DOCUMENT RESUME

ED 035 515 24 RE 002 349

AUTHOR	Lott, Deborah
TITLE	Visual Word Recognition: Its Implications for Reading Research and Instruction.
INSTITUTION	Southwest Regional Educational Lab., Inglewood, Calif.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
REPORT NO	TR-17
BUREAU NO	BR-6-2865
PUB DATE	7 Jul 69
CONTRACT	OEC-4-7-062865-3073
NOTE	59p.
EDRS PRICE	EDRS Price MF-$0.50 HC-$3.05
DESCRIPTORS	Language Skills, Optical Scanners, *Reading Instruction, Reading Processes, *Reading Research, Redundancy, *Visual Discrimination, Visual Perception, *Word Recognition

ABSTRACT

Little research has been done to explain just why words are recognized more easily than letters alone; although, this phenomenon has been accepted widely by educators. Therefore, a model of the processes involved in word recognition and suggestions concerning how these processes can be put to use in reading instruction are presented. The model describes word recognition as a feature-scanning process in which relevant cues, called distinctive features, are analyzed and synthesized. A description of the scanning process is given with its distinctive features defined. Explanations of how a skilled reader uses feature combinations to recognize letters and words and how such a reader uses the redundancy in a word or letter sequence are also offered. Graphs and a bibliography are included. A discussion of variables which influence the legibility of print is appended. (NH)

RE

BR-6-2865 TR17
PA-24
OE/BR

SWRL

SOUTHWEST REGIONAL LABORATORY FOR EDUCATIONAL RESEARCH & DEVELOPMENT

Visual Word Recognition: Its Implications for Reading Research and Instruction

TR17 7 JULY 1969

Published by Southwest Regional Laboratory for Educational Research and Development, a public agency supported as a regional educational laboratory by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the Office of Education, and no official endorsement by the Office of Education should be inferred.

VISUAL WORD RECOGNITION:
ITS IMPLICATIONS FOR READING RESEARCH AND INSTRUCTION

Deborah Lott

TABLE OF CONTENTS

VISUAL WORD RECOGNITION: ITS IMPLICATIONS FOR READING RESEARCH AND INSTRUCTION

Deborah Lott

Evidence that words are recognized more easily than letters alone has been available since the early studies of Cattell (1885; 1886). This phenomenon, while seldom questioned, has never been adequately explained. Educators have accepted the skilled reader's tendency toward total word recognition as a basic reading skill, but have made little attempt to identify the nature and development of this skill in reading acquisition.

This paper describes a model of the processes involved in word recognition and suggests how these processes can be put to use in reading instruction. The model describes word recognition as a feature scanning process in which relevant cues, called distinctive features, are analyzed and synthesized.

This paper specifically proposes: (1) to describe the scanning process; (2) to define distinctive features; (3) to describe how a skilled reader uses feature combinations to recognize letters and words; and (4) to describe how a skilled reader uses the redundancy in a word or letter sequence.

VISUAL SCANNING

Theories of reading rely heavily on the information available from such fields as visual perception, neurophysiology, psycholinguistics, and computer science. Because much of this information is incomplete, inconsistent, and has little relevance to the study of reading or word recognition, theories of reading tend to be highly intuitive and based on little empirical evidence. This paper attempts to develop a model for word recognition based on the best empirical evidence available.

Any reading theory must consider the nature of visual perception. Unfortunately, the various schools of psychology are no more in agreement on the nature of visual processes than on any other matter. One of the earliest theories of visual perception is that of the Gestalt psychologists, now frequently called whole-template matching.

Whole-Template Matching

Advocates of the whole-template matching model argue that recognition occurs when an entire stimulus in the external environment coincides with a complete stored image or whole-template in the nervous system of the perceiver. However, there are an infinite number of different stimuli in the environment, each of which may be transformed in an unlimited number of ways (e.g., perspective, orientation, size, location). The visual system could not possibly have a stored whole-template for each possible stimulus configuration in each possible transformation. Therefore, there must be canonical forms or idealized templates against which the different transformations of a particular stimulus object may be compared. Severely distorted and spatially transformed figures may or may not be recognized, depending upon the degree to which they coincide with these stored templates (Neisser, 1967, p.52).

But, were there a single canonical form or idealized template against which to compare the different transformations of a particular stimulus, the congruence of the stimulus in the external environment with the canonical form could not be assessed unless one of the forms could be superimposed on the other. How, then, can familiar patterns be recognized no matter where they happen to fall on the retina? It is generally argued that any familiar form must have fallen on every conceivable position of an adult's retina, thereby leaving behind so many templates that contact with one of them is inevitable. This line of argument resulted in the prediction that unfamiliar patterns in new positions on the retina would not be recognizable.

Wallach and Austin (1954) tested this prediction in a study examining the effect of retinal position on the recognition of an ambiguous figure. In this experiment, subjects identified a two-dimensional figure as a dog when it was horizontal, as a chef when vertical, and as a fairly balanced ambiguous figure when tilted 45 degrees. The experiment consisted of presenting the dog-version in one position on the retina and the chef-version in another. Then, the figure was presented in its ambiguous-version in either of the two positions on the retina. As predicted, the results revealed a significant tendency to recognize the ambiguous figure as that figure which had previously been presented in the same position on the retina.

The Wallach and Austin (1954) findings, however, could also be interpreted as the result of a formation of a learned association between the stimulus' presence in one position on the retina and a particular response, instead of the result of perceptual factors. In other words, subject performance could have been determined more by response bias effects than by perceptual effects.

Another study demonstrating the effect of retinal locus is reported by Miskin and Forgays (1952). These authors take the Hebbian (see pp.5-6) position that a particular perception occurs through the action of specialized neural cells assembled slowly by the repeated stimulation of a specific receptor matrix. In general, these neural cell assemblies are created in many positions across the retina. However, in reading, one is continually presented with words within a specific retinal locus and cell assemblies are created only within these limited areas of the retina. Thus, they predict that words will not be equally recognizable on different retinal loci.

Forty-eight eight-letter words were presented tachistoscopically to 16 adult subjects. For the first 24 words, the subjects were instructed to fixate a small traget in the center of the field; for the second 24 words, fixation was directed at random to one of four targets in the upper, lower, left, and right parts of the visual field, while the word always appeared in the center.

> The result is... that recognition below the fixation point was nearly twice as good as above, and recognition to the right was nearly two and a half times better than to the left.... It is clear that exposures of the same word in the left and right visual fields are not equivalent stimulus situations. However, it is possible that several factors other than selective retinal training may have been responsible, particularly since a recognition difference was also revealed in the control comparison (Miskin & Forgays, 1952, p. 44).

To determine whether subjects were selectively attending to certain areas of the field, a comparison of left and right field recognition of English words (by readers of English) was made concurrently with that of Yiddish words (by readers of Yiddish), in which the letters run in the reverse order. Recognition was 40% greater for English words to the right and 25% greater for Yiddish words to the left.

In a later study, Forgays (1953) argued that if separate parts of the receptor surface were individually trained in reading, there would be no gross differences for beginning readers in recognition thresholds for words presented to the right or left of fixation. This prediction was supported.

The results of these two studies (Miskin & Forgays, 1952; Forgays, 1953) were interpreted as evidence that only limited regions of the retina are trained during reading acquisition. However, the findings can also be explained by an internal scanning model of visual perception. Internal scanning models propose that stimuli are retained in short-term memory for a brief period of time and that it is this memory trace of the stimulus which is scanned and "read-out" (Sperling, 1963). The internal scanning process consists of a spatially sequential analysis of the persisting stimulus trace after the tachistoscopic exposure is terminated. Although internal scanning does not involve overt eye movements, it does entail those preparatory activities in the central nervous system which precede overt movements of the eyes. Furthermore, this internal scanning process is assumed to progress in a manner which corresponds to the sequence of eye fixations which would occur if the stimulus were actually present. Therefore, if the observer fixates on the center of the field and the word is presented to the left of the center, the visual system must first scan backwards to the beginning of the word before scanning appropriately from left to right, and processing time is lost (Harcum & Finkel, 1963).

Many whole-template models also consider the similarity of the stimulus in the external world to its stored canonical form: an observer's ability to recognize a stimulus is predicted by the distance between it and its canonical form (Posner, Goldsmith, & Welton, 1967). This interpretation of perceptual recognition as a function of degree of similarity can be explained in terms of pre-perceptual analysis--a cleanup of the input--which occurs in the visual system to make the stimulus more nearly approximate the whole-template or canonical form against which it is to be compared. The less marked the distortion, the more likely that this cleanup will result in recognition of the figure.

Some computer programs for pattern recognition use pre-perceptual cleanup prior to whole-template matching. In these programs, the stimulus is placed in front of an idealized whole-template and the computer assesses the percentage of area that the two have in common. Then, the stimulus can be rotated, magnified or reduced in size, or transformed in other ways to increase its congruence with the template (Uhr, 1966, p. 373).

Whole-template matching computer programs cannot tolerate even slight changes in position, orientation, and size of stimulus unless such cleanup measures are taken (Selfridge & Neisser, 1960, p. 64). This being the case, the applicability of a whole-template model to human perception in reading seems doubtful. It is possible to carefully limit the distortion of material fed into a computer by developing an ideal type style which allows only minor variation or only such variation as the computer is able to cope with. However, the skilled reader is continually presented with a wide variety of printed and handwritten messages across marked degrees of distortion (Uhr, 1963, p. 44), and he is able to read such material without too much difficulty. Even combinations of print variables which are far from optimal do not result in any marked reduction in one's ability to read (see Appendix). In fact, in an attempt to study more normal, rather than experimental, reading situations, Davenport and Smith (1965) found no consistent variations in the legibility of different print styles.

Smith (1969) and Smith, Lott, and Cronnell (in press) demonstrated that skilled readers can recognize words in which the letters alternate in case if letter size is held constant. The resulting distortion in text was quite marked and undoubtedly unfamiliar to the majority of readers, but the words were recognized as readily as normal text. It is unlikely that the subjects had stored images for total figures or whole-templates against which to compare the different distorted texts, and the notion of degree of similarity must be discarded since size distortion, which should be readily amenable to pre-perceptual cleanup, significantly retarded subjects' ability to recognize words.

Visual Scanning Model

The difficulties inherent in the whole-template approach to pattern recognition have led many theorists to look for other explanations. Feature scanning theories offer an alternative.

One of the first major theories relying on a feature analysis of visual perception is that of Hebb (see Dember, 1965; Neisser, 1967). Essentially, Hebb hypothesizes two major processes in visual perception: (1) perception of unity or the observer's ability to separate figure from ground; and (2) perception of identity. There are three levels of perception of identity: (a) perception of difference between two dissimilar figures, (b) perception of likeness between two similar figures, and (c) perception of a figure as a member of a particular category or class (Dember, 1965, p. 238).

Perception of identity occurs through analysis of the attributes of the pattern, or the lines and angles which compose it. Hebb hypothesizes the physiological existence of cell assemblies or feature analyzers which respond only to specific attributes. These cell assemblies are reduplicated across the input region of the visual system. Therefore, the image need not fall on any particular region of the retina to be recognized (see Neisser, 1967, pp. 51-78).

The main difference between Hebb's theory and later feature-oriented theories is that the only features Hebb describes are lines, angles, and contours. Thus, Hebb's theory, really a cross between feature theories and template theories, is based on filtering subtemplates or templates for parts of figures which cannot be consciously segregated from the whole.

Hunt (1962) also believes that template matching is not likely to be the basis of pattern recognition if templates must be matched to entire patterns, and that subtemplate matching provides a more plausible basis for visual perception.

> Dimensions and values play a role in a [sub]template-matching model. Each region of the projection matrix can be thought of as a dimension. The set of [sub]templates might appear in different dimensions and convey different information, since the information transmitted by a symbol is a function of the set of symbols from which it is drawn rather than a function of its own identity ... (Hunt, 1962, p. 126).

Subtemplate matching is frequently used in computer programs for pattern recognition. These programs use a scanning device which counts the number of intersections between the input and the lines or subtemplates defined by the machine (Uhr, 1963, p. 47).

The rationale underlying use of subtemplates in computer pattern recognition is described by Block, Nilsson, and Duda (1964, p. 78):

> If it is reasonable to assume that each pattern is composed of simpler patterns or features, then a matching scheme can still be used. The features are then building blocks of the complete patterns, and subtemplates matched to features can be used. The number of features is usually much smaller than the number of patterns that can be composed from them, and therefore, a subtemplate matching scheme could be an economical solution ....

A recent feature-based computer program for pattern recognition is Selfridge's Pandemonium model. Like the other subtemplate and feature theorists, Selfridge's basic assumption is that a pattern is equivalent to a function of its feature set, each member of which is individually common to several patterns and whose absence is also common to several other patterns (Selfridge, 1966, p. 341)

The process employed by Selfridge's Pandemonium program in pattern recognition has four levels. At each level, the stimulus is confronted by different demons or feature analyzers. The first level consists of the data demons; they serve merely to store and pass on the image to be recognized. The computational demons then check the image for the presence or absence of certain features. At the third level, the cognitive demons assign weights to the different features in terms of their contribution to particular patterns. Each cognitive demon then computes a shriek which relates how closely this weighted combination of features conforms to the pattern it represents. The decision demon merely selects the demon with the loudest shriek (Selfridge, 1966).

Quite similar to Selfridge's Pandemonium model for computer pattern recognition is Neisser's (1967) theory of human pattern recognition. Neisser's theory is described in some detail in his discussion of how an observer knows an A when he sees one. Recognition begins with the segregation of A from all other figures by preattentive processes. The mechanisms employed in this preattentive phase emphasize the global aspects of the stimulus. These preattentive mechanisms or analyzers are reduplicated across the input field.

The second phase in the recognition of A consists of directing focal attention towards it. During this phase, there is more extensive feature analysis and synthesis of the figure. Neisser agrees that a Hebbian analysis into lines and angles plays a role here, but suggests that more complex analyzers (such as concavity, symmetry, closure) are also involved.

Finally, there is an internal sequence of comparisons with stored records of earlier syntheses to determine the proper classification of the figure (Neisser, 1967, pp. 102-104).

Neisser reports a series of experiments designed to test his theory. All of these studies followed a general experimental design: subjects were asked to find particular target items which were embedded in a list of 50 items. The results revealed that time per item is not dependent on the number of possible targets, thus implying that the skilled reader processes different feature sets simultaneously. Subjects occasionally noted that they stopped searching without even knowing to which of the possible targets they had responded, suggesting that visual synthesis did not play a role in their responses. Instead, the subjects were able to develop preattentive recognition systems, sensitive only to key features and not dependent upon unique recognition and classification of the figures (Neisser, 1964).

It has been argued, however, that Neisser's visual search task may have limited relevance to reading or word recognition tasks since it establishes a specific response set (e.g., "Find the letter A.") which may not occur in reading. On the other hand, much of reading may consist of similar search behavior since the context of a passage allows the reader to anticipate what a particular word will be and the reader then needs only to check for confirmation.

