect of brightness from the Kuntz and Sleight data. Their Figure 4 suggests an increase in stroke-width optimum of perhaps three or four percentage points as illumination decreases from 31 to 3 foot-candles. The size factor could be assumed to operate throughout the illumination range (though perhaps not to exactly the same extent throughout) and therefore we can use the three or four percentage points to make illumination adjustments in other data. One additional qualification is necessary. It is more consistent with our tentative assumptions above, and with the remainder of the stroke-width data, to assume that the optima for black-on-white and white-on-black converge as illumination decreases. As will appear below, this calls for a somewhat larger adjustment for white-on-black than for black-on-white. We shall therefore arbitrarily allow four and one-half percentage points for the former and two and one-half for the latter.

If we assume 31 foot-candles to be visually equivalent to daylight, within the limits of accuracy that concern us, the adjustment of two and one-half percentage points can be applied to the black-on-white optimum at 12.5% of digit height indicated above. This would bring the optimum for this contrast relation at

3 foot-candles to about 15%.

Going down another step in the illumination scale, Soar (41) covered a wide range of stroke-widths at 1 foot-candle. Positions of optima by individual digits, though in most cases not conclusively established, differed widely from one digit to another. With all digits pooled the stroke-width effect seemed to be relatively slight. It is difficult to judge the implications of this for the pattern shown in other studies, because Soar's investigation was unique in a number of important respects.

b. White-on-black, reflected light

For white digits on a black ground under daylight, Berger (2) found an optimum at 7.7%, considerably narrower than for black on white. The 7.7% was confirmed by Berger in a night lighting situation but the actual illumination was probably several foot-candles. Loucks's results (27, 28) on white dial digits under incandescent light (intensity not stated) were not inconsistent with this finding.

The Brown, Lowery, and Willis data (11), at 40 and 80 foot-candles, point to an optimum somewhere above 11.6%, considerably higher than the Berger figure. This study, however, had the special feature of a very short exposure time (0.007 sec.), and there is a good possibility that this operated to favor the broader strokes.

The Kuntz and Sleight results (25) for white-on-black differed from Berger's in two respects. (1) The optimum was wider, which can be ascribed to over-all size as explained above, and (2) on the basis of a comparison within the experiment, the optimum was not found to be different for the two contrast relations. This is harder to account for. It might mean (1) that the difference was obscured by the confounding of stroke-width with size, (2) that the difference tends to disappear with reduction of illumination level, or (3) that there is no true difference for digits generally, Berger notwithstanding. We favor a combination of factors (1) and (2) as the most likely explanation.

If we estimate an optimum for white-on-black at 3 foot-candles, applying the Kuntz and Sleight factor of four and one-half percentage points arrived at above to Berger's original 7.7%, we get an optimum for this contrast relation of about 12%.

At a somewhat lower illumination level Reinwald (35) compared the AND, Mackworth, and Berger designs in white-on-black at various viewing angles. Brightness of white strokes was about 1.3 foot-Lamberts. The Berger digits were considerably less recognizable than the other two sets, most probably because of the Berger stroke-width 7.7% of height as compared with the AND and Mackworth 12.5%. The Berger optimum for daylight is apparently too narrow for the lower brightness, which is in accord with expectation.

Finally, at brightnesses below 0.01 foot-Lambert, Craik (14) recommended strokes which we have estimated to be about 20% of height for white digits on black.

c. Light-on-dark, transillumination

Berger's optimum strokes about 2.5% of height for transilluminated license plate digits constitute an extreme case in this category. Stroke brightness was not stated, but it may have been quite high, as the contours were cardboard cut-outs in front of 10-watt lamps. Contrast was presumably also quite high, and this combination of conditions could be expected to produce a very narrow optimum.

Brown, Lowery, and Willis (11) measured the effect of stroke-width at brightnesses from 0.33 to 3.35 foot-Lamberts, using exposures of 0.20 and 0.04 sec. In the upper part of this brightness range, an optimum estimated from their two digit styles and two exposure times pooled would be in the neighborhood of 12.5%. This is not far from the 12% estimated above for white-on-black under reflected light. Here again we have limited exposure times but the intervals are much longer than in the daylight part of the experiment. Some evidence for an effect of exposure time can be seen in the stroke-width curves for the transillumination conditions of their Plates 3 and 4. If we assume a small correction for exposure time, and arbitrarily set the value at two percentage points, the 12.5% figure drops to 10.5%, a little below the estimated optimum for white-on-black under reflected light. This is in good accord with our tentative assumptions.

As brightness goes down to 0.33 foot-Lamberts in the Brown, Lowery, and Willis study the optimum tends to shift toward a slightly higher value.

Looking at the still lower brightness region from 0.01 to 0.1 foot-Lamberts, the Loucks data are compatible with a broad optimum centering around perhaps 18%. Schapiro, at 0.011 foot-Lambert, found scores to improve with increasing stroke-width up to 20%. Our own results, which fail to show a difference between 11% and 16%, are a little out of line here, but we are not inclined to weight them heavily because of the small amount of data.

The stroke-width analysis can be summarized as follows, assuming unrestricted exposure time and allowing for a fairly broad indifference region around each optimum.

For black strokes on a white ground under reflected light at daylight illumination levels, the estimated optimum stroke-width is 12.5% of digit height; at 3 foot-candles the estimate is 15.0%, and at lower illuminations a still wider optimum can be predicted.

For white-on-black under reflected light at daylight illumination the estimated optimum is 7.7%, at 3 foot-candles, 12.0%, and at brightnesses below 0.01 foot-Lambert, 20%.

For light strokes on a dark ground with transillumination at very high stroke brightnesses the optimum is 2.5%, at 3 foot-Lamberts 10.5%, and at brightnesses in the region of 0.01 foot-Lambert, 18%.

Limited exposure time would tend to increase the estimated values. Some adjustments might be necessary for effects of digit width or style. We have taken Berger's experimental figures as points of departure, but they were determined on three digits out of the set of 10, and a correction might ultimately be required on this ground.

For such reasons we do not set great store by the exact figures arrived at. The analysis seems to demonstrate, however, that optimum stroke-width is a function of multiple variables, and that a set of not unreasonable assumptions makes it possible to reconcile most of the data. The figures should be regarded as first approximations, subject to further experimental checks.

3. Height/Width Ratio

Height/width ratio was investigated as a separate variable, area being held constant, by Soar (41) and by the present authors. The two experiments agreed in finding different results for different digits. They did not show good agreement on the details of the results by digits nor for the set of digits pooled. Our data pointed to an optimum for the set at a ratio in which the width was 55% to 60% of height, while for Soar's optimum the digits were broader. No explanation for these discrepancies is apparent, partly because the two experiments differed in many ways. Soar used a higher brightness, his test digits were in the AMEL style while ours were in the AND, in his data a considerable range of stroke-widths were pooled, and the height/width values used were different in the two cases, which made point-by-point comparisons on the performance curves difficult. A more analytic attempt to reconcile the two sets of data would therefore hardly be justified at this time. Two implications seem to follow from either experiment: (1) the interaction of factors determining such results is complex, and (2) there is a reasonable possibility of achieving a net gain in recognizability by adjusting widths digit by digit. Berger (3) made a somewhat analogous adjustment with his license plate digits.

4. Configuration

Scattered findings showing evidence for the superiority of one configuration over another by individual digits can be summarized. A number of these are from the Brown, Lowery, and Willis and Atkinson, Crumley, and Willis tests which included AND and Berger styles. It will be recalled that in these tests all styles were brought to the same digit-width, 60% of height; this narrowed the AND and broadened the Berger design.

