Recognition of sentences from prose

WILLIAM S. TERRY and HELEN M. MASON
University of North Carolina, Charlotte, North Carolina 28223

College students read 600-word passages and then were tested for recognition of verbatim sentences from the passages in comparison with paraphrased and distractor sentences. Verbatim sentences received higher recognition scores, even after a 20-min delay or with instructions to remember meaning rather than exact wording (Experiments 1 and 2). In a forced-choice test, subjects chose the verbatim sentences over two alternative forms of paraphrases (Experiment 3). Sentences containing more important information in the passages were better recognized, but more false recognition of their paraphrases also occurred.

Recent research has demonstrated substantial recognition memory for the exact wording of sentences from some materials. For example, college students showed significant verbatim memory for classroom material when tested after a lecture heard earlier (Goolkasian, Terry, & Park, 1979; Kintsch & Bates, 1977). These findings contrast with earlier reports of poor verbatim recognition memory for sentences (Bransford, Barclay, & Franks, 1972; Sachs, 1967, 1974). The differing results have been attributed to the use of nonlaboratory (and possibly more "natural") conditions in studies finding good sentence memory (Kintsch & Bates, 1977). Alternatively, the differences may be due to the use of unrelated sentences or brief prose paragraphs (Sachs, 1967), as opposed to the use of longer, connected prose materials (i.e., lectures; Kintsch & Bates, 1977). There is also the possibility that lectures or videotape presentations provide multiple modalities of information (i.e., spoken words, and also body and facial gestures) that make the material more memorable (Bates, Masling, & Kintsch, 1978).

Still other studies have shown varying amounts of verbatim memory, depending upon the depth of processing of the material (Graesser & Mandler, 1975) or the degree of similarity of the original sentences to the paraphrases presented in testing (Goolkasian et al., 1979).

In general, few studies have used written prose materials and found substantial verbatim memory. The purpose of the present research was to assess college students' recognition memory for verbatim sentences from lengthy prose passages. These studies incorporated various features from the studies reviewed above, but they differed in other aspects. The passages were written rather than presented in auditory mode. Connected

prose was used, rather than isolated or unrelated sentences. And the passages were considerably longer than those used in many studies of this type. Although allowing subjects to read the passages at their own pace does provide less experimenter control over certain variables (e.g., exact presentation rate), the past emphasis on auditory presentation with experimenter-controlled rate may have its own disadvantages. There may be individual differences in students' effective rates of attending to information. Also, subjects may process information differentially between auditory and visual presentation modes. Finally, since reading is a major source of information acquisition among college students, understanding their retention of written material may be of educational value.

In the first two experiments, recognition of verbatim sentences was tested as a function of retention interval and the instructions given the subjects. Target sentences were chosen that either were important to the central themes of the passages ("topic" sentences) or were less important ("detail" sentences). In Experiment 3, a different testing procedure was used to further assess the ability to discriminate verbatim sentences from paraphrased test items. The relative importance of the target sentences within the prose passage was also more explicitly varied.

EXPERIMENTS 1 AND 2

Method

Subjects. A total of 33 college students enrolled in courses in general psychology were used. Participation was through a subject pool, which provided extra course credit.

Materials. The two prose passages used were adapted from Meyer's (1975, pp. 195-201) "Anti-S Proteins and Schizophrenia" and "The Color of Parakeets" stories. These passages are reported to be comparable in reading level to *Scientific American* articles. Each passage is approximately 600 words in length, or two double-spaced typed pages.

The recognition tests consisted of a set of 21 sentences for each passage. Seven sentences in each set were verbatim sentences, which were identical reproductions of sentences from the prose passages. The originals were distributed throughout the passages, although none was the first or last sentence of a

passage. From each of the verbatim items, a paraphrase sentence was constructed. The paraphrases variously involved both word substitution and changed word sequences from the original sentences. The paraphrasing is comparable to that described in Kintsch and Bates (1977) and Goolkasian et al. (1979). (Paraphrases can involve changes in meaning, as well as the intended changes in surface form of the sentences. The form of paraphrasing was explicitly manipulated in Experiment 3.) To minimize stylistic differences between the verbatim sentences and our constructed paraphrases, the two versions of a sentence were randomly assigned to be the original or the paraphrased item. Finally, from each verbatim sentence, a distractor was generated. The distractors contained information that was factually untrue with respect to what was given in the passage, although the wording was kept similar to the sentences from which they were derived.

Of the 14 total verbatim sentences selected for testing, 7 were topic sentences and 7 were details (there were 3 or 4 from each passage). The topics were general statments on important points, usually introducing further information in subsequent sentences. The details amplified and elaborated on information in the passages, presenting specific instances or examples. To provide a somewhat more objective basis for determining the "importance" of individual sentences within the passages, Meyer's (1975) system of prose analysis was used. Within each passage, the topic sentences were all higher in the content structure (i.e., were of more central importance) than were the detail sentences.

