INSTITUTE FOR THE
STUDY OF
INTELLECTUAL
BEHAVIOR

Recognition Memory for Statements from a Classroom Lecture

Walter Kintsch and Elizabeth Bates

University of Colorado

Program* on Cognitive Factors in Human Learning and Memory Report No. 53

Institute for the Study of Intellectual Behavior

University of Colorado

Boulder, Colorado 80309

July, 1976

ABSTRACT

Research on memory for classroom lectures provides an interesting means for testing psychological theories of memory in a natural setting. Two experiments were undertaken on recognition memory for statements in a lecture, varying instructions and the response format. Three types of statements were tested: topic statements, details and such extraneous remarks as jokes and announcements. In both studies, memory for meaning was significant in all three categories. With a 2-day delay, there is still verbatim memory for all three types of statements; with a 5-day delay, there is verbatim memory only for extraneous statements. In both studies, extraneous remarks are remembered best. Contrary to predictions, there were no differences in memory for topics vs. details. The results are discussed from the perspective of macrostructures in text memory. The validity of recognition measures as a test for lecture memory is questioned.

---

* Supported in part by NSF Grant BNS 72 02084

Recognition memory for statements from a classroom lecture

Studies of memory for materials presented in a classroom lecture are important for their didactic implications. However, in addition to its applicability, such research is also of interest for psychological theory. First, it permits us to test laboratory results under natural conditions, where laboratory-specific strategies presumably play a lesser role. Second, such research provides an opportunity to investigate problems that are unique to the classroom situation. The two experiments reported here are an initial attempt to investigate memory for lectures.

There is a certain amount of material in the education literature concerning memory in the classroom. These studies have dealt primarily with such didactic issues as the effect of note-taking on recall, and the efficacy of various teaching methods. We were interested, instead, in two questions relevant to issues in memory. First, will students remember only the meaning of a lecture, or will they remember the meaning plus the actual words used? Secondly, is there a difference in the amount of memory for various types of statements? In particular, are topic statements remembered better than mere illustrative material, and is there preferential memory for extraneous statements (e.g. jokes, announcements) embedded in a lecture?

Memory for sentences and for text in general can be said to consist of traces from many different levels of processing: perceptual processing (visual or auditory), analysis of the words, phrases and sentences of the text, and construction of the meaning of the message, including its pragmatic implications. Traces from any of these activities may be retained in memory (for a review see Kintsch, 1977). Normally, however, a reader and/or listener is

concerned primarily with the meaning of the text. Hence memory for meaning tends to be stronger than memory for other aspects of processing. Other memory traces are less important, although they are rarely absent except in unusual situations. The precise conditions under which people remember not only the meaning of a text but also the exact words used are as yet poorly understood. In most classroom situations, there is no particular reason to remember any of the material verbatim. For example, in the situations about to be described students expected to receive an essay test over the material several weeks later. Is there verbatim memory under such conditions?

A second issue regards the nature of the lecture material itself. Most statements in a lecture are descriptive. Indeed, the speech act of lecturing is basically descriptive in nature. However, in every lecture there are statements that fultill a different pragmatic function: announcements about assignments, interactions with the audience at a personal level, joking comments to retrieve the attention of the listeners. These extraneous statements stand out from the lecture itself. Hence one could expect a von Restorff effect, such t h a t these unique items are remembered better. In the classroom this effect may even be augmented, since these extraneous statements tend to be particularly interesting to the listener, serving as a relief from the lecture itself. On the other hand, since the student knows that jokes will be of no use on the final, he is probably not intentionally trying to remember these statements.

Furthermore, not all descriptive statements made in a lecture are of equal importance. A lecturer usually works from a (not necessarily explicit) outline. He has a set of topic statements, and the purpose of the lecture is, above all, to get these ideas across to his audience. To each topic statement there corresponds a number of other, less important statments that illustrate and elaborate

the main point. These will be termed "detail" statements here. On the whole detail statements are judged to be less important by the lecturer than the topic statements. They are intended to promote understanding of the main point, but not to be remembered in themselves. Presumably, if a student is interested in mastering the course material, he should make a greater effort to remember the topic statements than the details.

In recent work on memory and comprehension of stories, recall could be related to the position of a statement or proposition in the overall structure of the story (e.g. Kintsch & van Dijk, 1975; Rumelhart, 1975). Propositions that are important in the macrostructure of a story are particularly likely to be recalled, and are frequently used in writing summaries. Lectures, like all text, have a macrostructure too, and hence one would expect that the corresponding macropropositions would be the ones most likely to be recalled. While much is known about the macrostructure of stories, practically nothing is known about that of lectures. Nevertheless, it appears reasonable to hypothesize that the statements that the lecturer designates as his topic statements correspond to macropropositions. If so, they should be remembered better than detail statements. Thus, by looking at which kind of statements students remember from a lecture, one can obtain some idea about how they organize the lecture in their memory. If the lecture is organized around the topic statements, one would expect these to be remembered better than detail statements, in analogy with the recall advantage of macropropositions in a story.

