COGNITIVE PSYCHOLOGY 11, 221–234 (1979)

The Role of Elaborations in Memory for Prose

LYNNE M. REDER
Carnegie–Mellon University

In two experiments, subjects read stories and were asked to make plausibility judgments about statements with respect to the stories. The inherent plausibility of the queried statements, the amount of attention subjects focused on information necessary for making a judgment, and the interval between presentation of the relevant story information and the test probe were varied orthogonally. The pattern of latencies obtained to make these judgments cast strong doubt on the notion that question answering is typically accomplished by searching for a single fact in memory. Rather, people seem to retrieve any relevant, available information and then use this to compute whether a statement seems true. The independent variables in these experiments can be interpreted according to whether they affect the retrieval or the judgment phase.

The experiments to be described set out to investigate how people make plausibility judgments about material that they have studied. One model of how assertions are judged plausible is that memory is searched for the proposition in question; if it cannot be found in memory, and only then, will an individual use other stored information to judge the probe's plausibility. This view of the plausibility judgment process can be discerned from the work of a number of theorists (e.g., Anderson & Bower, 1973; Kintsch, 1974).

A second model postulates that a person computes the plausibility each time a probe is presented. This model is more parsimonious in that one procedure is evoked in all cases. However the first model may be more efficient by being faster in those cases where direct retrieval can be used. Few reaction time studies have been conducted which ask subjects to judge plausibility. This explains the lack of data necessary to discriminate these two models.

The studies reported here contrast these models by using manipulations that are predicted to have different effects, depending on which model is a

This report is based on a thesis submitted in partial fulfillment of the Ph.D. requirements at the University of Michigan. I would like to thank the members of my dissertation committee: James G. Greeno, Gary M. Olson (cochairmen), David H. Krantz, Robert K. Lindsay, and Stephen Stich. I am also grateful to Karl Haberlandt and Jane Perlmutter for comments on the manuscript and especially to John R. Anderson for much encouragement and support.

This research was facilitated by a National Science Foundation graduate fellowship, a Horace H. Rackham Dissertation Grant, an NIMH Post Doctoral Fellowship, and National Science Foundation Grant BNS76-00959. Requests for reprints should be sent to Lynne Reder, Department of Psychology, Carnegie–Mellon University, Schenley Park, Pittsburgh, PA 15213.

0010-0285/79/020221-14$05.00/0

more accurate description of the true processes involved. The Logic of Additive Factors (see Sternberg, 1969) was used to interpret the pattern of latencies obtained. The subjects' task involved reading stories that were occasionally interrupted by questions that had to be answered about the story. These questions often required that the subjects draw inferences from the story. At the end of each story, the subjects were asked more questions about the story.

Three factors were manipulated, all thought to affect the processes used in plausibility judgments. The first factor was *type of inference* used to probe memory. The inferences were all plausible, but they varied in degree of plausibility as rated by an independent group of subjects. The three types of inferences used as probes were inferences of medium plausibility, inferences of high plausibility, and verb-based inferences. These types of probes will be described in detail later.

Type of inference queried might be expected to affect the time to make a plausibility judgment. This could be true for one of two reasons: (1) Some types of inferences may be spontaneously inferred more frequently than others. If a queried inference has been stored in memory, response time may be shortened because computation of plausibility can be avoided by direct retrieval; (2) alternatively, if no inferences are made in advance, or if probed inferences are not specifically searched for, then inference type could affect the time needed to *judge* plausibility.

A second factor of the experiment was the *delay* at which a question was asked. This could affect latencies in that information relevant to question answering may be lost with time, thereby slowing down search. On the other hand, delay might speed responses if subjects spontaneously generate inferences relevant to the question during the delay.

