Remembering operations*

PAUL A. KOLERS

University of Toronto, Toronto, Ontario, Canada M5S 1A1

Two commonplace assumptions about encoding are that sentences are encoded and recognized on the basis of their semantic features primarily and that information regarding form features such as typography is typically ignored or discarded. These assumptions were tested in the present experiment where, within a signal-detection paradigm, S sorted sentences according to whether he had seen them before or not (old vs new) and, if they were old, whether their reappearance was in the same typography as on the first occurrence or a different one. Of the two typographies, one was familiar and the other unfamiliar. Results show that a considerable amount of information regarding surface features is stored for many minutes and that ease of initial encoding is inversely related to likelihood of subsequent recognition: sentences in the unfamiliar typography were remembered better. The results are probably not due to time spent encoding; control tests suggest that time spent encoding a difficult typography does not by itself increase recognition of the semantic content embodied in the typography. Other control tests show that pictorial features or images of the sentences play no significant role in their subsequent recognition. One interpretation of the results is that the analytic activities or cognitive operations that characterize initial acquisition play a significant role in subsequent recognition.

We never perceive the world directly, hence what we remember can never be a direct copy of the world. Modern theorists emphasize the notion that perception and memory are selective: we perceive and remember features of the object or event that we have attended to. But how do we first perceive features? Features themselves are the results of perceptual encodings and operations; they are not "given" by the stimulus. They are useful descriptive categories, but must themselves have been created as perceptual events in some way. We may say, for example, that in reading a sentence we noticed this or that feature about it, its grammar, length, type size, and the like; in doing so we are using learned response categories to describe our experiences, but we are not defining anything basic about the processes of remembering. What we have done is to take the *results* of perceptual encoding—features—and sought to explain the *processes* of encoding in their terms; but features are not perceived directly any more than the object itself is. Hence, it means little to say that we recognize an object by virtue of recognizing its features, for the statement is a tautology.

A somewhat different view can be formulated as follows: What is recognized or remembered are the analytical operations themselves that have been used to transform an optical input into a perceptual experience. To put it another way, it can be alleged that we remember not only things or events (semantic features), but what we have done to acquire, apprehend, or attain

them. We remember in terms of the operations or activities of encoding as well as, sometimes, their results. This conjecture is tested in the present experiments.

EXPERIMENT I

EFFECTS OF APPEARANCE

Method

Materials

Thirteen different passages of connected discourse were taken from Miller (1962) to yield the source material. One hundred and twelve sentences ranging in length from 8 to 40 words (mean of 23 words) were culled from the source and prepared in several ways. Half of the sentences were prepared in normal typography, the other half inverted, i.e., the sentences were typed so as to appear rotated 180 deg on the horizontal axis (Fig. 3, Example I). Twenty-eight sentences of each kind were set aside for later use, and the remaining 56 were duplicated. Of the latter sentences, 28 were duplicated exactly and 28 were duplicated as to words but in the alternate typography. The result of these manipulations yielded 14 sentences that appeared twice in normal typography (N), 14 that appeared twice in inverted typography (I), 28 sentences that appeared once in N and once in I, 28 sentences that appeared only once in N, and another 28 that appeared only once in I. The 56 sentences that appeared only once were the even-numbered sentences taken from the passages; the 56 sentences that were duplicated were the odd-numbered sentences. A few exceptions to this rule were made to exclude very long and very short sentences.

The sentences were typewritten in a single font, one sentence to a page; they were enclosed in clear plastic envelopes and then were shuffled many times to destroy their natural order. For purposes of subsequent scoring, they were then grouped into units of eight sentences.

Procedure

The methods of signal-detection theory were utilized. The design required S to read a deck of 56 sentences, half appearing in N, the other half in I, and subsequently to read a deck of 112 sentences comprised of the 56 "old" sentences and 56 "new"

*This study was assisted under Grant 382 of the Ontario Mental Health Foundation and by the National Research Council of Canada (A 7655). The assistance of Linda Vettor and Stephen Israelstam in collecting and analyzing the data is also gratefully acknowledged. In addition, I thank Robert S. Lockhart and John C. Ogilvie for their helpful advice on aspects of design and analysis and J. Elisabeth Wells for the suggestion regarding *d'sem*. A briefer version of this paper was presented to the Eastern Psychological Association, April 1972.

