

The role of letter recognition in word recognition

MICHAEL J. COSKY

The University of Texas at Austin, Austin, Texas 78712

In a paradigm that avoids methodological problems of earlier studies, evidence was gathered addressed to the question of whether we read letter by letter. If word recognition involves letter recognition, then the difficulty of recognizing a word should vary with the difficulty of recognizing its letters. This was tested by assessing letter difficulty in two letter discrimination tasks and in a letter naming task, and then comparing 15 adult subjects' visual recognition latency to 72 easy-letter words and to 72 difficult-letter words. Word frequency and word length were also manipulated. Results indicated no effect for letter difficulty, although recognition latency reliably decreased with word frequency and monotonically increased with word length (21 msec/letter), suggesting that we do not read letter by letter, but that whatever plays a role in word recognition is smaller than the word and correlated with word length in letters.

Since Cattell (1885), numerous experiments have attempted to settle the question of whether we read letter by letter. The overwhelming majority of these studies have employed one of two basic paradigms: a comparison of word and letter recognition (e.g., Cattell, 1885; Erdmann & Dodge, 1898; Johnston & McClelland, 1973; Kolers & Katzman, 1966; Korte, 1923; Massaro, 1973; Mezrich, 1973; Reicher, 1969; Smith, 1969; Thompson & Massaro, 1973; Wheeler, 1970) or a comparison of word and nonsense recognition (e.g., Goldscheider & Muller, 1893; Massaro, 1973; Pillsbury, 1897; Reicher, 1969; E. E. Smith & Haviland, 1972; F. Smith, 1969). The rationale for these paradigms is usually straightforward. For example, in the word vs. letter paradigm, suppose that the time required to identify a single letter is measured to be $C + x$ msec; then, if we read letter by letter, a five-letter word should take $C + 5x$ msec to be read. Similarly, in the word vs. nonsense comparison, recognizing letters in nonsense should be no more difficult than recognizing letters in words if we do indeed read letter by letter. But, since letters and nonsense differ from words along many uncontrollable dimensions, both of these paradigms have serious methodological shortcomings, and in general the question of whether we read letter by letter has remained moot. Indeed, letter-by-letter models of reading have recently been proposed (e.g., Gough, 1972) and discussed (e.g., Brewer, 1972; F. Smith, 1971). The present study provides some

This study is partly based on results reported in a master's thesis completed at the University of Texas at Austin; the author wishes to acknowledge Philip B. Gough and David T. Hakes, whose support made this work possible. A report of part of this research was also presented at the meeting of the Midwestern Psychological Association, Chicago, May 1974. The author also wishes to thank Philip B. Gough for many helpful comments on earlier versions of this paper. The author is now at St. Olaf College, Psychology Department, Northfield, Minnesota 55057. Reprint requests should be sent to him at that address.

evidence relevant to the letter-by-letter question, evidence obtained in a novel paradigm—a comparison of two *kinds* of words.

If word recognition involves letter recognition, then the difficulty of recognizing a word should vary with the difficulty of recognizing its letters. This prediction can be tested by comparing the recognition latency of words composed of difficult letters with that of words composed of easy letters. Note that in this task recognition performance on a word is being compared with recognition performance on another word; the only component of the stimulus that is being manipulated is that aspect which is necessary to both classes of stimuli *as words*, i.e., the letters that make up the words. Thus, the recognition difficulty of the letters in a word should determine to some extent how difficult that word is to recognize.

METHOD

Assessments of Letter Recognition Difficulty

Letter recognition difficulty was first assessed in two letter discrimination tasks. In the first, 20 subjects (introductory psychology students at the University of Texas at Austin), tested individually, were asked to make a letter/nonletter judgment on each of the 26 uppercase letters and 13 distractors in each of eight presentations. The letters and distractors were presented in a different random order to each subject who was required to designate his letter/nonletter decision by pressing the appropriate button (left vs. right, counterbalanced across subjects) on the panel before him. The distractors were the following nonalphabetic characters: # ; &) * - = + / ?] < ←. The stimuli were presented singly on the screen of a Tektronix storage display unit: the letters and distractors were displayed by having the character generator program illuminate the appropriate pattern of points in a matrix 13 points high and 9 points wide; a single character subtended a visual angle of 20' x 16.7' at 24 in., the viewing distance that was employed here. See Figure 1 for the character font of the letters. A DEC PDP-8/1 computer recorded the latency of each discriminative response in milliseconds. The letters were then ranked according to mean discrimination latency (see Table 1, Rank Order 1).

A 26 by 8 within-subject analysis of variance was computed on the latency data. The main effects for letters and for trials were significant ($p < .001$): for letters, $F(25,475) = 7.21$ (error term



Figure 1. Enlarged version of the character font used in the word and letter experiments.

mean square = 17.633), and for trials, $F(7,133) = 8.63$ (error term mean square = 20.782). The Letters by Trials interaction was not significant, $F(175, 3325) = 1.07$, $p < .10$. The significant trials effect indicates a practice effect, since mean discrimination latency decreased monotonically from 523 msec on Trial 1 to 476 msec on Trial 8; the significant F for letters suggests reliable differences among the letters. Further, by ranking the letters for each subject, Kendall's coefficient of concordance (1948) could also be computed, $W = .32$, and its associated chi square was significant, $\chi^2(25) = 160.68$, $p < .001$. Although not large, this W does exceed others computed on similar letter recognition data, e.g., Derks (Note 1; Experiment 8). So reliable differences in discrimination time seem to exist for the 26 letters.