EXPERIMENT I

Subjects -- In the fall semester of 1975, thirty University of Colorado students in Psychology 451, History of Psychology (taught by one of the authors,

Walter Kintsch), attended both sessions during which the experiment was performed. The students were primarily seniors and psychology majors.

Materials -- A lecture on the history of the intelligence testing movement was delivered during a regular class period. The lecture was designed to appear indistinguishable to the students from the lectures normally delivered in that class. It was distinguished, however, in two ways: certain sentences were prepared in advance and delivered verbatim, and the lecture was designed to permit plausible distractor items to be used on a later recognition test.

Design of distractor items post hoc from a normal lecture leaves open the possibility that distractors are in some way different from statements that the student actually heard. Hence a student might eliminate a distractor not because of memory factors, but because he judges the distractor to be stylistically out of place. To control for this possibility, the whole lecture was designed beforehand in two strands, so that every prepared statement existed in two semantically distinct forms. While these paired forms differed in content, either could be substituted at a given point in the lecture without destroying the coherence of the lecture as a whole. One statement out of each pair was chosen at random, and was actually used in the lecture. The other alternative was later used as a distractor item. In this way, distractor sentences and actually spoken sentences did not differ systematically in style or in the type of content each contained.

For each of the statements actually selected for the lecture, two paraphrased versions were prepared in advance. One of these was chosen randomly for the lecture. The other served later as another type of distractor in the recognition test. All paraphrases were relatively natural restatements of the same point. No attempt was made to vary systematically particular syntactic or semantic elements (e.g. passives vs. actives).

In this manner, 21 pairs of statements were prepared (42 total), plus 21 paraphrases of the statements actually chosen. Eight of the pairs were designed as topic sentences, eight were details, and five were extraneous remarks (e.g. joking comments, announcements about assignments). Hence the lecture that the students heard contained eight designated topic statements, eight details, and five planned extraneous remarks. The eight topics, eight details, and five extraneous statements that were eliminated by random selection formed an alternative version of the lecture, which was not given. The three types of materials did not differ from one another systematically in length. There were long and short statements within each category. Only the experimenter's intuition determined what was called a topic or detail, but the point should be made that in making these assignments the experimenter acted precisely as he had in preparing numerous other lectures (except for the unusual care taken here, and the concern with verbatim statements). Hence the design of topics, details and extraneous statements reflects one teacher's intuitions about the macrostructure of a lecture. Examples of items, alternative statements and paraphrases are presented in the Appendix.

The 21 selected lecture statements were then ordered in an appropriate sequence and embedded in other lecture material. The extraneous statements (e.g. jokes and announcements) were also presented at predetermined times in the lecture.

The test booklet was constructed by assembling in random order all 63 statements: the 21 sentences actually used in the lecture ("old"), the corresponding paraphrases ("paraphrase"), and the 21 alternatives prepared but not spoken ("new"). Items from the same triplet were never presented adjacently on the test. Mimeographed test booklets were constructed, which requested subjects

to respond "yes" or "no" to each statement, and then provide a confidence judgment.

Procedure -- The lecture was given during a 75 minute class period on a Tuesday. Every attempt was made to make this appear to be a normal lecture. Most of the experimental sentences were spoken during the first hour of the class period. Notes were used to make certain that the test statements were delivered verbatim as planned.

The test was administered at the end of the class period on the following Thursday. The purpose of the test was briefly explained and subjects were given 20 minutes to fill out the test booklets. Subjects were instructed to answer "yes" only when they thought they had heard that particular sentence before verbatim, and to answer "no" when they thought that a sentence had not occurred at all, or was a paraphrase of a sentence actually spoken. Each response received a confidence rating on a three-point scale: "certain", "I think so", or "guess". Subjects were assured that their responses were anonymous and would not affect their grade.

Results

All responses were converted to a six-point scale, with 0 being assigned to a No-certain response, 1 to No-I think so, 2 to No-guess, 3 to Yes-guess, 4 to Yes-I think so, and 5 to Yes-certain. The average scores for the nine experimental conditions are shown in Table 1. Five orthogonal comparisons were performed among the means in Table 1. The results for the descriptive items and the extraneous items were treated separately. A number of subsidiary analyses were then performed by means of Scheffe's test, including comparisons between descriptive and extraneous items. The MS_{error} for the orthogonal comparisons was .50.

