COMPREHENSION AND RECALL OF SENTENCES

SAMUEL A. BOBROW and GORDON H. BOWER
Stanford University

Preliminary experiments showed that Ss better recall a noun pair if they generate their own linking sentence for the pair than if they merely read an equivalent linking sentence. Initial attempts to explain this effect in terms of memory search activities or idiosyncratically high-associative mediators proved unproductive in later experiments reported here. The hypothesis was then offered that the generate vs. read conditions differ in comprehension of the sentences and that comprehension aids retention. Subsequent experiments on incidental learning showed that recall is excellent when S is set to process a sentence in different ways designed to promote comprehension of its meaning, whereas equivalent exposure to or mouthing of the words in control sentences without comprehension produces relatively little recall.

The following experiments are concerned with the facilitation of paired-associate learning produced by embedding each word pair in a sentence. Rohwer (1966) found that an S who hears a linking sentence such as "The COW chased the BALL" will recall the COW-BALL pair better than a control S who simply studied the pair without a sentence context. In repeating some of Rohwer's paradigms, another phenomenon was uncovered which led into the present experimental series. The phenomenon is that Ss better remembered noun pairs embedded in sentences they generated than they did pairs embedded in sentences E gave them. At the time of study or input, Ss in the read condition read aloud a presented sentence (e.g., The COW chased the BALL), whereas those in the generation condition saw the pair COW-BALL and had to make up and say aloud a linking sentence. Although input times were controlled, later recall (of BALL when cued with COW) was about 25-30% higher in the generation condition. This result is quite reliable, having been replicated several times in the experiments reported subsequently.

Why does generating a linking sentence facilitate noun-noun recall more than does reading a similar linking sentence? Two hypotheses were considered initially. The first hypothesis presupposes that the second (object) noun, N2, can be recalled either directly to the cue of the first (subject) noun, N1, or failing that, N2 may occur via a mediate association from N1 to the verb connective, V, and thence from V to N2. The diagram in Fig. 1 illustrates the general idea, with associations labeled a, b, c. Interpreting the associative connections as probabilities, the probability of N2 recall to an N1 cue is c + (1 - c)ab, the two terms representing the direct link and the mediate association route, respectively.

With this as background, the first hypothesis to explain the advantage of self-generated sentences (which were mostly of the N1-V-N2 sort) is that S selects an idiosyncratically high-associative mediator, V, one which is his personal maximum in terms of associating N1 and N2. On this view, S recalls N2 to N1 better because his selected V on the average has higher a and b associations than does the V which E selected for him.

The second hypothesis supposes that the generation advantage is in some way related to the cognitive activity engendered by such pairs at input, rather than to the idiosyncratic result (V) of that activity. Several possibilities may be considered here: One is that the act of successfully searching for a sensible connective to link N1 to N2 parallels or is equivalent to the process of constructing
order to test an associative mediation account of the results of Exp. Ia.

**Experiment I**

**Method**

*Experiment Ia.*—The aim was to compare cued recall of \( N_1-N_2 \) pairs input when one of three activities was required of \( S \): (a) when presented with a pair, to make up and say a linking sentence—the generate condition; (b) when presented with a linking sentence, to read it aloud and study it—the read condition; or (c) when presented with a \( N_1-\#-N_2 \) triple, to retrieve the verb associated with the number and say a sentence using that verb to link \( N_1 \) and \( N_2 \)—the number condition. There were 15 items of each type, randomized in a mixed list and presented once each for 7 sec. on a memory drum. Instructions informed \( S \)s of the activity required for the three types of presentations and the nature of the impending recall test. After one study trial on the list, recall of \( N_2 \) to the \( N_1 \) cues was tested at a 7-sec. rate without feedback, with items tested in the same order as they had been presented. The 45 pairs were unrelated concrete nouns, and the sentences presented were sensible declaratives of the following form: article, noun, verb, article, noun; with the nouns in all capital letters. The assignment of particular noun pairs to the three presentation conditions was counterbalanced over \( S \)s.

The \( S \)s first learned a list of 15 number–verb pairs. The cues were the numbers 1–15, and the response terms were the appropriate verbs that would be used in the \( N_1-\#-N_2 \) items of \( S \)’s second list. Learning was by the anticipation method at a 3:3 sec. rate to a criterion of one perfect recitation plus three overlearning trials. After learning the number–verb list, \( S \) received instructions and then learned the noun-pair list. The \( S \)s were 24 paid Stanford University students attending summer sessions.

