

The distribution of information within letters

JOSEPH SHIMRON and DAVID NAVON
University of Haifa, Haifa, Israel 31 999

To study the relative importance of various letter segments for letter recognition, we presented each letter of two alphabets, English and Hebrew, preceded by a brief presentation of an intact or a mutilated version of it. Mutilations were done by eliminating a specific segment. It was reasoned that the more critical the eliminated segment is, the less the mutilated version activates the letter code in memory and, thus, the longer it takes to name the subsequently presented target letter. This procedure was successful in detecting significant differences consistent with our expectations. In further analysis, it was shown that the latency data were highly correlated with the distinctiveness of the mutilated segment, its uniqueness over the alphabet, its impact on the letter global shape, and its topography within the letter and other variables. The dependency of latency on the various factors varied considerably between alphabets. Some correlational analyses were done to evaluate the roles of the various factors.

Information is not evenly distributed along the printed line. Some words are more predictable from their context than are others, so they may be considered as carrying less information. Similarly, some letters in a word are more redundant than others. In a similar vein, some elements or characteristics of a single letter might be more critical or diagnostic for its identification than would others.

As early as 1879, Javal (see Huey, 1908/1968) noted that when an English text was presented in such a way that the lower half of each line was removed, readers could read it faster and more accurately than when the upper half was removed. Huey suggested that the reader could withstand the removal of one part better than another because the removed part probably contained less information. He concluded, thus, that the upper part of an English text was more informative than its lower part was. In a previous study (Shimron & Navon, 1980), we showed that, whereas reading the English text was impaired by mutilating the top part of the line, the reverse was found for the Hebrew text. This result was ascribed to the different ways in which information was distributed along the vertical axis of Roman and Hebrew letters. Kolars (1969) used the same rationale to suggest that the right halves of Roman letters were, on the average, more informative than the left halves.

This paper reports an attempt to study in more detail the relative importance of various letter parts as well as possible sources for it.

This work was supported by grants from the Ford Foundation received through the Israel Foundations Trustees and from the Israel Commission of Basic Research. Preparation of the manuscript was done while the second author visited at the Center for the Study of Reading of the University of Illinois at Urbana-Champaign. This report represents equal and shared contributions of both authors.

In some cases, the importance of a certain letter element is obvious. For example, the lower horizontal stroke of the letter E is necessary for distinguishing it from the letter F. In many other cases, a mutilation of a certain element does not transform the letter into another one, but still makes its identification more difficult. We reasoned that one way to study the informativeness of various letter elements (i.e., their contribution to letter identification) is to eliminate them one at a time and then to test the effect of those eliminations on recognition. Our question was to what extent the mutilated letter still has the perceptual effect of the intact one. In other words, how much does the mutilated version of a letter activate the internal representation of that letter in memory? To answer this question, we devised the following procedure: The subject was asked to name, as fast as possible, a letter presented visually. The letter was preceded by a prime that, in most cases, was a mutilated version of the same letter. Our rationale was that the more critical the eliminated part (or the properties to which it contributes) is for the recognition of the letter, the less (or the later) the mutilated version activates the letter code in memory, and, thus, the less facilitation of the subsequently presented intact letter is to be expected. That should be reflected in a longer latency to name the letter.

In order to attain more generality of our conclusions, we chose to investigate two alphabets. We used bold Hebrew letters and uppercase English letters.

EXPERIMENT 1: ENGLISH LETTERS

Method

Apparatus. The stimuli were presented via a three-field Gerbrands Harvard-type tachistoscope (Model T-3B-1). Viewing distance was 16 cm. The luminance of the fields was about 11.0 cd/m². A crystal microphone served to transmit the onset of the subject's vocal response to a voice-operated relay that terminated a digital

millisecond clock, started by the onset of the target letter. Latencies were recorded by means of a printer.