The first question of interest was whether there was evidence of memory for meaning, i.e. whether higher scores were assigned to old items and their paraphrases than to new items. For descriptive items, the mean recognition score for old items and paraphrases was 2.62, which was significantly higher than the mean score for new descriptive items 1.58, $F(1,232) = 78.76$, $p < .01$. For extraneous items, the two means were 2.55 and .79, respectively, $F(1,232) = 123.67$, $p < .01$. These F-values are for an analysis by subjects. If the analysis is performed by materials instead, equivalent results are obtained, with $F(1,51) = 21.88$ for descriptive items, and $F(1,51) = 24.93$ for extraneous items, $p < .01$ for both.

Verbatim memory was demonstrated by the fact that higher recognition scores were obtained for old items than for their paraphrases: for descriptive items, old items had a mean of 2.88 and paraphrases a mean of 2.35, $(F1,232) = 16.73$, $p < .01$; the means for the extraneous items are shown in Table 1, and are also significantly different, $F(1,232) = 68.75$, $p < .01$. However, these differences between old and paraphrase items are only marginally significant in an analysis by materials, $F(1,51) = 3.29$, $p = .076$, and $F(1,51) = 3.45$, $p = .069$, respectively.

The final orthogonal comparison concerned the question of whether memory for meaning is different for topic versus detail statements. As is apparent from Table 1, the superiority of old items and paraphrases over new items is about the same for topics and details, with $F < 1.0$ in both the subject and materials analysis.

Scheffé tests were used to analyze the differences between descriptive (collapsing topics and details) and extraneous statements. Overall, memory for meaning was much better for extraneous statements than for descriptives,

$F^* = 208.90$, $p < .01$ by subjects, and $F^* = 24.47$, $p < .01$ by materials. Further analyses revealed that the locus of this difference resided primarily in the better rejection of new extraneous items, rather than in performance on old items and their paraphrases. The F^* values obtained in the analysis over subjects for the difference between descriptive and extraneous items were $F^* = 7.27$, $p > .05$ for old items, $F^* = 24.74$, $p > .01$ for new items, and $F^* = 12.54$, $P > .05$ for paraphrases. The corresponding values for the analysis by materials are $F^* = 1.83$, $p > .05$, $F^* = 4.99$, $p > .05$, and $F^* < 1$, respectively.

The construction of the test instrument makes it possible for students to use a strategy of responding "no" to an item if they had responded "yes" to its paraphrase before, or vice-versa. In order to discourage such a strategy, students were given barely enough time to finish the recognition test, leaving very little time to cross-check items. Furthermore, the rather low correlation between recognition scores for old items and recognition scores for their paraphrases indicates that this response strategy was not an important factor in the present experiment: the average correlation for the 30 subjects was $r = -.26$. Nevertheless, in Experiment II we took measures to prevent any use of this strategy whatsoever.

The lack of a difference in recognition between topic statements and details led us to perform some post hoc analyses to check on some possible confoundings. For that purpose, two separate groups of students in other classes ($n = 15$ and $n = 19$) were asked to rate the sentences used in the lecture (presented in the same format as the test booklet) for both "interest" and "memorability". The ratings for the two groups of students correlated reasonably highly, $r = .78$. Topic statements were rated as less interesting and memorable than either details or non descriptive statements. However, neither the interest nor the

memorability ratings correlated significantly with the recognition scores, r_{IR} = .21, and r_{MR} = .41. The correlation between the interest ratings and the memorability ratings was also nonsignificant, r_{IM} = .33.

Discussion

Two days after listening to a classroom lecture, students showed significant memory for meaning, in that they were able to discriminate sentences actually spoken in the lecture or their paraphrases from control sentences that they had never heard before. In addition to this memory for meaning there was evidence for verbatim memory, in that students discriminated sentences that actually occurred in the lecture from paraphrases of those sentences. In fact, the amount of verbatim memory was relatively large when compared with memory for meaning: overall the difference between old items and paraphrases (verbatim memory) was 67% of the difference between old + paraphrase and new (the measure of memory for meaning used here).

As expected, extraneous sentences -- jokes, comments to the audience, and announcements -- were recognized better than descriptive statements. This is not surprising, insofar as extraneous statements by definition stand out in a lecture. Interestingly, however, this effect was largely due to the better ability to reject new extraneous statements, rather than better recognition of old items.

Contrary to predictions, topic statements were not remembered better than detail statements. Recognition scores for old (and paraphrased) topic sentences were slightly higher for topic sentences than for details. However, the false recognition rate for control topic sentences was also higher than that for the other two types of materials.