*Experiment Ib.*—In Exp. Ia, the verbs \( S \) used in linking the number items had been prefamiliarized by prior learning, so differential availability of the mediating verbs might have caused differential recall of the number vs. the read items. Experiment Ib replicated the read and number conditions, but used prefamiliarized verbs in both cases. The \( S \)s first learned 16 number–verb pairs as in Exp. Ia. In the noun-pair learning which followed, 8 of these verbs were used in \( N_1-\#-N_2 \) items and 8 appeared as the verbs in read sentences. The list had 24 other filler items, half of them sentences, which were testing auxiliary hypotheses of no concern here. Other procedural details were identical to those of Exp. Ia except for the input list being 40 pairs long. The \( S \)s were 24 paid Stanford University students attending summer sessions.

*Experiment Ic.*—This study was done to test the idea that the number items of Exp. Ib benefited in part because the \( N_1-\#-N_2 \) complex at input made...
possible some learning of Ni-#-Na associations, and these would supplement the sentence associative structures of Fig. 1 to aid recall of Ns to Ni. To this end, the Ni-#-Ns triples of Exp. Ia were replaced by Ni-Verb-Ns triples. For such items, S had to say the number associated with the verb, then say a sentence using that verb to link Ni and Ns. In terms of the associative mediation model, this verb-number condition should be similar to the prior number condition since the number is being emitted in the context of stimulation by Ni and Ns; hence, some Ni-#-Ns associations should be formed to aid recall of Ns to the Ni cue. The procedure in Exp. Ic was identical to that of Exp. Ia except for the aforementioned change in items. The Ss were nine paid university students.

Results and Discussion

Experiment Ia.—Pooling data over the 24 Ss X 15 items in each condition, the mean percentages of correct recall were read, 29%; number, 39%; generate, 58%. These means differ significantly, $F (2, 46) = 20.9, p < .001$. Newman-Keuls gap tests detect significant differences in all pairwise comparisons. For the number condition, there were 9% failures to recall the verb to the number during the time Ni-#-N2 was presented, so in these cases, S failed to come up with a linking sentence. Recall in these cases was only 13%, considerably poorer than for the other number items for which S came up with the linking sentence.

Since the generate items for some Ss were the read items for other Ss, those cases were considered in which S generated a sentence identical to that which other Ss read. There were 40 such coincidences in the generate condition, which yielded a recall of 68% compared to 26% for the same items by read Ss ($p < .001$). Thus, recall of generate items exceeded that of read items even when the sentence generated was the same as that read.

The main result of Exp. Ia is that recall of number items was intermediate between the generate and read items. A search for differential types of errors in the three conditions was uninformative since the overwhelming proportion (86%) of failures were omissions.

Experiment Ib.—This experiment controlled for prefamiliarization of the verbs used in the read and number condition on the grounds that differential availability of the mediating verb may have produced the number vs. read recall difference in Exp. Ia. The results discount this argument. In Exp. Ib, Ss recalled 58% correct with the number items vs. 35% with the read items, $t (19) = 4.28, p < .001$. Since the linking verbs were equally familiarized for these two cases, the recall advantage of number over read items cannot be explained in terms of differential verb familiarity.

Experiment Ic.—Mean correct recall percentages were generate, 55%; read, 37%; verb-number, 13%. These percentages differ reliably, $F (2, 16) = 24.4, p < .001$, and Newman-Keuls tests showed that all pairwise differences are significant. The generate vs. read difference in recall is replicated; more significantly, the verb-number items are markedly inferior to the read items. Thus, having to recall the number to the verb cue before saying a linking sentence retards learning, whereas recalling the verb to the number cue before saying the same linking sentence facilitates learning relative to read items (cf. Exp. Ia and Ib).

The poor recall of the verb-number items argues against any hypothesis regarding "active memory search" per se as the explanation for the generate-read difference in recall. The result shows that the mere act of searching for something in memory at the time of input of a noun pair is not the beneficial factor. Rather it appears that the memory search has to be relevant to constructing a relational bridge between the two nouns.