Stimuli. The 26 letters of the Roman alphabet served as target stimuli. They were made by applying Letraset Futura Bold letters (Sheet No. 103) on celluloid, duplicating on a white paper, and then pasting each of the duplicated letters at the center of a white tachistoscope card. They measured 15 mm (1.13 deg of visual angle) vertically. Mutilated versions used as primes were prepared in a similar manner, only the eliminated part was not rubbed off the Letraset sheet. Our criterion in producing the mutilated stimuli was to eliminate from each letter a fragment that consisted of either a 90-deg section of a curved segment or a straight segment that measured about half of the height or all of the width of a typical English letter and about half of the height or half of the width of a typical Hebrew letter.

All the stimuli are shown in Figure 1. A masking stimulus was prepared by cutting several letter segments and applying them haphazardly within a square with a side of 20 mm. A 19 x 18 mm rectangle circumscribing a cross with bars of the same width as the bars of the letter served as a neutral prime.

Design and Procedure. In each trial, a beep was played and the masking stimulus was shown for 800 msec simultaneously, followed after a 200-msec interval by a prime that could be either neutral (see above) or a mutilated version of the target letter. The prime was shown for 300 msec and was replaced by the masking stimulus, which was presented for 150 msec and was followed by a 500-msec presentation of the target letter. The subject had to name the target letter as quickly as possible, and his/her response and the latency from the onset of the target were recorded. Subjects were strictly warned not to try to respond before the target was presented by guessing it from the prime.

Each letter was presented twice with the neutral prime and twice with each of its mutilated versions. The experiment started by a block of 18 practice trials in which targets were Hebrew letters and primes were some mutilated versions of them. Then followed a block consisting of 130 trials in which the primes were mutilated letters. Each target-prime pair was presented just once in a random order. The third block consisted of a random presentation of 52 trials with the neutral prime. The fourth block was a replication of the second one, only the order of trials was changed by per-

muting five subblocks of 26 trials each. The particular order used for half of the subjects during the second block was used for the other half during the fourth one, and vice versa. Subjects received with the instructions a sheet displaying all the target letters as well as the mask. They were instructed to look at the field as soon as the beep was played and were supposedly encouraged to attend to the prime by its introduction in the instructions as "a clue to the identity of the subsequent letter."

Subjects. Twenty-four subjects were used. All were students at the University of Haifa and had been familiar with the Roman alphabet for at least 10 years. All had normal or corrected-to-normal vision.

Results

Errors were very scarce. Only in two cases were both replications of a certain target-prime pair in error. In those cases, the data of the letters in question (Q and J) were analyzed without the data of the subject who had erred. When one replication was in error, analysis was based on the other one. Mean latencies for each prime, including the neutral one, and for each letter were calculated. The difference in milliseconds between latency to name the target letter with a particular prime and with a neutral prime is given in Table 1 for each of the letters and each of its primes. It is called a facilitation score—but note that a negative score indicates facilitation and a positive one indicates inhibition. The primes themselves are presented in Figure 1 in the order in which they appear in Table 1—namely, arranged from short latencies on the left to longer latencies on the right. While inspecting these data and the following analyses, one should bear in mind that each prime latency is based on just two replications per subject.

Separate analyses of variance were conducted on the data of each letter to compare the effects of the various primes. Since only two replications per a given prime were used, the results might have been considerably affected by the positions those replications occupied within the sequence of trials. Therefore, in addition to the ordinary analysis using the interaction of subjects with primes as an error term, we calculated another F term in which replications within primes within subjects served as an error term; this term presumably reflects most of the variability due to sequence effects. The rightmost column in Table 1 presents the significance of min F' calculated on the basis of both types of F ratio (Clark, 1973). Newman-Keuls pairwise comparisons were done, using as an error term just the interactions of primes with subjects. The results of those comparisons are presented in Figure 1: Primes that are underlined by a common line are not significantly different from each other at the .05 level. For example, the leftmost prime for the A is significantly more facilitative than the neutral prime, but not significantly more facilitative than the second one from the left. The results of this experiment are discussed after the presentation of Experiment 2.



Figure 1. The English letters used in Experiment 1 and the various primes arranged by mean latency, with shortest latency on the left. The black rectangle stands for the neutral prime (see text). Horizontal lines show the results of post hoc comparisons.

