

Prose retention: Recognition test effects and style memory

ROBERT E. CHRISTIAANSEN
Vassar College, Poughkeepsie, New York 12601

D. JAMES DOOLING
Bell Laboratories, Holmdel, New Jersey 07733

and

THOMAS F. KEENAN
Kent State University, Kent, Ohio 44242

Seven minutes after reading a prose passage, subjects were given three different types of test sentences to recognize. Test items could be (1) identical to one read in the passage, (2) a paraphrase, or (3) thematic to the passage but not related to any of the individual input sentences. Test order effects were found for thematic test items, suggesting that the common practice of giving multiple recognition items may artificially inflate thematic effects obtained. Based on initial responses not affected by test order, the percentages of "yes" responses ordered themselves: identical > paraphrase > thematic. The three levels of responding indicated that sentence wording, sentence meaning, and passage theme were all represented in memory, but were differentially retained. Finally, a test for memory of style of the passage yielded no significant effects.

A prose passage is apparently encoded at multiple levels of abstraction, from the most superficial information such as sentence wording to the most abstract, usually called the theme of the passage (Dooling & Christiaansen, 1977; Kintsch, 1974; Meyer, 1975). The present experiment provides one method of testing for these various memory codes and investigates some methodological problems encountered in testing memory for prose material in general.

In our experiment, subjects read a passage once, and shortly afterward were asked to make recognition judgments on three types of test sentences. The test sentences could be (1) identical to one read in the passage, (2) a paraphrase of a passage sentence, or (3) thematic to the passage but otherwise different from any of the individual sentences read at input. By looking at the percentage of "yes" responses to each type of test sentence, we can infer what types of codes are available in memory. A higher percentage of yes responses to identical sentences compared to paraphrase implies that subjects had some memory for the sentence wording or format. A higher percentage of yes responses to paraphrase than to items which are merely thematic to the passage implies sentence gist memory.

We are grateful to David Payne and Heather Turnbull for assistance in conducting the experiment. Requests for reprints should be sent to Robert E. Christiaansen, Vassar College, Box 348, Poughkeepsie, New York 12601.

It is common practice to test memory with several recognition items. With prose material, such a procedure naturally leads to asking questions about similar topics, since the passage has one overall theme. In the present experiment, the possible effects of similarity among test items was assessed by giving each subject four test sentences, with the first and second items always similar in meaning to the third and fourth, respectively. If encountering similar items has an effect, there should be a distinct difference in performance on the first half of the test as compared to the second half.

We also investigated memory for another aspect of prose, the style of the passage. We explored memory for one possible style variable by using passages with either very general or very specific versions. Subjects who read either one of the homogeneous-style passages had the use of an additional cue on a recognition test to aid them in recognizing or rejecting a test sentence; a mixed-input group, who read the same passage with general and specific sentences intermingled, had no additional cue.

METHOD

Subjects

A total of 389 introductory psychology students were tested as subjects. Of these, we were able to use the data from 384, as four subjects failed to follow instructions and one was involved in a procedural error.

Materials

Three experimental passages of 10 sentences each were written for the experiment. We began by writing a neutral passage: "Man's curiosity had led him to the challenge of the unknown. Those in charge had made careful plans for the attempt. The equipment had been meticulously checked and double-checked. Only the best men were chosen for the task. It was necessary to consider many variables in calculating a precise time for the event. When the day arrived, the participants were made ready and transported to the site. The beginning went smoothly and quickly. Midway, an unexpected occurrence caused danger and was skillfully avoided. The last part required total concentration and expertise. There was an historic moment of triumph when the goal was finally reached."

The neutral passage could be given two different thematic interpretations, manipulated by presenting one of two titles: (1) *The Climbing of Mount Everest* or (2) *NASA Mission to the Moon*. For each thematic title, a more specific version of the passage was written by making appropriate lexical substitutions in each sentence. For example, the "unexpected occurrence" in the neutral passage became an "unexpected avalanche" in the specific version of the mountain passage and an "unexpected meteor" in the moon version. These materials, therefore, gave a single 10-sentence passage, with three versions for each sentence. All input materials and test sentences were drawn from the same pool of 30 sentences, as described below.