But, conscious of the role that the nature of the distractor can play in a discrimination task, it was feared that the rank order obtained in this first assessment might not be a reliable indicator of letter difficulty. Rather, this ordering might only reflect the letter's visual similarity or dissimilarity to the particular 13 distractors that were used. That is, it could be that the easy letters were merely less similar to the distractors than were the difficult letters; the rank order of letter difficulty would then be an artifact of the distractors used. So another assessment was made, this time having 19 of the original 20 subjects classify the 26 letters and 13 different distractors, again in each of eight random presentations; all other aspects were identical to the first assessment. The distractors were alphabetic inversions and mirror images: the seven inverted letters were A, J, L, R, U, V, and Y; the six mirror-image letters were B, E, F, G, K, and P. Again, the letters were ranked according to mean discrimination latency (see Table 1, Rank Order 2).

A 26 by 8 within-subject analysis of variance was computed on the latency data. The main effects for letters and trials were significant ($p < .001$): for letters, $F(25,450) = 5.72$ (error term mean square = 28.472), and for trials, $F(7,126) = 6.28$ (error term mean square = 22.163); and so was the Letters by Trials interaction, $F(175,3150) = 1.39$, $p < .01$. A practice effect was again evidenced, since mean discrimination latency decreased from 546 msec on Trial 1 to 499 msec on Trial 8. But this practice effect was not uniform across letters, as the significant interaction suggests. Using each subject's rank order of letter difficulty, Kendall's $W = .27$, and its associated chi square was significant, $\chi^2(25) = 128.4$, $p < .001$. Thus, for this second rank order of letter difficulty as well, reliable letter differences appear to exist.

In general, there was fair agreement between the two rankings ($\rho = +.67$, $p < .0004$, and $N = 26$), suggesting that the rank orders were probably not artifacts of the particular distractors used. The letters were then ranked according to mean discrimination

latency across the two assessments (see Table 1, Rank Order 3): "easy letters," with ranks 1-13 (D R B P E T K F H A M S L), and "difficult letters," with ranks 14-26 (N U O Q G I Y Z V X J C W).

This composite rank order of letter difficulty should be viewed with some caution. It necessarily applies only for the particular font of characters used here, so the generality of this assessment of letter difficulty is certainly limited. Yet, that lack of generalization poses no real problem for the present study, since the words to be presented for recognition were also displayed in this same character font. Due to the line of reasoning adopted in the hypothesis, the use of identical fonts in both the letter difficulty assessments and in the subsequent word recognition measure was crucial. So, at least within the framework of this study, this composite rank order of letter difficulty seemed an appropriate assessment of letter difficulty with which to test the question of whether we read letter by letter.

But objections can be raised as to the validity of this letter difficulty assessment—to be specific, (a) letter difficulty was measured on single characters not in the context of other letters (as they usually appear in words), and (b) the letter/nonletter discrimination task may require a subset of processes that are different from those that are involved in letter identification.

The first point merely questions whether an assessment of letter difficulty based on single, free-standing letters can be considered appropriate when letters in words are most frequently surrounded by other letters. For example, what are the possible lateral inhibition effects that might be operating on a letter's relative legibility when that letter is in close spatial proximity to the adjacent contours (the angles, bars, and curves) of its neighboring letters? It could be that the letter difficulty ranking given in Table 1 holds only for isolated letters, and that in alphabetic context this ranking simply would not obtain.

Table 1
Rank Orders of Letter Difficulty Based on Mean
Discrimination Latency (Milliseconds)

Letter	Rank Order 1*	Rank Order 2**	Rank Order 3†
A	11 (477)	10 (491)	10 (484)
B	8 (471)	1 (470)	3 (470)
C	26 (556)	22 (540)	25 (548)
D	1 (438)	4 (481)	1 (459)
E	4 (467)	2 (479)	5 (473)
F	6 (469)	9 (489)	8 (479)
G	16 (492)	20 (522)	18 (507)
H	15 (484)	6 (483)	9 (483)
I	24 (520)	13 (495)	19 (508)
J	23 (513)	25 (567)	24 (539)
K	7 (470)	8 (487)	7 (478)
L	9 (472)	17 (513)	13 (491)
M	12 (478)	11 (493)	11 (485)
N	13 (481)	15 (504)	14 (492)
O	22 (512)	7 (486)	16 (499)
P	2 (449)	12 (494)	4 (471)
Q	21 (506)	14 (501)	17 (503)
R	3 (454)	5 (482)	2 (468)
S	10 (474)	16 (505)	12 (489)
T	5 (468)	3 (480)	6 (474)
U	14 (482)	18 (516)	15 (498)
V	19 (503)	24 (546)	22 (524)
W	20 (505)	26 (617)	26 (560)
X	25 (552)	19 (517)	23 (535)
Y	17 (495)	21 (524)	20 (509)
Z	18 (496)	23 (541)	21 (518)

*Using nonalphabetic distractors and based on 160 observations per letter.