Table 1
Facilitation Scores for the Various Primes of Each English Letter (Experiment 1)

Letter	Facilitation Scores								Significance Level
A	-120	-116	-114	-94	-14	0			.001
B	-100	-95	-79	-66	-27	-6	-4	0	.01
C	-58	-23	-3	0	+37				n.s.
D	-86	-63	-41	0	+5				.05
E	-109	-97	-62	-23	0	+46			.001
F	-46	-15	-4	0	+24	+37			n.s.
G	-151	-129	-80	-48	0	+31			.001
H	-142	-108	-54	-37	-21	0			.001
I	-62	-19	0						n.s.
J	-72	-15	-11	0					n.s.
K	-217	-157	-142	-102	0				.001
L	-7	-3	0	+73					n.s.
M	-154	-149	-134	-132	-110	-107	-105	-104	0
N	-93	-79	-66	-57	-11	0			.01
O	-99	-85	-80	-50	0				.05
P	-123	-74	0	+14	+14	+39			.001
Q	-165	-161	-157	-119	0	+20			.001
R	-174	-122	-115	-91	-68	0	+90		.001
S	-140	-139	-127	-84	-29	0			.001
T	-70	-8	-3	0	+27				n.s.
U	-127	-121	-96	-37	0				.01
V	-160	-123	-51	-15	0				.001
W	-214	-204	-201	-200	-182	-165	-139	-139	0
X	-64	-44	-13	0					.05
Y	-133	-99	-44	0					.05
Z	-147	-114	-109	-87	-80	-36			.05

Note—The order within a line corresponds to the order within a respective line in Figure 1. A score is the difference in milliseconds between naming latency to that letter with that prime and with a neutral prime. Significance levels of min *F* ratios are given in the far right column.

EXPERIMENT 2: HEBREW LETTERS

Method

The same method as in Experiment 1 was used, only the target letters were Hebrew. Twenty-one of 22 letters in the Hebrew alphabet were used. One was not used because, in the type font employed (Letraset, Amit Sheet No. 12367), that letter consisted of just half of a vertical bar.

The second and fourth block of trials consisted of 119 trials in which the primes were mutilated letters. The trials were arranged in four subblocks of 24 each and one subblock of 23 trials. The third block consisted of 42 trials with the neutral prime. In the practice trials, subjects were presented with Roman letters.

Fifteen subjects were used, all of whom were students at the University of Haifa and were very familiar with the Hebrew alphabet.

Results

As in Experiment 1, errors were very scarce. One subject made errors in both replications of a certain target-prime pair. His data with regard to all appearances of that letter were not included in the analysis. When one replication was in error, analysis was based on the other one.

The data are presented in Table 2 and Figure 2, which are completely analogous to Table 1 and Figure 1, respectively.

GENERAL DISCUSSION

As can be seen in Figure 2, on the whole the primes were not equally facilitative. Some of them were, in fact, inhibitory; they delayed letter naming with respect to a neutral prime, probably because they resembled a letter that was different from the one to be named.

Looking across a whole set of mutilated letters as primes of differential facilitative effects, we can attempt to evaluate some general factors of letter discrimination. In order to simplify the analysis, we inspected for every letter the primes that produced the shortest and the longest latencies. We assumed that the most facilitative primes were those mutilated letters in which the missing element contributed very little to the recognition of that letter. By way of contrast, the least facilitative (or even inhibitory) primes were those mutilated letters in which the missing element was crucial for letter discrimination.

In analyzing the data further, we shall consider three types of variables that might have affected the process. The first type to be considered are topographic variables. The issue is whether there is a correlation between latency of recognition and the locus

Table 2
Facilitation Scores for the Various Primes of Each Hebrew Letter (Experiment 2)