Fig. 1. Relation of sentences in Pass 1 and Pass 2 decks.

ones. Call the first reading "Pass 1" and the second reading "Pass 2." Of the old sentences on Pass 2, half appeared in the same typography as on Pass 1, the other half in the alternate typography. The appearance of the old sentences on Pass 1 and Pass 2 may therefore be designated as normal-normal (NN), inverted-inverted (II), inverted-normal (IN), and normal-inverted (NI). The new sentences in Pass 2 are designated new-normal (WN) and new-inverted (WI), where "W" refers to "new" and the second letter indicates typography. The upper section of Fig. 1 shows the paradigm graphically.

After reading the 56 sentences in Pass 1, S was instructed to sort the 112 sentences of Pass 2 into one of three piles: new sentence; same words, same form; or same words, different form. Fifty-six sentences were properly sortable into the new category, and 28 sentences into each of the two same categories. The Ss were instructed that half of the Pass 2 deck sentences were new and half were old.

The 56 sentences of Pass 1 were scrambled into a different order for each S, in groups of 8 sentences each, and the sentences of Pass 2 were also scrambled, but into a still different order. One sentence of each of the four pairings (NN, II, NI, IN) was used to head the deck of 56, and another to end the deck. These were used to absorb primacy and recency effects and are not included in the scoring. The results, therefore, are based on S's sorting of 48 old sentences on Pass 2 and 56 new sentences.

Prior to the main task, each S was required to read aloud one page of normal text in less than 2 min. The Ss who failed to do so were disqualified. Successful Ss then read 10 lines of inverted typography aloud as practice. When S misread a sentence, he was required to reread it, both during the practice period and during the main test. A test session lasted approximately 1½ h, 30 min on Pass 1 and 60 min on Pass 2. The Ss were tested individually, reading all sentences aloud. Time taken to read each sentence was measured with a stopwatch.

Subjects

Twelve University of Toronto freshmen, six men and six women, were tested. They were paid for their services.

Analysis

An S's sortings were tallied as to hits and false alarms; these were converted to values of d' , which were subjected to analysis of variance. Two different false alarm rates are available for the computations. One is based on S's ability to tell an old sentence from a new one; for this analysis the new sentences, WN and WI, are used as false alarms. The second rate is based on S's ability to distinguish among old sentences with respect to typography; for this analysis, different form sentences (DF) are used as false alarms in the same form category and same form sentences (SF) are false alarms in the DF category. Figure 2A illustrates these conditions. In addition, different hit rates can be evaluated at different levels of analysis of performance. One level of analysis is old vs new. If a sentence is old, a second level of analysis is whether it is sorted into the SF or DF category. And among these sortings, a finer analysis can be carried out on the actual pairings of sentences. These different analyses evaluate separately a semantic and a typographic component in S's encodings. Figure 2B illustrates the logic of these analyses, the numbers therein showing the number of scorable sentences properly sortable at each level of analysis. However, analysis at the level of the second branching provides no information that is not also provided by analysis at the level of the third branching (pairs). Reported results are restricted therefore to the first and last branches of the tree.

Hits were tallied for the four pairings of sentences within different form and same form categories, and three values of d' were calculated. One was based on S's ability to tell old from new within a pairing, a semantic component of his judgment. Here, new sentences in N (WN) were false alarms for old (NN and IN), and new sentences in I (WI) were false alarms for old (II and NI). Second, S's ability to tell old sentences from each other was evaluated, a graphemic component of his judgment. For this analysis, same form sentences that S put into the DF category were taken separately as false alarms for different form sentences (IN and NI), and different form sentences that S put into the same form category were taken separately as false alarms for same form sentences (II and NN). These two analyses, one of semantic and the other of graphemic components, are indicated in the lower part of Fig. 1, where the values of d' found for segregating old from new sentences are indicated as d'_{new} and the values of d' based on segregating old sentences from each other are indicated as d'_{old} . The values of d'_{new} assess a semantic component to memory, but only within the correct graphemic category (SF and DF). A third analysis was carried out therefore to evaluate the semantic component irrespective of the graphemic category. An old sentence sorted into *either* the SF or DF category was taken as a hit, and the corresponding kinds of new sentences sorted into those categories were taken as

Fig. 2. Illustration of the task and of the analyses. (A = Sortings by S and the basis of false alarms. B = Levels at which different analyses were carried out.)

false alarms (i.e., the total of WN sentences in SF and DF was the false alarm rate for the total of NN and the total of IN sentences in those categories, and the total of WI was the false alarm rate for II and NI). This analysis, identified as d'_{sem} in the tables, is a broader and looser measure of the semantic component than is d'_{new} ; it assesses S's ability to tell old from new at the level of pairings (as does d'_{new}) but without regard to typography. [A corresponding estimate of S's sensitivity to the graphemic component independent of a semantic component cannot, obviously, be obtained from this experiment, but has been attempted in another (Kolers, in press).]