On the digit "2", Atkinson, Crumley, and Willis found the AMEL design to be better than the Berger style. The latter was mistaken for a "7" an excessive number of times under high reflected illumination, but not with the lower transillumination.

On the "3", the same investigators found some evidence that the AND style "3" was inferior to both the AMEL and the Berger style. The latter two are more open and more differentiated from the "5" in the upper half. The present authors found the experimental "3" of Figure 4 to be more recognizable than the AND "3"; the former is more open and has a round top. Aldrich also found a round-topped "3" to be better than one with a flat top.

On the "4", the evidence for the superiority of the closed over the open loop seems convincing. The closed Berger style "4" scored better than the open AND style for Brown, Lowery, and Willis and Atkinson, Crumley, and Willis. The latter found the AMEL closed "4" better than its AND counterpart, but here there was a small advantage in width. Error distributions from various studies indicate that the confusions between "4" and "6" are relatively fewer when the former has a closed loop.

For the digit "5", Atkinson, Crumley, and Willis found the AMEL design to be consistently better than either the AND or the Berger style. Inspection of the error distributions does not suggest a simple explanation for this.

On the digit "9", Brown, Lowery, and Willis found some evidence for superiority of the AND over the Berger style, and Atkinson, Crumley, and Willis some evidence for superiority of the AMEL over the Berger style. Again no simple explanation is apparent.

On designs of the "6", "7", "8", and "0", the literature discloses little of interest.

The digit "1" is in a marginal position because its dimensional and configurational characteristics cannot be separated. With this stipulation, it should be mentioned that Aldrich in his study of license plate numerals found the "1" with the foot piece definitely better than the "1" without.

The more conspicuous configurational characteristics afford a partial, but not complete, explanation for differences in recognition scores. If a design produces a high score, it does not necessarily follow that the design is intrinsically superior. It may be superior only in the context of the particular set with which it was tested.

Partly because of this interdependence it is very difficult to predict the effect of any particular configurational change. A digit element may be modified on the basis of confusion data and obvious similarity to some other digit, but it frequently develops on testing that new confusions have been introduced.

Prediction is made difficult, also, by such factors as number preferences, and the subject's expectation as to whether he will be shown ten digits or five digits or digits mixed with letters.

Responses may be determined to an unknown extent not only by the forms actually shown, but also by the subject's generalized conception of the forms in the set; for example, if a not quite decipherable test item suggests a "5", the subject may mentally compare it with an imagined "5" quite different from the "5" on which he is being tested.

Finally, it is very difficult to visualize, without making observational checks, what details will prove to be most salient, and will give a form its distinctive character, in a near-threshold situation.

5. Spacing and Grouping

These are not design variables in the strict sense, but factors with which design can interact.

In our height/width experiments a "4" of a particular design was very much more troublesome for two subjects when presented in a group than it was for any subject when presented alone. To cite an example from the more marginal situation of printed numerals, Tinker (43) found that grouping modified the rank order.

Extra spacing should tend to minimize these problems. Our data pointed to the superiority of spacing 16.2% of digit height over 5.4%. Wider spacings were not tested. Berger with his license plate numerals settled on a mean spacing about 22% of digit height.

6. Digits in Sets

Experimental comparisons between existing digit sets in different designs are complicated by differences in mean digit-width. With sets equated in height, the wider digits have an advantage in size which may be substantial. It is possible that this could be offset by a more favorable height/width ratio in the narrow digits, but on present evidence the height/width effect is not clear, and it seems probable in any case that it would be smaller than the size effect. This problem has sometimes been by-passed by equating the sets in width, but this results in a departure from the original designs. For some purposes the most meaningful comparison between sets in different designs would be on an equal area basis, but this has not been tried. It is necessary to take cognizance of these factors in evaluating evidence on the merits of different designs.

In our primary category of dial-type digits, Figure 1 shows the AMEL, the AND, the Craik, and the Leroy sets. This grouping is on the basis of general similarity rather than specific use.

For the Leroy, no comparative evaluations are available. It could be tentatively inferred that, in comparison with other designs on an equal area basis, the recurved tails of the "6" and "9" would cause trouble, as they appear to increase the general similarity of the rounded forms. The "5" also is more similar to the AND than the AMEL "5", the latter of which did well in an experimental test.

A modification of the Craik to cover a range of stroke-widths got a relatively unfavorable rating in the Schapiro experiment.

The AND in its original design held up relatively well in Reinwald's experiment, and the stroke-width variations of it did well in the Schapiro test. In a modified digit-width, Brown, Lowery, and Willis found the AND style not quite as good as a corresponding modification of the Berger, and in a similar comparison Atkinson, Crumley, and Willis found it definitely not as good as the AMEL. Some possible shortcomings of it are implied in the design principles formulated by Brown, Lowery, and Willis on which the AMEL set was based.

The AMEL set was designed for aircraft control panels. Atkinson, Crumley, and Willis explicitly avoided recommending it for dials, but promised a separate evaluation of it for that use. There is nothing to suggest that it would be less suitable for dials than the AND. It is the most recent of the designs in this general category. Besides being based on a systematic analysis of available evidence, it has held up well in experimental comparisons with the modified AND and Berger styles matched in digit-width. Two questions might be raised about it.

(1) Some of the evidence cited above favors a round-topped "3". Would such a "3" be good also in the context of the AMEL design?

(2) The AMEL stroke-width 12.5% of height was arrived at as the best compromise for a wide range of brightnesses above 0.1 foot-Lambert. If brightnesses well below 0.1 foot-Lambert should be contemplated on dials or elsewhere, it would be appropriate to ask for both the AMEL and the AND designs whether broader strokes would be a better compromise.

Turning to license plate digits, the main point of interest about Aldrich's study is the superiority of the Open design over the others, of which the Round of Figure 1 is a sample. Judged impressionistically, the Open set as a whole appears to reduce the similarity in digit outlines. Thought of in terms of the appearance of digits in test situations, the Open style digits tend to be differentiated by larger features of the outlines than the curved-in styles, i.e., by features which have a better chance of being discriminated under near-threshold conditions.

The Berger design was worked out in much detail, some general limits of style and height/width ratio having been assumed. In the Atkinson, Crumley, and Willis experiment the Berger style in greater digit-width than the original design was found to be less recognizable than the AMEL. This result is suggestive because with the digit-width adjustment the Berger style conforms to the AMEL design principles in most respects, angularity of contours being the main exception. It is possible that angularity per se is an unfavorable condition, or it is possible that angularity interacts with height/width ratio, being more suited to the original license plate ratio than to the modification in broader digit-width.

Evidence on the Mackworth digits is difficult to evaluate. As pointed out in the review of the literature, Mackworth's own experiment did not give a clear test of the digits as a separate group. Reinwald's experiment, in which he found the Mackworth digits insignificantly less recognizable than the AND when tested at various viewing angles, is subject to some of the same uncertainty because the test digits were interspersed with letters. In this respect the Schapiro study provided a simpler test situation, but one which was complicated by the inclusion of a range of stroke-widths. Also, as indicated in the review, it is difficult to reconcile the relatively good showing made by the Mackworth digits at 0.011 foot-Lambert with their disadvantage in over-all size. It is to be hoped that additional experimental evidence will be brought to bear on this problem.

The Ordnance digits shown in Figure 2 do not fall within our category of uniform height and alignment. They are illustrated here because of the somewhat related situations in which they are used, and because they bring into focus the question of the effectiveness of uniform as compared with non-uniform digit sets in dial situations generally. The relative merits of the two arrangements could be expected to depend in part on both space limitations and illumination conditions. A carefully planned experimental test should prove of interest.

C. Miscellaneous

1. Methodology

It may be helpful to recapitulate certain problems of technique.