The distinction made in this section among whole-template, subtemplate, and feature models of pattern recognition can be criticized as an artificial one. All three models assume that matching occurs between external stimuli and neural images. The distinction arises solely in terms of what is being matched: whole figures with whole-templates; lines and angles comprising a figure with subtemplates for lines and angles; or complex characteristics or features of the figure with correspondingly complex feature analyzers. Whole-figure to whole-template matching does not seem to be an economical process: too many whole-templates would have to be stored in the visual system. Both subtemplate and feature matching models are possible solutions; their differences lie mainly in terms of what characteristics of the figure are to be matched with internal subtemplates or feature analyzers. Subtemplates refer to actual segments of the figure; features refer to abstract characteristics of the total figure. At this time, however, the distinction between subtemplates and features is functionally irrelevant and both types of model will be included under the general discussion of feature models. What, then, are the features which must be detected in the stimulus and matched with their corresponding neural analyzers? Neurophysiological research attempts to answer this question.

Neurophysiological Basis for Feature Scanning Model

Like all behavioral phenomena, visual perception has a neurophysiological basis. Therefore, any description of visual perception should be feasible in terms of what is known about neural activity. To date, neurophysiological research has revealed that individual neurons are selectively responsive to specific patterns or stimulus characteristics. Furthermore, the retina does not appear to pass on an intact image of the stimulus to the visual cortex, but to pass on instead a highly summarized and reorganized account of the stimulus. Although not conclusive, these findings lend support to a feature system of analysis in visual perception while making any whole-template matching system less likely.

The first neurophysiological finding to consider is the all-or-none principle: If a neural impulse occurs at all, it occurs with its characteristic amplitude; this amplitude does not vary with the intensity of the stimulus but only with the diameter of the fiber.

Secondly, there are approximately 120,000,000 receptors in each eye, and only about 1,000,000 fibers in the optic nerve. But, even with this 120:1 reduction, there is a point-for-point mapping of the retina on the cortex (i.e., every point on the retina, when stimulated, will elicit a response at some point on the cortex) though the topological arrangment is not maintained (Kolers, 1968, p. 7).

The form of information conveyed by neural impulses is further restricted by the selective responsiveness of individual neurons to highly specific stimulus attributes. The first major division is among the _on_, _off_, and _on-off_ fibers: _on_ fibers fire when there is light; _off_ fibers fire when there is no light; _on-off_ fibers fire whenever there is any change in illumination (Mueller, 1965).

Recent research has revealed that individual neural fibers in the visual system play an even more selective role than the three types outlined above. Selective responsiveness of specific neural fibers results in extensive analysis and summarization of the stimulus within the retina, prior to the more complex analysis in the visual cortex.

By recording and examining the activity of single retinal ganglion cells of the frog in response to stimuli of different shades, sizes, and shapes, Maturana, Lettvin, McCullock, and Pitts (1960) were able to conclude that much stimulus processing is performed by the retina, and that a highly specific and summarized message is transmitted through the optic nerve fibers to the visual cortex. There is a natural separation of the retinal ganglion cells into five classes according to the operations that they perform on the visual image. Cells of one class measure light intensity; cells of the other four classes respond maximally to one or another quality, or configuration of qualities (sustained edge detection, convex edge detection, changing contrast detection, and dimming detection).[1] Each retinal ganglion cell performs

[1] It is useful to describe these five classes of ganglion cells in more detail. Class 1 performs an operation called "sustained edge detection." "These cells do not respond to general changes of illumination, whether a sudden on or off or just gradual increase or decrease of light intensity. On the other hand, the sharp edge of an object, lighter or darker than the background, . . . produces a burst of activity. . . This response to the moving or standing edge is independent of the shape of the object or the curvature of the edge. However, it is not entirely independent of size because large objects . . . give a response somewhat smaller than small objects . . . (Maturana et al., 1960, p. 148)."

The second class is that of "convex edge detection." The cells of this class do not respond to changes of the general illumination . . . They respond with a strong burst of activity to the movement of a small object darker than the background . . . exhibiting a sharp edge . . . (p. 149)."

Class 3 performs the operation of "changing contrast detection." These cells are highly sensitive to movement. "There is an optimal speed for a maximal response (p. 154)."

"Dimming detection" is the operation of the fourth class. "These units respond with a prolonged response to the off of light . . . This response may last for seconds, many minutes, or even indefinitely, according to the final degree of darkening that is reached (p. 157)."

Class 5 cells function as "dark detectors." "These units are continuously active, even under bright light, but their activity is inversely proportional to the light intensity and increases to a maximum in darkness (p. 159)."

only one of these five operations. Members of the five classes are uniformly distributed across the retina.

This evidence suggests that the frog's visual system makes a feature analysis of the stimulus relatively early in the recognition process. Moreover, the degree of specialization of nerve cells undoubtedly increases as the message is transmitted through successive synapses, resulting in many different types of neurons in the visual cortex, each type devoted to detection of exceedingly specific patterns or changes in illumination.

Much of the current knowledge on the neurophysiology of the visual cortex stems from the work of Hubel and Wiesel (1962). Working with the cat, Hubel and Wiesel studied receptive fields of individual cells in the visual cortex (i.e., that region of the retina over which the firing of a particular cell in the cortex is influenced). They found that while circular spots were the most effective stimuli for activating ganglion cells and lateral geniculate cells, they were ineffective at the cortical level. Instead, the shape, orientation, position, direction, and velocity of the stimulus were found to be the crucial variables in producing an optimal cortical discharge.

Although the neurophysiological research described above supports the existence of highly specific feature analyzers in the visual system, it is perhaps only of limited relevance to the study of human pattern recognition. Whether or not there are actual differences in neural mechanisms, there are differences in the discrimination abilities of different species (Sutherland, 1957; 1963).[2] Species-specific discrimination abilities are probably a function of learning, i.e., those discriminations necessary for the survival of a species being more likely to be learned by members of that species. Human beings, then, present a unique example since human beings must learn to read and, in learning to read, they learn to make discriminations which are necessary for reading.

---

[2]Sutherland's research (1957) revealed that octopuses can be taught to discriminate between vertical and horizontal rectangles but not between diagonal (or oblique) rectangles. These results imply that the octopus uses a relatively simple feature system, consisting mainly of vertical (and possibly horizontal) lines.

Sutherland (1963) later repeated this experiment with cats. The cats, like the octopuses, were trained to discriminate between horizontal and vertical rectangles and between two oblique rectangles. However, unlike the octopuses, the cats were able to perform as well with the oblique rectangles as with horizontal-vertical ones. These findings suggest that different species may use different sets of features, perhaps reacting to those features critical to the survival of the species.

Developmental Aspects of Feature Scanning in Human Perception

If discriminative ability develops as the result of learning, it is not surprising that the discriminative ability of humans is largely a function of age. For example, Rudel and Teuber (1963) found that children aged 3 to 5 had no difficult in learning to discriminate between vertical and horizontal lines or between ⊔ and ⊓ shaped figures, but failed to learn to discriminate between pairs of oblique lines and between ⊐ and ⊏ shaped figures. For 5-year-olds, discrimination of horizontal from oblique lines was significantly more difficult than discrimination of vertical from oblique lines. Although performance in both oblique and right-left open figure discrimination improved radically at the age of 6½, by 8½ years the subjects still had not attained as much proficiency with these figures as with the horizontal-vertical lines and the top-bottom open figures.

The results of the Rudel and Teuber (1963) study suggest that as children learn to read they learn to make discriminations which they will need in reading. It is important to note, however, that oblique discriminations do remain more difficult than horizontal-vertical ones even as the children grow older, at least up to 8½ years.

Mandes and Ghent (1962) offer specific information on the feature scanning process utilized by children and adults. Geometric figures in which members of a family differed only on one side (distinguishing feature) were presented tachistoscopically. The figures appeared with the distinguishing feature either at the top, bottom, left, or right of the figure and the subject was given a multiple choice array in which the figures were presented in these four orientations.

The results demonstrated that for both 6-year-olds and adults, recognition was better when the distinguishing feature was at the top than at the bottom. Children revealed a significant tendency to attend to the right of the figure; adults revealed a significant tendency to attend to the left.

The results of the Mandes and Ghent (1962) study conform rather closely to those of Miskin and Forgays (1952) and Forgays (1953) described earlier (see pp. 3-4). It appears that there is probably an early tendency to scan from right to left and from top to bottom. English, however, is written from left to right (probably as a result of writing ease) so the skilled reader of English must have acquired appropriate scan direction. The further finding of Miskin and Forgays (1952) that readers of Yiddish tend to scan from right to left, as do the children in the Mandes and Ghent (1962) study, supports the conclusion that age differences, at least in direction of scan, are due to learning rather than to maturation.[3]

Gibson (1965) sought information on the feature systems used by children of different ages by having children aged 4 to 8 match standard figures against all transformations and copies of them. The children were to select only identical copies. Total discrimination improved from age 4 to 8, but this improvement occurred at different rates for the different transformations. All aged children had little difficulty discriminating open from closed figures; rotation, reversal, and line-to-curve transformations improved consistently from 4 to 8 years; and perspective transformations remained quite difficult throughout, although some improvement did occur at ages 7 and 8. It appears that those transformations which are critical in reading are learned quickly whereas noncritical transformations, such as perspective, do not show substantial improvement until later. Gibson's (1965) conclusions concerning the rates at which abilities to make different transformations are acquired are weakened, however, by a study (Schaller & Harris, 1969) using stimuli with greater differences in perspective transformation. In this study, fewer errors occurred at all ages and younger children were able to achieve asymptotic performance.

Bruner (1965) studied children's performance in a letter discrimination task. The children were shown two boards with letters outlined by lighted bulbs. A third board was unlighted but was wired so that bulbs forming one of the two letters could be lighted by touching the appropriate bulbs. The children's task was to discover which of the two letters was available on the third board. The results showed that 4-year-olds pushed random bulbs. The older children first pushed non-discriminating bulbs (i.e., bulbs shared by the two letters) and later pushed bulbs relevant to discrimination.

The results support Bruner's idea that learning to read is to a large extent learning to distinguish <u>information space</u> from <u>image space</u>, i.e., learning to attend to the critical features which allow the reader to discriminate between two letters rather than to attend to those parts of letters which are shared.

---

[3]Bryden (1966) argues that cerebral dominance is the primary factor producing left-right differences in the recognition of single-element material while directional scanning becomes important only with multiple-letter arrays. He presented tachistoscopically to adult subjects arrays of single letters or of three letters to the left or right of fixation. The results revealed that the left-right difference was about seven times greater for the three-letter arrays and that there was zero correlation between the left-right differences on the two tasks.

A later study by Olson (1967) is directly concerned with the conceptual strategies which children of different ages use to distinguish information space from image space. In this study, Bruner's bulb-board was used but the stimuli were patterns rather than letters. The performances of different aged children in free and constrained conditions were compared: in the free condition, the children were permitted an unrestricted choice of bulbs to solve the problem; in the constrained condition, the children were permitted to press only one bulb at a time--after each press the experimenter asked each child if he knew the correct pattern.

Three major strategies were used by the children: (1) Search Strategy a nonrandom or quasi-systematic search for the bulbs that would light (this search was independent of the examples provided, and was the strategy used by the youngest children); (2) Successive Pattern-Matching Strategy an almost total concentration upon on-pattern bulbs, redundant and informative alike (this strategy started to develop with the 5-year-olds); and (3) Information-Selection Strategy an increasing percentage of informative bulbs (this strategy developed at about the age of 7).

> In effect, the older the child, the more likely he is to solve the problem directly upon achieving the minimum information necessary for that solution . . . . Constraint improves this likelihood strikingly. Note too that, to put it figuratively, a five-year-old operating with constraints imposed will perform in an informationally more efficient fashion than will a seven-year-old operating freely. This suggests . . . that perhaps the effect of years is to internalize informational constraints (Olson, 1967, p. 143).

Wright (1964) studied cue strategies used by different aged children in haptic discrimination tasks where tactual-proprioceptive cues were used. The children explored blocks which differed in shape and texture, but with only one dimension designated as relevant. An observer rated their hand movements on a 5-point scale of relevance. The results show that the younger children spent more time exploring the irrelevant dimension and even within homogeneous age groups, relevance scores were negatively correlated with trials to criterion. Although preschool children did learn distinctive observing responses to different cue dimensions, and did eventually discriminate dimensional relevance, they continued to make ritualistic responses to irrelevant cues. Older, school-aged children stopped making irrelevant attending responses, but only with overtraining (Wright, 1964, p. 9).

The many differences between child and adult performance in discrimination tasks raises the question of whether these differences arise from varying degrees of skill or from basic differences in perceptual strategies. One hypothesis, claiming basic differences in perceptual strategies, is that children rely more heavily on auditory mediators than adults, probably as a result of the emphasis on oral reading in elementary school and the emphasis on swift silent reading in adulthood (Goodman, 1968). Gibson and Yonas (1966b) tested this hypothesis in an experiment in which children and adults scanned down a list of 30 strings of four letters each, looking for a designated letter. The authors found that a highly confusable visual context (i.e., highly similar features available in both target and nontarget letters) reduced scanning rate significantly for both children and adults, but that when letters which were highly confusable aurally with the targets were played over earphones, the performance of neither group was affected. These results do not support the hypothesis that visually presented letters are encoded to acoustic representations before they are recognized, for either children or adults.

Another hypothesis is that adults process visual information in a parallel manner whereas children process visual information in a serial manner. Gibson and Yonas (1966a) report a study in which they attempted to demonstrate this developmental difference in information processing. In this study, the performances of children in second, fourth, and sixth grades and of college sophomores were compared in a visual search and scanning task under three experimental conditions. In Condition I, there was a single target to be found in a list of letters with low visual confusability; in Condition II, two target letters were sought but only one appeared on the list; in Condition III, a single target letter was sought in a list of highly visually confusable letters. It was predicted that if parallel processing increased with age, age would interact with number of targets. The results revealed that search time decreased with age in all three tasks, searching for two targets was no harder than searching for one, and a highly confusable visual context increased search time at all age levels. These results imply, then, that both children and adults process visual information in a parallel manner and that the main difference in performance at different ages is in level of skill.

The studies reported in this section support the existence of a feature scanning process in human visual perception. They also suggest that as children learn to read, they learn to make more efficient use of the features of the writing system. Furthermore, this increase in skill is probably best described in quantitative, rather than qualitative, terms.

Feature Scanning: The Skilled Reader

Although there are few qualitative differences in the word recognition processes of children and adults, there are large quantitative differences. The skilled reader makes efficient use of visual information: he scans appropriately from left to right and top to bottom; he attends mainly to the critical features of the stimulus.

Because of the left-to-right scan pattern, skilled readers rely more heavily on information available from the beginning of a word. Huey (1968, pp. 96-98) studied the importance of the first and last halves of words by having subjects read passages from which the first or last half of each word was deleted. His findings demonstrated that more words were recognized, in less time, when the first halves were presented. Similarly, Huey (1968, p. 99) describes research by Javal in which subjects were found to perform better in reading material which the bottom half of the letters had been deleted than in reading material in which the top half had been deleted.