Letter	Facilitation Scores										Significance Level
א	-54	-50	-48	-13	0	+47					n.s.
ב	-86	-51	-47	-43	-37	-25	0	+33			n.s.
ג	-102	-51	-40	-21	0						n.s.
ד	-76	-70	0	+20	+30	+44					n.s.
ה	-105	-80	-52	0	+47	+63					.01
ו	0	+19	+37								n.s.
ז	0	+19	+19	+261							n.s.
ח	-61	-56	-41	-28	0	+11	+55				.01
ט	-70	-39	-13	-3	0	+4	+11	+36			n.s.
י	-78	-14	0	+13	+21	+32	+62				.05
כ	-122	-96	-53	-27	-24	0	+10				.05
ל	-81	-77	-59	-38	-19	-19	0	+24	+48		.05
מ	0	+28	+34	+159							n.s.
נ	-80	-54	-45	-22	0	+9	+11	+39			n.s.
ס	-48	-45	-13	-10	-1	0	+7	+52			n.s.
ע	-134	-118	-96	-70	-66	-37	0	+11			.05
פ	-170	-105	-93	-64	-27	0					.01
צ	-168	-127	-125	-124	-94	-92	-90	0			.01
ק	-42	0	+4	+4	+58						.05
ר	-115	-98	-80	-77	-75	-63	-20	-2	0		.05
ש	-131	-128	-123	-110	-108	-108	0	+60			.001

Note—The order within a line corresponds to the order within a respective line in Figure 2. A score is the difference in milliseconds between naming latency to that letter with that prime and with a neutral prime. Significance levels of min F^2 ratios are given in the far right column.

of the missing elements with regard to the two major axes of the letter matrix (right-left, up-down). If such a correlation exists, what is its source?

Second, we asked about the relationship between recognition latency and other possible sources of element informativeness that are unrelated to the location of the element with respect to the major axes of the letter. Those sources have to do with the relationship of elements to other elements within the letter that presumably affect the likeness of the mutilated letter to its template or with the presence of those elements in other letters that presumably affect their informational value.

We considered two types of informational variables: distinctiveness and uniqueness.

Distinctiveness

Distinctiveness of a letter element is the extent to which it defines a difference between one letter and others, which is determined by the extent to which the other nonmutilated features of the letter constitute or are subsumed in the set of features of one (or more) of the letters. For example, the lower horizontal stroke of the letter E is perfectly distinctive because, in its absence, the letter would look exactly like an F. Similarly, the diagonal stroke of the letter R is perfectly distinctive because, in its absence, the rest of the figure looks exactly like a P. The upper diagonal of the letter K is fairly, although not perfectly, distinctive because the rest of the features constitute a subset of the letter R. In contrast, the upper horizontal stroke of the letter E is not distinctive at all because, even in its absence, the remaining pattern is not compatible with any other letter.

More formally, if each letter j in the alphabet is conceived of as a set of elements E_j , and the perceptual contribution of elements is represented by a salience function f (see Tversky, 1977), then the distinctiveness of a certain element e for a given letter k may be defined as the maximum of the term $f(E_k - e) / f(E_j)$ over all letters of the alphabet other than k that satisfy: $(E_k - e) \cap E_j = \phi$.

Distinctiveness, as it is defined here, may be construed as the criticality of the feature for differentiating between the letters and other letters in the alphabet.



Figure 2. The Hebrew letters used in Experiment 2 and the various primes arranged by mean naming latency, with shortest latency on the left. The black rectangle stands for the neutral prime (see text). Horizontal lines show the results of post hoc comparisons.

We computed distinctiveness by a method congenial with the above definition. Our measure was the ratio between the number of line segments in the mutilated letter and the number of line segments in the intact letter of which the mutilated version was a part. If the mutilated letter could not be entirely subsumed in any letter of the alphabet, the distinctiveness value of the mutilated element was zero. If the mutilated letter was identical with another letter, the distinctive value of the mutilated element was one. Our count of number of line segments was, of course, fairly arbitrary, but we believe that it must be monotonous with the measure that would have resulted had we known the features by which letters are analyzed.

Uniqueness

The uniqueness of an element for a given letter is inversely related to the number of other letters of which it is a part. For example, if we superimpose all letters in an alphabet, one upon the other, some letter parts will overlap more than will others. Some letter elements may appear in just one letter. This may be the case with the small diagonal of the letter Q. We shall say that its uniqueness for the letter Q is very high.