The surprising finding of no difference between topics and details contrasts strongly with results obtained with stories. It is possible that instructions to recognize verbatim sentences from a lecture two days before are rather unnatural, and hence distorted the results of the present experiment in some way. Experiment II was designed to provide a replication of the first experiment with a somewhat different methodology, and to determine whether the emphasis on verbatim memory was responsible for the results.

EXPERIMENT II

Subjects -- In the spring semester of 1976, seventy-one University of Colorado students in Psychology 468, Developmental Psychology (taught by one of the authors, Elizabeth Bates), attended and participated in both sessions during which the experiment was performed. Students were primarily juniors or seniors, and psychology majors.

Materials -- The topic selected for the experimental lecture was an introduction to Freudian theory. The method of preparation for the lecture was similar to the one used in Experiment I. The instructor outlined two days worth of lecture notes, including 18 possible topics (roughly analogous to an opening statement for a one or two paragraph passage), 18 possible details, and 12 possible joking comments, announcements or other extraneous material. For each of these 48 possible items, two paraphrased versions were written, yielding 96 possible sentences. Nine topics, nine details, and six extraneous comments were randomly selected from this list. Hence within each type of material, all statements had an equal probability of occurring in the lecture. Those topics, details and extraneous comments which were not selected were later used as distractor items for recognition memory tests. (The alternative material excluded by this process was presented to the students in a later lecture, after the experiment.)

Of the 24 items chosen, one of the two possible paraphrases for each was randomly chosen for actual inclusion in the lecture. The other version then served as a paraphrase in the verbatim memory test. These 24 items were built into a 75-minute lecture containing a great deal of other material as well. In some cases, particularly for topic statements, there was some overlap in content (though not in wording) between experimental sentences and other material in the lecture. It was simply impossible to create a coherent lecture without such overlap. However, for the most part items were kept maximally independent in content as well as in wording.

As in Experiment I, paraphrases were fairly natural restatements of the same point. No effort was made to vary systematically particular syntactic or semantic elements. Also, as in Experiment I, there were no systematic differences among the three types of material in length; there were long and short statements within each category. The division of the materials into topics, details and extraneous comments were based entirely on the instructor's intuitions concerning the lecture outline, reflecting the sort of lecture macrostructure that the instructor normally uses for material in this course. For examples of statements, alternatives and paraphrases, see the Appendix.

Two types of recognition memory test booklets were prepared for the second session. One emphasized verbatim memory in a multiple-choice format. The other emphasized memory for meaning, with a Yes-No and confidence rating format.

The Yes-No memory for meaning format contained a total of 52 items. The 24 original experimental sentences, 22 of the 24 alternative statements, and 6 extemporaneous sentences selected afterwards from the lecture were included. (Two of the alternative detail statements were excluded accidentally by the typist, resulting in a reduction from 24 to 22 in the alternative items.) The extemporaneous

items were included as a control for the possibility that the prepared items were delivered in an unnatural manner that somehow cued the students to remember those statements. The old experimental sentences were in all cases those that were actually spoken during the lecture. Items were checked against an audiotape of the lecture to be certain that this was the case. Since the instructions in this part of the experiment stressed memory for meaning only, we did not want to set off a verbatim memory strategy by including both originals and paraphrases of the experimental sentences. For the distractor items, one of the two possible paraphrases was randomly selected for inclusion in the test booklet. Students were instructed to place a rating after each item: 0 = sure that it did not occur, 1 = think it probably did not occur, 2 = might not have occurred, 3 = might have occurred, 4 = think it probably did occur, 5 = sure that it did occur. Students were asked to skip no items, and to guess if they were not sure. The instructions, stressing memory for meaning and discouraging verbatim memory, were as follows:

> The following are phrases which either did or did not occur in last Thursday's lecture. Any one of them could have occurred, according to the lecture outline, but not all of them were actually directly discussed. You might have the feeling that one or two words in the sentence are different than they actually were in class. That much of a difference isn't important. We want to know if you can remember whether or not that statement, or a statement almost exactly like that, was in the lecture.