In the studies mentioned above, the scanning phenomena are clearly physical and involve overt movements of the eye. However, in the later studies employing tachistoscopic presentations in which overt eye movements cannot be made, the scanning effects described above were maintained. Bruner and O'Dowd (1958) prepared versions of 90 common English nouns with typographical reversals at the beginning, middle, or end. Tachistoscopic recognition was more disturbed when the error was in the beginning of the word than when in the middle or end. Similarly, Miskin and Forgays (1952) found better tachistocopic recognition of figures to the right of fixation and Mandes and Ghent (1962) found better tachistoscopic recognition of figures with their distinguishing features on the top and left of the figure.

The results of the tachistoscopic studies described above are explainable in terms of an internal scanning process in short-term memory. Scanning without overt eye movements is possible because the visual input is stored briefly in short-term memory. Before the visual memory decays, it can be <u>read out</u> or scanned in the same manner as if the stimulus were still in view (Sperling, 1963; Neisser, 1967, pp. 15-45). The internal scanning interpretation of tachistoscopic recognition is supported by research demonstrating a clear correlation between post-exposure eye movements and accuracy of report (Bryden, 1961).

DISTINCTIVE FEATURES

Up to this point, a great deal has been said about the distinctive features of the writing system, but there has been no attempt to define precisely what these features are. This lack of definition is not an oversight, but the result of no one having offered an acceptable definition.

Distinctive features are closely related to what are called, in the psychological literature, <u>cues</u>. A cue is any discernible aspect of a stimulus event which varies sufficiently from one event to at least one other event that it can be used as the basis for discrimination between the two events (Bruner, Goodnow, & Austin, 1956, p. 26). Thus, a cue has some range of values, whether discrete (e.g., "yes-no," or "A-B-C-D...") or continuous (e.g., "0 to 100%") in nature. The term distinctive feature refers to a special type of cue. Only those cues which are used, rather than could be used, to discriminate between events are called distinctive features.

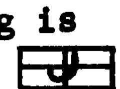
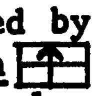
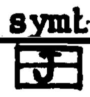
Historically, "distinctive feature" is a linguistic term in phonology. It was first introduced by Jakobson, Fant, and Halle (1963). In phonology, distinctive features are speech elements (mainly articulatory) which make up sounds as well as words. English speech sounds can be described by about a dozen distinctive features; speech sounds of all languages can be described by 30 to 40 different features. Experimental work with phonological distinctive features has supported the analysis of Jakobson et al. (1963). For example, sounds agreeing in all but one feature are consistently judged more alike than those differing by two features (Greenberg & Jenkins, 1964). However, analysis of the phonological feature system in terms of the weights which should be assigned to the differnt features is not yet complete.

The analysis of the distinctive features of the writing system is not to the point where features can be as clearly identified as for phonology. What can be said with some assurance is that, individually, orthographic features are visual cues, properties, or elements which combine in various ways to produce letters, letter sequences, words, and word sequences. These elements are distinctive because they contain information which reduces the set of alternatives that the configuration (word or letter) might be. A distinctive feature cannot be present in all possible letters and words.

Gibson (1965) was one of the first to attempt to define the distinctive features of written language. Feature selection was based on four criteria: (1) the distinctive features must be critical ones which reduce uncertainty (i.e., present in some letters but not in others); (2) the distinctive features must be invariant under perspective and size transformations; (3) the distinctive features must yield a unique pattern for each grapheme; and (4) the list of features must be an economical one (Gibson, Osser, Schiff, & Smith, 1963). Gibson's examination of printed capital letters presented in isolation resulted in a list of features which included straight and curved lines, intersection, symmetry, and discontinuity. From this feature list, Gibson then predicted that certain letters would be more confusable than others. This prediction was supported. There is no means, however, of identifying which of the proposed features contributed to the higher confusability of certain letters. For example, if only one or two of the proposed features were relevant, the predicted difference would probably still have been obtained.

A recent study (Dunn-Rankin, 1968) supports several of Gibson's findings. The performance of 315 second and third grade children was studied in a relative discrimination (modified matching-to-sample) task. Presented with a target letter to the left of five pairs of letters in normal orientation, each child was to circle the letter of each pair that looked most like the target letter. By analyzing the cumulative choices of the children, sets of linear scale values were assigned to the lower case letters, describing their relative similarity to every other letter of the alphabet.

The advantages of Dunn-Rankin's analysis stem from the fact that a fairly reliable measure of similarity is produced and this quantification allows some insight into the relative importance of the different features. Unfortunately, there is no evidence that a relative discrimination task provides much information about confusions made in reading. By requiring the children to make a response concerning the similarity of two letters (that were obviously not identical) under unlimited time conditions, the experimenter created a cognitive problem-solving task in which decisions were made concerning which dimensions were relevant. Thus, the choices available to the subject may have biased his response. As Trabasso and Bower (1968) point out, noticing and using a cue for the specific purpose of naming or identifying that cue is quite different from noticing, learning, and using a cue for the purpose of naming or identifying a total stimulus pattern (Trabasso & Bower, 1968, pp. 85-86). For example, rotational errors were probably encouraged since the child had enough time to rotate the figures mentally (if not physically) before responding. This task probably also encouraged the child to respond to items on the basis of a few cues or features which might not even be critical ones in reading and word recognition tasks.

Another, rather dissimilar, feature system has been proposed by Eden and Halle (1961) and by Eden (1962) for use in computer recognition of handwritten messages. In this system, all cursive writing is described by four primitive symbols: (1) bar  ; (2) hook  (3) arch  ; and (4) loop . By certain transformations about the horizontal and vertical axes, a larger number of features is generated. These different features or strokes are then combined according to specified rules, and continuous lines are drawn between them. This feature system has the advantage of being precisely defined and therefore readily amenable to empirical investigation. The system works well in computer recognition. There is no evidence, however, that these primitive symbols are the features used by skilled readers.

To date, most attempts at feature definition have relied on confusion matrices (e.g., Gibson et al., 1963). There is evidence, however, that this method is far from adequate. In a recent study, Fisher, Monty, and Glucksberg (1969) obtained letter confusion matrices at two different exposure durations. These two matrices were then compared with one another and with two matrices compiled by other researchers. Virtually no correspondence was found between the resulting patterns of confusions from the four matrices. Fisher et al. concluded that confusion matrices are a function more of procedure and technique than of underlying perceptual mechanisms.

It appears, then, that definition of the distinctive features of written language will have to be delayed until more adequate means of investigation are developed.

FEATURE COMBINATIONS

Features have been described both as descriptive attributes abstracted from letters (Gibson, 1965) and words and as discrete segments or subtemplates of letters and words (Eden, 1962; Eden & Halle, 1961). Both definitions imply that individual letters and letter sequences are composed of some number of different features. Furthermore, it is clear that it is not the existence of these features in themselves that makes a form recognizable, but the total arrangement of the features in the whole configuration. As Huey (1968, p. 75) noted: "Thus –'ↄ is not recognized as 5 nor <| as K, although the constituent parts are present." Discrimination of the different features within a configuration is crucial to recognition, but without knowledge of the rules or relations by which these features are combined, recognition cannot occur.

Criterial Sets of Features

Those combinations or sets of features which uniquely determine the identity of a particular letter or word configuration are termed criterial sets of features. The skilled reader probably does not visually process all the available features, but only those which are sufficient for recognition, i.e., a criterial set. Those features which are not members of a criterial set are redundant: they offer no new information to the reader concerning the identity of the letter or word. Recognition occurs through the isolation of significant or criterial features from irrelevant background detail (Stevens, 1961).

There are often a number of possible criterial sets defining a particular configuration: these different sets are termed alternative criterial sets of distinctive features. Different criterial sets may be used by different readers on a particular configuration, or even by the same reader when presented with the same configuration on different occasions. Which criterial set will be used by a reader on a particular occasion depends largely on his expectancies concerning what the configuration will be.

A set of features describes a particular configuration. Within this set, there are one or more subsets of features which are sufficient for recognition. These subsets are the alternative criterial sets. A skilled reader can use any one of the different alternative criterial sets to recognize a configuration just as the average observer can use any one of many different sets of cues to recognize a particular make of car.

Similarly, the average observer is often presented with a wide variety of stimulus events to which he can react with a limited number of responses. Because there are usually more stimulus events than response categories, a particular response is often elicited by more than one stimulus event. Thus, there are classes of equivalent stimuli with respect to certain responses (Lawrence, 1963, p. 180). For example, one often responds with "dog" whether the stimulus happens to be a German Shepherd or a Pekinese.

The skilled reader is often presented with letters and words occurring in a variety of print conditions (e.g., different cases, sizes, type styles) and these different forms of the same letter or word are often composed of quite different feature sets. Recognition of these configurations occurring in different print forms as members of a single category is possible because of the functional equivalence of their feature sets. Having learned that these different configurations are functionally equivalent, the reader scans simultaneously for any number of equivalent criterial sets.

Furthermore, when an observer recognizes an object as a member of a particular equivalent set, it is often not necessary that he continue his analysis to the point where he can uniquely recognize the object as a specific member of this set, even if it is possible for him to do so. For example, after responding "dog" to a particular furry object, the average observer probably would not examine the animal more closely to determine that this particular dog was a "Pekinese." Similarly, the skilled reader would not continue his analysis of the word dog to determine that the stimulus consisted of the word "dog" printed in lower case IBM delegate type, even if he could do so were such a response necessary. In this sense, then, the reader is unaware of the style in which a word he has recognized appears, e.g., whether it is in capital or lower case print. When presented with the word hat or with the word HAT, the visual system's scan of the word reveals the features that comprise one of the equivalent feature sets for "hat." The reader could, on closer analysis, identify the print case, but this in itself is irrelevant to the recognition process.

Feature Processing

It is assumed that the reader processes functionally equivalent sets of distinctive features simultaneously rather than serially.[4]

[4]It is important to note the distinction between simultaneous processing of visual information and simultaneous scanning for visual information. By simultaneous processing, Neisser (1964) refers to the ability of the visual system to check simultaneously for the presence of different features or stimulus attributes within a limited foveal area. By simultaneous scanning, Eriksen and Lappin (1967) refer to the ability of the visual system to independently scan total configurations presented simultaneously on different foveal locations. Although there is evidence that both simultaneous scanning and simultaneous processing occur in the human visual system, this paper's main concern is with the phenomenon of simultaneous processing of visual information.

Neisser's (1964) research is pertinent here. Neisser demonstrated that the visual system does simultaneously process more than one feature set, since it takes no longer for the skilled reader to search for several items in a list than to search for one item. It seems probable than Neisser's subjects placed the alternative correct items into a single equivalent category, "correct." For this reason, they were able to respond to the presence of any one of the possible correct items without being aware of which particular item it was. In simultaneously processing the equivalent sets of the different correct items, the subjects probably did not process the additional features needed for unique recognition of particular members and, therefore, were often unaware of the specific identity of the item to which they had correctly responded.

Sanford (1887, p. 426) provides further evidence that functionally equivalent items are recognized as members of a category rather than as particular members of that category. Sanford notes that when upper case A is reduced in size and is presented at a distance which does not allow recognition, it is most frequently mistaken for lower case a. The subject has enough information to recognize the letter as a member of the category "a" but not enough information to identify it more specifically in terms of case.

Smith, Lott, and Cronnell (in press) and Smith (1969) have shown that the skilled reader's ability to recognize words printed in different type styles is due to his ability to search simultaneously for a large number of equivalent sets of features. In one study (Smith, Lott, & Cronnel, in press), subjects searched for words in passages of text, some of which were printed in normal upper and lower case, others with alternate letters varying in case, or with alternate letters varying in size. The number of words recognized in a given time for the different conditions supported the experimental hypothesis that alternation of size rather than alternation of case accounted for the differences between conditions. These results support the view that different forms of the same letter sequence are treated as functionally equivalent by the skilled reader. Furthermore, although subjects were aware of the peculiarity of the print in several of the conditions, many were unable to identify the scheme which this distortion followed.

A series of recent studies by Posner et al. (Posner, Boise, Eichelman, & Taylor, 1969; Posner & Keele, 1967; Posner & Mitchell, 1967) cite results contrary to the notion of functional equivalence. In these studies, subjects were shown letters (either simultaneously or successively) which were either of same or different case. The experimenter measured the length of time required for a subject to respond whether or not the two letters had the "same name." The results indicated that subjects responded with "same" more quickly when the two letters were of the same case than when they were of different cases (e.g., AA versus Aa). While this finding is not predicted by the functional equivalence hypothesis, it may not be relevant to it. Posner and his associates were studying equivalence in a matching task and the process involved may have little similarity to word recognition processes. Research by Eriksen, Munsinger, and Greenspon (1966) suggests that quite different processes are involved in same-different discrimination than are involved in recognition.

The role of functional equivalence in word recognition is somewhat complicated by the redundancy which pertains to continuous text. For example, in the stimulus configuration THE CHT, H is recognized as H in THE and A in CHT (Neisser, 1967, p. 47).

REDUNDANCY

> It seems certain that in the longer reading the parts most distant from the fixation point are not clearly seen except with the mind's eye; they are filled in mentally by suggestion from what can actually be seen, somewhat as we recognize a friend from a glimpse of his hat and cane or of his bowed form (Huey, 1968, p. 63).

Printed English is not composed of a random sequence of language elements: some elements are more likely to occur than others; some elements are more likely to occur within the context of other elements. The extent of this nonrandomness or predictability is estimated to be about 50% (Shannon, 1951). If a skilled reader is shown a sequence of letters from actual text, he can predict the next letter with far better than chance accuracy (Carson, 1961). Research has demonstrated that subjects can read passages in which up to 30% of the text has been deleted (Chapanis, 1954).

Redundancy describes the extent to which language is nonrandom or predictable as a result of its statistical properties or constraints. There are many levels at which redundancy may operate: interword or contextual redundancy, intraword or letter redundancy, and featural redundancy. At each of these levels, redundancy may be specifically described in terms of: (1) distributional constraints or the redundancy due to the fact that language elements are not used equally in the language as a whole, and (2) sequential constraints or the redundancy due to the probabilities that certain language elements will be preceded or followed by certain other elements. Although both distributional and sequential constraints must be considered to obtain an accurate estimate of language redundancy, the amount of redundancy due to distributional constraints is small relative to that due to sequential constraints (Garner, 1962, p. 252).

Contextual Redundancy

Certain words occur in the language with markedly higher frequencies than other words; certain words are more likely to occur in certain passages and in certain positions within a passage. Context serves, then, to decrease the set of alternative words (Miller, 1962) and enables the skilled reader to approach a passage of normal text with certain expectations concerning what a word will or will not be. As a result, the skilled reader can sometimes identify the word without visually processing it. When this happens, the unprocessed word is totally redundant.

It is possible to determine the degree of predictability or constraint placed on a word by its surrounding context by deleting the word from its context and measuring the subject's ability to guess the word. Aborn, Rubenstein, and Sterling (1959) studied subject's ability to replace a single word deleted from sentences of 6, 11, or 25 words in length. In each case, the word was omitted from one of four positions: sentence initial, early medial, late medial, and sentence final. Analysis considered sentence length, word position, and word class (noun, verb, adjective, adverb, pronoun, or function word). The results showed large differences in the predictability of words of different classes, with function words the most predictable. Words in medial positions were more predictable than words in initial or final positions, and predictability increased with sentence length regardless of position (up to about 11 words).