(1) It is reasonable to expect an increase in variability with a decrease in control of fixation and accommodation. Conditions related to such control may also affect contrast and level of visual sensitivity. It is necessary to work out arrangements of the visual field best suited to the different, and sometimes conflicting, requirements in the particular experiment.

(2) In the type of function being investigated, data from small groups are subject to normal variability, and are also liable to marked biases resulting from fixations and other idiosyncrasies of individual performance. The optimum number of subjects must be decided in relation to the goals of the project. For conditions such as those of this study, four to eight subjects make a small group, suitable for exploratory tests. Twelve to twenty might be adequate for more critical comparisons. Other things being equal, it would be safer to spread a given amount of data over many subjects with few repetitions per subject than the reverse.

(3) Experimental results, particularly the kinds of confusions which occur, will depend on the subjects' expectations with respect to the class of characters on which they are going to be tested. In digit experiments, the most convenient class for reference and cross comparison is that which contains all ten digits and nothing else. If circumstances dictate the use of a sub-class, a possible loss of comparability with data from other experiments should be recognized. Subjects' expectations will depend not only on what they are told but also on their cumulative experience with the sequence of trials.

(4) Testing procedures in which it is possible for the subject to control the balance between speed and accuracy are more naturalistic than tachistoscopic techniques, but sometimes result in shifting adjustments which confuse the data. Confusions of this type can be minimized if speed and accuracy scores are evaluated in relation to each other.

Christensen reported a study (13) in the not unrelated area of dial reading which included certain variables of experimental methodology. Exposure time was found to have strong interactions with other variables. This tends to confirm some of the more specific assumptions of the present study, and directs attention to the general problem of adapting experimental technique to the question the research is intended to answer.

2. Special Operational Situations

Digits are used in physical and psychological settings. Recognizability may be a function of many aspects of the settings. Some have already been discussed at length. A few other examples can be cited.

(1) On dials, position of a numeral is one cue for recognition. It was suggested by the Loucks data that if the numbers are too large to be readily associated with the correct scale marks, more may be lost for this reason than is gained from size alone.

(2) Dial pointers sometimes overlap the scale numbers. The tip of a pointer may merge with a contour in such a way as to perceptually change one digit into another. If this type of conflict is not otherwise avoidable, it might be possible to modify the digit configuration to minimize the confusion.

(3) Digits sometimes occur in limited groups. On a given dial face the numerals may be multiples of two, five, or ten. A digit in the second place may be flanked on the right always by "0" or "5", or by "0" only. The fewer the alternatives, the easier the discrimination in any case, but it might be possible to improve the discrimination still further by adapting the designs to the particular context; with some digits eliminated, designs could be manipulated more freely to enhance the differentiation among those retained.

3. Unsolved Problems

An experiment can seldom duplicate an operational situation exactly. Therefore applications of research data usually involve assumptions and approximations. This process is less likely to miscarry if effects of the mechanisms involved are understood in broader terms than those of the specific operation. Research directed toward this goal is likely to require the exploration of sets of related variables. A few areas within the field of digit design where such exploration might prove fruitful can be mentioned.

a. Height/width ratio

The effect of height/width ratio under conditions of constant area appears to vary with the pattern of the individual digit, and perhaps also with other variables. Existing data cannot easily be systematized, but they can serve as a reference point for further work.

b. Relation of design to space limitations

If space limitations did not exist, many problems of design could be by-passed by making the digits larger. The problem is to achieve maximum recognizability in the space available. This is related to the problems of height/width ratio, inter-digit spacing, non-uniformities in size, and extensions above or below the line.

c. Configuration

The two preceding areas can be thought of as presenting problems of systematic measurement. Configuration is in a different class because the psychological dynamics seem to be elusive. The need here is to isolate the major variables and describe them in meaningful terms. Several lines of attack suggest themselves.

(1) An analysis of a variety of empirical data only recently available, such as confusion tables and the results of our study of defective contours.

(2) A coordination of digit recognition data with data on the mechanisms of visual discrimination. This is the approach taken by Berger, particularly in his more recent work (6, 7).

(3) An examination of the actual appearance of digits in near-threshold situations. The first step would be descriptive. The purpose would be to determine which characteristics of design are most distinctive. It might be discovered that the relative importance of different characteristics changed with the nature of the test situation.

d. Novel designs

This is closely related to the more general problem of configuration. A less analytic and more empirical attack might pay in the early stages, but not indefinitely. The problem would be to work out more recognizable designs on novelty level 3, where certain conventional restrictions are abandoned but no great burden of new learning is imposed. In both areas c and d it might be desirable to take cognizance of production technology. For example, the engraving processes sometimes used for dial-type numerals are not well suited to non-uniform strokes.

e. Evaluation of standard designs

This is a different type of problem area from the four above. The experimental comparisons of different digit styles have served the purposes of particular investigations, but have often been inconclusive with respect to the merits of standard designs as originally specified. A possible approach here is a comparison of designs on an equal area basis.

4. Supplementary Note

As the final typing of this manuscript was nearing completion an article by Lansdell (47) came to hand reporting an evaluation of a set of digits designed on our novelty level 3. The new digits are more angular than their conventional counterparts, have non-uniform strokes, and the loops are filled-in. They look quite different from any familiar forms, but could be learned in one trial. They were experimentally compared with a somewhat broadened set in the Mackworth style. On the basis of somewhat limited data, the angular design proved better when presented singly, but not when presented in groups.

SECTION V

CONCLUSIONS

For transilluminated digits in the Air Force-Navy Aeronautical style at very low brightness, the experimental results reported here justify the following conclusions.

(1) Over-all size is important for recognition, which must be

taken into account in comparing different digit designs.

(2) It is easier under these conditions to measure the effect of dimensional changes than of configurational changes. Dimensional variables in this context are height/width ratio and stroke-width, not over-all size.

(3) With area held constant, height/width ratio has effects which vary from digit to digit and cannot easily be generalized.

(4) Very narrow stroke-widths and very close inter-digit spacing are disadvantageous.

(5) An open style, round-topped "3" is more recognizable than a flat-topped "3".

(6) The types of confusions that occur both with digits having intact contours and with digits having small defects show reasonable consistencies which should be of value in further analyses of problems of design.

(7) Observations on the perceptual phenomena associated with digit recognition and on the performance of individual subjects can throw light both on the psychological mechanisms involved and on problems of experimental procedure.

Certain additional conclusions can be drawn from a collation of the literature.

(1) Many interactions affect digit recognition, involving both variables of digit design and variables of experimental procedure.

(2) The stroke-width data reported by different investigators show a number of superficial contradictions, but also a general trend in the direction of an increase in optimum strokes as brightness decreases. With the aid of certain plausible assumptions about the effects of contrast relation, degree of contrast, and exposure time, most of the contradictions can be resolved and the data fitted into a coherent pattern.

(3) Comparative tests of different digit styles have usually involved modifications of original designs or special features of the testing situation which it is necessary to take into account in interpreting results.

(4) Among dial-type numerals, the AMEL set (Aeronautical Medical Equipment Laboratory, Navy) has been most systematically validated, but up to this time specifically for use on aircraft control panels.

BIBLIOGRAPHICAL REFERENCES

1. Aldrich, M. H. Perception and visibility of automobile license plates. Proc. 17th annual Meeting Highway Res. Bd., 1937, 393-412.

2. Atkinson, W. H., Crumley, L. M., and Willis, M. P. A study of the requirements for letters, numbers, and markings to be used on trans-illuminated aircraft control panels. Part 5 - The comparative legibility of three fonts for numerals. TED No. NAM EL-609 Part 5. United States Navy, Naval Air Material Center, Naval Air Experiment Station, Philadelphia, Pa., June, 1952.