The set of test sentences for a given passage was divided across three pages in a booklet, with one version of each basic sentence (verbatim, paraphrase, or distractor) placed on each page. Thus, direct comparison of the three versions was prevented.

Procedure. The students were run in groups of 10-12 in a classroom. They were informed that verbatim memory for sentences would be tested (except in Experiment 2). The students were told that, after reading the prose passages, they would be asked to rate sentences based on the two stories, using the following scale: (a) sure the sentence did not occur before (exactly as worded), (b) almost sure the sentence did not occur, (c) unsure if the sentence did or did not occur, (d) almost sure the sentence did occur, and (e) sure the sentence did occur. This scale appeared at the top of each page of test items.

Each subject received a booklet containing (1) a cover page of instructions and an example of the procedure using the "Galileo" paragraph from Sachs (1967), (2) the two prose passages, and (3) the two sets of recognition test sentences. The passages and tests were presented in counterbalanced sequence. The students were asked to read the passages at a normal speed and not to go back and repeat already read sections. The subjects proceeded through reading and testing at their own pace.

In Experiment 1 (n = 20), subjects read the two passages and completed the recognition test form for one passage. After this, they watched a 20-min film on infant development and then completed a test form for the alternate passage.

In Experiment 2 (n = 13), the instructions were changed to state that a true-false test would be given and only the recognition of correct information would be required during the test phase. The practice items were true-false questions, to emphasize the instructional manipulation. After reading the two passages, the verbatim sentence recognition instructions were given, and testing of both passages occurred immediately.

Results

The first analyses to be presented used the combined topic and detail sentences, with the latter breakdown being described subsequently.

Test sentence ratings were converted to a 5-point scale, with 1 being assigned to "sure did not occur before" and 5 to "sure did occur." For each subject, a mean recognition score was derived for all the verbatim, paraphrase, and distractor items, averaging across sentences of each type within an experimental condition. (All statistically significant effects use $p < .05$.)

The results of Experiment 1, in which verbatim memory was assessed either immediately or after a brief retention interval, are shown in Table 1. As can be seen, verbatim items received higher recognition scores than did paraphrase items, which in turn received higher scores than did distractors. However, recognition did not vary across retention interval, as immediate and delayed tests produced nearly identical scores.

An analysis of variance of these results used the two between-groups factors, resulting from the counterbalancing of story read first and story tested first, and the two within-subjects factors of test delay and test item type. The difference among test items was highly significant $[F(2,32) = 61.39]$. The only other significant effect was an interaction among the four factors $[F(2,32) = 5.46]$. This appeared to be due to a smaller (but still significant) verbatim-paraphrase difference on the parakeet story items than on the schizophrenia story items. (This story effect is not readily interpretable, since it could be due to the stories themselves or to our test items on the stories.)

The students' ability to discriminate verbatim from paraphrase items was confirmed in a separate analysis of only these two test-item types. Again, the difference was significant $[F(1,19) = 20.28]$. The above analyses are by subjects. Analysis by items (the 14 verbatim sentences matched with their corresponding paraphrases and distractors) also showed a reliable difference in comparison across all three item types $[F(2,26) = 54.5]$ and, importantly, between verbatim and paraphrase items alone $[F(1,13) = 7.5]$.

The results of Experiment 2, in which the instructions to the subject were changed to emphasize retention of meaning rather than exact wording, are also shown in Table 1. The changed instructions had little effect on the large verbatim-paraphrase difference found in Experiment 1, as subjects readily differentiated the two. Within-subjects analyses showed significant differences

Table 1
Mean Recognition Scores for Test Sentences
in Experiments 1 and 2

Exper-iment		Test Sentence Type		
		V	P	D
1	Immediate Test	3.49	3.09	2.07
	Delayed Test	3.41	3.11	2.11
2	"Meaning" Instructions	3.67	3.20	1.97

Note—V = verbatim, P = paraphrase, D = distractor.

among the three item types [F(2,24) = 47.35] and between verbatim and paraphrase items alone [F(1,12) = 11.5]. Analysis by items also showed significant differences, both across all three item types [F(2,26) = 51.05] and between verbatim and paraphrase items [F(1,13) = 6.56]. There were no substantial differences as a function of the counterbalancing factors, which were therefore excluded from these analyses.