The verbatim memory booklets were in a multiple-choice format. There were a total of 46 items (by typist error, the same two alternative detail items omitted in the Yes-No form were also omitted here). Extemporaneous items were not included here because, for the reasons expressed earlier, we felt that

paraphrases written after the fact might be in some way stylistically different from statements produced extemporaneously. The items included 24 experimental items (9 topics, 9 details and 6 extraneous comments) and 22 distractor items (9 topics, 7 details, and 6 extraneous comments). Each item was presented in a multiple-choice form with three choices: (a) the experimental sentence that was actually spoken (or a randomly selected paraphrase of a distractor item), (b) the paraphrase of the experimental sentence excluded by earlier random selection (or the other randomly-selected paraphrase of a distractor item), (c) neither. The 46 items were presented in random order, both between and within items. The instructions, stressing verbatim memory, were as follows:

> The following phrases are items that either did or did not occur word for word in last Thursday's lecture. We want to know how well you can remember verbatim sentences taken from a lecture when you do not know that verbatim memory will be tested in advance. All items are forced choice. Choose either sentence (a), sentence (b), or check (c) for neither. If you check (c), it means that you don't think that sentence occurred in the lecture at all. If you are not sure about an item, please guess anyway. Don't skip any items.

Procedure

Prior to the lecture, the instructor memorized the 24 test sentences verbatim. Each item was written on a separate index card, in proper order of mention. While lecturing, the instructor would glance briefly at the cards several sentences prior to the appropriate point in the lecture, prepare the context for the next test statement, and then make the statement in as natural a manner as possible to avoid cuing students to these sentences in any obvious

way. Two informants who were aware of the experiment in advance were present in the classroom. They report that they were generally unable to discern test sentences in the flow of the lecture. The lecture was audiorecorded by the teaching assistant on a portable cassette recorder.

The experimental lecture was given on a Thursday, the same day that questions were distributed for a take-home essay exam covering material up to but not including the Freud lecture. In some respects, this decision may have depressed the level of attention that students paid to material that was clearly not relevant to the weekend exam. However, such timing also virtually insured that students would not be reviewing notes or rehearsing the material in the period prior to memory testing.

The next class was held on the following Tuesday, yielding a five-day delay in comparison with the two-day delay in Experiment I. At this time students were asked to participate voluntarily in a test of memory for the previous lecture. Test booklets were distributed, and those who wished to participate were given 35 minutes to respond. 30 students completed the multiple-choice booklet, and 29 completed the yes-no booklet. An additional 12 students were asked simply to write down everything they could possibly remember from the lecture, including jokes or seemingly trivial material. (These recall data will be discussed only briefly in this paper.) Students were assured that their responses were anonymous and would not affect their grade.

After test booklets had been collected, the experiment was explained in detail to the students, and a discussion was held concerning their reactions to the items and their memory for the materials. Later in the course, when data analysis was complete, this information was also given to the students in class.

Results

Yes-no recognition -- Recognition scores were calculated as in Experiment I on a six-point scale with the endpoints 0 (no-certain) and 5 (yes-certain). Average scores are presented in Table 2. Orthogonal comparisons showed that old and new items were discriminated from each other for all three item types, with $F(1,196) = 325.7$, 75.4, and 63.13 for extraneous statements, details, and topics, respectively in the analysis by subjects ($MS_{error}=.45$. The corresponding F-values in the analysis by materials are $F(1.45) = 12.42$, 7.83, and 8.21. All F's are significant at the .01 level.

Scheffé tests yielded a significant F* for the difference between the descriptive and extraneous statements, $F^* = 63.15$, $p < .01$. In the analysis by materials the same comparisons did not reach significance, however ($F^* = 8.45$, $p < .05$). As in Experiment I, the main difference between descriptive and extraneous items was in the better rejection of new extraneous statements ($F^* = 32.21$, $p < .01$ by subjects, $F^* = 1.13$, $p < .05$ by materials).

The six extemporaneous statements that were selected afterwards from the lecture for inclusion in the test were included to determine whether the prepared statements were in some way different from the sentences spoken spontaneously. These extemporaneous statements had a mean recognition score of 3.57, which is almost equal to the mean of the old descriptive items in Table 2, $F^* < 1$.

Multiple choice tests -- In Tables 3 and 4 the proportion of choices of the three response alternatives available to a subject for each sentence are shown. For new sentences there is no meaningful distinction between the two paraphrase forms for each test sentence and hence the data have been combined.

Memory for the meaning of a sentence is demonstrated by choosing either the sentence itself or its paraphrase. Table 3 shows that memory for meaning was substantial for both descriptive and extraneous sentences, with the latter being somewhat higher. A much more pronounced difference between these two types of statements appeared in the responses to new sentences. Subjects were much better able to reject new extraneous statements than new descriptive statements, the proportions of correct rejections being .81 and .52 respectively. If memory for meaning is indexed by the difference in the proportion of choices of old sentences or their paraphrase versus the choice of new sentences, this difference is more than twice as large for extraneous statements (.68, with a 95% confidence interval from .60 to .76) than for descriptive statements (.32, with a 95% confidence interval from .26-.38).