3. Berger, C. Stroke-width, form and horizontal spacing of numerals as determinants of the threshold of recognition. I and II. J. appl. Psychol., 1944, 28, 208-231, 336-346.

4. Berger, C. Some experiments on the width of symbols as determinant of legibility. Acta Ophthal., 1948, 26, 517-550.

5. Berger, C. Experiments on the legibility of symbols of different width and height. Acta Ophthal., 1950, 28, 423-434.

6. Berger, C. Legibility of letters. Nordisk Medicin, 1951, 45, 788ff.

7. Brown, F. R. A study of the legibility of trans-illuminated markings in aircraft cockpits. TED No. NAM EL 600 Part 2. United States Navy, Naval Air Material Center, Naval Air Experiment Station, Philadelphia, Pa., February, 1949.

8. Brown, F. R. A study of the requirements for letters, numbers, and markings to be used on trans-illuminated aircraft control panels. Part 4 - Legibility of uniform stroke capital letters as determined by size and height to width ratio and as compared to Garamond bold. TED No. NAM EL-609 Part 4. United States Navy, Naval Air Material Center, Naval Air Experiment Station, Philadelphia, Pa., March, 1953.

9. Brown, F. R., and Lowery, E. A. A study of the requirements for letters, numbers and markings to be used on trans-illuminated aircraft control panels. Part 1 - The effect of stroke width upon the legibility of capital letters. TED No. NAM EL-609 Part 1. United States Navy, Naval Air Material Center, Naval Air Experiment Station, Philadelphia, Pa., September, 1949.

10. Brown, F. R., and Lowery, E. A. A study of the requirements for letters, numbers and markings to be used on trans-illuminated aircraft control panels. Part 2 - A survey of pilot preferences for markings for rheostat controls. TED No. NAM EL-609 Part 2. United States Navy, Naval Air Material Center, Naval Air Experiment Station, Philadelphia, Pa., February, 1950.

11. Brown, F. R., Lowery, E. A., and Willis, M. P. A study of the requirements for letters, numbers and markings to be used on trans-illuminated aircraft control panels. Part 3 - The effect of stroke-width and form upon the legibility of numerals. TED No. NAM EL-609 Part 3. United States Navy, Naval Air Material Center, Naval Air Experiment Station, Philadelphia, Pa., May, 1951.

12. Chapanis, A., Garner, W. R., and Morgan, C. T. Applied experimental psychology: Human factors in engineering design. New York: John Wiley & Sons, 1949.

13. Christensen, J. M. Quantitative instrument reading as a function of dial design, exposure time, preparatory fixation, and practice. WADC TR 52-116. United States Air Force, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, September, 1952.

14. Craik, K. J. W. Instrument lighting for night use. FPRC 342. Great Britain, Air Ministry, Flying Personnel Research Committee (Psychological Laboratory, Cambridge), July, 1941.

15. Crook, M. N. Printed materials, maps, and charts. Chapter 2 in Human factors in undersea warfare. Washington: National Research Council, Committee on Undersea Warfare, 1949.

16. Crook, M. N., and Baxter, F. S. Recognition time for dial-type numerals as a function of size and brightness. AF TR 6465. United States Air Force, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, March, 1951.

17. Crook, M. N., Harker, G. S., Hoffman, A. C., and Kennedy, J. L. Effect of amplitude of apparent vibration, brightness, and type size on numeral reading. AF TR 6246. United States Air Force, Air Materiel Command, Wright-Patterson Air Force Base, Ohio, September, 1950.

18. Crumley, L. M., and Willis, M. P. A study of the requirements for letters, numbers and markings to be used on transilluminated aircraft control panels. Part 6 - A survey of pilot preferences for special attention plate markings. TED No. NAM EL-609 Part 6. United States Navy, Naval Air Material Center, Naval Air Experiment Station, Philadelphia, Pa., September, 1951.

19. Dunlap, K. Report of the Highway Research Board, National Research Council, Division Office, App. E., p. 3, article 4, Washington, D. C., 1932.

20. Forbes, T. W., and Holmes, R. S. Legibility distances of highway destination signs in relation to letter height, letter width, and reflectorization. Proc. 19th annual Meeting Highway Res. Bd., 1939, 321-325.

21. Forbes, T. W., Moscowitz, K., and Morgan, G. A comparison of lower case and capital letters for highway signs. Proc. 30th annual Meeting Highway Res. Bd., 1950, 355-373.

22. Gleason, J. G. The design of numerals for use in counter-type instruments: A review of the literature. No. 166-1-39. United States Navy, Special Devices Center, Port Washington, N. Y., December, 1947.

23. Kappauf, W. E. Studies pertaining to the design of visual displays for aircraft instruments, computers, maps, charts, tables and graphs: A review of the literature. AF TR 5765. United States Air Force, Air Materiel Command, Wright-Patterson Air Force Base, Ohio, April, 1949.

24. Kappauf, W. E. The design and use of instruments. Chapter 3 in Human factors in undersea warfare. Washington: National Research Council, Committee on Undersea Warfare, 1949.

25. Kuntz, J. E., and Sleight, R. B. Legibility of numerals: The optimal ratio of height to width of stroke. Amer. J. Psychol., 1950, 63, 567-575.

26. Loucks, R. B. Legibility of aircraft instrument dials: The relative legibility of tachometer dials. Project 265, Report No. 1. United States Army Air Forces, School of Aviation Medicine, Randolph Field, Texas, May, 1944.

27. Loucks, R. B. Legibility of aircraft instrument dials: A further investigation of the relative legibility of tachometer dials. Project 265, Report No. 2. United States Army Air Forces, School of Aviation Medicine, Randolph Field, Texas, October, 1944.

28. Loucks, R. B. Legibility of aircraft instrument dials: The relative legibility of various climb indicator dials and pointers. Project 286, Report No. 1. United States Army Air Forces, School of Aviation Medicine, Randolph Field, Texas, November, 1944.

29. Loucks, R. B. Legibility of aircraft instrument dials: The relative legibility of manifold pressure indicator dials. Project 325, Report No. 1. United States Army Air Forces, School of Aviation Medicine, Randolph Field, Texas, December, 1944.

30. Luckiesh, M., and Moss, F. K. Reading as a visual task. New York: D. Van Nostrand, 1942.

31. Mackworth, H. N. Legibility of raid block letters and numbers. FPRC 423(s). Great Britain, Air Ministry, Flying Personnel Research Committee (Psychological Laboratory, Cambridge), April, 1944.

32. McLaughlin, S. C., Jr. Configuration and stroke-width in numeral legibility, 1948 (unpublished thesis, Tufts College Library, Medford, Mass.).

33. Paterson, C. C., Walsh, J. W. T., and Higgins, W. F. An investigation of radium luminous compound. Proc. phys. Soc. London, 1917, 29, 215-249.

34. Paterson, D. G., and Tinker, M. A. How to make type readable. New York: Harper and Bros., 1940.

35. Reinwald, F. L. Legibility of symbols of the AND10400, Mackworth, and Berger type-faces at vertical and horizontal angles of presentation. Scientific Report No. 1. Prepared under Contract AF 30(602)-212 for Rome Air Development Center, Rome, N. Y., by the Department of Psychology, Colgate University, 1953.

36. Rock, M. L. Visual performance as a function of low photopic brightness levels. J. appl. Psychol., 1953, 37, 412-427.

37. Roethlein, B. E. The relative legibility of different faces of printing types. Amer. J. Psychol., 1912, 23, 1-36.

38. Roush, R. G., and Hamburger, F., Jr. An electronic chronograph for measurement of voice reaction-time. Amer. J. Psychol., 1947, 60, 624-628.