The uniqueness of an element *e* for a given letter *k* may be conceived of as its diagnosticity $p(e | k) / p(e | \bar{k})$, where \bar{k} is the set of all other letters in the alphabet.¹

To score uniqueness, the location of the mutilated element was defined within the common matrix for all uppercase letters of the type we used. Then, each one of the letters was superimposed upon that matrix, and the number of times that location was covered by other letters of the alphabet was counted. Uniqueness is inversely related to this measure of line segment overlap; thus, it was defined as its negative.

Likeness

So far, we have conceived of the letter primes as if their only function is to differentiate one letter of the alphabet from the others. Thus, attention was given to the differences between letters. But a prime may fail to facilitate letter recognition, not because it suggests other letters, but, rather, because it does not suggest very strongly the image of the target letter. For example, the pattern resulting from mutilating the upper horizontal of the letter E is uniquely different from any other letters of the alphabet, but it makes it quite dissimilar with the stored image of an E. It seems that the damage would be considerably less when the mutilation is at the vertical stroke. That still leaves this pattern quite similar to the typical E. Whereas the previous measures were functions of the other members in the stimulus ensemble (namely, the letters of the alphabet), the variables that we sub-

sume under the heading "likeness variables" are to some extent independent of the range of alternative stimuli. Conceivably, even if all element combinations had existed so that all mutilations had been equally disruptive from an informational point of view, some mutilations would still have resembled the prototype less than others.

By relating to letter envelope and edge as likeness variables, we do not wish to convey that mutilations of inner elements cannot destroy the Gestalt of the letter. For example, mutilating the horizontal stroke of an H may be quite harmful. Our taxonomy is based on our intuition that the effects of mutilations of letter envelope or edge are fairly independent of the range of alternatives.

Variables

To gain some insights about the role of the factors mentioned above on letter recognition, we correlated the latency facilitation score associated with the primes (see Tables 1 and 2) with the following variables:

Element Variables

Topographic Variables

Whether the element is at the left or at the right half of the letter.

Whether the element is at the lower or at the upper half of the letter.

Informational Variables

The distinctiveness of the element

The uniqueness of the element

Likeness Variables

Whether the absence of the element changes the envelope of the letter

Whether the element constitutes the edge of a stroke

Letter Variables

The number of line segments that constitute the letter.

To simplify the analyses, we analyzed only primes associated with the shortest and longest latencies for a given letter. In Table 3, we present percentages or mean scores of primes with the shortest and with the longest latencies, in Hebrew and English separately, according to the above variables.

We also computed Pearson product-moment correlation coefficients between every one of the above

Table 3
Percentages of Primes Associated With Shortest and Longest Target Naming Latencies in Hebrew and in English that Satisfy the Condition at the Left (a, b, e, f), or Mean Scores of Variables (c, d) for These Primes

Condition or Variable	English			Hebrew		
	Shortest	Longest	p <	Shortest	Longest	p <
(a) At the Right Half	15.40	61.50	.025*	76.20	23.80	.01*
At the Left Half	69.20	23.10	.025*	14.30	66.70	.01*
(b) At the Upper Half	50.00	23.10	.05*	57.10	47.60	n.s.*
At the Lower Half	46.20	65.40	.05*	42.90	52.40	n.s.*
(c) Element Distinctiveness	.16	.44	.01†	.21	.61	.001†
(d) Element Uniqueness	-4.81	-3.58	.10†	-8.38	-3.14	.001†
(e) Mutilation Changes Letter Envelope	61.50	65.40	n.s.*	19.00	71.40	.025*
(f) The element Constitutes the Edge of a Stroke	34.60	61.50	.052*	19.00	76.20	.025*

Note—Variable g is not included in the table since the number of line segments is the same in the shortest and longest target naming latencies. For Conditions a and b, percentages do not add up to 100 because some elements could not be located at either of the sides. *In a McNemar test. †In a matched-pairs t test.

variables, including the latency facilitation scores and all other variables (see Table 4).

Finally, we ran a stepwise multiple regression with the latency facilitation scores as the dependent variable and all other variables as independent variables.