Results and Discussion

Overall performance in segregating new sentences from old was .83 hits and .14 false alarms, average $d' = 2.21$. Further analysis was carried out on correct segregation of sentences according to their pairing in order to evaluate the components of the overall effect.

(1) Semantic encoding. The hit rates for d'_{new} and d'_{old} are the same; the two calculations of d' differ only by virtue of the false alarm rate used, as explained under *Analysis*. The hit rates were: NN = .49, NI = .56, II = .79, and IN = .91. The corresponding values of d' are shown in Table 1. Analysis of variance for d'_{new} yielded $F(3,30) = 10.3$, $p < .01$, with neither sex nor any of the interactions being significant. The corresponding analysis for d'_{old} yielded $F(3,30) = 18.8$, $p < .01$. A third ANOVA was carried out on the difference scores between d'_{old} and d'_{new} because, as Table 1 shows, the values of d'_{new} are everywhere greater than the values of d'_{old} . This ANOVA yielded a nonsignificant F , showing that the two distributions of d' scores are similar as to pattern; but the mean of the difference scores is significantly greater than zero [$F(1,30) = 16.6$, $p < .01$]. Hence, the four pairings NN, II, NI, and IN differ significantly among themselves irrespective of which false alarm rate is used, although values of d'_{new} are significantly greater than values of d'_{old} .

A more lax criterion of hits can be used. Let the sentences be assessed only on the basis of a semantic criterion without regard to typography. The resulting values of d' (labeled d'_{sem} in Table 1) yielded $F(3,30) = 24.7$, $p < .01$, indicating that even when a loose semantic-only criterion is used, appearance of the sentence (as given by its typography) affects the results. The Ss remember more than the semantic content of the sentences they have read.

(2) Pictorial encoding. The four pairings were examined with respect to whether the sentences of Pass 2 appeared in the same typography as in Pass 1

Table 1
Values of d' for Different False Alarms

	Pairing				SEM*
	NN	NI	IN	II	
d'_{new}	1.85	1.58	3.03	2.69	.21
d'_{old}	1.58	1.18	2.60	2.17	.14
d'_{sem}	1.79	1.88	3.34	3.16	.17

*Standard error of the mean

(pictorially identical) or in the alternate typography (pictorially different). Evaluating same form (NN + II) against different form (NI + IN) yielded nonsignificant values of F for d'_{new} , d'_{old} , and d'_{sem} . Hence, pictorial duplication of the sentences plays no significant role in the results.

(3) Initial encoding. Comparing (II + IN) against (NN + NI) evaluates the role of initial appearance of the sentences. This comparison yielded $F(1,30) = 28.7$ for d'_{new} and $F(1,30) = 48.1$ for d'_{old} , $p < .01$ for both tests.

(4) Final encoding. The role of appearance of the sentence in Pass 2 was evaluated similarly by comparing (NN + IN) against (II + NI). This comparison yielded $F(1,30) = 2.04$ for d'_{new} and $F(1,30) = 8.20$ for d'_{old} ; the first of these is not significant, but the second has $p < .01$. The initial encoding accounts for more than 90% of the conditions sum of squares in both calculations; the final encoding accounts for a few additional percent.

(5) Correlation. The rank-order correlation between d'_{old} and d'_{new} was calculated for each of the four pairings. Three of the four values of ρ ranged from $\rho = .14$ to $\rho = .45$, $p > .05$; the fourth (for II) was $\rho = .65$, $.05 > p > .01$. As the hit rates are the same for the two calculations of d' , the nonsignificance of the correlations is due to variations in the false alarm rate. The nonsignificant correlations imply that sensitivity to semantic and graphemic characteristics is mediated by different perceptual properties or cognitive skills: The Ss sensitive to graphemic variables are not necessarily similarly sensitive to semantic variables, and vice versa.

(6) Adjusted scores. All of the sentences that S misread either in Pass 1 or Pass 2 were eliminated from the scoring. The maximum effect of this correction was to change the value of d' by 6% in one pairing; all others were less. All of the statistical tests reported above were carried through again on the adjusted scores. No change in the significance levels or in the pattern of results was found.