39. Sanford, E. C. The relative legibility of the small letters. Amer. J. Psychol., 1887, 1, 402-435.

40. Schapiro, H. B. Factors affecting legibility of digits. WADC TR 52-127. United States Air Force, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, June, 1952.

41. Soar, R. S. Height-width proportion and stroke-width in numeral visibility, 1952 (unpublished thesis, University of Minnesota Library, Minneapolis, Minn.).

42. Tinker, M. A. The relative legibility of the letters, the digits, and of certain mathematical signs. J. gen. Psychol., 1928, 1, 472-496.

43. Tinker, M. A. The relative legibility of modern and old style numerals. J. exp. Psychol., 1930, 13, 453-461.

44. Uhlaner, J. E. The effect of thickness of stroke on the legibility of letters. Proc. Iowa Acad. Sci., 1941, 48, 319-324.

45. United States, Army-Navy Aeronautical design standard. Numerals and letters - standard form of aircraft instrument dial. AND10400, October, 1944.

46. United States, Air Force-Navy Aeronautical design standard. Numerals and letters - aircraft instrument dial, standard form of. AND10400, July, 1948.

ADDENDUM

47. Lansdell, H. Effect of form on the legibility of numbers. Canad. J. Psychol., 1954, 8 (2), 77-79.

TABLE I

MEAN SIZE THRESHOLD SCORES FOR THREE DIGIT-WIDTHS
(8 SUBJECTS)

DIGIT	SIZE THRESHOLD (digit height in inches) NARROW (90 per cent of standard width)	STANDARD (AND design)	WIDE (110 per cent of standard width)
2	.181	.172	.162
4	.182	.177	.174
5	.186	.193	.163
8	.180	.178	.172
mean	.182	.180	.168

TABLE II

MEAN SIZE THRESHOLD SCORES FOR FOUR HEIGHT/WIDTH RATIOS, AREA HELD CONSTANT. HEIGHT/WIDTH EXPERIMENT 1, SINGLE DIGITS
(12 SUBJECTS)

		SIZE THRESHOLD (digit height in inches) HEIGHT/WIDTH RATIO (width as per cent of height)				(per cent of score at AND ratio) HEIGHT/WIDTH RATIO (width as per cent of height)			
PART[1]	DIGIT	50%	57%	67%	80%	50%	57%	67%	80%
1	2	.229	.227	.229	.236	100.0	99.1	100.0	103.1
	5	.229	.232	.240	.240	95.4	96.7	100.0	100.0
	8	.244	.248	.261	.264	93.5	95.0	100.0	101.1
2	0	.234	.234	.246	.240	95.1	95.1	100.0	97.6
	6	.229	.231	.251	.240	91.2	92.0	100.0	95.6
	7	.210	.219	.218	.225	96.3	100.5	100.0	103.2
3	3	.212	.223	.219	.218	96.8	101.8	100.0	99.5
	4	.240	.231	.236	.232	101.7	97.9	100.0	98.3
	9	.219	.218	.223	.214	98.2	97.8	100.0	96.0
	mean	.227	.229	.236	.234	96.2	97.0	100.0	99.2

[1]Three different groups of 12 subjects each served in the three parts.

TABLE III

MEAN RECOGNITION TIME SCORES FOR FOUR HEIGHT/WIDTH RATIOS, AREA HELD CONSTANT. HEIGHT/WIDTH EXPERIMENT 2, SINGLE DIGITS (12 SUBJECTS)

	RECOGNITION TIME (seconds)				RECOGNITION TIME (per cent of score at AND ratio)			
	HEIGHT/WIDTH RATIO (width as per cent of height)				HEIGHT/WIDTH RATIO (width as per cent of height)			
DIGIT	50%	57%	67%	80%	50%	57%	67%	80%
0	1.58	1.32	1.48	1.41	106.8	89.2	100.0	95.3
2	1.34	1.34	1.36	1.37	98.5	98.5	100.0	100.7
3	1.54	1.65	1.59	1.50	96.9	103.8	100.0	94.3
4	1.94	1.65	1.65	1.57	117.6	100.0	100.0	95.2
5	1.51	1.56	1.73	1.63	87.3	90.2	100.0	94.2
6	1.68	1.47	1.76	1.62	95.5	83.5	100.0	92.0
7	1.14	1.09	1.18	1.31	96.6	92.4	100.0	111.0
8	1.48	1.47	1.50	1.84	98.7	98.0	100.0	122.7
9	1.28	1.27	1.34	1.50	95.5	94.8	100.0	111.9
mean	1.50	1.42	1.51	1.53	99.3	94.0	100.0	101.3

TABLE IV

MEAN RECOGNITION TIME SCORES FOR TWO HEIGHT/WIDTH RATIOS, AREA HELD CONSTANT. HEIGHT/WIDTH EXPERIMENT 3, GROUPED DIGITS (12 SUBJECTS)

	RECOGNITION TIME (seconds)		RECOGNITION TIME (per cent of score at AND ratio)	
	HEIGHT/WIDTH RATIO (width as per cent of height)		HEIGHT/WIDTH RATIO (width as per cent of height)	
DIGIT	57%	67%	57%	67%
0	1.28	1.25	102.4	100.0
1	1.33	1.52	87.5	100.0
2	1.14	1.16	98.3	100.0
3	1.34	1.31	102.3	100.0
4	2.91	1.25	232.8	100.0
5	1.25	1.30	96.2	100.0
6	1.24	1.33	93.2	100.0
7	1.08	1.08	100.0	100.0
8	1.41	1.43	98.6	100.0
9	1.18	1.22	96.7	100.0
mean	1.42	1.28	110.9	100.0
mean[1] (N = 8)	1.24	1.26	98.4	100.0

[1] Mean was recomputed without the "1" because it was not included in the other height/width experiments and without the "4" because of extreme scores (see text).

TABLE V

MEAN SIZE THRESHOLD SCORES FOR FOUR STROKE-WIDTHS.
STROKE-WIDTH EXPERIMENT 1. AND STYLE DIGITS. (8 SUBJECTS)

	SIZE THRESHOLD (digit height in inches)			
	STROKE-WIDTH (per cent of height)			
DIGIT	11.2%	12.5%	13.7%	16.1%
4	.257	.259	.274	.266
5	.285	.289	.283	.268
7	.219	.229	.227	.223
8	.285	.281	.285	.289
mean	.262	.264	.267	.262

TABLE VI

MEAN SIZE THRESHOLD SCORES FOR THREE STROKE-WIDTHS IN A NARROW DIGIT STYLE AND ONE STROKE-WIDTH IN THE AND DESIGN. STROKE-WIDTH EXPERIMENT 2. (8 SUBJECTS)

	SIZE THRESHOLD (digit height in inches)			
	NARROW DIGIT STYLE			
	STROKE-WIDTH (per cent of height)			
DIGIT	8.3%	10.6%	12.7%	AND DESIGN
2	.212	.199	.189	.199
5	.225	.195	.193	.210
8	.227	.212	.202	.218
mean	.221	.202	.195	.209

TABLE VII

MEAN RECOGNITION TIME SCORES FOR TWO DIGIT SPACINGS. (4 SUBJECTS)

	RECOGNITION TIME (seconds)		
	SPACING (per cent of digit height)		
DIGIT	5.4%	16.2%	DIFFERENCE
0	1.44	1.44	0
1	1.78	1.36	+.42
2	1.39	1.54	-.15
3	1.52	1.34	+.18
4	1.54	1.59	-.09
5	1.74	1.35	+.39
6	1.46	1.41	+.05
7	1.62	1.13	+.49
8	1.72	1.55	+.17
9	1.48	1.38	+.10
mean	1.57	1.41	+.16