**Using alphabetic inversions and mirror images as distractors and based on 152 observations per letter.

†Across both assessments; based on 312 observations per letter.

Table 2
Rank Orders of Letter Difficulty Based on Mean
Recognition Latency (Milliseconds)

Letter	Rank Order 1*	Rank Order 2**	Rank Order 3†
A	1 (433)	1 (416)	1 (425)
B	4.5 (439)	11 (429)	6.5 (434)
C	23 (483)	24 (473)	23 (478)
D	14 (454)	16 (442)	16 (448)
E	3 (438)	5.5 (423)	4 (431)
F	2 (434)	2 (421)	2 (427)
G	21 (480)	21 (470)	21.5 (475)
H	13 (449)	13 (430)	12 (439)
I	11 (448)	5.5 (423)	9 (435)
J	20 (479)	23 (472)	21.5 (475)
K	16 (456)	15 (438)	15 (447)
L	4.5 (439)	3.5 (422)	4 (431)
M	6.5 (440)	3.5 (422)	4 (431)
N	6.5 (440)	11 (429)	9 (435)
O	11 (448)	14 (432)	13 (440)
P	11 (448)	8.5 (428)	11 (438)
Q	25.5 (524)	26 (494)	26 (509)
R	9 (442)	11 (429)	9 (435)
S	8 (441)	7 (427)	6.5 (434)
T	19 (465)	19 (451)	19 (458)
U	18 (463)	8.5 (428)	14 (445)
V	22 (482)	20 (458)	20 (470)
W	24 (487)	25 (478)	25 (483)
X	17 (458)	17 (444)	17 (451)
Y	15 (455)	18 (450)	18 (453)
Z	25.5 (487)	22 (471)	24 (479)

*Letters presented in pseudo-letter context; based on 200 observations per letter.

**Letters presented without context; based on 200 observations per letter.

†Across both context conditions; based on 400 observations per letter.

The second possible objection speaks to a large body of evidence (e.g., Brand, 1971; Posner, 1970) which seems to indicate that classification or categorization is not influenced by experimental manipulations that affect identification, suggesting that classification may occur without, or prior to, identification. The obvious implication here, of course, is that the subject in the letter discrimination task did not have to identify the stimulus letter in making his letter/nonletter decision, that he could have been using a different set of operations relative to what might be required in letter identification—in the kind of letter identification that a letter-by-letter model of reading would supposedly entail. So both of these arguments can make it difficult to accept the letter difficulty assessment—and hence, the proposed test of the letter-by-letter hypothesis.

In order to answer these kinds of arguments, a further assessment of letter difficulty was undertaken. Subjects were required to name the letters aloud, and the latency of their response was measured; the use of this *naming* task hopefully circumvented the second objection. And, by presenting the letters both singly and in a pseudoletter context, it was hoped that the two resulting letter difficulty rankings would reveal whether any systematic changes in letter difficulty occurred as a function of the letter-like context, thereby attempting to address the first argument.

So each of the 20 subjects (introductory psychology students at the University of Texas at Austin who had not participated in any of the earlier experiments) was presented a different random order of 10 trials of each of the 26 uppercase letters displayed in the same character font that was used before. On five of the letter's trials, it appeared singly, as in the earlier letter difficulty assessments; on the other five trials, the letter was flanked by two pseudoletter figures which were spaced from the letter at the same distance that

they would be if the stimulus string were a three-letter word. The pseudoletter character subtended the same maximum visual angle as a single letter and was displayed on the screen by having the character generator program simultaneously present, at one character space, all the points that are used for all the letters of the alphabet. Each subject was required to name each letter aloud as quickly as possible, and the latency of response was measured in milliseconds using the same apparatus as before.

Thus, with the factorial combination of 2 levels of letter context, 26 letters, and 5 trials, each subject received 260 randomized trials, effecting a completely within-subject design.

The letters were ranked according to mean recognition latency as a function of pseudoletter context (see Table 2, Rank Orders 1 and 2); a composite rank order, computed across both levels of letter context, was also compiled (Table 2, Rank Order 3). A 2 by 26 by 5 within-subject analysis of variance was computed on the latency data. The main effects for letters and contexts were significant ($p < .001$): for letters, $F(25,475) = 15.95$ (error term mean square = 5.911); and for context, $F(1,19) = 92.67$ (error term mean square = 3.532). No other main effects or interactions were significant ($p > .05$).

The main effect for letters indicates that there are reliable differences among the letters in the letter naming task; and the effect for context indicates that its presence reliably increased letter recognition time: mean response time to letters in context is 460 msec compared with 442 msec for free-standing letters. But, of particular importance here is that context and letters did not interact, indicating that the *relative* difficulty of the letters did not vary as a function of whether other letter-like contours were flanking the letter. A correlational measure also demonstrates the apparent inability of the pseudoletter context to affect relative letter difficulty: a rank order correlation coefficient computed between Rank Orders 1 and 2 in Table 3 was significant ($p < .0001$) and accounted for a substantial portion of the variance; $\rho = +.91$, $N = 26$. So it seems as though adjacent contours do not affect relative letter difficulty, implying that the first objection to the earlier letter difficulty assessment may not be damaging.¹

Due to the high degree of correspondence between letter naming Rank Orders 1 and 2, it seemed reasonable to collapse across them and use that composite rank order of letter difficulty (based on letter naming times) to compare with the earlier assessment of letter difficulty (based on letter discrimination times). This comparison would address the second objection raised earlier, regarding the validity of the letter discrimination task.