(7) Reading time. Time taken to read the sentences varied greatly for normal and inverted typographies. The average times are shown in Table 2, misread and reread sentences eliminated. The data are organized according to pairing as a matter of convenience. (They could be presented as the average of N and I on the two passes.) Learning effects in reading I are distinguishable (Pass 1 I vs WI), but the main fact is the sizable difference

Table 2
Average Reading Time for Sentences According to Orientation (Minutes)

Pass	Pairing				WN	WI
	NN	NI	II	IN		
1	.15	.14	.89	1.00		
2	.16	.59	.55	.17	.15	.64

Table 3
Values of d' for Reversed Sentences

	Pairing				SEM*
	NN	NI	IN	II	
d'_{new}	1.52	1.38	2.27	2.90	.21
d'_{old}	.27	.82	1.46	1.86	.16
d'_{sem}	1.61	1.66	3.21	3.30	.21

*Standard error of the mean

between I and N. It is shown below that these differences are not by themselves the basis of the difference on the classification test (d' scores).

The results show, therefore, that the initial appearance of a sentence is preserved in memory as an encoded feature, but not as a pictorially encoded feature. That is, S remembers something related to the appearance of the words of the sentence, but not the appearance itself as a pictorial representation or image. This information is preserved in memory not quite as well as semantic information ($d'_{\text{new}} > d'_{\text{old}}$), but very well nevertheless.

The data of Table 1 reveal an interaction between encodings on Pass 1 and Pass 2. Recognition of a sentence that first appeared in normal type is better when it reappeared in normal type than when it reappeared inverted (i.e., $NN > NI$), but recognition of a sentence that first appeared inverted is better when it reappeared in normal type (i.e., $IN > II$). Moreover, in all cases S read aloud and took substantially longer to read inverted sentences than to read normally oriented sentences. Two control tests studied first whether the interaction between typography and performance was due to some hidden feature of the sentences themselves that affected their salience or recognizability and second whether the overall finding depended upon reading aloud. A third control experiment evaluates the significance of the difference in reading time on recognition.

EXPERIMENT II

REVERSAL OF SENTENCES

Method

Everything remained the same except that the sentences forming IN and NI were exchanged. By placing the appropriate sentences from the Pass 2 deck in the Pass 1 deck and the corresponding sentences of the Pass 1 deck in the Pass 2 deck, IN became NI and NI became IN. Six new Ss were tested.

Results

Overall performance was poorer (more false alarms), but the same pattern of results and the same significance levels were found. For example, analysis of d'_{new} yielded $F(3,12) = 6.9$, analysis of d'_{old} yielded $F(3,12) = 19.7$, and analysis of d'_{sem} yielded $F(3,12) = 18.9$,

$p < .01$ in all cases. The average values are shown in Table 3. Hence, the results of Experiment I are not due to characteristics of the particular sentences used.

EXPERIMENT III

SILENT READING

Method

The sentences as used for Experiment I were used again in this experiment. All conditions remained the same as before, except that in this case S read silently. As a consequence, E was unable to monitor S's reading as to its accuracy or special examination of parts. Twelve freshmen were tested, half men.

Results

The pattern of results was again the same. The main effect of conditions (pairings) was $F(3,30) = 16.3$ for d'_{new} and $F(3,30) = 7.44$ for d'_{old} , $p < .01$ for both tests. Eliminating typography as a variable of concern (d'_{sem}) yielded $F(3,30) = 21.7$, $p < .01$. The obtained values of d' are shown in Table 4. The difference found in Experiment I between NN and NI is preserved in Experiment III, but the difference between II and IN is not. When the d' values are assessed pairwise, the obtained values are $F(1,30) = 54.1$ for initial encoding ($NN + NI$ vs $II + IN$), $F(1,30) = 10.0$ for final encoding ($NN + IN$ vs $II + NI$), both of which are significant at $p < .01$, and $F(1,30) = 1.21$, $p > .05$ for the "pictorial" encoding ($NN + II$ vs $NI + IN$).

EXPERIMENT IV

DISTINCTION BETWEEN GRAPHEMIC AND SEMANTIC ENCODING

The marks on this page are elements in a symbol system that skilled readers have learned how to interpret. Concatenated into words, the letters have "psychological reality" at the semantic level; but they are also, still, marks on the page. When they are read by a skilled reader, can their encoding as marks be distinguished from their encoding as words? In other words, can we distinguish visual analysis and encoding at the level of graphemes from analysis and encoding at the level of comprehension of the message embodied in the graphemes?