TABLE VIII. ERROR FREQUENCIES ON DIGITS WITH DEFECTIVE STROKES (16 SUBJECTS)[1]

TEST DGT.	DE-FECT	1	2	3	4	5	6	7	8	9	?	sum
0	A	-	-	-	-	-	4	-	4	4	1	13
	B	-	-	-	-	1	9	-	2	2	-	14
	C	-	-	-	-	-	3	-	3	2	2	10
	D	-	-	-	1	-	6	-	3	2	1	13
	E	-	-	-	-	-	3	1	2	2	2	10
	F	-	1	-	-	-	4	-	1	6	1	13
	G	-	1	1	2	1	5	1	1	9	1	22
	H	-	-	-	3	-	7	-	1	1	1	13
	sum	0	2	1	6	2	41	2	17	28	9	108

TEST DGT.	DE-FECT	0	1	3	4	5	6	7	8	9	?	sum
2	A	-	-	-	-	-	1	1	-	-	-	2
	B	-	-	-	1	-	-	1	-	2	1	5
	C	-	-	1	2	-	-	-	2	-	-	5
	D	1	-	-	2	1	1	1	2	-	1	9
	E	-	-	1	-	-	-	-	1	-	-	2
	F	-	-	-	-	1	-	1	1	-	-	3
	G	-	-	-	-	-	-	4	1	1	2	8
	H	-	1	-	-	-	1	4	-	5	-	11
	sum	1	1	2	5	2	3	12	7	8	4	45

TEST DGT.	DE-FECT	0	1	2	4	5	6	7	8	9	?	sum
3	A	-	-	-	2	12	3	-	2	-	1	20
	B	-	-	-	1	1	-	-	1	1	3	7
	C	-	-	-	-	8	2	-	2	-	1	13
	D	-	-	-	-	4	2	-	-	-	2	8
	E	-	1	2	1	5	1	2	1	-	-	13
	F	1	1	2	-	3	1	1	2	1	1	13
	G	-	1	-	1	1	1	-	-	-	2	6
	H	-	-	-	-	12	-	1	1	1	2	17
	sum	1	3	4	5	46	10	4	9	3	12	97

TEST DGT.	DE-FECT	0	1	2	3	5	6	7	8	9	?	sum
4	A	3	-	-	-	3	5	1	-	-	2	14
	B	-	-	1	-	-	6	-	-	-	3	10
	C	-	-	1	1	2	3	1	-	-	-	8
	D	-	-	-	-	1	13	-	1	-	1	16
	E	-	-	1	-	2	8	2	-	-	-	13
	F	-	-	1	-	-	4	1	-	-	2	8
	G	1	-	3	-	-	6	-	-	-	2	12
	H	1	-	1	-	1	3	1	-	1	1	9
	sum	5	0	8	1	9	48	6	1	1	11	90

TEST DGT.	DE-FECT	0	1	2	3	4	6	7	8	9	?	sum
5	A	-	-	-	1	-	5	-	-	1	2	9
	B	-	-	1	1	-	8	-	4	-	-	14
	C	1	-	-	2	1	5	-	4	2	-	15
	D	2	-	-	1	-	1	-	-	2	-	6
	E	-	-	-	-	-	3	-	5	-	-	8
	F	-	-	-	-	1	2	-	6	2	-	11
	G	1	-	-	3	1	2	-	3	-	-	10
	H	-	-	-	2	-	4	-	-	5	-	11
	sum	4	0	1	10	3	30	0	22	12	2	84

TEST DGT.	DE-FECT	0	1	2	3	4	5	7	8	9	?	sum
6	A	2	-	-	1	1	2	1	3	-	2	12
	B	3	-	-	-	-	-	-	2	-	-	5
	C	-	-	-	1	4	3	-	2	-	1	11
	D	-	-	-	-	-	6	1	-	1	1	9
	E	1	-	-	-	1	1	1	3	-	-	7
	F	-	-	1	-	2	2	-	2	1	1	9
	G	1	1	2	-	3	-	-	-	-	1	8
	H	1	-	-	-	3	-	-	1	-	1	6
	sum	8	1	3	2	14	14	3	13	2	7	67

TEST DGT.	DE-FECT	0	1	2	3	4	5	6	8	9	?	sum
7	A	-	2	-	-	-	-	-	-	-	2	4
	B	-	2	-	-	-	-	1	-	1	-	4
	C	1	-	-	-	-	-	-	1	-	2	4
	D	-	-	2	-	-	-	-	-	2	-	4
	E	1	-	-	1	-	-	-	-	1	-	3
	F	-	-	-	1	1	-	-	-	2	1	5
	G	-	-	2	2	-	1	-	-	-	2	7
	H	-	-	-	1	-	-	-	-	3	-	4
	sum	2	4	4	5	1	1	1	1	9	7	35

TEST DGT.	DE-FECT	0	1	2	3	4	5	6	7	9	?	sum
8	A	-	-	1	-	2	5	1	-	-	1	10
	B	1	-	-	-	1	2	10	-	1	1	16
	C	3	-	1	-	1	-	1	-	3	1	10
	D	3	-	2	-	1	2	2	-	4	3	17
	E	-	-	1	-	-	1	2	-	2	2	8
	F	2	-	-	3	-	2	3	-	7	1	18
	G	1	-	-	6	-	4	3	-	1	1	16
	H	-	-	-	3	-	1	8	-	-	1	13
	sum	10	0	5	12	5	17	30	0	18	11	108

TEST DGT.	DE-FECT	0	1	2	3	4	5	6	7	8	?	sum
9	A	2	-	1	-	-	-	-	2	2	1	8
	B	-	-	1	-	-	-	2	2	1	1	7
	C	1	-	-	1	-	-	1	-	-	2	5
	D	-	-	-	1	1	1	2	-	-	1	6
	E	1	-	-	-	1	-	2	-	1	1	6
	F	3	-	-	-	-	2	-	1	1	2	9
	G	3	-	2	-	-	-	2	1	2	-	10
	H	2	1	-	-	-	-	3	-	1	-	7
	sum	12	1	4	2	2	3	12	6	8	8	58

[1] Each test digit with stroke defect in each location presented 32 times. Locations shown in Fig. 5.

TABLE IX. PER CENT ERRONEOUS IDENTIFICATIONS ON AND STYLE DIGITS[1]

A. PER CENT ERRORS, SIZE THRESHOLD DATA. 60-300 TRIALS PER DIGIT

TEST DIGIT	RESPONSE DIGIT 0	1	2	3	4	5	6	7	8	9	total
0		-	1.3	-	3.9	-	6.5	-	2.6	2.6	16.9
1	-		-	-	1.7	-	-	-	-	1.7	3.4
2	-	.8		-	2.0	.8	.4	1.2	.8	1.6	7.6
3	-	1.6	-		1.2	6.0	.4	-	.8	1.6	11.6
4	-	-	1.6	-		-	7.2	-	.4	-	9.2
5	.5	-	1.0	5.5	1.5		3.5	-	1.5	1.0	14.5
6	6.8	-	-	-	3.6	1.6		-	1.6	.4	14.0
7	-	1.2	-	1.2	-	.8	-		.4	1.2	4.8
8	1.8	-	-	1.8	1.2	-	1.2	-		.6	6.6
9	5.4	-	-	.6	-	.3	.9	.3	.3		7.8
total	14.5	3.6	3.9	9.1	15.1	9.5	20.1	1.5	8.4	10.7	