The mean recognition scores for topic and detail sentences from both experiments are shown in Table 2. Since there were no differences across retention intervals in Experiment 1, the data are pooled across the delay conditions. Topic sentences generally received higher recognition scores, both in correct recognition of verbatim sentences and in false recognition of paraphrases and distractors. In both experiments, the topic sentences generated a smaller verbatim-paraphrase difference than did the detail sentences. Data for the verbatim and paraphrase items in the two experiments were pooled into a single analysis. There was a significant topic-detail difference overall [F(1,12) = 5.7] and a marginally significant interaction of this factor with that of the test-item types [F(1,12) = 4.12, p < .10] in the analysis by items.

EXPERIMENT 3

The purpose of Experiment 3 was to test the generality of the verbatim-paraphrase discrimination found in the first two studies. Therefore, different prose passages were used and a different method of recognition testing was employed. In addition, the form of the paraphrasing was more explicitly varied. It is possible that the results of the first experiments are attributable to gross differences between the verbatim and paraphrase items. Paraphrases that more closely approximate the wording (and meaning) of the original sentences should be less discriminable (Goolkasian et al., 1979).

The present study again assessed the "importance" of the target sentences in the prose structure on subsequent recognition. Information that is high in the "content structure" (Meyer, 1975) or "macrostructure" (Kintsch, 1976) of the passage could be expected to be retained better. However, variations between sentences (as in Experiments 1 and 2) also involve differences in wording, syntactic structure, position in the prose passage, and so on, which could be determining performance. One method that avoids such differences is to use a common target paragraph that is embedded in different prose passages, so as to have more or less importance within the overall passage. Britton, Meyer, Simpson, Holdredge, and Curry (1979) found substantially better free recall of a paragraph when it was high, rather than low, in the content structure of the text. In the present experiment, this procedure was used, now testing verbatim recognition memory.

Method

The subjects were 54 students enrolled in a general psychology course. The experiment was conducted during a class period.

The prose passages used were the "Sources of Nuclear Energy" and the "Future Sources of Energy" stories from Meyer (1975, pp. 190-193). Each passage had a common target paragraph embedded approximately halfway through the passage. This paragraph's information was high in the content structure of the nuclear energy passage (i.e., it was a central theme) and low in the structure of the future sources passage (i.e., describing only one of several future sources). These passages were used because they have been found to produce reliable differences in the free recall of the target paragraph as a function of their location in the content structure (Britton et al., 1979).

The test forms consisted of five multiple-choice questions derived from the target paragraph. Each question presented (1) the verbatim copy of one of the sentences, (2) an order paraphrase, in which the word sequence was changed from the original sentence by shifting the location of phrases or making active-passive voice changes, (3) a word paraphrase, in which synonyms were substituted in the original sentence, and (4) a distractor, which was worded similarly to the original sentence, except that specific facts or the truth value of the information was changed. The sequence of presentation of these four test-item types was varied across the five questions.

Subjects were instructed that verbatim memory would be assessed, although they were not told that only one paragraph would actually be tested. Each subject completed a practice example, read the prose passage, and then completed the test form. Half of the students read the passage in which the target was high in the content structure (Group High), and the remaining students read the passage in which the target was low (Group Low).

Results

For each subject, the proportion of each test-item type chosen, across the five test questions, was computed. The mean proportions of choice for each test-item type are shown in Table 3. As can be seen, verbatim items were chosen more frequently than the two paraphrase versions, and the latter were chosen more frequently than the distractors. This overall pattern is present within both the high low groups. However, the manipulation of the relative importance had an effect on the distribution of recognition errors: The high-importance group made more choices of order-change paraphrases, whereas the low-importance group erred more in choosing word-change paraphrases and distractor items.

Table 2
Mean Recognition Scores for Topic and Detail Sentences
in Experiments 1 and 2

Exper-iment		Test Sentence Type		
		Verbatim	Paraphrase	Distractor
1	Topic	3.60	3.41	2.24
	Detail	3.33	2.78	1.94
2	Topic	3.80	3.69	2.14
	Detail	3.49	2.70	1.77

Table 3
Mean Proportions Choice of Each Test Sentence Type as a Function of the Target Material's Having Been High or Low in the Content Structure of the Prose Passage

	Verbatim	Order Paraphrase	Word Paraphrase	Distractor
High	.615	.281	.089	.015
Low	.555	.179	.185	.081

These data were first analyzed using a correlated t test, combining both groups of subjects, to show that the verbatim sentences were indeed chosen more frequently than the next most frequent choice, the order paraphrases [t(53) = 7.78]. Thus, a robust verbatim-paraphrase discrimination was present, replicating the findings of Experiments 1 and 2.