Shepard (1963) measured contextual constraint by the rate at which subjects generated alternatives to replace omitted words. He assumed that if the guessing rate is high, the constraints on the word are low; if the guessing rate is low, the constraints on the word are high. Shepard found a monotonic increase in average rate of listing words as a function of informational uncertainty of the omitted word (as measured by the Shannon technique--see p. 23).

In another experiment (Morton, 1964), subjects read out loud as quickly as possible 200-word passages of zero, first through sixth, and eighth order approximation to English (as determined by the Shannon method). Morton assumed that when a subject reads passages with varying degrees of constraint, his reading speed increases up to the passage which has the amount of constraint normally used by the reader. The results showed that reading speed per syllable increased up to the fifth order approximation.

These experiments demonstrate that contextual constraints do exist. However, when a single word is deleted from a sentence, skilled readers can replace it with an accuracy of only 40 (Aborn et al., 1959) or 50% (Morrison & Black, 1957). This finding is of interest because it reveals that words are considerably less predictable than letters (Garner, 1962, p. 261).

Intraword Redundancy

Since words are less predictable than letters and since words can readily be broken into sequences of letters and spaces, most studies of redundancy use the letter as the unit of analysis. Furthermore, according to Garner (1962, p. 241), less than 15% of the total constraint among letters is due to influences which extend across word boundaries and fairly accurate estimates of language redundancy may be made in terms of letters within word units.

Intraword redundancy has been estimated at between 50 and 60%, the exact amount depending largely on whether unilateral or multivariate estimates are made. The major unilateral technique for estimating intraword redundancy is Shannon's (1951) "guessing game." A subject is given samples of English and asked to guess the next letter of the sample. Guessing continues until the correct response is given; then the procedure is repeated with the next letter in the sequence. Using the Shannon technique, Burton and Licklider (1955) found intraword constraint of English to be about 50% when a sample of 12 or more consecutive letters is studied.

Multivariate estimates of intraword redundancy tend to yield somewhat higher estimates of constraint, around 60% (Carson, 1961). When the multivariate approach is employed, a letter is deleted from any position within a word and the subject is required to fill in the missing letter. Research using this technique has the advantage of being able to study predictability of any letter position within a sequence and thus offers valuable information concerning the relative degrees of constraint in various positions within the word (Garner, 1962, pp. 224-239).

Featural Redundancy

Redundancy can also be described in terms of the constraints due to relations between features. Featural redundancy, like other forms of redundancy, refers to the availability of more features than are necessary and sufficient to define a particular configuration, i.e., more features than one criterial set.

Featural redundancy is increased whenever the amount of information about the configuration is increased. This information may come from a number of sources. For example, a series of letters may provide redundant information concerning the relative size of the letters. When the relative size of lower case letters is distorted, a subject's ability to recognize words is significantly reduced (Smith, 1969; Smith, Lott, & Cronnell, in press).

Featural redundancy is further increased in words or letter sequences adhering to English spelling patterns, which are generally defined in terms of sequential letter probabilities rather than in terms of sequential feature probabilities. However, it seems likely that there would be a high correlation between letter probabilities and feature probabilities. Since there is no reliable evidence at this time concerning what the features actually are, precise information on featural constraints is impossible to obtain and sequential letter probabilities must be used as the best predictor of degree of featural constraint. Still, the two forms of constraint are not equivalent. Featural constraints refer to the probability that certain features will be followed or preceded by certain other features rather than the probability that certain letters will be preceded or followed by other letters. An example of sequential letter constraints is the probability that *a* will be followed by *t*; an example of featural constraints is the probability that the feature set which defines *a* will be followed by an ascender.

Since featural redundancy is reflected in measures of sequential letter constraint, the probability that a certain feature will be preceded or followed by certain other features is related to the probability that a letter will be preceded or followed by certain other letters. Therefore, the skilled reader's knowledge of English spelling patterns enables him to predict from the context of certain features which other features are likely to occur, and words and non-words comprised of frequently ocurring spelling patterns are recognized on the basis of fewer features (Smith, 1967).

Because the visual system attends to the word more or less as a unit, the relations between features across letters are more important in word recognition than the relations between features within letters. When these relations are disrupted, as in <|C—/\, the average reader is unable to recognize the word Key. On the other hand, this attention to relations between features across letters often results in mistakes in recognition of individual letters in the word due to combining parts of adjacent letters (Huey, 1968, p. 94).

Smith (in press) demonstrated the sensitivity to features across letters by comparing the contrast thresholds for recognition of letters in words and of letters in isolation. When letters were in words or nonwords with high sequential letter constraint, they were recognized at significantly lower contrast thresholds than when they were in isolation. Although the observer did not have enough information to recognize the word, the relations between certain features within the word allowed him to recognize parts, or letters, of the configuration.

Similarly, Kolers (1965) found that subjects can recognize letters in a word that has been transformed temporally or spatially and still be unable to recognize the word itself. In this case, the reader does not have sufficient information about the relations between features across letters to recognize the word.

Use of Redundancy

There is little doubt that the skilled reader uses redundancy in word recognition tasks, and that he has lower recognition thresholds for highly redundant sequences than for less redundant ones. It is also clear that a reader's knowledge of the rate of co-occurrence of letters and words in the language improves his ability to recognize words. Because the skilled reader "knows" that some words or letters are not likely to occur at certain points in a sequence, the amount of uncertainty concerning that word or letter is reduced and the word or letter can be recognized from a very brief exposure. The ability to recognize configurations from limited visual information is attributed to a "filling-in" process based on learned patterns of sequential co-occurrence or redundancy (Wiener & Cromer, 1967). The basis for this knowledge of language redundancy, however, has not yet been identified. Among the possibly pertinent factors are meaningfulness, pronounceability, word frequency, letter frequency, and sequential probability of letters.

Gibson, Bishop, Schiff, and Smith (1964) attempted to analyze the effects of meaningfulness and pronounceability on word recognition by measuring contrast thresholds for recognition of three types of trigrams: (1) trigrams which were pronounceable as monosyllables, according to rules of English pronunciation; (2) trigrams which were meaningful, as defined by semantic reference, but not pronounceable; and (3) control trigrams which were low in both meaningfulness and pronounceability. Both the highly meaningful and the highly pronounceable syllables were recognized at lower contrast thresholds than the control trigrams, and the highly pronounceable syllables were recognized at lower contrast thresholds than the highly meaningful ones. But the two variables cannot really be compared.

Highly pronounceable letter sequences are clearly more readily recognized than unpronounceable sequences, but is this increased recognizability due to their high pronounceability, or is it due to the high sequential probability of their letters? Anisfeld (1964) argues the case for sequential probability. He notes that in the Gibson et al. (1964) study the frequencies of digrams and trigrams for the pronounceable and unpronounceable sequences were not controlled. Using the Underwood and Schulz (1960) digram-frequency tables, he summed the frequencies of the successive digrams of each word. This analysis revealed a significantly higher recognition score for words with high digram frequency. However, since pronounceability and digram frequency co-vary, it is difficult to identify the source of the observed effects in this study.

A later study (Thomas, 1968) examined the roles of pronounceability and consonant-vowel order in children's tachistoscopic recognition of three-letter words and pseudo words. His central finding was that even with sizable differences in pronounceability, there were no differences in recognition for CCV sets. He concluded that the Gibson et al. (1964) findings could be accounted for by consonant-vowel order as well as by pronounceability.

The answer to the question of whether visual recognition is influenced by pronounceability <u>per se</u> or by the underlying spelling patterns is not available. The answer is difficult to obtain because spelling is intricately involved in the pronunciation of written sequences (Venezky & Weir, 1966). There are clear advantages to making empirical analyses on the basis of spelling patterns since spelling patterns are quantifiable in terms of sequential probabilities. Moreover, there is evidence that pronounceability *per se* is not a pertinent factor in word recognition, since Gibson's pronounceability effects are just as great for deaf children as for hearing children (Gibson, Shurcliff, & Yonas, 1966). Since it is unlikely that pronounceability plays a role in the visual word recognition of deaf children, the similarity of the spelling patterns of pronounceable trigrams to English is probably the basis of their improved recognizability.

Another hypothesis relates recognizability to the spoken frequency of words in the language. Postman and Rosenzweig (1956) examined the effects of spoken frequency on visual recognition by giving either visual or auditory training on selected trigrams to adult subjects and then measuring the contrast level at which these trigrams were recognized by the two groups. They found that when visual training was used, frequency of prior exposure significantly influenced recognition thresholds, but when auditory training was used, there were only small and insignificant changes in threshold as a function of frequency. Thus, it seems that spoken frequency of words has little influence on visual recognition.

The Postman and Rosenzweig (1956) study supports the hypothesis that written frequency of words affects recognition, with more frequent words being recognized more readily. Howes and Solomon (1951) also report data from two experiments in which exposure duration threshold, as measured tachistoscopically by the ascending method of limits, was found to be an approximately linear function of the logarithm of the relative word frequency, as determined by the Thorndike-Lorge count (1944).

However, in some cases, non-words are recognized more easily than, or nearly as easily as, actual words (Smith, 1967). Also Neisser (1967) claims that even rare words have low thresholds for recognition if no word with a similar configuration exists in the language. If word frequency were the only factor involved, non-words (those having zero frequency) and rare words (those having low frequency) would be much more difficult to recognize than frequent ones. Moreover, word frequency is often confounded with word length and syntactic function, although these factors are more relevant dimensions of the recognition process in connected discourse.

Frequency of individual letters in written English is not likely to be the only variable affecting word recognition thresholds. If it were the only significant variable, letters in isolation would be recognized as readily as letters in words and there is evidence that this is not the case (Smith, 1967). Thus, intraword redundancy must be at least partly a function of the relations between letters within words.

The lack of definitive evidence regarding the relative influence of one variable over another on word recognition thresholds is probably due to the fact that not one but several variables interact to determine the predictability or recognizability of a particular stimulus.

An example of such complex interactions is found in the work of Owsowitz (1963). Owsowitz found, contrary to general expectations, that words with low digram frequencies were recognized more readily than words with high digram frequencies. A study was later conducted by Biederman (1966) in an attempt to replicate Owsowitz's findings. In the Biederman study, words varying in digram frequency and word frequency were presented tachistoscopically and the responses (if any) were recorded at each exposure level until correct recognition occurred three times in a row. Performance was measured by number of incorrect trials before criterion was reached. The results of Biederman's study failed to replicate those of Owsowitz: The digram frequency effect was not significant, while the word frequency effect and the digram frequency by word frequency interaction effect were significant. Biederman (1966, p. 209) concludes that:

> WF [word frequency] appears to 'wash out' any DF [digram frequency] effects at <u>high</u> WF, while <u>high</u> DF removes any significantly greater number of trials to criterion . . . . <u>High</u> DF, low WF stimuli are not significantly harder to recognize than high DF, high WF stimuli . . . , while at <u>low</u> DF the expected relationship obtains; high WF stimuli are recognized easier than low WF stimuli.

Broadbent and Gregory (1968) further investigated the word frequency by digram frequency interaction effect. They compiled three lists of 24 five-letter words, half with counts of 100 or more per million and half with counts from 5 to 25 per million, according to the Thorndike-Lorge (1944) count. Within each word frequency class, half of the words had high digram frequencies and half had low digram frequencies, according to the Baddeley, Conrad, and Thomson (1960) rates. The words were presented tachistoscopically. The results showed that the usual word-frequency effects occurred for words of high digram frequency. However, the usual digram frequency effects did not occur among the low frequency words, since those of low digram frequency were markedly more often recognized than those of high digram frequency. This supports the existence of a rather complex interaction between word frequency and digram frequency.

Research on the effect of word frequency on word recognition is complicated further by evidence that frequency of prior exposure affects response bias rather than perceptibility. In a study by Goldiamond and Hawkins (1958), subjects were first exposed to nonsense words with differing degrees of frequency. Subjects were then told that the training words would be flashed subliminally at regular intervals on the screen and that they were to respond whether they saw the word or not. In this experiment, however, all the flashes were blanks and the experimenter merely mimicked the ascending method of limits, with one response always predetermined as "correct."

The results show an association of 1.00 between the order predicted by the frequency of prior exposure and the obtained order.

> The results of this study can be interpreted as challenging a perceptual interpretation of the relationship between word-frequency and recognition-intelligibility, where word-frequency can be placed under laboratory control. Perception was not involved in this study, yet the logarithmic recognition-frequency curves were obtained . . . . We assume that frequency as a variable does not affect perceptibility . . . . Stating that frequency does not affect perception, but does affect response bias, eliminates the contradiction as well as explaining the data (Goldiamond & Hawkins, 1958, pp. 462-463).

The statistical properties of language affect performance in tasks other than purely perceptual ones. It is helpful, therefore, to examine response effects in non-perceptual tasks in which the experimenter can clearly define the stimulus available to the subject. Smith, in an unpublished paper, used a non-perceptual technique to examine digram frequency and word frequency effects. The stimuli used in this study were 40 of the words used by Broadbent and Gregory (1968) in their tachistoscopic evaluation of word frequency and digram frequency effects on perception. From each of these words, two letters were deleted from all possible position, resulting in 400 different stimuli. The stimuli were then typed on cards with blanks indicating letter deletions. Each subject was given unlimited time to list as many words as possible that could be generated by the sequence of letters and blanks.

Smith predicted several ways in which the digram frequency and word frequency of the target word could affect performance: (1) by differences in the probability of a hit (i.e., a correct response); (2) by differences in the number of non-target words (i.e., noise); and (3) by differences in the serial order in which target and non-target words are generated.

Results showed that hits were lowest for low word frequency with high digram frequency items. Both word frequency and digram frequency main effects on hits were significant. Digram frequency was the major variable in the generation of non-targets or noise items; there was a significant word frequency by digram frequency interaction for noise rate; and noise varied as a function of position of deletion. Furthermore, there was no indication that word frequency affected the order in which responses were produced, i.e., there was no tendency for high word frequency words to be produced before low word frequency words.

A possible interpretation of these word frequency and digram frequency interaction effects can perhaps be made in terms of (1) the uncertainty of the stimulus (the number of alternatives that can be generated by a sequence of letters and blanks), and (2) the decodability of the stimulus (the relative frequency of the target word within this list of alternatives).

The greater the decodability of a word, the more likely it is to be given as a response (especially if the uncertainty of the stimulus is low); the greater the uncertainty of the stimulus, the less likely the target is to be given as a response (especially if the decodability of the target is low). A study by Broerse and Zwaan (1966) supports this analysis:

> In summary it may be concluded that high frequency of the solution word and high redundancy [i.e., low uncertainty] of the missing word part both facilitate the identification of the word. A large number of alternatives provided by the given n-gram and high redundancy [i.e., low uncertainty] of this word part lengthen solution time, although it cannot be decided whether these factors are to be distinguished or not (Broerse & Zwaan, 1966, p. 444).