B. PER CENT ERRORS, RECOGNITION TIME DATA. 96 TRIALS PER DIGIT

TEST DIGIT	RESPONSE DIGIT 0	1	2	3	4	5	6	7	8	9	total
0		x	x	x	x	x	x	x	x	x	x
1	-		-	-	-	-	-	-	-	1.0	1.0
2	-	-		-	1.0	-	-	1.0	-	1.0	3.0
3	-	1.0	-		2.0	9.4	-	-	-	-	12.4
4	-	1.0	1.0	-		1.0	8.4	-	-	-	11.4
5	-	-	-	1.0	-		1.0	-	2.0	1.0	5.0
6	3.2	-	1.0	-	6.2	-		-	-	-	10.4
7	-	-	-	-	-	-	-		-	-	0
8	-	-	1.0	-	-	-	-	-		-	1.0
9	-	-	-	-	-	-	-	-	-		0
total	3.2	2.0	3.0	1.0	9.2	10.4	9.4	1.0	2.0	3.0	

C. PER CENT ERRORS, DATA OF TABLE VIII. 256 TRIALS PER DIGIT, ALL DEFECTS POOLED

TEST DIGIT	RESPONSE DIGIT 0	1	2	3	4	5	6	7	8	9	?	total
0		-	.8	.4	2.3	.8	16.0	.8	6.6	10.9	3.5	42.1
1	x		x	x	x	x	x	x	x	x	x	x
2	.4	.4		.8	2.0	.8	1.2	4.7	2.7	3.1	1.6	17.7
3	.4	1.2	1.6		2.0	18.0	3.9	1.6	3.5	1.2	4.7	38.1
4	2.0	-	3.1	.4		3.5	18.8	2.3	.4	.4	4.3	35.2
5	1.6	-	.4	3.9	1.2		11.7	-	8.6	4.7	.8	32.9
6	3.1	.4	1.2	.8	5.5	5.5		1.2	5.1	.8	2.7	26.3
7	.8	1.6	1.6	2.0	.4	.4	.4		.4	3.5	2.7	13.8
8	3.9	-	2.0	4.7	2.0	6.6	11.7	-		7.0	4.3	42.2
9	4.7	.4	1.6	.8	.8	1.2	4.7	2.3	3.1		3.1	22.7
total	16.9	4.0	12.3	13.8	16.2	36.8	68.4	12.9	30.4	31.6	27.7	

[1]The figure in a cell shows per cent of trials in which the test digit was incorrectly identified as the digit at the head of the column. x's mean the indicated test digit was not used.

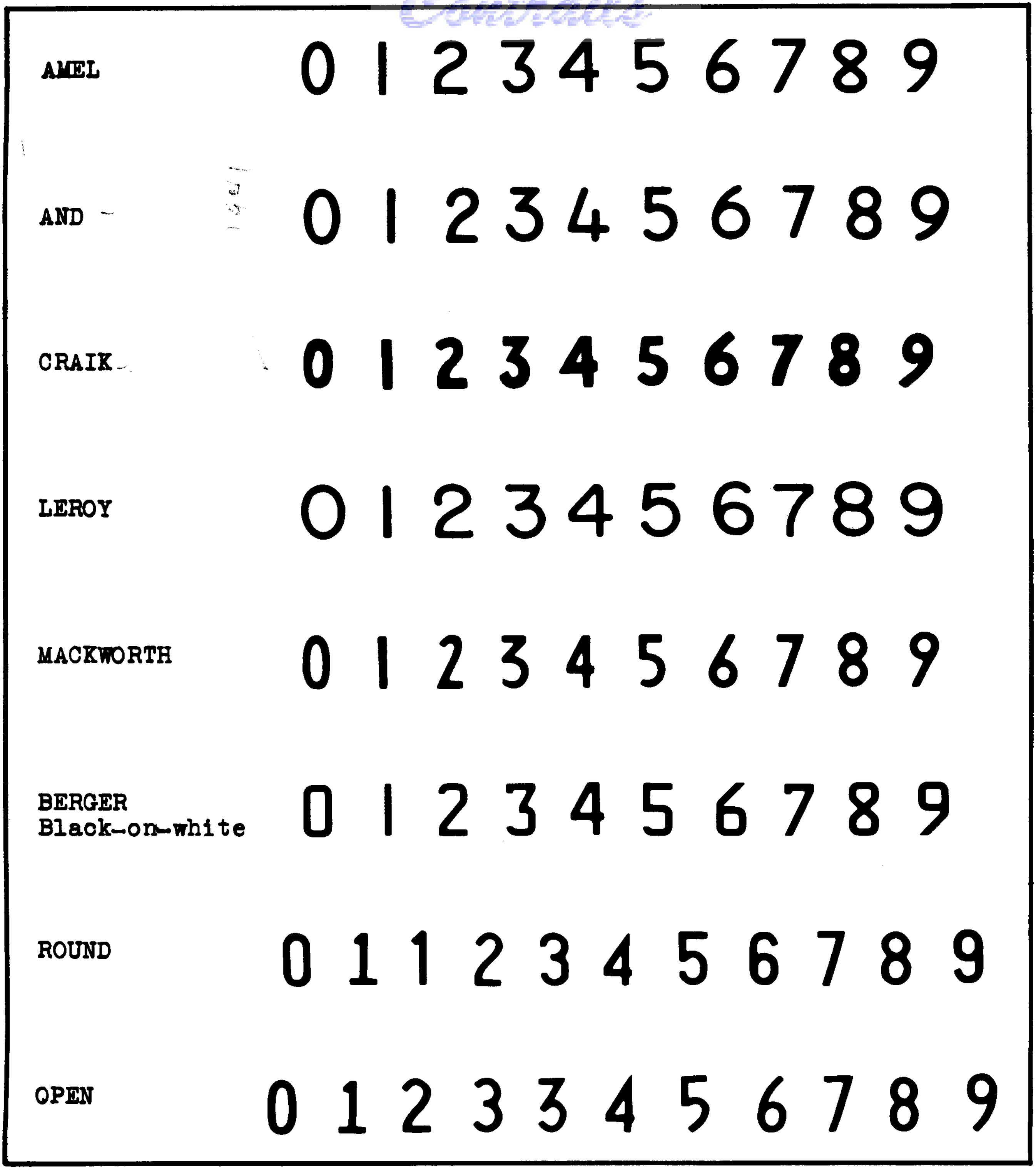

FIGURE 1. Digit sets studied by different investigators. The top four are treated in this report as dial-type numerals; the bottom four were intended for viewing at a distance. "Round" and "open" are license plate numerals studied by Aldrich. Reproduction of dimensions may not be precise, but stroke-widths are intended to represent those specified in the original designs except for the Leroy, in which the stroke-width is arbitrary. The Berger black-on-white has a wider stroke than the white-on-black usually shown.

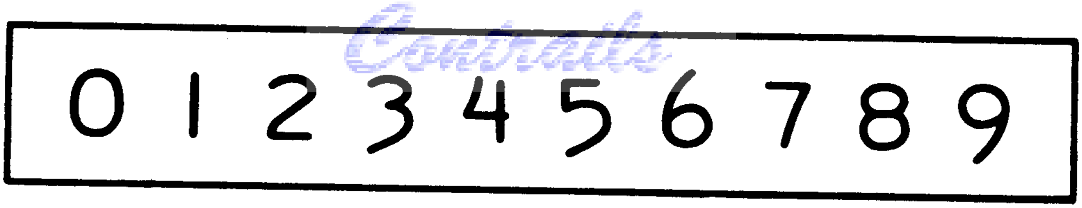

FIGURE 2. The Ordnance digits. Specifications provide for a range of digit-widths and stroke-widths.