A second analysis compared the false recognitions by Groups High and Low across the two paraphrase forms and the distractors. The overall difference across test-item types was significant [F(2,104) = 21.18], as was the Groups by Item Type interaction [F(2,104) = 6.46]. Post hoc comparisons showed that Group High subjects more frequently chose order paraphrases than did Group Low subjects, whereas the latter group more frequently chose word paraphrases and distractors [ts(52) = 2.31, 2.39, and 2.75, respectively].

The above analyses are by subjects. Comparable results were obtained in analysis by items. Overall, verbatim items were chosen more often than order paraphrases [t(4) = 3.36]. There was also a significant interaction of the order/word paraphrase and distractor items with Groups High and Low [F(2,8) = 6.60].

The reliability of the present verbatim-paraphrase discrimination is strengthened by the results of a pilot study. Students (n = 33) read the schizophrenia passage of the first experiments but were tested using procedures like that of Experiment 3. The five target sentences selected for testing were different from those of the first experiments. The results, in terms of mean proportion selection of verbatim, order paraphrase, word paraphrase, and distractor items were .63, .20, .13, and .04, respectively.

DISCUSSION

The major finding of the present experiments is the accurate recognition of verbatim sentences from prose read previously. While such a finding could be due to memory processes, the results might instead be due to the particular materials and procedures used in any given study. For this reason, it is important to note the replicability of the verbatim-paraphrase discrimination across prose passages, test sentences, and methods of assessing recognition. The present results also showed that the superiority of verbatim over paraphrase items was maintained across a 20-min delay and occurred in the absence of explicit instructions for a verbatim memory test.

These results may not seem surprising when compared with other findings of verbatim memory for statements from classroom lectures after 5-day delays (Kintsch & Bates, 1977). However, the present results do contrast with reports of forgetting across delays of seconds and minutes (e.g., Sachs, 1967, 1974).

The studies that have or have not found memory for the exact wording of prose differ procedurally in a number of ways. Such differences could explain the discrepant findings. For example, Bates et al. (1978) suggested that coherent discourse (as opposed to a series of unrelated sentences) may enhance the ability to reconstruct how a sentence should have been worded at a particular point in the prose. Thus, recognition of verbatim items may reflect not simply retention of the actual sentence structure, but also the presence of a reconstructive process that occurs during retrieval (testing). This view suggests that reconstruction would be less viable if the prose consisted of unrelated statements.

Alternatively, different task requirements may promote different processing strategies. Thus, studies using brief retention intervals may encourage only surface processing of the sentences (so-called "maintenance rehearsal") across the short delay until the time of the test (Craik & Lockhart, 1972). When longer retention intervals are expected, the information may be more deeply processed (so-called "elaborative rehearsal"), leading to better retention over longer delays.

Previous studies of the relative importance of the to-be-remembered sentences have shown mixed results (Goolkasian et al., 1979; Kintsch & Bates, 1977). In the present case, sentences that were higher in the content structure of the passages were more accurately recognized than sentences lower in the prose structure. However, the paraphrases (Experiments 1 and 2) and order paraphrases (Experiment 3) of the higher sentences elicited more false recognition. Thus, although the more important information in the prose may be retained better, retention may not be in a form conducive to verbatim testing. Possibly some abstraction of the more important sentences does occur (Bransford et al., 1972), as this material forms the basis of the passage's organization in memory.

REFERENCES

BATES, E., MASLING, M., & KINTSCH, W. Recognition memory for aspects of dialogue. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, 4, 187-197.

BRANSFORD, J., BARCLAY, J. R., & FRANKS, J. J. Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, 1972, 3, 193-209.

BRITTON, B. K., MEYER, B. J. F., SIMPSON, R., HOLDREDGE, T. S., & CURRY, C. Effects of the organization of text on memory: Tests of two implications of a selective attention hypothesis. *Journal of Experimental Psychology: Human Learning and Memory*, 1979, 5, 496-506.

CRAIK, F. I. M., & LOCKHART, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 671-684.

GOOLKASIAN, P., TERRY, W. S., & PARK, D. C. Memory for lectures: Effect of delay and distractor type. *Journal of Educational Psychology*, 1979, 71, 465-470.

GRAESSER, A., & MANDLER, G. Recognition memory for the meaning and surface structure of sentences. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, 1, 238-248.

KINTSCH, W. Memory for prose. In C. N. Cofer (Ed.), *The structure of human memory*. San Francisco: Freeman, 1976.

KINTSCH, W., & BATES, E. Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, 3, 150-159.

MEYER, B. J. F. *The organization of prose and its effect on memory*. Amsterdam: North Holland, 1975.

SACHS, J. S. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, 1967, 2, 437-442.

SACHS, J. S. Memory in reading and listening to discourse. *Memory & Cognition*, 1974, 2, 95-100.