Uncertainty is defined as the size of the set of alternatives from which a particular symbol is drawn (Miller, Bruner, & Postman, 1954): the larger the set of alternatives, the more uncertain and the less redundant the symbol is; the smaller the set of alternatives, the less uncertain and the more redundant the symbol is. Highly redundant sequences are more predictable because they are drawn from a smaller set of possible alternatives. Highly redundant sequences are more readily recognized at least partly because they are more predictable. When the exact stimulus available to a subject is not clearly defined (as in a word recognition task), it is not always possible to determine the uncertainty of the stimulus perceived by the subject. In such situations, the best predictor of uncertainty is probably the average digram productivity of the stimulus word, where digram productivity is defined as the number of different words in which a particular two-letter sequence occurs in English. Digram frequency, the variable usually employed as an estimate of uncertainty, is not really an adequate predictor because it is too heavily influenced by word frequency and thus reflects decodability effects as well as uncertainty effects.

The size of the set of possible alternatives (<u>the uncertainty</u>) is not the only pertinent variable in word recognition. Some alternatives occur more frequently than others and are, therefore, more likely to be reported. Thus, within each set of possible alternatives, each member must be weighted according to its frequency of occurrence, or its <u>decodability</u>. Highly decodable words with low uncertainty should be easiest to recognize and highly uncertain words with low decodability should be most difficult. As is the case with uncertainty, it is not always possible to obtain an accurate measure of decodability in word recognition tasks, since the exact members of the set of alternatives generated by a stimulus cannot always be identified. In such cases, the best estimate of decodability is the relative frequency of the target stimulus in relation to the language as a whole.

Before specific predictions may be made by the uncertainty-decodability interpretation of redundancy usage, however, further research is needed.

A TENTATIVE MODEL OF THE FEATURE SCANNING PROCESS

The information derived from the literature discussed earlier resulted in the development of a model of the visual processes involved in word recognition. This model, like Neisser's (1967) and Smith's (1967), describes the word recognition process as one in which the distinctive features of the stimulus are analyzed and synthesized within the visual system of the observer.

It is assumed here that in stimulus recognition the visual system scans the total configuration (letter or word) for its distinctive features. Unless sufficient features are analyzed during this initial scan, those distinctive features which are discriminated lead the visual system to anticipate and check for other features which are criterial for recognition. This check can result in: (1) unique recognition (the criterial features anticipated are found); (2) ambiguous recognition (the criterial features anticipated are not found, but their absence reduces the number of alternatives as to what the configuration might be); and (3) invalid recognition (the criterial features anticipated are not found and the check fails to reduce the number of alternatives, i.e., the check results in no new information).

Several ambiguous or invalid checks may be made on a single configuration, each leading to another loop in the process of visual recognition, before the letter or word is recognized. This entire analysis and synthesis occurs within the visual system of the reader, prior to conscious recognition of the configuration and prior to acoustic or semantic analysis of the stimulus (Cohen, 1968).

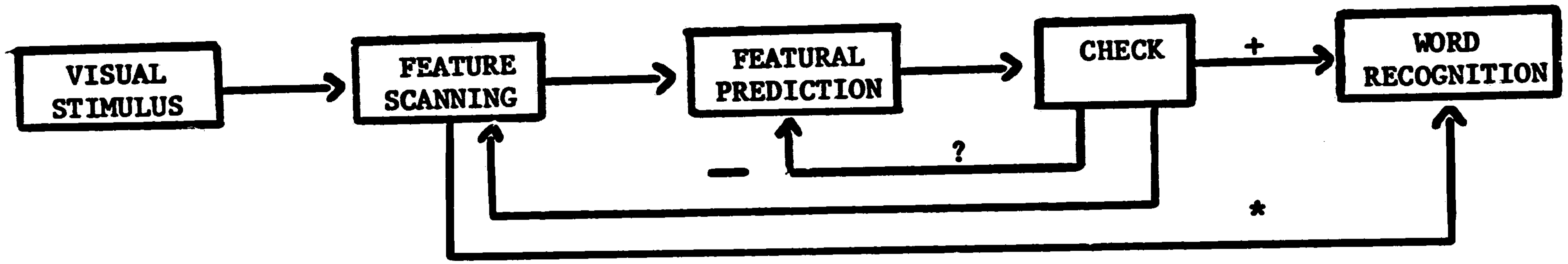
The word recognition process described here is depicted in Figure I.

The model presented in Figure I is complicated by the redundancy which pertains to normal text. The reader does not approach the stimulus initially without some expectations concerning what the configuration will be. These expectations are called the environment of the stimulus; they include the observer's knowledge of the distributional and sequential constraints of the language as a whole and of the passage under consideration in particular. If contextual constraint is 100%, the reader should be able to move directly from the environment of the stimulus to recognition, without visually processing the word at all. If contextual constraint is high (but less than 100%), the observer should be able to move from the environment to the feature prediction stage, without making the initial feature scan of the stimulus. Figure II illustrates the model when environmental constraints are considered.

The environment-stimulus interaction model described above can be seen in Huey's (1968, p. 82) description of the reading of the blind:

> The reading of the blind. . . . seems to illustrate this combination of methods of perceiving words. A practiced

FIGURE I



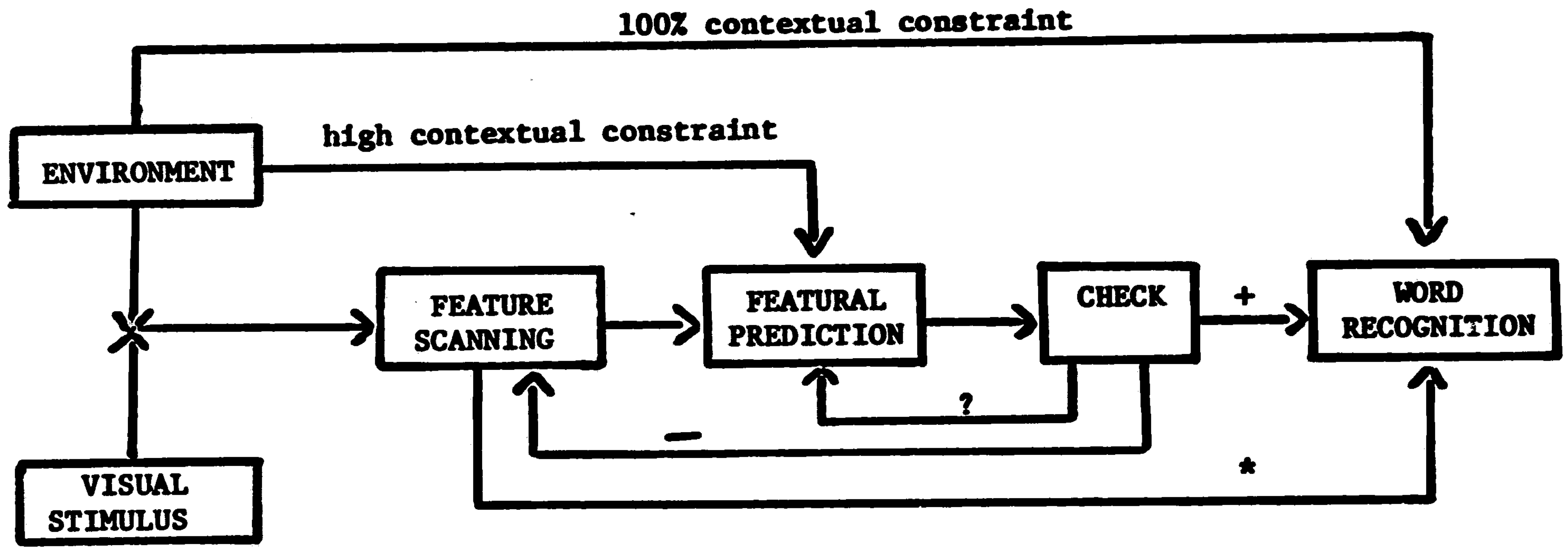
Key: + = valid predictions

? = ambiguous predictions

- = invalid predictions

* = sufficient features analyzed during initial scan for recognition

FIGURE II



Key: + = valid predictions

? = ambiguous predictions

– = invalid predictions

* = sufficient features analyzed during initial scan for recognition

> reader of the raised-letter page goes ahead with the fingers of the right hand to examine the general outline of the word, while a finger of the left hand follows, gliding successively over the letters. Ordinarily, however, only a part of the letters are examined, while the finger passes over the others without touching the points.

As Huey notes in the passage above, the initial scanning proceeds generally from left to right; it also moves from top to bottom (Mandes & Ghent, 1962). Thus, features from different areas of the stimulus have differing degrees of importance in the recognition process, with initial and upper features being more likely to be criterial in recognition than lower and final features (Huey, 1968, p. 99; Bruner & O'Dowd, 1958; Garner, 1962, p. 219). Features may be sampled randomly across the entire sequence and then assigned differential weights according to their position, or the features may be selectively sampled so that the sampled set is likely to contain more features from certain positions (i.e.c top and initial). Selective sampling is more efficient since the more heavily weighted or criterial features are more likely to be analyzed initially, and there is evidence that it is the process generally employed (Mandes & Ghent, 1962).

Although the scan moves generally from left to right, certain areas of the stimulus are scanned simultaneously. Eriksen and his colleagues (Eriksen, 1966; Eriksen & Lappin, 1967; Eriksen & Spencer, 1969) estimate that stimuli separated by an angle of one degree or more can be scanned simultaneously. Thus, the reader receives simultaneous feature information from several areas of the stimulus configuration. This simultaneous feature extraction is independent, that is, the probability of extracting one feature is not related to the probability of extracting another feature.

Still, the total feature analysis process can hardly be called independent. English is comprised of letter sequences with quite marked degrees of constraint. Knowledge of a feature, letter, or word enables the reader to anticipate what other features, letters, and words are likely to occur. Thus, in the word recognition process, there is a great deal of environmental interaction on all levels. Features may be extracted simultaneously from different segments of the word (with probably more features extracted from initial and top parts since they are higher in information value), but the extraction of each feature determines which other features may be "filled in" more or less automatically without further feature checks, and which features are criterial in reducing the number of alternatives to one and must be checked before recognition can occur. For example, extraction of _X_ features allows "fill-in" of _Y_ features and predicts the presence of _Z_ features. Check of _Z_ features results in total "fill-in" and recognition of the configuration. Which features will be checked depends on which features are originally extracted and on the degree of constraint available. Thus, feature checks may vary greatly from one reader to another or from one occurrence of the word to the next for the same reader.

The presence or absence of certain distinctive features in a particular configuration and the relations between these features lead the visual system to check for other features which are criterial to recognition, i.e., those features which are necessary and sufficient for recognition. Which features will be considered criterial depends to a great extent on the amount of redundancy in the configuration to be recognized. Thus, individual letters in different words or even in different positions within the same word will not necessarily be recognized on the basis of the same criterial sets. For example, the criterial features for recognition of the letters o and n are not necessarily the same as the criterial features for the recognition of the word on nor are the criterial features for on necessarily the same as those for no. The criterial features used in the recognition of the may be quite different depending on which the is referred to in the passage: The top of the table.

The word recognition process may be seen as a complex interaction between the characteristics of the stimulus (its distinctive features) and the characteristics of the environment (the constraints) in which it is placed. The skilled reader must be able to make use of both stimulus and environment if the word recognition process is to progress efficiently. The beginning reader, then, must learn to attend to the criterial aspects of the stimulus while taking into account the information available outside of the stimulus itself and within the environment which surrounds it.

EDUCATIONAL IMPLICATIONS

Reading is the process whereby a person translates alphabetic symbols on the printed page into meaning (Gibson, 1965). Beginning readers must be taught how to make this translation in the most efficient manner possible. But before children can make this translation, they must be able to visually discriminate the printed letter from other letters.

Visual differentiation of letters and words is the first step in learning to read (Gibson, 1965). However, it is not uncommon to find beginning readers with marked difficulties in their ability to visually discriminate among similar letters (Wheelock & Silvaroli, 1967). These difficulties are not due only to inadequate figure-name correspondences, but are also often due to difficulties in children's ability to visually process information. For example, children have difficulty in discriminating between different diagonal lines and between right and left figures (Rudel & Teuber, 1963); they scan visual stimuli from right to left rather than left to right, as is required in reading English (Mandes & Ghent, 1962; Forgays, 1953); they attend to irrelevant cues in attempting to discriminate between stimuli (Bruner, 1965; Wright, 1964; Olson, 1967).

As children grow older, these difficulties in visual discrimination tend to disappear (Rudel & Teuber, 1963; Mandes & Ghent, 1962; Bruner, 1965; Wright, 1964; Gibson, 1965). Moreover, those skills which are most relevant to reading are acquired early; perceptual skills with little relevance to reading are acquired much later (Gibson, 1965). This implies that changes in the discrimination ability of children are not maturational alone, but that children are taught to process visual information efficiently (Olson, 1967).

If children can be taught to efficiently process visual information, how should instruction be given to achieve this aim? The instructional program should have six major areas of concern: (1) perceptual training; (2) training in use of distinctive features of letters and words; (3) training in use of criterial feature sets of letters and words; (4) training in recognition of and use of alternative criterial sets; (5) training in recognition of and use of functionally equivalent criterial sets; and (6) training in acquisition of knowledge and use of redundancy. Although these areas have been separated and serially ordered for discussion, there is obviously much interaction among them.

The first phase of the instructional program, perceptual training, is essentially that provided by current reading readiness programs in which children are taught perceptual strategies which will later be used in reading. Since children tend to scan from right to left (Mandes & Ghent, 1962) and since English must be scanned from left to right, the first perceptual strategy to be mastered is direction of scan.

Another perceptual strategy to be strengthened is learning to attend to relevant cue dimensions in a stimulus, i.e., those cue dimensions along which different stimuli may be compared and discriminated. Olson (1967) and Bruner (1965) describe a discrimination technique which could readily be employed for instruction at this stage. The child is provided with two (or more) patterns or models and is forced to choose the "correct" one from information regarding the presence or absence of a single cue in the target. For example, the child may be shown two patterns formed by unlighted and lighted bulbs arranged on a matrix. The child is then given another bulb-board which has been rigged so that the bulbs forming one of the patterns will light when touched. After touching each bulb, the child is asked which of the two model patterns is available on the rigged board. To complete this task in the fewest trials possible, the child must attend to those dimensions of the stimulus along which the different patterns may be discriminated.

Perceptual training with nonsense figures should, however, probably be kept at a minimum. Although there is evidence (Cowles, 1969) that such training significantly improves performance in reading readiness tests, Gates, cited by Wheelock and Silvaroli (1967), found low correlations between a child's ability to discriminate between geometric figures and his ability to read.

Training in use of the distinctive features of letters and words is the second phase of instruction. A possible instructional technique at this phase is reproduction training. It has been hypothesized that since a subject must recognize and use more features to reproduce a form than to discriminate it from other forms, reproduction training should be more beneficial to recognition processes than discrimination training. A study by Gibson and Osser (1963) found that reproduction training facilitated letter learning. However, since reproduction was achieved by typing the letters, it seems likely that the obtained improvement was due more to motivational and attentional factors than to any advantage of this form of reproduction itself. A recent study by Williams (1968), using a more conventional definition of reproduction, found that reproduction training was not as effective for letter learning as discrimination training.

The discrimination technique developed by Olson (1967) and Bruner (1965) could be used again in the third phase of instruction, training in use of criterial feature sets of letters and words. At this stage, the same techniques as those employed during perceptual training would be used, but the stimulus patterns would be letters and words rather than patterns. Discrimination training should begin with letter pairs which are highly discriminable--i.e., greatest number of features not shared--and progress to less discriminable pairs--i.e., greatest number of features shared (Coleman, 1967; Dunn-Rankin, 1968).