Passages homogeneous in style could be constructed from the pool of sentences, just as described, by using all specific or all neutral sentences. Four such passages were used, mountain specific, mountain neutral, moon specific, and moon neutral. Mixed versions were also constructed, where specific and neutral sentences alternated throughout. That is, if given the moon title, all sentences were either neutral or moon specific. Finally, each subject was given only 9 of the 10 possible sentences to read at input. One of the sentences from Serial Positions 2-9 of the passage was omitted without the subject's knowledge. The passages were constructed such that omissions did not disrupt the flow of the text.

The test items consisted of four single sentences typed on separate pages of a booklet. An identical sentence was exactly the same as one read in the passage. A paraphrase sentence represented a target sentence that was in the passage, but in the alternative format. Thus, if a subject read the target sentence in its neutral form at input, the paraphrase test sentence was in the specific form for that particular theme (mountain or moon). A thematic sentence was the sentence that was omitted at input for that subject. It could be in either the same or a different format than it would have had if it had been present. The two thematic test conditions were thematic-same and thematic-different, where same-different refers to the format congruence between the omitted sentence at input and the test sentence.

The test sentences were given in four different orders, such that each type of test item occurred first for one fourth of the subjects, second for another fourth, etc. It should be emphasized that the four test items given each subject actually referred to only two possible target sentences. One target sentence was IN the passage, and both the identical and paraphrase test items referred to this target sentence. The other target sentence was omitted, or OUT, of the passage at input, and both the thematic-same and thematic-different test items referred to it. The target sentence could be any sentence from Positions 2-9 of the passage, except that the IN and OUT targets were always four sentences apart. The test orders were constrained such that Items 1 and 2 were always different along the IN-OUT dimension. Test Items 3 and 4, therefore, referred to the same target sentences as Items 1 and 2, respectively, but were in a different format.

Procedure

Subjects were tested in groups of about 50. They were told that the experiment concerned language comprehension and that they should give the passage one reading in their normal manner. Following this self-paced input, subjects listened to a comedy record for 7 min (Carlin, 1972). They then went through the test booklet one page at a time and judged whether or not each test sentence had been in the passage by circling yes or no. They also rated their confidence in each decision as low, medium, or high.

Design

There were four major between-subjects variables: (1) test order, one of four; (2) style of the passage, homogeneous vs. mixed; (3) format of the target sentence at input, specific vs. neutral; (4) passage theme, mountain vs. moon. The within-subjects variable was test sentence: identical, paraphrase, thematic-same, and thematic-different. Each of the eight sentences from Positions 2-9 served as the target sentence under all conditions, but Sentences 2-5 were used twice as often as 6-9.

RESULTS AND DISCUSSION

The dependent measure was simply the percentage of yes responses, since adding the confidence ratings did not affect the results. It should be pointed out that a yes response to an item other than an identical test sentence was a false alarm.

Recognition Test Effects

The first analysis was done using all four recognition responses per subject. Two significant effects were obtained. Test sentence was significant [$F(3,1056) = 170.20$, $MS_e = .18$, $p < .001$], with the test sentences ordering themselves as expected. The percent yes responses were: identical, 83.0; paraphrase, 35.9; thematic-same, 22.6; thematic-different, 24.5. But the interaction of Test Sentence by Test Order was also significant

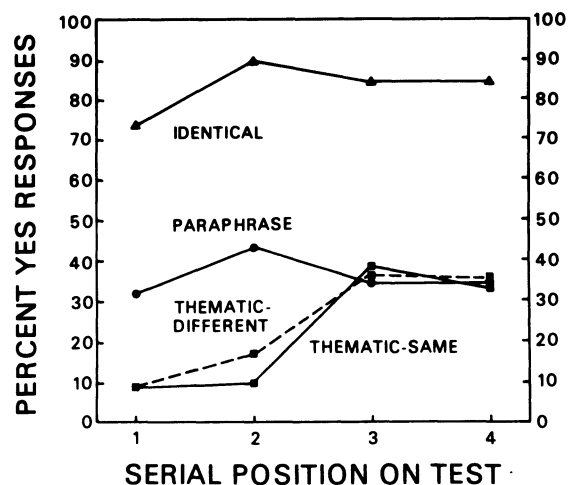


Figure 1. Mean percent yes responses for the Test Sentence by Test Order interaction, plotted as a function of serial positions on the test.