The association model of Fig. 1 is also informed by the verb-number result since that model expected about equal facilitation of recall by the number and the verb-number conditions. Further evidence against the association model of Fig. 1 was reported by Rohwer and Lynch (1967). Their Ss learned a list of eight Ni-N2 pairs each embedded in a sentence where, for different Ss, either two, four, or eight unique verbs were used to link the eight noun pairs. The mediation model of Fig. 1 would suppose that the V term would become associated to four, two, or one different N2 terms, respectively, in the three cases; consequently, correct N2 recall to Ni should increase the fewer the sentences in which a given V occurs. How-
ever, Rohwer and Lynch found no recall differences among the three groups, although all of them exceeded recall by control Ss who studied the word pairs without sentence contexts.

**Experiment II**

Having become disenchanted with the associative mediation hypothesis (which appears wrong) and the memory search hypothesis (which is simply too vague), the authors came on another hypothesis which appeared most plausible. The idea is simply that recall is facilitated by S "comprehending the meaning" of the sentence, and this occurs more reliably when one has to make up a linking sentence than when one merely reads off a linking sentence that occurs in a long list of other sentences and pairs. Assuredly, the notion of comprehending the meaning of a sentence is one of those grandly vague concepts about which there is much controversy in linguistics and analytic philosophy. Two particular views of comprehension suggested the two experiments which follow. One, due to Katz and Fodor (1964), supposes that comprehension is coordinate with providing a semantic interpretation for the sentence so that, e.g., the sense intended for ambiguous words can be specified. The second view, from a paper by Kochen, MacKay, Maron, Scriven, and Uhr (1967), supposes that comprehension of a sentence predisposes one to alter his cognitions about the items named in the sentence, so that he is prepared to "go beyond" the sentence in asserting plausible implications of it or to answer questions requiring use of the information in the sentence, etc.

The hypothesis is twofold: First, comprehension of a sentence aids its retention; and second, with respect to the generate-read difference, discovering a plausible relation and generating a sentence linking N1 to N2 assures better or more reliable comprehension than does simply reading a sentence. The first part of this hypothesis is tested in Exp. II. Regarding the second part, only plausible arguments can be marshalled alongside the generate–read recall difference. For example, it is a common experience that one can read (or listen) without comprehension, whereas it is infrequent that one would generate sentences without understanding what he is saying.

The aim of Exp. II was to induce Ss to process (presented) sentences in different ways, one way designed to promote comprehension of its meaning, the other way not, and then to test later for recall of the key nouns. An incidental learning procedure was used to reduce deliberate attempts at memorization.

**Method**

**Experiment IIa.**—Eight undergraduate Ss were assigned to the disambiguation condition and eight Ss to the spelling condition. The material was 30 declarative sentences of the following form: article, concrete noun, verb, article, adjective, concrete noun. All sentences contained an ambiguous word; the ambiguous word was equally often the subject, verb, and object. For the spelling condition, in half of the sentences, one word (N1, V, or N2) was obviously misspelled (e.g., balll).

Each sentence was presented by a memory drum for 5 sec. and S read it aloud. The drum then turned to a query about the sentence just read. In the disambiguation condition, the query consisted of presenting the ambiguous word and two very brief (1–2 words) identifications of its different meanings. The S was to select and say aloud the intended meaning of the word as it had been used in the prior sentence. For example, if the sentence were "The cow chased the rubber ball," he would be queried whether "ball" in that context referred to a society dance or a small spherical object. (No S made any "mistakes" on such queries.) In the spelling condition, the query consisted of the single ambiguous word (correctly spelled), and S indicated whether or not this word had been misspelled in the prior sentence. For a quarter of the sentences, the correct answer to this query was "Yes" and for the rest "No." In both conditions, Ss were permitted 5 sec. to read and answer the query, then the memory drum turned to the next sentence.

The Ss were misinformed about the purpose of the experiment: they were told that E was investigating how easily Ss could disambiguate (or detect spelling errors in) words depending on subtle syntactical differences in the sentences.