Table 4
Values of d' Obtained for Silent Reading

	Pairing				SEM*
	NN	NI	IN	II	
d'_{new}	1.91	1.21	2.66	2.51	.16
d'_{old}	1.64	1.12	2.12	2.14	.18
d'_{sem}	1.82	1.18	2.76	2.45	.14

*Standard error of the mean

Fig. 3. Examples of the eight transformations. The asterisks show where to begin reading.

This question is an important control on the interpretation of the results of Experiment I. One might think that those results were obtained because, in coping with the difficult typography, S necessarily formed a better internal representation of the semantic features of the sentences. It is necessary to ask, therefore, whether variations in difficulty of encoding at the graphemic level, and variations in time so spent, necessarily improve encoding at the semantic level. Here two tests are described whose outcome shows that visual analysis of difficult graphemes can be carried forth without necessarily enhancing interpretation of the semantic features of the text being read.

Method

Materials

Eight passages from Miller (1962), each about 300 words long, were transformed into each of the eight arrangements illustrated in Fig. 3, making a total of 64 test pages. These were encased individually in clear plastic envelopes, an arrow indicating on each page the direction in which it was to be read. These 64 pages were the reading materials.

To test "comprehension" of the materials, 20 questions were made up for each passage, only half of which could be answered on the basis of the contents of the passage. The S's task was to mark with a plus (+) those questions that he thought were answerable on the basis of the passage and with a minus (-) those he thought were not. This kind of comprehension test is designed to assess whether information regarding the general sense of the passage was acquired rather than to test S's

short-term memory for details (Kolers, 1966). In order to eliminate the problem of how to test for visually presented material without biasing the results with visual presentations, the questions were delivered by means of a tape recording to two groups of Ss (auditory groups). For a third group, however, the questions were presented in print, in normal orientation (visual group).

For purposes of analysis, correct identifications were taken as "hits," appropriate misidentifications as "false alarms," and the resulting values of d' were subjected to analysis of variance.

As the questions of many "comprehension tests" actually provide information about the materials being read (Kolers, 1968a), the 20 questions for each passage used here were pretested by finding how many could be identified correctly by Ss who did not read the passage at all. After suitable modifications of the original battery were made, fresh Ss scored at little better than the chance level. In all cases, key words that Ss might have used as clues occurred about equally often among right and wrong answers. Thus, in order to answer correctly at better than the chance level, S would have had to retain some sense of the semantic content of the passage; memory of key words alone would not suffice. For example, of the following six sentences taken from a test, three are answerable on the basis of the passage read and three are not. The reader can try them out. (1) Did they discover anything when they examined the symptoms? (2) Was there an interruption in the medical studies? (3) Was there a spiritual crisis any time in his life? (4) Did they know that his ailments were not organic? (5) How did the situation abroad compare with the one at home? (6) Did the psychologists maintain morale?

Pretesting of the comprehension test was carried out by allowing Ss about 4 min to study and mark all 20 questions of a test, presented in written form. In the experiment proper, however, the questions were presented singly to the auditory group (thus hampering comparison); a 2-sec pause intervened

Table 5
Average Reading Time (Minutes) for Three Groups of Ss

Ss		Transformation								SE _M [†]
		N	R	I	M	rN	rR	rl	rM	
Auditory	Square 1*	1.58	9.69	15.87	15.04	9.44	17.78	13.31	11.04	1.58
	Square 2	1.65	5.04	11.90	10.62	6.57	14.34	10.40	8.19	1.35
Visual		1.61	5.23	11.76	11.14	7.46	11.70	8.80	7.56	1.18

*Highly inflated by scores from one S.

[†]Standard error of the mean.

between the end of one and the beginning of the next, during which S gave his answer. The visual group was allowed 1 min to read and answer its test pages. Performance at better than chance levels again required that S retain some sense of the semantic component of the passage read. (Of the six questions illustrated, Nos. 2, 3, and 5 are the right answers.)

Procedure

Testing was carried out in the order dictated by a 8 by 8 Latin square constructed so that every transformation type (Fig. 3) preceded and followed every other one exactly once (Bradley, 1958). Therefore, each S read one passage in each of the eight transformations illustrated, the transformation imposed on a passage being different for each of the eight Ss. Immediately after reading a passage, the auditory group listened to its series of 20 questions and indicated after each question whether it was answerable on the basis of the passage just read. The Ss in the visual group read the same passages, but then saw, in normal orientation, all 20 questions for each passage.