Regarding verbatim memory, even after a five day retention interval students choose old sentences more frequently than their paraphrases. Verbatim memory was substantial for extraneous statements, with the mean difference between old and paraphrases being .39 (with a 95% confidence interval from .28 to .51). For descriptive statements this difference was reduced to .06, with a 95% confidence interval from .00 to .12.

Table 4 shows that the differences in the responses to topic and detail statements were small. Old topic sentences were somewhat better recognized than old detail sentences, but this small advantage was more than balanced by a substantially greater false recognition rate for new topic sentences. Indeed, for new items the likelihood that one or the other version of a topic control sentence was chosen was greater than that of a correct rejection (choosing the "Neither" category). In part this inflated false recognition rate can be explained by the overlap between some control topic sentences and other material touched upon in the lecture. If the new topic sentences are classified into sentences for which

there was some overlap with other lecture material and sentences for which no overlap existed, the false recognition rate for the later type of sentences was reduced to .47. This value is still larger than the false recognition rate for details but is no longer discrepant.

The 12 recall protocols collected at the same session were not analyzed in any detail and need not be discussed here. It is sufficient to say that with a couple of exceptions, these protocols were extremely scant, containing only 3 or 4 sentences each, and stood in marked contrast to the significant amount of recognition memory shown by students in the same experiment.

Discussion

Five days after hearing an apparently normal classroom lecture, students demonstrate both memory for meaning (in two different response formats) and verbatim memory for statements taken from that lecture.

In the yes-no format, with instructions stressing memory for meaning and discouraging verbatim memory, old items were remembered better than new items within all three types of materials (topics, details and extraneous comments) and across materials. The same preference for old items (in this case old and paraphrase) over new items was also demonstrated in the multiple-choice format stressing verbatim memory, both for extraneous remarks and for descriptive statements. Also, in both types of recognition test, memory for meaning was better for extraneous material than for descriptive material. However, again in both recognition tests, the better memory for meaning for extraneous material was primarily due to an ability to reject new distractor items, rather than to recognition of old items.

In both types of recognition test, there were no significant differences in memory for meaning for topics vs. details. This finding is contrary to predictions

based on story recall data. In Experiment I we had initially suspected that this surprising finding might be due to the somewhat unnatural emphasis on verbatim memory. However, the results from Experiment II suggest that the effect is not due simply to instructions, since the instructions for the two types of recognition test differed markedly, with one explicitly discouraging verbatim memory.

Regarding the amount of verbatim memory shown, after a five-day delay old items were still preferred over their paraphrases in the multiple-choice test. However, this effect was significant only within the category of extraneous statements. For descriptive statements, the difference was quite small although still in the predicted direction.

GENERAL DISCUSSION

There are three main issues that need to be considered in light of the findings from Experiments I and II. The first regards the surprising strength of verbatim memory in a situation where presumably none is called for. The second concerns memory for the three types of materials tested here (topics, details and extraneous remarks), and the unfortunate implications the findings may have for our knowledge of the macrostructure of lectures, and for the question of ecological validity in classroom memory processes. Third, there are some questions about the nature of recognition memory and its appropriateness in the study of memory for lectures.

Verbatim memory was significant in all three types of materials in Experiment I, after a two-day delay. Although the evidence for verbatim memory is not overwhelming, and seems to be decreasing from 2 to 5 days after the lecture, we were somewhat surprised to find that it existed at all. Some recent research on recognition memory for sentences embedded in prose (Bransford, Franks and Barclay 1972;

Sachs 1967) suggests that in conditions facilitating semantic integration, memory for surface form is markedly reduced. The experiments presented here took place in a natural setting, where students were not expecting anything like a verbatim memory test, after a delay of several days. Hence one might predict that the semantic integration effect, with accompanying loss of surface form, might be stronger than in laboratory investigations of the same processes. It is true that our paraphrases differed rather unsystematically from originals, often varying several syntactic and lexical elements. Sachs, and Bransford, Franks and Barclay, varied fewer elements and in a more systematic fashion in constructing their paraphrases. Hence the "distance" between our paraphrases may be greater, facilitating memory for surface form. Nevertheless, even with a greater number of "clues" to surface form, one might well question the ecological validity of memory processes that retain so many verbatim traces under conditions in which presumably only the semantic core will be required for future use. Keenan (1975) has suggested that such surface traces may be useful in the reconstruction of main points, serving as mnemonics that permit retrieval of more substantive material. This hypothesis, though plausible as a general principle, does not explain the finding that memory for surface form is stronger for jokes and other extraneous material than it is for the substantive material in a classroom lecture.