A rank order correlation coefficient was computed between the composite rank orders of letter difficulty based on letter discrimination (Rank Order 3, Table 1) and on letter naming (Rank Order 3, Table 2): $\rho = +.61$, $N = 26$, $p < .005$. There seems to be a marked tendency for letters to be similarly ranked in both letter difficulty assessments, although the magnitude of the ρ value leaves a good portion of the variance unaccounted for. But, on closer inspection of these two rank orders, it can be seen that of the 13 letters that were classified as "easy letters" as a result of the letter discrimination assessment, 10 (A B E F H L M P R S) are also classified as easy letters via the letter naming measure. So, even though a number of letters may change their ranks somewhat between the two rank orders—and consequently reduce the computed ρ —it seems that the two letter difficulty assessments produced generally similar letter difficulty results. The import of these similar results is that perhaps the subjects in the letter naming and letter discrimination tasks were using the same sorts of operations and that the argument against the originally employed letter difficulty measure may not be tenable for these kinds of stimuli.

But, before we can further consider these letter naming data, note that they may be confounded with what can be called the production latency of the letters. Production latency is the amount of time it takes to begin to say a letter's name after the letter has been recognized. If any of the letters systematically differ along this variable, then any obtained difference in naming latency would clearly be confounded (see Stewart, James, & Gough, Note 2). This

possible source of artifact was tested for by having 10 subjects (none of whom participated earlier) pronounce each letter aloud as soon as possible in response to an independent light signal, and the latency of each response was measured in milliseconds, using the same apparatus as before. The interval between the presentation of the letter and the onset of the light signal varied randomly between 1.0 and 2.5 sec in order to allay anticipatory effects and to assure that the letter had been recognized before it had to be vocalized. Each subject received 130 completely randomized trials consisting of five trials of each letter. Grand mean production latency² for the 26 letters was 430 msec, and a 26 by 5 analysis of variance (within-subject) revealed a significant effect for only the Letters by Trials interaction, $F(100,800) = 1.41$, $p < .01$. A thorough description of the nature of this interaction would be quite complex (and probably uninteresting), but it generally indicates a slight practice effect for some of the letters. There were no other significant interactions or main effects, indicating that the letter naming results were not confounded with the letters' production latencies.

Thus, as assessed in both the letter discrimination and letter naming tasks, it seems that reliable and similar differences in recognition difficulty exist among the letters.

Stimulus Materials and Design

In view of the agreement between the letter naming and letter discrimination assessments of letter difficulty, it was decided to use the rank order of letter difficulty based on the letter discrimination times (Table 1, Rank Order 3) to construct the easy- and difficult-letter words which would then be used to test the letter-by-letter hypothesis. From this rank order of letter difficulty, letters with ranks 1-13 (D R B P E T K F H A M S L) were used to make up the easy-letter words; letters with ranks 14-26 (N U O Q G I Y Z V X J C W) were used for difficult-letter words. This dichotomy could not be strictly maintained; some of the easy letters had to be used for the difficult-letter words, and vice versa. But the proportion of such instances was small; of the total number of letters used in making up both kinds of words, only 11% came from the "inappropriate" half of the letter difficulty ranking. Taking the median letter of each easy and difficult word, no overlap was found between the two resulting distributions (the median letter distribution for easy words ranged from ranks 1-10, and the range of the median letter distribution for difficult words included ranks 13-25); also, the medians of these two distributions were 5 and 16, respectively. Further, the grand mean discrimination latency for the 396 letters used in the easy-letter words was 475 msec; for the difficult-letter words, this grand mean was 506 msec. All of this of course suggests that the easy- and difficult-letter words were constructed from considerably different sections along the letter difficulty continuum.

In constructing a stimulus word, two other variables were manipulated: word length and word frequency. Word length varied from 3 to 8 letters. A word's frequency was determined by the number of genres in which that word occurred in the Kučera and Francis (1967) word count: high-frequency words, 11-15 genres; medium-frequency words, 6-10 genres; and low-frequency words, 1-5 genres. As an independent check of this method of classification, a word's frequency—high, medium, or low—via this method was compared to that word's frequency level as determined by a procedure adopted by Rubenstein, Lewis, and Rubenstein (1971). Using the Lorge Magazine Count (Thorndike & Lorge, 1944) in their procedure, words were classified as high-frequency if occurring 184-552 times, as medium-frequency if occurring 61-183 times, and as low-frequency if occurring 20-60 times. Of the items that appeared in both counts, over 75% were classified in the same frequency level by the two procedures, suggesting that the method used here is reliable.

Thus, with 2 levels of letter difficulty, 6 word lengths, and 3 levels of word frequency, there were 36 cells in the design; four exemplars were constructed for each cell. The resulting 144 words (see Appendix for listing) were presented in a different random order to each of 15 subjects who were required to read each word aloud, and the recognition latency to each word was recorded in milliseconds.

Subjects

Fifteen subjects, none of whom took part in the letter difficulty assessments, participated; all were native speakers of English, all had normal vision (or corrected to normal), and all were participating as part of an undergraduate psychology course requirement at the University of Texas at Austin. All were naive with respect to the purpose of the experiment.