Prior to undertaking these tests, Ss read aloud as rapidly as they could one passage of 310 words in normal orientation. If they were unable to do so in less than 2 min, they were disqualified from further testing. The Ss who did qualify then read two 100-word passages (in normal orientation) and took a 15-item comprehension test for each as practice. Time taken to read each of the eight test passages was measured with a stopwatch. The Ss were tested individually.

Subjects

Sixteen Ss participated in the auditory experiment (one replication of the 8 by 8 square), and another eight participated in the visual experiment. All were right-dominant male undergraduates, all but one at the University of Toronto, and were paid for their participation. Because of the presence of an exceptional S in one square of the auditory group, the data for the experiment are presented as three separate Latin squares.

Results and Discussion

As expected, time scores differed radically among the eight transformations, consistent with data acquired previously (Kolers, 1968b). The average time to read each of the eight transformations is shown in Table 5 separately for the three groups of eight Ss. Analysis of variance was carried out on the time scores, yielding significant F ratios in all cases. The greatest part of the difference between the values of Square 1 and Square 2 (auditory group) is attributable to one S in Square 1 who met our criterion of reading N in less than 2 min but took between two and four times his group's average to read the transformations: for example, 32.75 min for R, 38.01 min for I, and the like. Eliminating his scores

brings the average of the remaining seven Ss into line with the times measured in the other two groups; but even with his scores retained, the two squares do not differ significantly, for the pattern of performance is the same despite large individual differences: $F(7,98) = 20.6$ for the difference between transformations, $p < .01$. Values of F for squares and for order of tests were not significant.

A measure of reliability of performance can be obtained in another way, by calculating coefficients of concordance (Kendall's W) for reading times. The eight transformations were ranked according to the speed with which they were read by the eight Ss of a square: $W = .69$ and $W = .73$ for the two squares of the auditory group, and $W = .70$ for the visual group, $p < .01$ in all cases. Thus, Ss agree markedly on the order of difficulty of the transformations as expressed by the amount of time required to read them.

Average values of d' are shown in Table 6 separately for the two groups of Ss who heard the questions (auditory group) and the group of Ss who read them (visual). ANOVA performed on these scores yielded $F(7,42) = 1.36$ for Square 1, $F(7,42) = 0.58$ for Square 2, and $F(7,42) = 1.13$ for the visual group, $p > .05$ in all three cases. Thus, no difference is found in performance on the comprehension tests despite large differences in reading time. This is the case both for tests whose questions were presented by ear and for tests whose questions Ss read.

The lack of relation between time taken to read and performance on the comprehension tests is shown in another way by the correlation of these scores. The 16 Ss of the auditory group were ranked in respect to performance on the comprehension tests; that is to say, their values of d' were ranked individually and the coefficient of concordance was computed for the resulting 16 by 8 matrix. The value found was $W = .07$, $p > .05$. The matter was assessed further by finding for each of the 16 Ss his average time to read the transformations and his average performance on the comprehension tests and correlating these (Spearman's rho). The result yielded $\rho = -.08$, $p > .05$.

The results reveal therefore that graphemic encoding can profitably be distinguished from semantic encoding and that variations in one can leave the other unaffected. Typographic complications severely retard S's ability to read text but do not necessarily affect his

Table 6
Values of d' in the Comprehension Test

Ss		Transformation								SE_M^*
		N	R	I	M	rN	rR	rI	rM	
Auditory	Square 1	2.73	2.95	2.04	2.26	3.36	2.33	2.52	2.71	.36
	Square 2	2.89	3.11	2.31	2.60	2.48	2.84	2.42	2.56	.36
Visual		2.97	3.00	2.45	3.18	3.54	2.71	2.68	3.34	.34

*Standard error of the mean

comprehension of it. Typographic complications thus are able to create problems for S in a way that requires extensive visual analysis; his ability to read the transformed text shows that he has solved the problems, but doing so does not enhance his understanding of the material he has read. Hence, the graphemic and the semantic can be examined and conceptualized as separate levels of functioning: Ss can operate on the graphemic patterns as patterns, and doing so does not involve any special appreciation of the patterns as words.