In the stepwise multiple regression in English, distinctiveness and left/right were the only significant variables ($p < .001$). They accounted for 40% and 21%, respectively, of the total variance. The same analysis in Hebrew revealed a more complex picture. The uniqueness variable entered first in the equation and in itself accounted for 32% of the variance ($p < .001$). The variables of distinctiveness, number of elements, and left/right entered next in this order, with marginal contributions to the variance accounted for 6%, 5%, and 7% ($p < .10$, $p < .10$, and $p < .05$), respectively. However, in the equation having all four variables, the beta weights were .28, .22, $-.38$, and $-.33$, respectively. A more detailed discussion of these results, and of various partial correlations we calculated, follows.

Topographic variables: Right/left part of the letter. Kolers (1969) observed that the most helpful clues in an English letter appear on its right. But, to the best of our knowledge, no experimental evidence has been so far presented to support this claim.

The retinal position and the angular width of the letters were such that laterality effects or reading habits that might affect scanning direction were most likely eliminated.

Our data indicate that mutilation of line segments at the left and line segments at the right of the letters does not have the same effect on recognition, and, also, that the English alphabet is diametrically different in this sense from the Hebrew. Subjects who identified English letters (see Table 3) appear to have identified very well letters mutilated in their left part but not in their right part. The opposite was true for subjects who identified Hebrew letters, although the effect was somewhat smaller. The correlations be-

tween the horizontal position and the latency facilitation score were .57 for the English alphabet and $-.43$ for the Hebrew alphabet.

The marked difference between the effects observed within the two alphabets indicates that these topographic effects cannot be attributed to any inherent advantage of one sector of the stimulus or of the visual field over the other. This is best illustrated by the fact that, whereas mutilating the right portion of the English letter O was more disruptive than mutilating its left portion (Figure 1), the reverse was true of the Hebrew letter Samech (third from top in the right column of Figure 2), which is very similar to an O.

An apparent explanation for this interaction of alphabets with horizontal position is that, whereas most of the information of English letters resides at the right (11 of them are right-facing, and only one, J, is left-facing), the information in Hebrew letters is located mostly at their left side (14 of them are left-facing, and only one is right-facing). This explanation is supported by the fact that the correlation between horizontal position and the latency facilitation score in Hebrew letters is considerably reduced ($r = .28$) and becomes just marginally significant ($p < .10$) once the variable of uniqueness is partialled out. However, this does not recur in the English alphabet. There the variable of horizontal position accounts for roughly 30% of the variance of facilitation scores, regardless of whether either, all, or none of the other variables is partialled out. As we comment later, we believe that this is due to the fact that, because of the versatility of curvature in the English font used, we did not find the right way to measure uniqueness. However, our data evidently are not incompatible with the possibility that the greater contribution of the right side of English letters to their identification is *not* due just to the concentration of unique or distinctive features in this side.

Topographic variables: Lower/upper part of the letter. We recently found (Shimron & Navon, 1980) that mutilation of the top of a whole line of text

Table 4
Correlation Coefficients Between the Variables Tested

	English Letters								Hebrew Letters							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
(1) Facilitation Score	1.00								1.00							
(2) Left/Right	.57	1.00							-.43	1.00						
(3) Upper/Lower	-.28	-.34	1.00						-.06	.21	1.00					
(4) Distinctiveness	.63	.18	-.22	1.00					.27	-.37	-.16	1.00				
(5) Uniqueness	.12	.37	-.10	-.07	1.00				.57	-.37	-.15	.07	1.00			
(6) Change of Envelope	.10	.12	-.07	.21	-.03	1.00			.40	-.20	-.19	.30	.44	1.00		
(7) Mutilation of Edge	.31	.13	-.06	.28	-.03	.65	1.00		.41	-.25	-.14	.37	.49	.95	1.00	
(8) Number of Line Segments	-.26	-.22	.16	-.22	-.02	-.39	-.35	1.00	-.37	-.23	-.16	.23	-.41	-.22	-.18	1.00

was more harmful than mutilation of its bottom in mixed-case English, but not in Hebrew, in which the opposite was true. We attributed this mainly to the presence of informative features at the top of mixed-case English letters and at the bottom of Hebrew letters.