Apparatus

A DEC PDP-8/I computer was programmed to present the words on the screen of a Tektronix storage display unit (Type 601). A Grason-Stadler throat microphone (E 7300 M) and voice-operated relay (E 7300A-1) were used to detect the onset of the subject's voice as he read the word. The latency of each identification response was recorded by the computer, and a printout of the stimulus words and their corresponding latencies (in milliseconds) was obtained for each subject.

A word subtended from 1°2' of horizontal visual angle for three-letter to 2°45' for eight-letter words. The words were presented in the identical font that was used in the letter difficulty measures.

Procedure

Each subject was seated in a booth (viewing distance 24 in.) and was tested individually. He was instructed that the stimuli to be presented were common English words and that his task was to read each word aloud as soon as possible after it was presented. The subject was told the function of the throat microphone and shown how to wear it. A 7½-W night-light provided dim overall illumination for the experimental setting. Each subject received the same random order of 18 practice items, consisting of 3 words at each of the 6 word lengths; none of the practice words were used in the experiment proper. The experimenter remained in an adjacent booth to record any errors, i.e., when the throat microphone was not triggered (e.g., the subject spoke too softly) or when it was triggered accidentally (e.g., false starts, coughs, etc.). After instructing the subject, the experimenter initiated the program: a fixation point was provided on the screen and, after a random delay of 1.0-2.5 sec, a word was presented and the subject read it aloud. His response erased the screen and it remained blank as the Teletype (in an adjacent room) printed out the preceding word and its latency. During this printout interval (1-2 sec), the subject was allowed to ask the experimenter to stop the program if necessary (e.g., for the subject to clear his throat), since the voice-operated relay could not be activated during the printout. At the end of this interval, the fixation point reappeared, signaling the subject to be ready, and after a random delay of 1.0-2.5 sec, another word was presented. An entire session lasted about 20 min.

RESULTS

The error rate was 3.05% per subject; the mean recognition latency for trials on which errors occurred was 748 msec. These error latencies were not included in the statistical analyses. Rather, each subject's error latency was replaced by an estimate which was the mean response latency of the other three words in the same cell in which the error occurred. The occurrence of errors was randomly distributed across the 15 subjects and throughout the 144 words.

The latency data are presented in Table 3; mean visual word recognition latency is first shown as a function of word length and word frequency, and then as a function of word length and letter difficulty. Recognition latency increased monotonically with word length (21 msec/letter) and decreased consistently with word frequency. After Clark (1973), an

Table 3
Mean Visual Word Recognition Latency (Milliseconds)
as a Function of Word Length and Word Frequency* and
as a Function of Word Length and Letter Difficulty**

	Word Length (Letters)						Mean
	3	4	5	6	7	8	
Word Frequency							
Low	542	569	599	614	642	730	616
Medium	539	536	541	583	619	600	570
High	508	528	543	545	547	582	542
Letter Difficulty							
Low	514	543	566	588	601	647	577
High	545	545	555	575	604	628	575
Mean	530	544	561	581	602	637	

*Each cell mean is based on 120 observations.

**Each cell mean is based on 180 observations.

analysis of variance was computed on the latency data, treating subjects and words as random effects, and letter difficulty, word length, and word frequency as fixed effects; the resulting completely within-subject design had four words nested in each of its 36 cells. Each subject's recognition latency to each of the 144 words was used in the analysis, except for errors. (Mean visual recognition latency of each word is also given in the Appendix.)

The main effects for subjects, word length, and word frequency were significant ($p < .001$): for subjects, $F(14,1512) = 220.0$ (error term mean square = 9.522); for word length, $F'(5,178) = 8.96$ (error term mean squares = 64.517); and for word frequency, $F'(2,105) = 15.13$ (error term mean squares = 66.680). But there was no significant difference between the means of the easy and difficult-letter words ($F' < 1$, error term mean squares = 43.672); in fact, the small observed difference was not even in the predicted direction. Mean recognition latency for easy-letter words was 577 msec, and for difficult-letter words, 575 msec. Further, no interaction involving letter difficulty was reliable. Several two-way interactions were significant: subjects interacted with word length and with word frequency, as did word length with word frequency: for Subjects by Word Length, $F(70,1512) = 2.76$, $p < .01$, and for Subjects by Word Frequency, $F(28,1512) = 2.99$, $p < .001$ (for both, error term mean square = 9522); and for Word Length by Word Frequency, $F'(12,175) = 1.96$, $p < .05$ (error term mean squares = 50,581).

Examination of the Subjects by Word Length interaction revealed that the slope of the word length function ranged from 4 msec/letter to 46 msec/letter among the 15 subjects. Similarly, for the Subjects by Word Frequency interaction, the negative slope of the word frequency function was not constant across subjects, ranging from a -3- to a -66-msec/frequency level.

The significant Word Length by Word Frequency

interaction was probably caused by some peculiarity of the eight-letter, low-frequency words. From Table 3, note that the recognition latency for this cell is nearly 90 msec greater than any other cell mean, and this probably accounts for a substantial portion of the variance contributing to the marginally significant interaction.