The extent of letter discrimination training, however, should be limited so that letter-by-letter reading will not be encouraged, and instruction should move quickly to word discrimination tasks. Word discrimination should progress on the same basis as letter discrimination, beginning with readily discriminable words differing along many feature dimensions and progressing to more difficult discriminations.

The fourth phase of the program, training in the use of alternative criterial sets, is begun during the third phase because in making discriminations between many of the letter and word pairs, different features are criterial in the different discrimination pairs. The use of alternative criterial sets can be further enhanced by asking the child if he can solve the problem in another way, using different features or cues.

In training children to use functionally equivalent criterial sets of features, the fifth stage of the program, the emphasis is on developing generalization of a single response to a variety of stimuli. Since equivalence training is best accomplished by attaching a single label or verbal mediator to the different equivalent stimuli (Cantor, 1955; Vanderplas, 1963), the child should be exposed frequently to letters and words in differing type styles and cases while asked to verbally label each stimulus.

The sixth stage, training in the use of redundancy, is perhaps best accomplished by displaying the sequential and distributional properties of written English through use of a controlled vocabulary in reading. Frequently occurring spelling patterns should be introduced early and should occur with high frequency in the program. Although children learn to use redundancy quite early in reading instruction (Lott & Cronnell, 1969; Gibson, Osser, & Pick, 1963; Amster & Keppel, 1968), use of redundancy would probably be enhanced through use of a controlled vocabulary.

The instructional program outlined in this section is speculative due to the lack of empirical evidence relevant to an instructional setting. The basic principles outlined here are drawn from the available research in word recognition. Research is needed at this time to determine how well these principles can be translated into instructional programs in the classroom.

APPENDIX

VARIABLES INFLUENCING THE LEGIBILITY OF PRINT

For many years it was believed that ease or difficulty in reading could be explained in terms of the typographic factors of the printed page. Within the last two decades, this belief has been discarded and the effects of typography on reading behavior no longer remain a popular area of research. Perhaps, as Davenport and Smith (1965) point out, the neglect is somewhat unwarranted since knowledge of this area remains far from complete. To this point, the majority of research has demonstrated only slight variation in legibility as the result of more marked variations in typographic structure. Probably the finding that skilled readers are able to perform quite successfully on materials printed in a wide variety of typographic combinations is much more significant and enlightening than the discovery of wide variability would have been.

Of the many articles and books studying the various factors related to the legibility of print, perhaps the most exhaustive and empirically sound is How to Make Type Readable by D. H. Paterson and M. A. Tinker (1940). Using almost exclusively the Chapman-Cook Speed of Reading Test--a test measuring both speed and comprehension in reading--the authors compare the effects on legibility--i.e., the ease of recognition of word, phrases and sentences as a result of varying typographic factors (Tinker, 1966)--of varying a number of print factors. Among the variables discussed are: styles of type face, size of type, width of line, size of type in relation to width of line, leading, leading and line width in relation to type size, spatial arrangements of the printed page, and printing surfaces.

For each variable there appears to be an optimal condition, any departure from which results in reduced legibility. Thus it may be predicted with some degree of assurance that each variable will fall along a bell-shaped legibility continuum. But, even so, it is impossible to determine optimal legibility in any particular instance because legibility is the product of the combination of all variables, rather than of the condition of any one variable. Further complicating the situation is the possibility that a whole series of highly legible combinations may exist in any given instance. The use of different styles of type face appears to have little significant effect on reading rate.

In the experiment from which the above results were drawn, all the different styles of type face were set up with 10 point type on a 19 pica line. The tendency of the authors to generalize their findings beyond those materials also set up with 10 point type on a 19 pica line cannot be accepted. It appears that use of different type sizes and different line widths might result in marked differences in the relative legibility of particular type faces. This point is supported by the data in Figure I: while there is no difference in the legibility of Granjon and Scotch Roman type when 10 point type is used, the legibility of the two type faces differs somewhat more at other type sizes.

Patterson and Tinker note that while the different type styles do not markedly affect the legibility of print, the use of different cases certainly does. Italic print produces a relatively small retarding effect on reading rate. Use of material printed entirely in capitals, however, reduces reading rate by 10 to 12 per cent. In explanation, the authors cite the increased printing space required for capitals, the fact that word forms composed entirely of capitals are less characteristic than lower case word forms, and the reader's lack of reading habits with capitals. It should be noted that the research by Paterson and Tinker in comparing lower case with upper case print was conducted with materials printed in 10 point "Old Style." Whether these results would necessarily have been forthcoming if another type face had been used is not clear. Another interesting point is that Paterson and Tinker do not present experimental evidence in support of the reasons they offer for the observed phenomenon. It seems that by simply reducing the size of the capital print to cover the same area of the page as the lower case print, the authors could have determined whether the increased size of capital letters was an important factor. Similarly, it seems possible that use of other sizes of print could have resulted in a reversal in the respective legibilities discovered. Also, perhaps their results were due to use of an optimal size for the lower case letters than optimal size for the capitals.

The heaviness of individual letters is another interesting variable influencing legibility of print. Paterson and Tinker note that bold face type can be read at a greater distance and almost as fast as ordinary lower case print and conclude that the heaviness of the print does not appear to play an important role in legibility. However, the authors do not specify the size of the print, the leading, or the style of type face which was used in this study. Possibly, legibility would be affected by letter heaviness in other type styles, in other sizes of print, or in print having different leading. In other words, there probably is an optimal condition for the heaviness of letters, although this variable may be less critical than others.

Although Paterson and Tinker note that type size is not as important a factor in legibility as previous authors had believed, it is obvious from their data that the variable of type size follows to some degree a non-linear pattern.

It is especially interesting to note that the shape of the curve differs with use of different styles of type (Figure I) and may even be multi-modal. Unfortunately, it is impossible to determine from the data for Scotch Roman type whether use of this type face would also result in a bimodal legibility curve if it were examined in the 9 point and 11 point conditions. Made relatively clear by the two studies, however, is that Granjon type tolerates increased size (at least up to 12 point type) somewhat better than does Scotch Roman type.

A glance at Figures II and III reveals the many complications involved in a discussion of such a variable as the length of a line of print. Again, a legibility plotting of line length reveals a somewhat bell-shaped curve.

This research has revealed, moreover, that the optimal length for a line of type will vary considerably as other print factors are altered. For example, 40 pica lines were found to have a high degree of legibility when used with type set solid, but when the same 40 pica lines were set up with 2 point leading, a fairly marked retarding effect on reading was noted. (Figure III)

The fact that legibilities of print size and line length are interrelated is perhaps somewhat more expected. This interaction is apparent in the fact that a 30 pica line significantly retarded reading rate when 10 point type was used, but was well within the range of optimal line lengths when 12 point print was used. To quote Paterson and Tinker, "...neither size of type nor line width, as separate factors, can be relied upon as final determinants of legibility. Both factors (and perhaps others as well) work hand in hand and must be properly balanced to produce a printed page which will promote a maximum reading rate." (1940, p. 59)

One of the variables which must be considered along with length of line and size of type is leading--as evidence in Figure III. The different effects of leading for 8 point, 10 point, and 12 point type when a 19 pica line is used are readily apparent from Figure IV. Paterson and Tinker offer a great deal of information regarding various combinations of leading, line width, and type size. The results of this research can be perhaps best summarized in terms of the optimal combinations of leading and line width for different sizes of type. Thus, the optimal combinations are summarized below:

8 point type	2 (or 4) pt. leading	14 (or 21) pica line
10 point type	2 point leading	19 pica line
11 point type	1 point leading	25 pica line
12 point type	4 point leading	25-33 pica line

Other print variables affecting legibility are discussed, but not researched in detail in the book. Among these are such spatial arrangements of the printed page as: size of full page, size of printed page, margins, single vs. double columns, and paragraphing. An interesting point here is that material printed with no margins is slightly easier to read than material printed with conventional margins. Other variables are the printing surface and the degree of illumination.

As is readily obvious from this discussion, the many variables involved in producing legible print are extremely complicated and interrelated. At this time, a truly exhaustive study of these interrelations has not been made. One might predict that, were such a study conducted, the data would reveal that as one particular factor--such as size--is varied, the legibility curves of other variables--such as type face--would assume new shapes, perhaps intersecting, thereby producing drastic differences in the rank ordering of the data. In short, the legibility of any one factor is relative to the condition of all the other typographic factors (Figure V).

As mentioned earlier, perhaps the most remarkable observation which can be made at present is that even when combinations of print variables which are far from optimal are presented, there are no truly marked reductions in one's ability to read. In fact, in an attempt to study more normal, rather than experimental, reading conditions, Davenport and Smith (1965) demonstrate that even these slight variations in legibility tend to disappear.

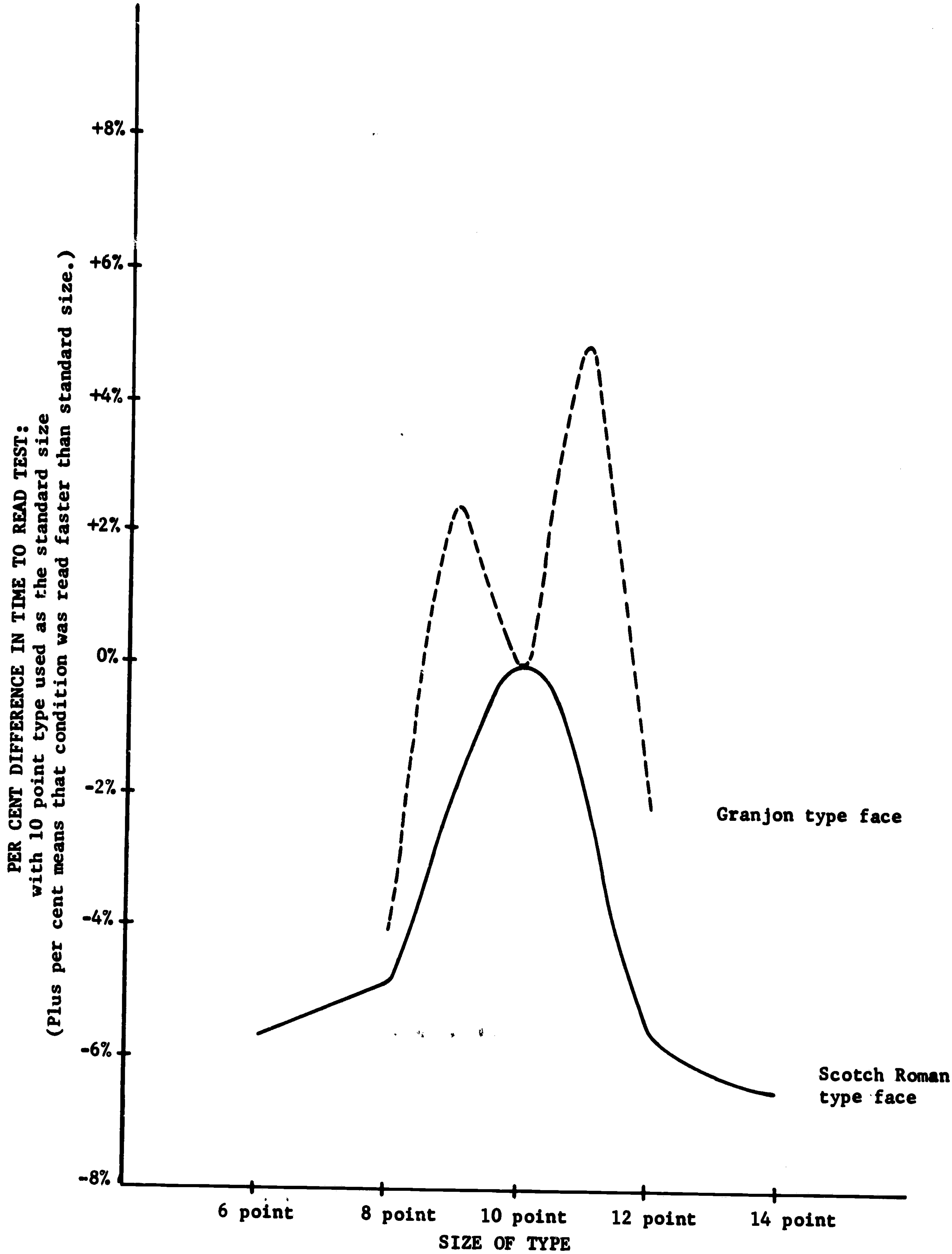


FIGURE I: Influence of size of type on legibility of print (From the data of Paterson & Tinker, 1940, p. 34-35)

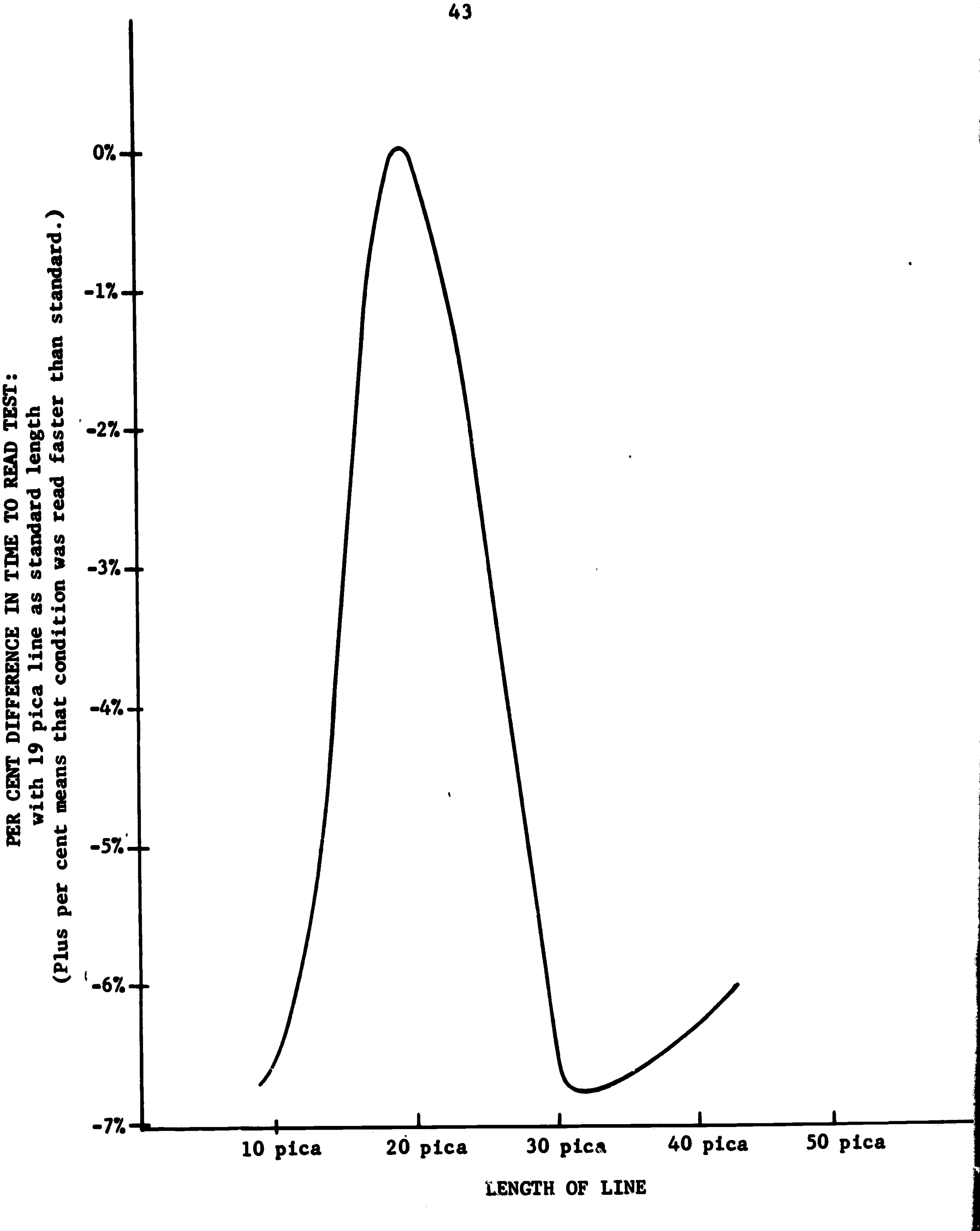


FIGURE II: Influence of line length on legibility of print (Material printed in 10 point print set solid--type face not specified.)
(From data of Paterson & Tinker, 1940, p. 44)