After the 30 sentence-query pairs had been presented once, S was informed that sentence retention would be tested by cueing with N1 (subject noun) for recall of N2 (object). Items were tested in their order of input at a 5-sec. rate without feedback information. Questioning of Ss at the end of the test turned up none who admitted to expecting a retention test of the input material.
COMPREHENSION AND SENTENCE RECALL

Experiment IIB.—This was similar to Exp. IIA and used the same sentences, except input time was 7 sec. per sentence and recall test time was 7 sec. per item. Seven S's (undergraduates) were assigned to the continuation condition and 7 to the repetition condition. In the repetition condition, S first read the presented sentence aloud and then had to make up and say a sentence that was a sensible continuation, elaboration, or implication of the action or state of affairs stated in the sentence. The S's were given several illustrations of how to do this. For example, if the sentence read were "The farmer discovered a diamond," a sensible continuation might be "He sold it to a jeweler and used the money to buy a tractor." This read-then-continue activity was always done within 7 sec., after which time the memory drum turned to the next sentence. The S's in the repetition condition were told that the study concerned speed of speaking; they were instructed to read each sentence aloud three times as rapidly as they could and were led to believe that their response times were being recorded. Following one input trial of the 30 sentences, recall instructions were given and the recall tests followed (recall N2 to N1 cue).

Results

Experiment IIA.—The results strongly supported the comprehension hypothesis. Recall was 49% for the disambiguation S's, but only 18% for the spelling S's, $t$ (14) = 3.45, $p < .001$. Recall in the spelling condition was further analyzed for those 15 sentences which contained a spelling error vs. the 15 sentences which did not. Recall was 14% for sentences containing a spelling error and 23% for those which did not, $t$ (7) = 3.00, $p < .01$. Both recall percentages are significantly below that for the disambiguation S's. Comparing these same sentences for S's in the disambiguation condition, there was no difference in recall.

This ancillary finding, of poorer recall by spelling S's of sentences with misspelled words, admits of several interpretations. A plausible view is that it is a simple repetition effect. It may be presumed that if S fails to find a spelling error on a first reading of a sentence, he reads it again, whereas detection of a spelling error stops S's processing of the sentence. Thus, sentences without errors enjoy the memorial benefits of a second reading.

Recall in the disambiguation condition was analyzed according to whether the queried (ambiguous) word was N1, V, or N2. Differences in recall were relatively small and cannot be taken seriously in any event since the specific word pairs could not be counterbalanced over this variable. Comparing the disambiguation and spelling S's on the types of errors made, the disambiguation S's had a higher rate of either synonymic or associatively intrusive identifications or the correct word at the recall test. The disambiguation and spelling S's on the types of errors made, repetition S's had a much higher proportion of omission errors (92% vs. 55%). About 19% of the errors for continuation S's consisted either of synonyms of N2 or key words generated in his continuation of that sentence. In only 15% of the continuation items did S reuse the N2 word in his continuation, but N2 recall in these cases was only slightly higher than in the remaining continuations which did not use the N2 word.

In summary, Exp. IIA and IIB have shown that a set to process sentences so as to disambiguate words or to continue the action of the sentence produces better incidental recall than does three rapid recitations of the sentence or searching for spelling errors in it. This advantage occurred despite the fact that the queried identifications or the verbal continuations provided a multitude of extraneous words that could have competed with the correct word at the recall test. The incidental learning produced by these 'comprehension' sets is about equivalent to that which has been obtained in read-sentence conditions under comparable circumstances but with intentional learning instructions. In two unpublished experiments involving comparable list length (30), temporal intervals, and similar sentences, the authors have had intentional learning S's read a sentence once (5 sec.), then immediately repeat it from short-term memory (5 sec.). Later long-term recall by these read-repeat S's was 41% in one study and 52% in the other.
TABLE 1

PERCENTAGE OF RECALL OF N₂ TO N₁ CUE FOR
GROUPS T AND NT FOR ITEMS SHOWN
IN DIFFERENT TYPES OF SENTENCES

<table>
<thead>
<tr>
<th>Sentence type</th>
<th>Group NT</th>
<th>Group T</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declarative</td>
<td>.24</td>
<td>.25</td>
<td>.24</td>
</tr>
<tr>
<td>Negative</td>
<td>.21</td>
<td>.17</td>
<td>.19</td>
</tr>
<tr>
<td>Question</td>
<td>.23</td>
<td>.19</td>
<td>.21</td>
</tr>
<tr>
<td>Negative question</td>
<td>.25</td>
<td>.23</td>
<td>.24</td>
</tr>
</tbody>
</table>

These figures are comparable to the 49% and 43% incidental recall of the disambiguation and continuation Ss. These results tend to bolster a central conclusion from the literature on incidental learning (e.g., Mechanic, 1962; Postman, 1964); viz., the important factor for learning is how S processes the material at the time of input, not whether he has an intention to memorize.