The only other significant effect was for words nested within the Letter Difficulty by Word Length by Word Frequency interaction, $F(108,1512) = 4.01$, $p < .001$; but, since words were only nested alphabetically within each cell, no very meaningful interpretation of this effect can be made, i.e., other than to say that the four words within a cell differed among themselves.

Although the analysis of variance indicated no effect of letter difficulty, such an effect could be hidden by the dichotomization and subsequent lumping of the data into easy- and difficult-letter word groups. That is, the analysis of variance would not be able to tease out any subtle effect of letter difficulty that might be present; but a regression analysis would surely allow a more sensitive test of the letter-difficulty effect if one indeed exists in the word recognition data.

So mean word recognition latency for each of the 144 words was first regressed on the three letter-difficulty rankings based on the letter discrimination tasks (i.e., for each ranking in Table 1, mean discrimination latency of the letters that made up the word was used as the predictor); word length and word frequency were also entered into the regression analysis. Sequential F tests (Draper & Smith, 1966) of the resulting regression equations indicated no reliable contribution of letter difficulty to word recognition. All Fs evaluating the contribution of any letter difficulty measure (or any combination of them) were not significant ($p > .05$). In fact, of all possible combinations of letter difficulty predictors, the contribution of letter difficulty as assessed in Rank Orders 1 and 2 combined effected the largest R^2 , and it was only .0008! Clearly, letter difficulty had no influence on word recognition latency.

Of course, both word length and word frequency did have reliable effects: word length added 22 msec per letter to word recognition time and word frequency 38 msec per decreasing frequency level. Both sequential F tests regarding word length and word frequency were significant ($p < .001$): for word length, $F(1,142) = 47.92$; and for word frequency, $F(1,142) = 30.83$. And, when both word length and word frequency were entered together in the regression equation, their contribution was again reliable: $F(2,141) = 53.34$, $p < .001$ (and $R^2 = .431$). Thus, both statistical analyses indicate no effect of letter difficulty, but robust effects of word length and word frequency.

In spite of the fact that the letter difficulty assessments based on letter discrimination and letter naming seemed to produce such similar letter difficulty results, it was decided to test again, at least statistically, the role of letter recognition in word recognition by regressing mean word recognition latency for each of the 144 words on the three letter-difficulty predictors based on the letter naming times (i.e., for each rank order in Table 2, mean naming latency of the letters that made up the words was used as the predictor); again, word length and word frequency were also entered as predictors. Sequential F tests of the resulting regression equations indicated no reliable ($p > .05$) contribution of any letter difficulty measure (or any combination of them); and the R^2 values for the letter difficulty predictors were all less than .017. So, again, letter difficulty, as measured by the letter naming task, does not predict word difficulty.

Before moving to a discussion of these findings, again the question arises whether the word naming latency results are confounded with the production latency of the stimulus words. If any of the words in a given condition systematically trigger or activate the timing apparatus sooner or later than those in another condition, then the word naming results would merely be an artifact of the acoustic or articulatory properties of the initial segments of the words. To test this, each of 10 subjects (none of whom participated earlier) were presented the 144 words in a different random order and asked to pronounce each word aloud in response to an independent light signal—an asterisk displayed on the screen; using the same apparatus as before, the latency of response was measured in milliseconds. All other aspects of this experiment were identical to those of the production latency test of the letters (described above). Grand mean production latency for the 144 words was 379 msec, and a 2 by 3 by 6 within-subject analysis of variance of the data revealed no significant main effects or interactions, suggesting that the word naming latency results were not an artifact of the production latency of the words.

GENERAL DISCUSSION

The results taken in total indicate that letter difficulty, as measured here, does not influence word recognition, strongly suggesting that letter recognition does not play a role in word recognition processes. And this, of course, implies that reading may not take place letter by letter. Although these conclusions are based on an essentially null result regarding the influence of letter difficulty on word difficulty, and even though arguing from such data may seem unconvincing, the facts that similar and reliable letter difficulty rankings were obtained across subjects, tasks, and (in all but one case) trials, and that these same differences repeatedly failed to surface in the

word recognition data (regardless of statistical test), adequately support the conclusion that letter recognition does not play a role in word recognition.

It is possible, however, to conceive of a letter-by-letter model of reading and word recognition that would not necessitate letter recognition difficulty to predict word recognition difficulty; and consequently, these data would not rule out such a letter-by-letter interpretation. This would be especially true for a model that incorporates a letter-by-letter component that is not sensitive to differences among letters in recognition difficulty, and more specifically for a letter-by-letter component that is insensitive to the relatively small difference—30 msec—that existed between the mean letter recognition latencies of the letters used to derive the easy- and difficult-letter words. But, at the very least, these findings need to be accommodated by letter-by-letter models if they are to be retained as plausible characterizations of reading and word recognition processes.