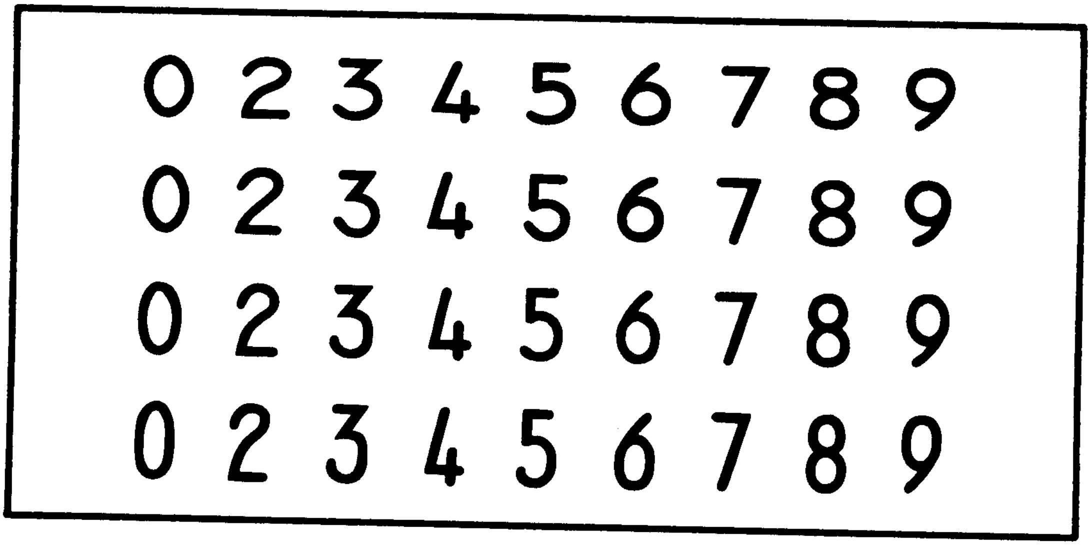

FIGURE 3. Four height/width ratios; area held constant. Mean widths 80%, 67%, 57%, and 50% of height. AND style.

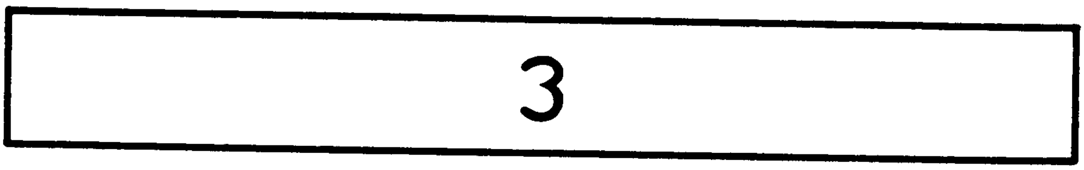

FIGURE 4. Experimental digit "3", found to be more recognizable than the AND "3".

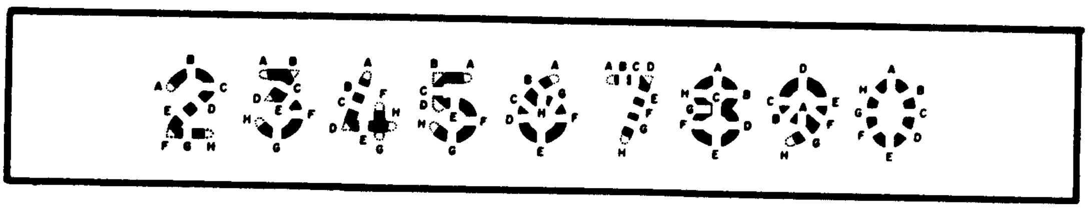

FIGURE 5. Experimental stroke defects. One missing segment tested at a time.

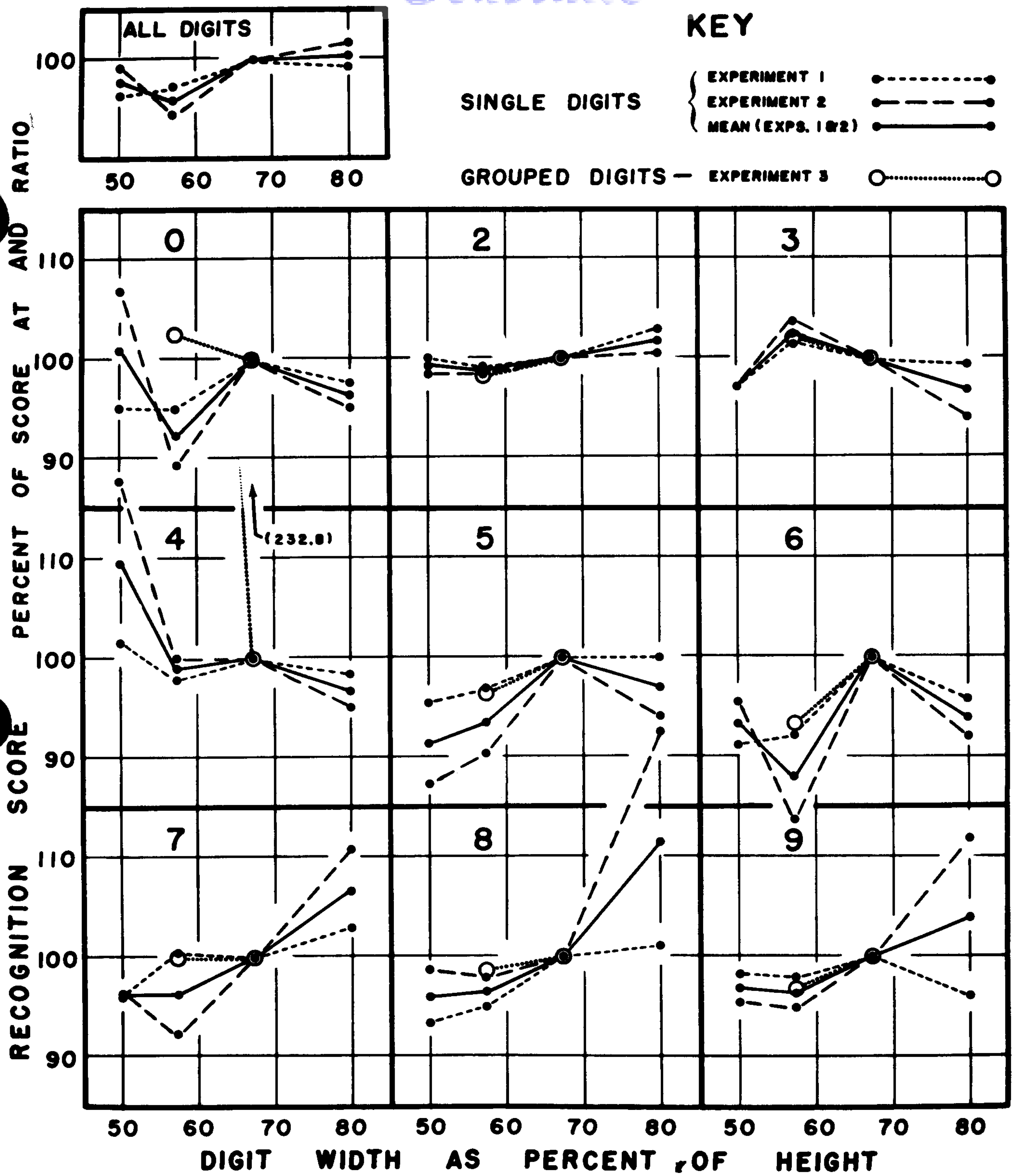

FIGURE 6. Digit recognition as a function of height/width ratio, area held constant. Scores plotted as percentages of score at the AND ratio (67% of height). Experiment 1, size threshold method; Experiments 2 and 3, recognition time method. Experiment 3 is omitted from the graph for all digits because of the distortion introduced by extreme scores on the "4".