This result demonstrates, therefore, that variations in reading time such as have been found in other tests do not necessarily induce a better semantic representation of the text read. Notice, however, that this is *not* a demonstration of serial ordering (graphemic analysis prior to semantic analysis). Indeed, such a conceptualization may be meaningless in respect to perception and reading (Kolers, 1972). It is rather a demonstration of relative independence, or dissociability of function, and says nothing about ordering.

It has been shown previously that S can perceive the letters of a familiar word without having a perception of its "wordness" or semantic content (Kolers & Katzman, 1966). In the present experiment, that finding has been extended to show that S's efforts to identify unfamiliar graphemes as words do not necessarily affect his perception at the semantic level of the words or of the sentences they form.

GENERAL DISCUSSION

If after reading a sentence S remembered its semantic content primarily, without regard to surface features such as typeface, size, and the like, then all of the combinations of sentences used in Experiment I would have yielded equivalent levels of recognition. The reason is that all of the sentences are formally similar as semantic or even semantic and syntactic items. This result was not found; rather, highly significant differences in recognition of sentences were found as a function of their initial and final appearance. These differences were preserved in tests that examined S's ability to tell old sentences from new ones (d'_{new} and d'_{sem}) and to tell old sentences apart from each other (d'_{old}). How should these results be explained?

Time Differences

One explanation could be based on differences in

time. Inverted sentences take considerably longer to read than do normally oriented sentences. The corresponding "explanation" therefore is that S, taking more time to puzzle out a sentence in I, actually is studying its content; and a more studied sentence will be remembered better than a less studied one. (Pictorial memory of the sentence could also be enhanced by extended study, as discussed below.) Alternatively, one might say that S has to work harder to extract the semantic content of what he has read in an inverted sentence than in a normal one. Having worked harder to attain its semantic representation, S carries that representation away in a stronger and more certain way. Both of these plausible hypotheses are probably incorrect, as the following shows.

In another study (Kolers, *in press*), it was found that increasing the amount of time for which S read a passage in a foreign language does not by itself increase his memory of its graphemic characteristics. Ss in a control group who spent two to three times more time reading a passage in a language they did not know than did other Ss did not thereby recall its graphemic characteristics any better. Hence, time by itself does not enhance encoding at the graphemic level. On the other hand, bilingual Ss who read paragraphs made up of sentences that were sometimes in English, sometimes in French, and sometimes in a mixture of the two remembered with better than chance success the form of a sentence's appearance. In this case, differences in time to read were trivial. Thus, surface features of sentences were remembered despite virtual equivalence in time of their reading.

Experiment IV, moreover, was undertaken as a control on the very question of time and semantic encoding; there it is shown that different transformations can require substantially different amounts of time to be read and yet are not necessarily comprehended any better for it. The geometric transformations are therefore shown to exert their influence at the level of the patterns as graphemes, and S's difficulties in analyzing them at that level do not necessarily affect his semantic interpretation of the text. In two experiments, therefore, we show that large variations in time required to process sentences at the graphemic level do not necessarily create a corresponding improvement in memory, while near-equal amounts of time spent reading leave S sensitive to the appearance of sentences. Although time spent reading must be important at the limits of

performance—other things remaining equal *S* will do more poorly with a long passage exposed for 1 sec than with a long passage exposed for 1 min—variations in reading time do not by themselves account for the results of Experiments I-III. An alternative must be sought.

Pictures and Images

If memory for sentences in these experiments is not primarily of their semantic content, what is its source? Perhaps *S* remembers the semantic content of the sentence plus something else, and the "something else" are the pictorial features of the sentences or images of them. Or perhaps *S* remembers the pictorial features primarily. If the pictorial hypothesis were correct, in either of the two forms just expressed, performance on like-paired sentences would be better than performance on unlike-paired sentences: (NN + II) must be recognized better than (NI + IN). This result was not found, however. Thus, we disconfirm the temporal, the purely semantic, and the pictorial explanations of the data.

A Hypothesis

An alternative invited by the results considers the activities involved in reading an unfamiliar typography. An inverted sentence creates more problems for *S* than a normal sentence; it requires more analysis and more conjecture than the familiar typography does. The large influence on performance created by initial encoding of the sentences suggests that *S* may remember the sentences not necessarily by virtue of remembering the words themselves, but by remembering the operations he performed to acquire or encode the words. Differences in encodability of familiar and unfamiliar graphemes could affect subsequent recognition in that the greater the number of operations *S* had to perform to acquire a sentence, the greater the number of events he would have available as aids to memory at the time of recognition. The encoding activities themselves would provide the basis for the observed difference in performance on the various pairings.