Although an implication of these data is that reading is not accomplished letter by letter, the finding that word recognition latency reliably increased with word length strongly argues that we don't read word by word either. That is, the evidence suggests that whatever plays a role in word recognition is (a) necessarily smaller than the word, and (b) correlated with word length in letters. Models of word recognition that consider featural information (Smith, 1971) or "spelling patterns" (Gibson, 1970) to be the building blocks of the reading process obviously satisfy both of these constraints. Similarly, those models that employ a phonological component at higher levels of the process are also tenable, since both phonemes and, to a lesser extent, syllables meet these requirements. Clearly, such models deserve and are receiving more attention. A number of studies, employing various tasks and paradigms, have reported findings relevant to the role of the syllable in word recognition (Eriksen, Pollack, & Montague, 1970; Forster & Chambers, 1973; Henderson, Coltheart, & Woodhouse, 1973; Klapp, 1971, 1974; Klapp, Anderson, & Berrian, 1973; Spoehr & Smith, 1973) but have not produced data that permit a generally acceptable statement about the effect of syllables; and there is some evidence (Cosky & Swinney, Note 3) regarding the role of phonemes that suggests that number of surface phonemes predicts word naming latency better than number of letters. But, at present, further empirical tests and/or theoretical reformulations are needed before any clear picture emerges.

Presumably it is clear that the underlying word vs. word comparison that was used here is well suited to test predictions of word recognition models, since the major pitfall of the letter vs. word and nonsense vs. word comparisons appears to be avoided. Note that,

in these latter paradigms, one is forced to compare entirely different entities and in doing so to make the implicit assumption that similar perceptual strategies and response outputs operate for the three kinds of verbal items. That assumption seems tenuous (e.g., Mezrich, 1973). Thus, the data resulting from studies employing the letter vs. word and the nonsense vs. word paradigms cannot be unequivocally interpreted, and testing a model using these paradigms in fact does little to demonstrate the model's relevance to the processes underlying reading and word recognition.

So, although two earlier findings were replicated, i.e., the increase in word recognition latency as a function of word length in letters (Stewart, James, & Gough, Note 2) and the decrease in word recognition difficulty with the increasing word frequency—the “word frequency effect” (e.g., Howes & Solomon, 1951)—the major result of this study showed no effect of letter recognition difficulty on word recognition. To the extent that the present measures of letter recognition difficulty are valid, this result argues that the role of letter recognition in word recognition is probably far less than letter-by-letter models of reading would have us believe.

APPENDIX

Mean Visual Recognition Latency (milliseconds) of Stimulus Words

Easy-Letter Words

Low frequency: dam 542, ebb 552, pad 524, par 544; drab 535, pert 550, reap 563, trek 591; baker 555, brake 556, debut 675, tread 634; dapper 697, dealer 569, dreary 626, pebble 613; branded 606, dreamer 567, feather 644, rafters 633; deadened 946, radiated 686, reversal 630, teardrop 713.

Medium frequency: bet 520, mad 466, pat 485, tap 548; bare 549, deaf 545, fate 554, trap 545; beard 567, draft 542, error 521, sheep 587; draped 725, farmer 543, melted 514, stream 644; broader 719, dressed 599, drifted 553, pretend 672; basement 567, decrease 621, dreadful 655, probable 651.

High frequency: add 498, bed 505, eat 495, met 492; left 496, part 524, rate 512, team 553; bread 531, dream 531, there 557, trade 540; attend 554, better 536, father 532, market 502; dropped 523, earlier 525, federal 588, present 585; interest 559, pressure 610, problems 561, research 563.

Difficult-Letter Words

Low frequency: coy 565, gin 574, wig 499, zoo 534; claw 572, jinx 636, oily 576, quiz 526; gowns 682, lousy 533, quill 559, slugs 598; jovial 668, joyous 587, waxing 566, zoning 582; coiling 704, jocular 754, noxious 606, nozzles 618, conjugal 813, cowering 747, virtuous 726, vocation 580.

Medium frequency: cow 608, gun 553, Jim 545, sin 591; coin 503, lazy 482, owns 602, wing 510; cling 542, joint 543, laugh 516, vocal 508; convey 624, gazing 584, lively 516, voices 511; columns 595, consist 621, vicious 630, victims 564; catholic 584, critical 589, junction 601, location 534.

High frequency: ill 533, joy 508, now 522, win 507; join 555, long 507, slow 559, will 514, shown 597, voice 515, which 559, young 513; giving 555, motion 506, window 588, within 585; contact 588, knowing 545, nothing 507, victory 513; economic 685, occasion 548, politics 564, position 567.

REFERENCE NOTES

1. Derks, P. L. *Complexity, similarity, and letter recognition.*

Paper presented at the meeting of the Psychonomic Society, St. Louis, 1971.

2. Stewart, M. L., James, C. T., & Gough, P. B. *Word recognition latency as a function of word length.* Paper presented at the Midwestern Psychological Association Convention, Chicago, 1969.

3. Cosky, M. J., & Swinney, D. A. *Is reading accomplished phoneme-by-phoneme?* Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1975.