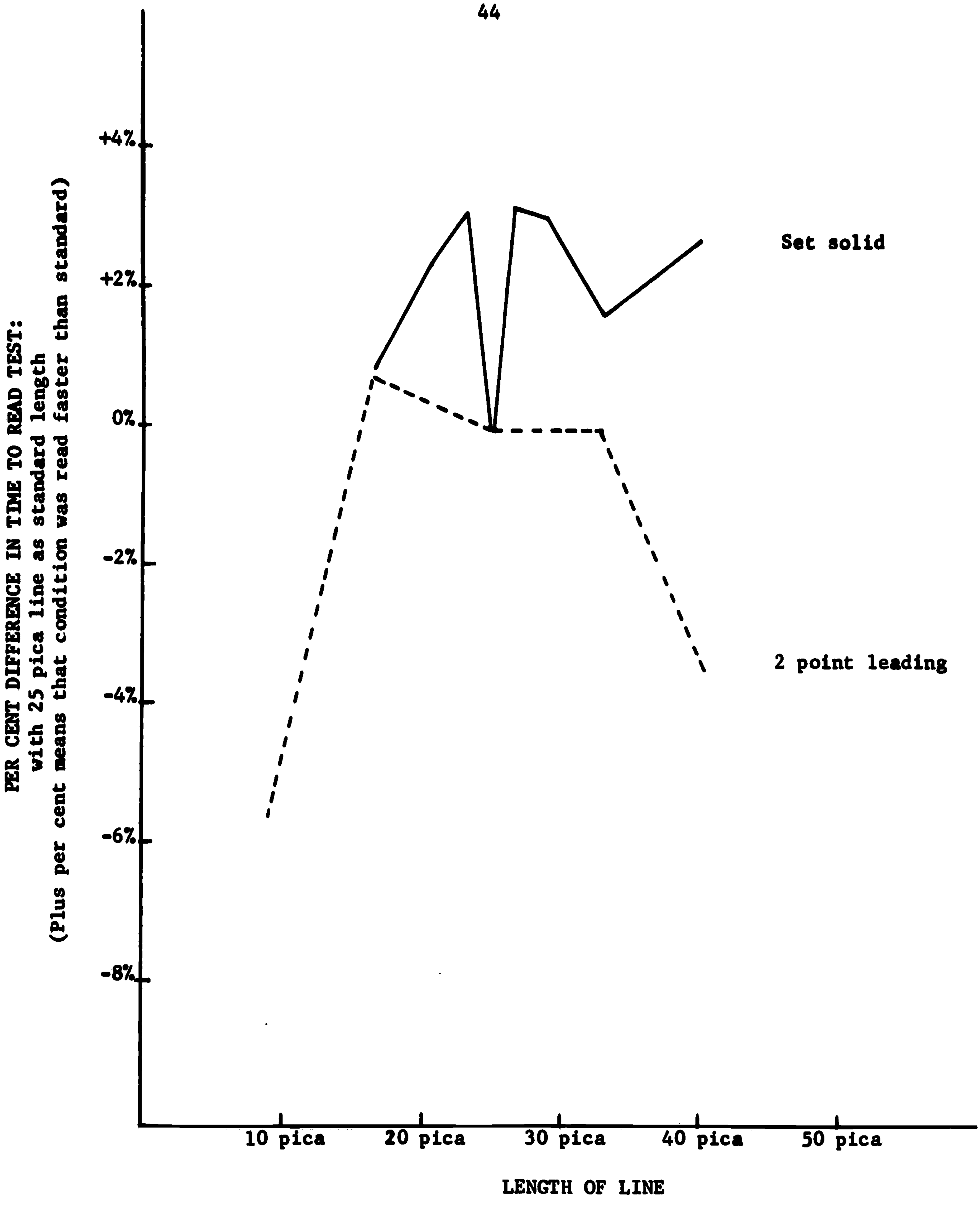


FIGURE III: Influence of line length on legibility of print. (Material printed in 12 point print--type face not specified.) (From data of Paterson & Tinker, 1940, p. 46 & 48)

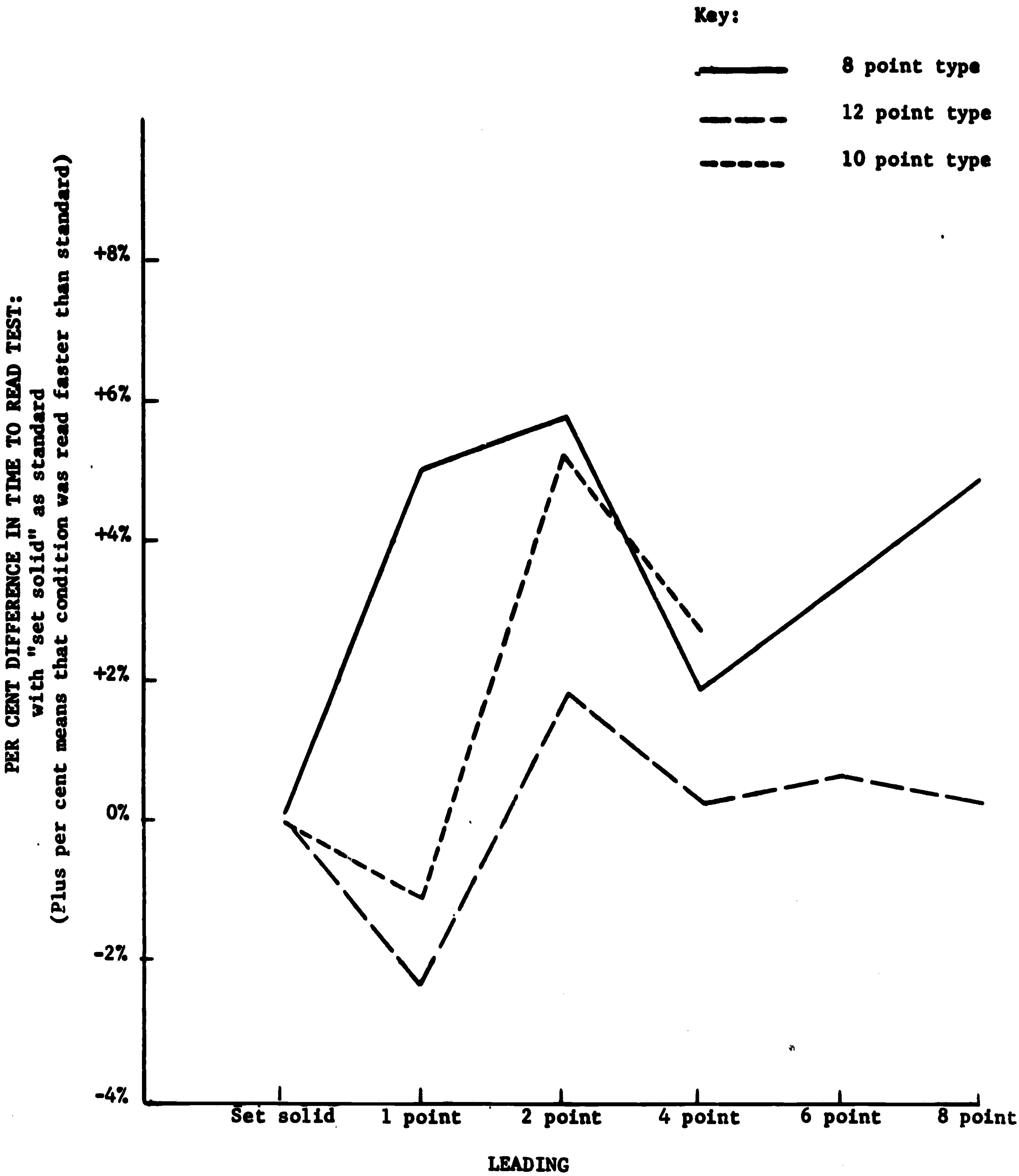


FIGURE IV: Influence of amount of leading on legibility of print. (Material printed in 19 pica line--type face not specified.) (From data of Paterson & Tinker, 1940, p. 65, 66, & 68)

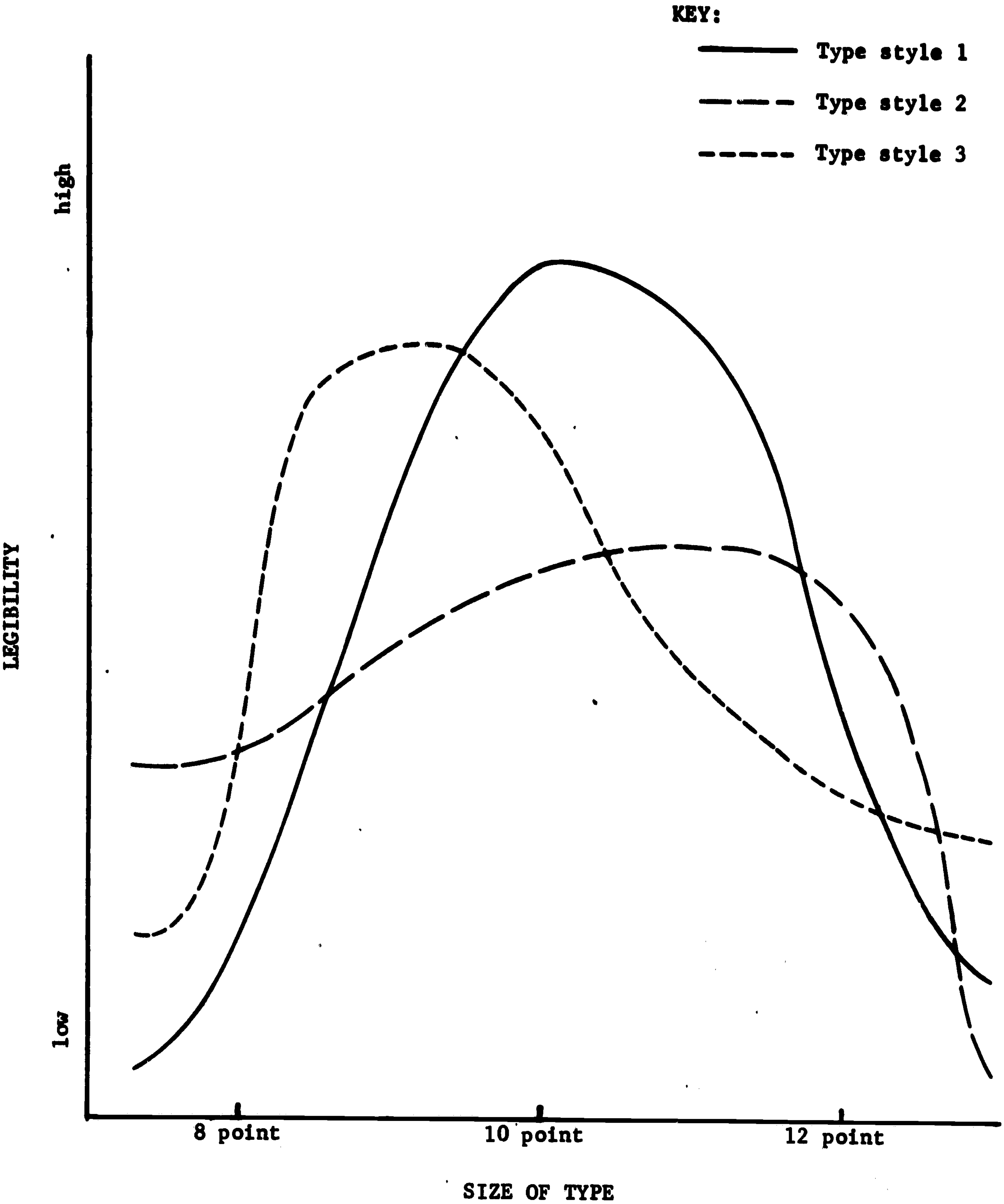


FIGURE V: A hypothetical representation of the interrelationship between type size and type style.

Rank order of legibility of type faces for:

8 point type: style 2 > 3 > 1
10 point type: style 1 > 3 > 2
12 point type: style 2 > 1 > 3

REFERENCES

Aborn, M., Rubenstein, H., & Sterling, T. D. Sources of contextual constraint upon words in sentences. Journal of Experimental Psychology, 1959, 57, 171-180.

Amster, H., & Keppel, G. Letter-sequence habits in children. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 326-332.

Anderson, I.H., & Dearborn, W. F. Psychology of teaching reading. New York: Ronald Press, 1952.

Anisfeld, M. A comment on "The role of grapheme-phoneme correspondence in the perception of words." American Journal of Psychology, 1964 77, 320-321.

Baddeley, A. D. Sequential dependencies among letters and the learning of nonsense syllables. Paper presented at the meeting of the Experimental Psychology Society, London, England, January 1964.

Baddeley, A. D., Conrad, R., & Thomson, W. E. Letter structure of the English language. Nature, 1960, 186, 414-416.

Biederman, G. B. Supplementary report: The recognition of tachistoscopically presented five-letter words as a function of digram frequency. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 208-209.

Block, H. D., Nilsson, N. J., & Duda, R. O. Determination and detection of features in patterns. In J. T. Tou and R. H. Wilcox (Eds.), Computer and information sciences. Washington, D. C.: Spartan Books, 1964, pp. 75-110.

Breland, K., & Breland, M. K. Legibility of newspaper headlines printed in capitals and in lower case. Reading Research Quarterly, 1944, 28, 117-120.

Broadbent, D. E., & Gregory, M. Visual perception of words differing in letter digram frequency. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 569-571.

Broerse, A. C., & Zwaan, E. J. The information value of initial letters in the identification of words. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 441-446.

Bruner, J. S. Eleven hypotheses in search of a method. In H. Levin (Ed.), Planning for a reading research program. Ithaca, New York: Cornell University, 1965. pp. 35-37.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. A study of thinking. New York: John Wiley & Sons, 1956.