Here, however, the correlation between lower/upper and facilitation scores in English was only $-.28$ ($p < .10$), and its marginal contribution to the variance of latency facilitation as indicated by the stepwise multiple regression was negligible. Lower/upper effect in Hebrew letters was even weaker.² This was somehow surprising since, as mentioned above, we found the lower part of Hebrew letters to be significantly more informative in a task of reading lines of text. One reason for the difference between the two studies may have to do with the differences in the tasks and conditions in the two experiments. Mutilating a complete half or third, as we did in the other study, may have a stronger effect than mutilating a single line segment.

Informational variables: Distinctiveness. There are not too many letters in the alphabet in which mutilation of one line segment would change the letter identity. But, as could be predicted, when it happened it was almost always associated with the longest latency in recognition. With one exception, such primes were never associated with shortest latencies. Indeed, it can be seen from Table 3 that the mean score for distinctiveness in both Hebrew and English was three times higher among primes associated with longest latencies than among those associated with shortest latencies. For the English letters, this variable accounted for the greatest percentage of the variance. For the Hebrew letters, it entered second in the stepwise regression and accounted for 6% of the variance.

Informational variables: Uniqueness. As mentioned above, it was predicted that absence of elements that are diagnostic in that they appear just in the target letter or in a few more letters would be more disruptive than absence of elements that are present in many letters.

The correlation coefficients between the facilitation scores and uniqueness in Hebrew letters was .57.

By itself, it explained 32% of the variance in the multiple regression and was the first factor in the equation. We calculated partial correlations between the facilitation score and uniqueness while holding constant, one at a time, each of the other independent variables. These partial correlations were never smaller than .46, which indicates that none of the other independent variables can, in itself, account for the uniqueness effect on latencies. On the other hand, no other variable correlated significantly ($p < .05$) with the facilitation score when uniqueness was held constant.

However, in English the correlation coefficient between uniqueness and latency was nonsignificant. We believe that this difference has something to do with the variety of segment types in English and in Hebrew. Most Hebrew letters fit a design of a square block. There are many fewer curves and diagonals in Hebrew as compared with English letters, at least in the type style we used. Also, the measure of uniqueness we used was positively related to the diversity of segment types. If letters do not overlap much, as it happens in English, many segments may be judged as unique, although in fact they are functionally equivalent. Also, perhaps diversity makes the relation of the segment to the rest of the character more crucial than its mere presence. It may be that, for that reason, our method of rating uniqueness failed to capture the psychological variable that was so compellingly captured by it in the Hebrew alphabet.

A strong effect of uniqueness supports models of letter recognition in which features are not just counted but, rather, weighted by their diagnosticity across the alphabet (e.g., Rumelhart & Siple, 1974).

Likeness variables: Change of letter envelope. Bouma (1971) defined letter envelope as the "smallest enclosing polygon without indentations." The concept is particularly helpful in understanding why most letter confusions occur within groups that are easily characterized by the common envelope of the group members. For example, more or less the same envelope characterizes the lowercase letters: a, s, z, and x; e, o, and c; y, v, and w, etc. (cf. Lupker, 1979).

The correlation coefficients between change of envelope and latency facilitation scores was .10 in English and .40 in Hebrew. In the multiple regression, none of them was found to have a significant contribution. The main reason is probably the high correlation of envelope change with the variable of edge mutilation. At least one of these variables might not have any independent causal role in recognition.

Likeness variables: Mutilation of an edge of a stroke. We made a distinction here between mutilations of an unconnected edge of a stroke and others that are either applied to the middle of a line or to an edge connected with some other stroke.

Mutilation of an edge of a stroke was significantly more harmful in both English and Hebrew. Does this factor have any independent contribution? The multiple regressions suggest that it does not. This variable, for obvious reasons, correlated very highly with the envelope change variable (.65 and .95 in English and Hebrew, respectively). Its correlation with latency facilitation became small (.12) and non-significant, once the envelope change variable was partialled out, for the English but not for the Hebrew letters. So, it is not completely clear, without further experimental investigation, which variable assumes here a more important causative role.