This leads us to the second point concerning memory for the three different types of materials. Both memory for meaning and verbatim memory were stronger for extraneous material than for descriptive statements. Contrary to expectations, topics and details did not differ significantly. Indeed, there was a tendency for stronger false recognition of new topics than new details.

The better memory for extraneous material is not surprising, since these items were designed expressly to stand out from the descriptive statements in the lecture. We can expect, then, a von Restorff effect, in which unique elements are

remembered better than non-unique elements. Such an effect might be enhanced by the fact that jokes in particular provide relief from the heavy information load of a standard lecture. Hence the extraneous items in these experiments are unique both in content and in accompanying affect, in comparison with descriptive items.

The results for topics vs. details are more surprising. Research on recall of stories leads us to predict that macropropositions will be remembered better than details. One possible explanation of our failure to find such an effect is that a macroproposition from a teacher's perspective may not be discriminated as such by a student -- despite the frequent presence of key, summarizing words and such non-verbal cues as pauses before beginning a new topic. Indeed, this discouraging possibility was one of the main motivations behind our decision to replicate the findings with at least two instructors. Another possibility is that, for the students, the "macroconcept" is not isomorphic with the topic statement introducing that concept. A student may recognize a main idea when he hears one, but discard the particular phrase introducing that concept. A third possibility is that topic statements are necessarily more abstract than details, and because they are abstract they are connected to more nodes in memory. As a result, topics may be more easily confused with other general points, whereas details are more concrete and more distinct from one another. The higher false recognition rate for distractor topics in both experiments supports such an interpretation.

Whatever the explanation for our findings regarding types of lecture statements, one unfortunate implication remains. Most instructors would probably hope that the topics he selects are the most memorable points in a lecture, followed by informative details, with extraneous material coming in a poor third.

Our results suggest that the opposite sequence is more likely, at least for recognition memory.

This brings us to the third issue, regarding the appropriateness and nature of recognition memory in such research. It is possible that recognition memory is not a valid indicator of the memory processes involved in understanding lectures. We assumed at the outset that the topic-detail effect in recognition memory should by analogy support the macroproposition-subordinate proposition effect demonstrated in recall of stories. This prediction was not based on parallel recognition memory findings for stories. McKoon (1975) has demonstrated that for short prose passages, recognition memory for superordinate propositions (i.e. propositions sharing elements, explicitly or implicitly, with many propositions in a paragraph) is superior to that for subordinate propositions. However, it may be that the same recognition memory effect does not hold for larger text units. More research on recognition memory for a variety of large text units is required before we can attribute our findings to some peculiar aspect of recognition memory for lectures.

Furthermore, our experiments suggest that recognition memory processes include not only memory in the strict sense (i.e. retrieval of stored information), but also "metamemory" processes. The largest differences between extraneous and descriptive material was found not in recognition of old items, but in rejection of new items. We have suggested that false recognition of topics may be due to an inability to distinguish general, abstract statements that share many nodes with other aspects of memory. However, details and jokes were generally equally concrete. Yet students were much better at rejecting jokes that they had not heard than details that they had not heard. In addition rejection of new jokes was clearly better than recognition of old jokes. Hence this difference cannot be due simply to a match-to-storage procedure, since the rejection is better than the

demonstrated storage itself. It seems plausible that students know a great deal about their own memory processes, so that upon seeing a new joke or announcement, they know immediately that if they had heard it, they would have remembered it. Indeed, in the class discussion following Experiment II, some students made precisely that point. This would suggest that recognition memory involves not just recognition, but reconstruction or reasoning from knowledge about lectures and knowledge about one's own memory processes.

In summary, we suggest that further investigations of memory for lectures include measures to determine whether the students' macrostructure for the lecture matches the one assumed by the lecturer. This might include ratings of various statements by the students in terms of their relative importance to the course material, and summaries of the lecture in outline form, constructed immediately after listening. The few recall protocols collected in Experiment II were not particularly encouraging to us. However, recall protocols collected after a shorter delay may tell us more about the structure of memory for course material.

TABLE 1

Mean recognition scores for sentences actually spoken in the lecture (old), paraphrases of sentences actually spoken, and control sentences not presented (new), as a function of sentence type.

	OLD	PARAPHRASE	NEW
TOPIC	2.92	2.42	1.78
DETAIL	2.85	2.29	1.37
EXTRANEOUS	3.31	1.79	.79

TABLE 2

Mean recognition scores for old and new sentences as a function of sentence type (Yes-No format).

	OLD	NEW
TOPIC	3.76	2.36
DETAIL	3.42	1.89
EXTRANEOUS	4.32	1.14

TABLE 3

Mean Proportion of choices of each response class for descriptive and extraneous statements. (Standard errors in parentheses).