Bruner, J. S., & O'Dowd, D. A note on the informativeness of parts of words. Language and Speech, 1958, 1, 98-101.

Bryden, M. P. The role of post-exposure eye movements in tachistoscopic perception. Canadian Journal of Psychology, 1961, 15, 220-225.

Bryden, M. P. Left-right differences in tachistoscopic recognition: Directional scanning or cerebral dominance. Perceptual and Motor Skills, 1966, 23, 1127-1134.

Burton, N. G., & Licklider, J. C. R. Long-range constraints in the statistical structure of printed English. American Journal of Psychology, 1955, 68, 650-653.

Cantor, G. N. Effects of three types of pretraining on discrimination learning in preschool children. Journal of Experimental Psychology 1955, 49, 339-342.

Carson, D. H. Letter constraints within words in printed English. Kybernetik, 1961, 1, 46-54.

Cattell, J. Über die Zeit der Erkennung und Benennung von Schriftzeichen Bildern und Farben. Philosophy Studies, 1885, 2, 630-635. Translation, in Cattell (James McKeen Cattell: Man of Science, 1860-1944. Lancaster, Penn.: Science Press, 1947, I, 13-25).

Catell, J. The time taken up by the cerebral operations. Mind, 1886, II. Translation, in Cattell (James McKeen Cattell: Man of Science, 1860-1944. Lancaster, Penn.: Science Press, 1947, I, 41-97).

Chapanis, A. The reconstruction of abbreviated printed messages. Journal of Experimental Psychology, 1954, 48, 496-510.

Chung, Shin-Ho Neurophysiology of the visual system. In P. A. Kolers & M. Eden (Eds.), Recognizing Patterns. Cambridge, Mass.: MIT Press, 1968, pp. 82-99.

Cohen, G. A comparison of semantic, acoustic and visual criteria for matching of word pairs. Perception and Psychophysics, 1968, 4, 203-204.

Coleman, E. B. Collecting a data base for an educational technology: I. Evolving a psycholinguistic reading program. Southwest Regional Laboratory, 1967.

Cowles, J. D. An experimental study of visual perceptual training and readiness scores with certain first-grade children. Paper presented at AERA meeting, Los Angeles, 1969.

Davenport, J. S., & Smith, S. A. Effect of hyphenation, justification, and type size on readability. Journalism Quarterly, 1965, 42, 382-388.

Dember, W. N. Psychology of Perception. New York: Holt, Rinehart and Winston, 1965.

Dodge, R. Visual perception during eye movement. Psychology Review, 1900, 7, 454-465.

Dunn-Rankin, P. The similarity of lower-case letters of the English alphabet. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 990-995.

Eden, M. Handwriting and pattern recognition. Institute of Radio Engineers (now IEEE) Transactions on Information Theory, 1962, IT-8, 160-166.

Eden, M., & Halle, M. Characterization of cursive handwriting. In C. Cherry (Ed.). Information theory: Proceedings of 4th London Symposium. London: Butterworth Pub., 1961, pp. 287-299.

Eriksen, C. W. Independence of successive inputs and uncorrelated error in visual form perception. Journal of Experimental Psychology, 1966, 72, 26-35.

Eriksen, C. W., & Lappin, J. S. Independence in the perception of simultaneously presented forms at brief durations. Journal of Experimental Psychology, 1967, 73, 468-472.

Eriksen, C. W., Munsinger, H. L., & Greenspon, T. S. Identification versus same-different judgment: An interpretation in terms of uncorrelated perceptual error. Journal of Experimental Psychology, 1966, 72, 20-25.

Eriksen, C. W., & Spencer, T. Rate of information processing in visual perception: Some results and methodological considerations. Journal of Experimental Psychology Monograph, 1969, 79, 1-16.

Fisher, D. F., Monty, R. A., & Glucksberg, S. Visual confusion matrices: Fact or artifact. Journal of Psychology, 1969, 71, 111-125.

Forgays, D. G. The development of differential word recognition. Journal of Experimental Psychology, 1953, 45, 165-168.

Garner, W. R. Uncertainty and structure as psychological concepts. New York: Wiley & Sons, 1962.

Gibson, E. J. Learning to read. Science, 1965, 148, 1066-1072.

Gibson, E. J., Bishop, C. H., Schiff, W., & Smith, J. Comparison of meaningfulness and pronounceability as grouping principles in the perception and retention of verbal material. Journal of Experimental Psychology, 1964, 67, 173-182.

Gibson, E. J., Osser, H., & Pick A. D. A study of the development of grapheme-phoneme correspondences. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 142-146.

Gibson, E. J., Osser, H., Schiff, W., & Smith J. An analysis of critical features of letters, tested by a confusion matrix. Cornell University Cooperative Research Project: A basic research program in reading, No. 630 (Final Report), Cornell University, Ithaca, New York, 1963.

Gibson, E. J., Pick, A., Osser, H., & Hammond, M. Role of grapheme-phoneme correspondence in perception of words. American Journal of Psychology, 1962, 75, 554-570.

Gibson, E. J., Shurcliff, A., & Yonas, A. The role of pronounceability in perception of pseudo-words by hearing and deaf subjects. Project Literacy Report, No. 7, Cornell University, Ithaca, New York, 1966, 62-72.

Gibson, E. J., & Yonas, A. A developmental study of visual search behavior, Perception and Psychophysics, 1966, 1, 169-171, (a).

Gibson, E. J., & Yonas, A. A developmental study of the effects of visual and auditory interference on a visual scanning task. Psychonomic Science, 1966, 5, 163-164, (b).

Gibson, J. J., & Osser, H. A possible effect of learning to write on learning to read. In H. Levin (Ed.), A Basic Research Program for Reading, Cornell University, Ithaca, New York, 1963.

Goldiamond, H., & Hawkins, W. F. Vexierversuch: The log relationship between word-frequency and recognition obtained in the absence of stimulus words. Journal of Experimental Psychology, 1958, 56, 457-463.

Goodman, K. S. The psycholinguistic nature of the reading process. In K. S. Goodman (Ed.), The psycholinguistic nature of the reading process. Detroit: Wayne State University Press, 1968, pp. 13-26.

Greenberg, J. H., & Jenkins, J. J. Studies in the psychological correlates of the sound system of American English. Word, 1964, 20, 157-177.

Harcum, E. R., & Finkel, M. E. Explanation of Miskin & Forgays' result as a directional reading conflict. Canadian Journal of Psychology, 1963, 17, 224-234.

Howes, D. H., & Solomon, R. L. Visual duration threshold as a function of word-probability. Journal of Experimental Psychology, 1951, 41, 401-410.

Hubel, D. H., & Wiesel, T. N. Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. Journal of Physiology, 1962, 160, 106-123.

Huey, E. B. The psychology and pedagogy of reading. Cambridge, Mass.: MIT Press, 1968 (originally published in 1908).

Hunt, E. B. Concept learning. New York: John Wiley & Sons, 1962.

Jakobson, R., Fant, G., & Halle, M. Preliminaries to speech analysis. Cambridge, Mass.: MIT Press, 1963.

Kolers, P. A. Reading as a perceptual skill. In H. Levin (Ed.), Planning for a reading research program. Cornell University, Ithaca, New York, 1965, pp. 19-21.

Kolers. P. A. The recognition of geometrically transformed text. Perception and Psychophysics, 1968, 3, 57-64.

Kolers, P. A. Some psychological aspects of pattern recognition. In P. A. Kolers and M. Eden (Eds.), Recognizing patterns, Cambridge, Mass.: MIT Press, 1968, p. 4-61.

Lawrence, D. H. The nature of stimulus: Some relationships between learning and perception. In S. Koch (Ed.), Psychology: A study of science. Vol. V. New York: McGraw Hill Book Co., 1963, p. 179-212.

Lott, Deborah, & Cronnell, B. The use of redundancy by beginning readers: Technical Report No. 13, 1969. Southwest Regional Laboratory, Inglewood, California.

Mandes, E., & Ghent, L. The effect of stimulus orientation on the recognition of geometric forms in adults. Paper presented at the American Psychological Association. Philadelphia, Penn., 1962.

Maturana, H. R., Lettvin, J. Y., McCulloch, W. S., & Pitts, W. H. Anatomy and physiology of vision in the frog. Journal of General Psychology, 1960, 43, 129-175.

Mayzner, M. S., & Tresselt, M. E. Individual differences in stored digram frequencies and the immediate serial recall of letter and number strings. Psychonomic Science, 1967, 7, 359-360.

Miller, G. A. Some psychological studies of grammar. American Psychologist, 1962, 17, 748-762.

Miller, G. A., Bruner, J. S., & Postman, L. Familarity of letter sequences and tachistoscopic identification. Journal of General Psychology, 1954, 5, 129-139.

Miller, G. A., & Selfridge, J. A. Verbal context and the recall of meaningful material. American Journal of Psychology, 1950, 63, 176-185.

Miskin, M., & Forgays, D. G. Word recognition as a function of retinal locus. Journal of Experimental Psychology, 1952, 43, 43-48.

Morrison, H. M., & Black, J. W. Prediction of missing words in sentences. Journal of Speech Disorders, 1957, 22, 236-240.

Morton, J. The effects of context upon speed of reading, eye-movements, and eye-voice span. Quarterly Journal of Experimental Psychology, 1964, 16, 340-354.

Mueller, C. G. Sensory Psychology. Englewood Cliffs, New Jersey: Prentice-Hall, 1965

Neisser, U. Visual search. Scientific American, 1964, 210, 94-102.

Neisser, U. Cognitive psychology. New York: Appleton-Century-Crofts, 1967.

Olson, D. R. On conceptual strategies. In J. Bruner, R. Olver, & P. Greenfield, (Eds.), Studies in cognitive growth. New York: Wiley & Sons, 1967. pp. 135-153.

Owsowitz, S. E. The effects of word familiarity and letter structure familarity on the perception of words. Rand Corporation Publications, 1963, No. P-2820.

Paterson, D. G., & Tinker, M. A. How to make type readable. New York: Harper & Brothers, 1940.

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor, R. L. Retention of visual and name codes of single letters. Journal of Experimental Psychology Monograph, 1969, 79, 1-16.

Posner, M. I., Goldsmith, R., & Welton, K. E. Perceived distance and the classification of distorted patterns. Journal of Experimental Psychology, 1967, 73, 28-38.

Posner, M. I., & Keele, S. W. Decay of visual information from a single letter. Science, 1967, 158, 137-139.

Posner, M. I., & Mitchell, R. F. Chronometric analysis of classification. Psychology Review, 1967, 74, 392-409.

Postman, L., & Rosenzweig, M. R. Practice and transfer in the visual and auditory recognition of verbal stimuli. American Journal of Psychology, 1956, 69, 209-226.

Rosenzweig, M. R., & Postman, L. Frequency of usage and the perception of words. Science, 1958, 127, 263-266.

Rudel, R. G., & Teuber, H. L. Discrimination of direction of line in children. Journal of Comparative and Physiological Psychology, 1963, 56, 892-898.

Sanford, E. C. The relative legibility of the small letters. American Journal of Psychology, 1887, 1, 403-435.

Schaller, M. J., & Harris, L. Children's same-different judgments of various transformations of letter-like forms using the method of paired comparisons: Effects of transformation and age of child. Paper presented at SRCD, Santa Monica, California, 1969.

Selfridge, O. G. Pattern recognition and learning. Methodos, 1956, 8, 163-176.

Selfridge, O. G. Pandemonium: A paradigm for learning. In L. Uhr (Ed.), Pattern recognition. New York: John Wiley & Sons, 1966, pp. 339-348.

Selfridge, O. G., & Neisser, U. Pattern recognition by machine. Scientific American, 1960, 203, 60-68.

Shannon, C. E. Prediction and entropy of printed English. Bell System Technical Journal, 1951, 30, 50-64.

Shepard, R. N. Production of constrained associates and the informational uncertainty of the constraint. American Journal of Psychology, 1963, 76, 218-228.

Smith, F. Effects of high and low digram frequency in the identification of words. Unpublished manuscript, Ontario Institute for Studies in Education, Toronto, undated.

Smith, F. The visual identification of words. Unpublished doctoral dissertation, Harvard University, 1967.

Smith, F. Familarity of configuration versus discriminability of features in the visual identification of words. Psychonomic Science, 1969, 14, 261-263.

Smith, F. The use of featural dependencies across letters in the visual identification of words. Journal of Verbal Learning & Verbal Behavior, 1969, in press.

Smith, F., Lott, Deborah, & Cronnell, B. The effect of type size and case alternation on word identification. American Journal of Psychology, 1969, in press.

Sperling, G. The information available during brief visual presentations. Psychology Monographs, 1960, 74, (498).

Sperling, G. A model for visual memory tasks. Human Factors, 1963, 5, 19-31.

Stevens, M. E. Automatic character recognition: A state-of-the-art report. National Bureau of Standards Technical Note, No. 112, May 1961.

Sutherland, N. S. Visual discrimination of orientation by octopus. British Journal of Psychology, 1957, 48, 55-71.

Sutherland, N. S. Cat's ability to discriminate oblique rectangles. Science, 1963, 139, 209-210.

Thomas, H. Children's tachistoscopic recognition of words and pseudowords varying in pronounceability and consonant-vowel sequence. Journal of Experimental Psychology, 1968, 77, 511-513

Thorndike, E. L., & Lorge, I. The teacher's word book of 30,000 words. New York, Bureau of Publications, Teacher's College, Columbia University, 1944.

Tinker, M. A. Bases for effective reading. Minneapolis: University of Minnesota Press, 1966.

Trabasso, T., & Bower, G. H. Attention in learning: Theory and research. New York: John Wiley & Sons, 1968.

Uhr, L. Pattern recognition computers as models for form perception. Psychology Bulletin, 1963, 60, 40-73.

Uhr, L. (Ed.) Pattern recognition. New York: John Wiley & Sons, 1966

Underwood, B. J., & Schulz, R. W. Meaningfulness in verbal learning. New York: J. B. Lippincott Co., 1960.

Vanderplas, J. M. Associative processes and task relations in perceptual learning. Perceptual and Motor Skills, 1963, 16, 501-509.

Venezky, R. L., & Weir, R. H. A study of selected spelling-to-sound correspondence patterns. Final Report, Cooperative Research Project No.3090, Stanford University, 1966.

Wallach, H., & Austin, P. Recognition and the localization of visual tracers. American Journal of Psychology, 1954, 67, 338-340.

Wheelock, W. H., & Silvaroli, N. J. An investigation of visual discrimination training for beginning readers. Journal of Typographic Research, 1967, 1, 147-156.

White, B. W. Recognition of familiar characters under an unfamiliar transformation. Perceptual and Motor Skills, 1962, 15, 107-116.

Wiener, M., & Cromer, W. Reading and reading difficulty: A conceptual analysis. Harvard Educational Review, 1967, 37, 620-643.

Williams, J. Effects of discrimination and reproduction training on ability to discriminate letter-like forms. Paper presented at AERA, February, 1968. Eric No. ED 017418.

Wright, J. C. Acquired relevance of cues and ritualistic attention to irrelevant cues in children's learning. Paper presented at American Psychological Association Conference, Los Angeles, 1964.