EXPERIMENT III

The last experiment to be reported was mainly exploratory: the question was whether S would remember sentences better if he were required to be active in transforming the syntax of sentences he reads at the time of input. The thought was that requiring S to process the syntax of a sentence—to rearrange or delete words or alter the tense of verbs—and come out with a specified syntactical form might produce greater attention to and learning of the key-word associations than would merely reading the sentences without syntactically transforming them.

Method

Two groups of 12 Ss (undergraduates) had three input-output cycles on a list of 28 noun pairs embedded in sentences. There were seven instances of four sentence types: simple declaratives (The COW kicked the BALL), negatives (The COW did not kick the BALL), questions (Did the COW kick the BALL?), and negative questions (Did the COW not kick the BALL?). Instances of the four types were scrambled in blocks of four in the input list. The noun pair exemplifying particular sentence types was varied over Ss. Each sentence was presented for 5 sec. The Ss in Group NT (nontransformers) were instructed to read aloud each presented sentence and study it. Those in Group T (transformers) were instructed to look at the sentence presented; if it was not a simple declarative, then to transform it to the simple declarative form and to say aloud the simple declarative. A quarter of the presented sentences were simple declaratives, and these were simply read aloud by the T Ss. The Ss were thoroughly informed of their task and the nature of the impending recall task (recall N₂ to N₁ cue). Items were tested in the order of their input at a constant lag of 28, but the input order varied over the three trials.

Results

The recall percentages for Trial 1 are shown in Table 1. The most noteworthy thing about the data in Table 1 is that they reveal no differences whatsoever, neither due to sentence type nor to the T-NT treatment. The condition means were similarly equal on Trials 2 and 3, so those data are not presented. Thus, the conclusion from this study is essentially negative: with respect to associations of the key words (subject and object nouns), these sentence types differed very little, nor did it matter whether S read off a sentence as opposed to transforming it before he said it. Possibly, other syntactical forms or requiring different transformations of S (e.g., to passives) may require greater degrees of sentence processing and might improve memory for the transformed material.

Discussion

These experiments began with attempts to explain the generate vs. read difference in recall. This recall difference probably results from more reliable comprehension when sentences are generated than when they are read. Experiments Ila and lib showed that instructional sets which promoted sentence comprehension produced better incidental recall than did instructions for rapid recitation of the sentence or searching for spelling errors in it. Further evidence relevant to the comprehension hypothesis is that noun-pair recall is retarded relative to uninstructed control Ss if at input the key nouns are embedded in a semantically anomalous sentence (e.g., the BOOK sang the tasty WINDOW) for which there is no sensible interpretation. This deleterious effect of anomalous sentences on N₁-N₂ recall has been reported by Rohwer and Levin (1968) and has been replicated in the present laboratory. The absence of an effect on memory of requiring S to transform the syntax of a sentence may
or may not be relevant to the comprehension and memory hypothesis. Its relevance depends on whether one believes that S must first comprehend the meaning of a sentence before he can carry out a proper transformation of its syntax. The declarative transforms used here do not require a prior semantic interpretation since Ss can readily transform nonsense sentences like “Didn't the DAX flumox the WUG?”

Of relevance to the generate vs. read difference in recall, in other work the authors have found a similar (though smaller) recall difference comparing instructions to read a sentence then repeat it verbatim, as opposed to read a sentence then change it to produce a new and different linking sentence. If the read-then-generate instructions severely restrict the search for the new sentence (e.g., the verb of the generated sentence must be synonymous with the verb of the sentence just read), then the read-generate condition is only marginally better at recall than a read-repeat condition. The poor recall produced by such restrictions on the generation process may result from S confining his memory search to verb synonyms rather than understanding the relational linkage between the two key nouns.

The authors are still puzzled by the results of the number condition in their earlier experiments: Why does recalling the linking verb from a number (before saying the linking sentence) facilitate recall relative to just reading the same sentence? The associative mediation account of this effect seems not to be valid (see Exp. Ic). The only explanation suggested by the comprehension hypothesis is that recalling the verb from the number makes S more interested in seeing how the verb fits in with the two nouns, whether it makes a sensible sentence. Possibly Ss monitor the accuracy of their number-verb recall by first checking to see whether the candidate verb recalled produces a comprehensible sentence linking the two nouns. By this means, then, number items would more often lead to comprehension than read items, and hence to better recall. Experimental tests of this hypothesis are clearly needed.

REFERENCES


(Received July 29, 1968)