REFERENCES

- BRAND, J. Classification without identification in visual search. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 171-186.
- BREWER, W. F. Is reading a letter-by-letter process? In J. F. Kavanaugh & I. G. Mattingly (Eds.), *Language by ear and by eye*. Cambridge, Mass: M.I.T. Press, 1972. Pp. 359-365.
- CATTELL, J. McK. Ueber die Zeit der Erkennung und Benennung von Schriftzeichen, Bildern und Farben. *Philosophische Studien*, 1885, **2**, 635-650. (Translated as: On the time required for recognizing and naming letters and words, pictures and colors.) In A. T. Poffenberger (Ed.), *James McKeen Cattell: Man of science* (Vol. 1) *Psychological research*. Lancaster, Pa: Science Press, 1947. Pp. 13-23.
- CLARK, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 335-359.
- DRAPER, N. R., & SMITH, H. *Applied regression analysis*. New York: Wiley, 1966.
- ERDMANN, B., & DODGE, R. *Psychologische Untersuchungen über das Lesen auf experimenteller Grundlage*. Halle: Neimeyer, 1898.
- ERIKSEN, C., POLLACK, M., & MONTAGUE, W. Implicit speech: Mechanism in perceptual encoding? *Journal of Experimental Psychology*, 1970, **84**, 502-507.
- FORSTER, K., & CHAMBERS, S. M. Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 627-635.
- GIBSON, E. J. The ontogeny of reading. *American Psychologist*, 1970, **25**, 136-143.
- GOLDSCHIEDER, A., & MULLER, R. F. Zur Physiologie und Pathologie des Lesens. *Zeitschrift für Klinische Medizin*, 1893, **23**, 131-167.
- GOUGH, P. B. One second of reading. In J. F. Kavanaugh & I. G. Mattingly (Eds.), *Language by ear and by eye*. Cambridge, Mass: M.I.T. Press, 1972. Pp. 331-358.
- HENDERSON, L., COLTHEART, M., & WOODHOUSE, D. Failure to find a syllabic effect in number-naming. *Memory & Cognition*, 1973, **1**, 304-306.
- HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, 1951, **41**, 401-410.
- JOHNSTON, J. C., & MCCLELLAND, J. L. Visual factors in word perception. *Perception & Psychophysics*, 1973, **14**, 365-370.
- KENDALL, M. G. *Rank correlation methods*. London: Griffin, 1948.
- KLAPP, S. Implicit speech inferred from response latencies in same-different decisions. *Journal of Experimental Psychology*, 1971, **91**, 262-267.
- KLAPP, S. Syllable-dependent pronunciation latencies in number naming: A replication. *Journal of Experimental Psychology*, 1974, **102**, 1138-1140.
- KLAPP, S., ANDERSON, W., & BERRIAN, R. Implicit speech in reading, reconsidered. *Journal of Experimental Psychology*, 1973, **100**, 368-374.
- KOLERS, P. A., & KATZMAN, M. T. Naming sequentially presented letters and words. *Language and Speech*, 1966, **9**, 84-95.
- KORTE, W. Über die Gestaltauffassung im indirekten Sehen. *Zeitschrift für Psychologie*, 1923, **93**, 17-82.
- KUČERA, H., & FRANCIS, W. N. *Computational analysis of*

- present-day American English*. Providence, R.I: Brown University Press, 1967.
- MASSARO, D. W. Perception of letters, words and nonwords. *Journal of Experimental Psychology*, 1973, **100**, 349-353.
- MATTHEWS, M. L., & HENDERSON, L. Facilitation of foveal letter recognition by metacontrast. *Psychonomic Science*, 1971, **23**, 153-154.
- MEZRICH, J. J. The word superiority effect in brief visual displays: Elimination by vocalization. *Perception & Psychophysics*, 1973, **13**, 45-48.
- PILLSBURY, W. B. The reading of words: A study in apperception. *American Journal of Psychology*, 1897, **8**, 315-393.
- POSNER, M. I. On the relationship between letter names and superordinate categories. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 279-287.
- REICHER, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 1969, **81**, 275-280.
- RUBENSTEIN, H., LEWIS, S. S., & RUBENSTEIN, M. Homographic entries in the internal lexicon: Effects of systematicity and relative frequency of meanings. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 57-62.
- SMITH, E. E., & HAVILAND, S. E. Why words are perceived more accurately than non-words: Inference versus unitization. *Journal of Experimental Psychology*, 1972, **92**, 59-64.
- SMITH, F. The use of featural dependencies across letters in the visual identification of words. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 215-218.
- SMITH, F. *Understanding reading*. New York: Holt, Rinehart and Winston, 1971.
- SPOEHR, K. T., & SMITH, E. E. The role of syllables in perceptual processing. *Cognitive Psychology*, 1973, **5**, 71-89.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University, 1944.
- THOMPSON, M. C., & MASSARO, D. W. Visual information and redundancy in reading. *Journal of Experimental Psychology*, 1973, **98**, 49-54.
- WHEELER, D. D. Processes in word recognition. *Cognitive Psychology*, 1970, **1**, 59-85.

NOTES

1. The finding that *relative* letter difficulty in this task is not influenced by adjacent contours can also be seen in data reported by Matthews and Henderson (1971), who presented tachistoscopic exposures of target letters (A, E, O, or U) alone or flanked by two Os, two Xs, or two Hs. From their data, it can be seen that no systematic changes in the relative letter difficulty of the target letter occurred as a function of the kind of letter context (irregular, circular, straight, or none) in which the target was viewed.
2. One subject's data were discarded because his overall mean response time was over 700 msec (nearly *twice* that of the other subjects), and it was inferred that his task performance probably represented a fundamentally different kind of response process relative to that of the other subjects. But discarding his data did not result in any change in the computed p levels of any of the Fs.

(Received for publication July 8, 1975;
revision accepted August 5, 1975.)