Table 1
Test Orders Used in the Experiment for Four Independent Groups of Subjects

Test Order	Serial Position on the Test			
	1	2	3	4
A	Identical	Thematic-Same	Paraphrase	Thematic-Diff
B	Paraphrase	Thematic-Diff	Identical	Thematic-Same
C	Thematic-Same	Identical	Thematic-Diff	Paraphrase
D	Thematic-Diff	Paraphrase	Thematic-Same	Identical

[$F(9,1056) = 7.97$, $MS_e = .18$, $p < .001$]. This interaction is best illustrated by plotting performance on each test sentence as a function of the four serial positions in the test booklet, as shown in Figure 1. As can be seen in the figure, identical and paraphrase test sentences were only slightly affected by their serial position in the test. On the other hand, both types of thematic test sentences were greatly affected. The percentage of yes responses for both thematic test items clearly increased from Serial Position 2 to 3. These differences, for both thematic tests, were significant at the .01 level by Newman-Keuls tests. None of the serial position differences for the identical or paraphrase test was significant.

Table 1 shows how the test sentence conditions were ordered for four independent groups of subjects. Counterbalancing was such that the first two test sentences always referred to different sentence ideas. The last two test sentences always repeated the meanings of the first two, but in a different format. The dotted line in Table 1 marks the point in the test sequence where the test sentence ideas begin to repeat. The serial position differences shown in Figure 1, therefore, indicate that subjects were more willing to say yes to a thematic test sentence when they had seen it in a different format, earlier in the test. Apparently, giving subjects two sentences similar in meaning causes confusion. Presumably, if we had given subjects several related test items, instead of just two, the difference between paraphrase and thematic tests would have been entirely eliminated when collapsed across all responses. Our conclusion would then have been that there was no specific memory for sentence gist independent of theme memory. Such a conclusion is certainly incorrect, as will be shown in the next section.

An example of a study where test effects, similar to those just described, may have played a role is the well-known study by Bransford and Franks (1971). In their experiments, from 6 to 12 sentences in the recognition test were related to the same complex idea, and in most cases subjects had to go through the test list twice. Such a procedure may have greatly inflated the amount of thematic responding and obscured subjects' ability to recognize individual sentences.

Levels of Encoding

To investigate encoding, and not the testing effects,

subsequent analyses used only the first response from each subject. Using these data, there were two significant effects. The effect of test sentence was significant [$F(3,352) = 60.46$, $MS_e = .15$, $p < .001$]. The mean percent yes responses were: identical, 73.9; paraphrase, 32.2; thematic-same, 9.4; thematic-different, 9.4. These are the same means shown in the first serial position of Figure 1. Newman-Keuls tests, at the .01 significance level, confirmed that (1) identical > paraphrase, (2) paraphrase > thematic, and (3) the two thematic tests did not differ. Thus, sentence gist was retained independent of the overall theme of the passage, and even sentence format was retained in this task.

Memory for Style

Referring to the same analysis of variance as reported in the last section, no significant effect was found for the style variable (homogeneous vs. mixed), nor did it enter into any significant interaction. The means important to the style comparisons are shown in Table 2. Although identical sentences were slightly better recognized and paraphrases more often rejected in the homogeneous condition, neither of these differences was significant by planned comparisons.

The pattern of results suggests that memory for the passage style had little effect on performance in this experiment. Our failure to find reliable style effects does not necessarily imply that memory for style is typically poor. It may be that the particular style dimension we chose to manipulate is not a very salient one.

One final significant effect was obtained in the analysis of the first response data: the main effect of format of target sentence at input [$F(1,352) = 4.53$, $MS_e = .15$, $p < .05$]. Specific targets led to a higher percentage of yes responses, 35.4, than neutral targets, 27.1.

Table 2
Percent Yes Responses for Tests of Style Memory

Style of Passage	Test Sentence			
	Identical	Paraphrase	Thematic-Same	Thematic-Diff
Homogeneous	75.0	27.1	8.3	10.4
Mixed	72.9	37.5	10.4	8.3

REFERENCES

- BRANSFORD, J. D., & FRANKS, J. J. The abstraction of linguistic ideas. *Cognitive Psychology*, 1971, 2, 331-350.
- CARLIN, G. *George Carlin FM & AM. (Son of Wino)* New York: Atlantic Recording Corporation, 1972. (Record)
- DOOLING, D. J., & CHRISTIAANSEN, R. E. Levels of encoding and retention of prose. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 11). New York: Academic Press, 1977.
- KINTSCH, W. *The representation of meaning in memory*. Hillsdale, N.J: Erlbaum, 1974.
- MEYER, B. J. F. *The organization of prose and its effects on recall*. New York: North Holland, 1975.

(Received for publication February 24, 1978.)