In this conjecture, variations in time spent on the task reflect the difference in amount of work or encoding activity required by a transformed sentence. In other words, a difficult set of graphemes requires more work of analysis, and the additional work required takes more time. Time on a task does not enhance performance directly; rather, unfamiliar combinations of activities take more time to organize and run off and, in this experiment, are remembered better than familiar ones. This additional time spent on graphemic analysis does not necessarily enhance understanding or interpretation at the semantic level.

Meaning and Analysis

Do the coding operations that *S* remembers actually carry the semantic relations of the sentence? This question arises naturally from the foregoing. One could argue this view by asserting that the meaning of a sentence is obtained from an analysis of the relations of its constituent words, so that remembering the analysis automatically recovers the meaning. Attractive as this view is, it is not wholly supported by the data. If only the operations were remembered, then sentences inverted on both Pass 1 and Pass 2 (II) would be recognized better than sentences inverted on Pass 1 and normal on Pass 2 (IN), for the operations are identical in the first case but not in the second. Unfortunately for this argument, the reverse is found in one experiment (Table 1), although not in two others (Tables 3 and 4). It is possible, of course, that it is only the operations themselves that are remembered and that the difference between IN and II reflects differences in skill requirements that transformations of text elicit (Kolers & Perkins, 1969). It is also possible that *S* remembers encoding operations *and* their result, the semantic items that he encoded. Thus, it is not yet clear how best to conceptualize this particular ambiguity. The main demonstration of the role of encoding activity in subsequent recognition is, of course, unaffected by this subsidiary question.

Other Studies

These results are of the same import as those on modality effects in recognition. The modality by which information is acquired (Margrain, 1967; Murdock & Walker, 1969) and even the language of its appearance (Kolers, 1963) are on occasion retrievable from memory. If the nervous system does different work on information coming in through the ears and through the eyes, and even, perhaps, on information coming in via one language or another, then these differences in operations (the present results suggest) are enough to explain the differences obtained in those data.

A final consideration is the realization that these results argue against a theory of memory in terms of rapidly decaying icons that preserve all of the features of a stimulus for a short interval. The time between reading and test in the present experiments averaged about 45 min, more than enough for an icon to decay; nevertheless, form features of the sentences that are altogether irrelevant to their semantic encoding were preserved in memory remarkably well. Their preservation may be thought to reflect the activities and operations that went into their encoding, the work *S* engaged in to acquire the sentences, rather than time-based aspects of memory.

REFERENCES

- Bradley, J. V. Complete counterbalancing of immediate sequential effects in a latin square design. *Journal of the American Statistical Association*, 1958, 53, 525-528.
- Kolers, P. A. Interlingual word associations. *Journal of Verbal Learning & Verbal Behavior*, 1963, 2, 291-300.
- Kolers, P. A. Reading and talking bilingually. *American Journal of Psychology*, 1966, 79, 357-376.
- Kolers, P. A. Reading temporally and spatially transformed text. In K. S. Goodman (Ed.), *The psycholinguistic nature of the reading process*. Detroit: Wayne State University Press, 1968a.
- Kolers, P. A. The recognition of geometrically transformed text. *Perception & Psychophysics*, 1968b, 3, 57-64.
- Kolers, P. A. Some problems of classification. In J. F. Kavanagh and I. G. Mattingly (Eds.), *Language by ear and by eye*. Cambridge, Mass: M.I.T. Press, 1972.
- Kolers, P. A. Remembering trivia. *Language & Speech*, in press.
- Kolers, P. A., & Katzman, M. T. Naming sequentially presented letters and words. *Language & Speech*, 1966, 9, 84-95.
- Kolers, P. A., & Perkins, D. N. Orientation of letters and errors in their recognition. *Perception & Psychophysics*, 1969, 5, 265-269.
- Margrain, S. A. Short-term memory as a function of input modality. *Quarterly Journal of Experimental Psychology*, 1967, 19, 109-114.
- Miller, G. A. *Psychology, the science of mental life*. New York: Harper & Row, 1962.
- Murdock, B. B., & Walker, K. D. Modality effects in free recall. *Journal of Verbal Learning & Verbal Behavior*, 1969, 8, 665-676.

(Received for publication December 18, 1972;
revision received March 1, 1973.)