RESPONSE CLASS	TYPE OF MATERIAL	
	DESCRIPTIVE	EXTRANEOUS
Old-item-verbatim	.43 (.03)	.63 (.04)
Old-item-paraphrase	.37 (.03)	.24 (.03)
Old-item-combined	.80 (.03)	.87 (.02)
Neither (Old)	.20 (.03)	.13 (.02)
New items-combined	.48 (.05)	19 (.04)
Neither (New)	.52 (.05)	81 (.04)

TABLE 4

Mean Proportion of choices of each response class for two types of descriptive statements. (Standard errors in parenthesis)

RESPONSE CLASS	TYPE OF MATERIAL	
	TOPIC	DETAIL
Old item-verbatim	.46 (.03)	.40 (.03)
Old item-paraphrase	.37 (.03)	.37 (.02)
Old item-combined	.83 (.03)	.77 (.03)
Neither (Old)	.17 (.03)	.23 (.03)
New item-combined	.58 (.05)	.39 (.04)
Neither (New)	.42 (.05)	.61 (.04)

REFERENCE NOTES

Keenan, J. M. The role of episodic information in the assessment of semantic memory representations for sentences. Unpublished doctoral dissertation, University of Colorado, 1975.

McKoon, G. Surface and semantic information in text memory. Unpublished doctoral dissertation, University of Colorado, 1975.

REFERENCES

Bransford, J. D., Barclay, J. R., and Franks, J. J. Sentence memory: a constructive vs. an interpretive approach. Cognitive Psychology, 1972, 3, 193-209.

Kintsch, W. Memory, Language, and Thinking. New York: John Wiley and Sons, 1977.

Kintsch, W. and T. A. van Dijk. Comment on se rapelle et on resumé des histories. Langages, 1975, 40, 98-116.

Rumelhart, D. E. Notes on a schema for stories. In D. G. Bobrow and A. Collins (eds.), Representation and Understanding. New York: Academic Press, 1975.

Sachs, J. Recognition memory of syntactic and semantic aspects of connected discourse. Perception and Psychophysics, 1967, 2, 437-444.

Footnotes

This study was supported by a grant from the National Institute of Mental Health MH15872 to Walter Kintsch.

APPENDIX

Examples of Old Statements, Paraphrases and Distractors

Topics

	Experiment I	Experiment II
Old:	The doctrine of natural biological evolution formed the rationale for Galton's study of the eminent families of Britain.	The closed energy model is still critical for the psychoanalytic approach to therapy.
Para:	Galton compared the eminent families of Britain with the natural biological variations that figure so preeminently in the doctrine of evolution.	The psychoanalytic approach to therapy still depends critically on the concept of a closed energy system.
New:	The inheritance of human intellect implied for Galton the practicability of supplanting inefficient human stock by better strains.	The concept of a limited energy system explains a great deal about neurotic development.
Para:	(None in Experiment I)	The development of neurosis can be explained in large measure by the concept of a limited energy system.

Details

Old:	Galton was the brilliant younger cousin of Darwin	Around 1887, Freud was working with Joseph Breuer, studying the method of free association.

Appendix, Cont.

Para:	Darwin was the older cousin of the extremely intelligent Galton.	Freud learned the method of free association from Joseph Breuer around 1887.
New:	The phrenologists had tried to do the same thing before but failed.	In 1885, Freud spent time with Jean Charcot studying hypnosis as a clinical method.
Para:	(None in Experiment I)	Freud studied hypnosis as a clinical method under Jean Charcot in 1885.

Extraneous Statements

Old:	Isadora Duncan suggested to George Bernard Shaw that they should combine her beauty and his intelligence; Shaw however objected that the child may turn out with his looks and her brains.	Oh, speaking of anxiety, that reminds me. Marcia and I will not be able to answer question between now and next Tuesday.
Para:	Isador Duncan told Bernard Shaw that she wanted a child from him in order to combine her beauty and his intelligence; Shaw, however, was afraid the child might get her brains and his looks.	Oh, speaking of anxiety, I forgot to mention that Marcia and I won't be answering questions until the exams are in on Tuesday.
New:	The Spartans purposefully bred their strongest warrior with their most beautiful maidens, but in the	In case I didn't mention it, Marcia and I will try to have the papers back to you a week from Tuesday.

Appendix, Cont.

New, Cont. end they became just as decadent as the Athenians who had more fun all along.

Para: (None in Experiment I)

Oh, if I didn't tell you before, Marcia and I plan to give you back the papers a week from Tuesday.