Furthermore, the effect of this variable may be due, not to the significance of edges or envelopes, but, rather, to an artifact of the distribution of informativeness over the letter space. That this might be the case is suggested by the fact that the correlation of the edge mutilation variable with the facilitation scores for Hebrew letters decreased from .41 to .19 ($p < .25$) when uniqueness was partialled out. This was not the case for English letters, but that might be because, as conjectured above, the uniqueness variable was poorly defined with respect to English letters.

Letter variables: Number of line segments. We predicted that the number of line segments in a letter will be negatively correlated with the latency facilitation scores. It was reasoned that, other things being equal, the more line segments there are in a letter, the less its recognition will be affected by mutilation of a single element. The expected correlations were found in both English and Hebrew, although the former did not quite reach significance. It should be pointed out, however, that those correlations became smaller and nonsignificant when some other variables were partialled out. In Hebrew, that occurred when uniqueness was held constant ($-.19, p < .25$), and in English it occurred when each of the other variables, except for uniqueness, was held constant. In Hebrew, number of line segments did contribute considerably to the prediction of facilitation scores, as indicated by the multiple regression analysis. However, the contribution of this variable was nonsignificant ($p = .087$)

when it was added to uniqueness and distinctiveness, which were already in the equations as the first two variables.

SUMMARY

The paradigm of priming letters with a mutilated version of themselves for the purpose of evaluating the diagnostic value of each line feature was found to be sensitive to a number of variables expected to play a role in letter recognition.

Some of the variables studied played their roles differently in the two alphabets investigated. This may serve as a reminder that studies of letter recognition should not be excessively anglocentric.

A more substantive lesson is that letter recognition depends on the variety of alternatives in each language. Two of the informational variables (uniqueness and distinctiveness) were found to have a significant effect—the first, just in Hebrew, and the second, in both languages. Together, they represent the only group that had a salient effect in both alphabets.

Likeness, as defined in this study, was not found to have a significant independent effect in either alphabet. Our findings lead us to suspect that some of its effect might be mediated by its covariation with informational variables.

We expected the topographic variables to have an effect, but we also expected informational or likeness variables to account for topographic effects. However, the strong effect of horizontal position in the English alphabet appears to indicate that the topography of features is important in its own right. However, as we pointed out above, we believe that a better operational definition of informational variables might be able to show that the topographic effect is reducible to an informational account.

In summary, we managed to map out the relative importance of various letter segments in two alphabets and to show with a considerable degree of certainty that it is greatly mediated by their informativeness—namely, by their value for distinguishing between the target letter and other letters in the alphabet. A more conclusive statement about the sources of the differential criticality of the various segments will have to await a well-designed experimental study with *contrived* stimulus material rather than natural alphabets.

REFERENCES

- BOUMA, H. Visual recognition of isolated lower-case letters. *Vision Research*, 1971, 11, 459-474.
- CLARK, H. B. 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.
- HUEY, E. B. *The psychology and pedagogy of reading*. Cambridge, Mass: M.I.T. Press, 1968. (Originally published, 1908.)

- KOLERS, P. A. Clues to a letter's recognition: Implications for the design of character. *Journal of Typographic Research*, 1969, 3, 145-167.
- LUPKER, S. J. On the nature of perceptual information during letter perception. *Perception & Psychophysics*, 1979, 25, 303-312.
- RUMELHART, D. E., & SIPLE, P. Processes of recognizing tachistoscopically presented words. *Psychological Review*, 1974, 81, 99-118.
- SHIMRON, J., & NAVON, D. The distribution of visual information in the vertical dimension of Roman and Hebrew letters. *Visible Language*, 1980, 14, 5-12.
- TVERSKY, A. Features of similarity. *Psychological Review*, 1977, 84, 327-352.

NOTES

1. Note that one may try to define distinctiveness, as well, in terms of diagnosticity. It may be considered as inversely related to the diagnosticity of the *rest* of the features of the given letter—namely, of $E_k - e$. We did not pursue this definition further since, for intuitive reasons, we preferred our own.

2. Nevertheless, we counted eight letters—the naming of which was most facilitated by a prime mutilated at the top and was least facilitated (or inhibited) by a mutilation at the bottom—and only four letters in which the reverse was true.

(Manuscript received March 24, 1981;
accepted for publication June 29, 1981.)