The third factor called *prior exposure*, manipulated the attention subjects gave to probe-related material prior to judging the probe. The three levels of prior exposure involved either: (a) presenting, as part of the story, the statement that was later to be judged; (b) priming the to-be-judged statement by asking a second, highly related question; or (c) neither presenting, nor priming the statement. The prior exposure factor, too, was expected to affect the time to judge plausibility. It could affect the likelihood of the queried statement being stored in long-term memory during reading of a passage. If a statement is stored in memory, and retrieved at test, the latency could be reduced for a "plausibility judgment" by avoiding the computation of plausibility.

EXPERIMENT I

Method

Procedure and Design

Subjects read 11 stories, the first one practice. They read each story at their own pace. A story was initiated by a button press. The first sentence of the story was then displayed on a computer-controlled video terminal. The subject pressed the button again after reading the

sentence in order to have it immediately replaced by the following sentence of the story. In this way, times to read sentences were also surreptitiously collected. Periodically, the story was interrupted by one or two questions. Instead of the next sentence appearing, a warning appeared, indicating that a question followed: "QUESTION—TRUE OR FALSE?" A quarter second later, the question was displayed below the warning. The probe and warning statements were displayed at a different position on the screen from the sentences of the story so that they could be easily distinguished from lines of the story. Subjects were told that they were to respond true if the statement seemed consistent with the story and false otherwise. After the subject responded true or false by pushing one of two buttons, feedback was given as to whether the response was considered correct. The story then resumed. After the story was completed, the subject was advised that more questions would follow. Feedback was also given for the delayed questions. After these were answered, the screen advised the subject to push a button when ready to initiate the next story.

The experiment used a factorial design with three factors: (1) inference type (verb-based, high-plausible, and medium-plausible statements); (2) prior exposure of inferences (presented, primed, or neither); and (3) delay of test (immediately and after the story). Examples of each type of inference and each type of prior exposure are provided in Tables 1 and 2, respectively. More discussion of inference types and the priming manipulation is given below under Materials. The prior exposure condition in which information was neither primed nor presented was called the "not-presented" condition.

Materials

Ten stories written by five different authors were selected for the study. (One additional story was selected for "practice," however, the following information does not apply to it.) An example of the stories used is given in the Appendix. The stories were all of approximately equal length. Since the probe sentences and priming sentences were the critical materials, their construction and selection will be discussed in detail.

A. Types of inference statements or probes. Prior to Experiment I, 20 different subjects read the same 10 stories and, for each story, generated continuations or elaborations at specified points. Approximately 9 or 10 of the sentences in each story were followed by a mark that indicated to the subject to write one inference or elaboration based on the input preceding the mark. These generated inferences (generations) were sorted into groups that expressed the same basic idea and each basic idea was scored on its frequency of production.

A certain class of inferences that tends to be generated infrequently, verb-based inferences, was constructed by the experimenter prior to scoring.[1] For example, "the teacher contacted Mary's face with her hand" is a verb-based inference that follows from "the teacher slapped Mary." It seemed unlikely that subjects would write down this type of inference. Nonetheless, a tally was taken to see how often the experimenter's inferences were generated by others and note was taken of any other verb-based inferences generated.

From the pool of subject-generated statements and verb-based statements, a subset was selected for rating: one statement of high generation frequency, one statement of moderate frequency, and one verb-based inference for each predetermined stopping point in each story. The verb-based statements of highest frequency within a story were always selected, although occasionally a verb-based statement had to be selected that was generated by the investigator only.

Another group of 20 subjects rated these selected statements on their plausibility with respect to the same 10 stories. They read each story twice. On the second reading the subject stopped at the appropriate points and rated the plausibility of each of the three selected statements.

From the pool of generated inferences that had been rated, a subset was selected for use in Experiment I. Nine statements were selected for each story, three verb-based, three highly

[1] For motivation for this class of inferences, see Reder (1976).

TABLE 1

AN EXAMPLE FROM EACH LEVEL OF INFERENCE TYPE

I. Verb-based inferences
From the line: "The heir told his father he wanted no part of his greasy food fortune."
Inference: "The heir communicated with his father."

II. High-plausible inferences
From the lines: "The heir decided to join Weight Watchers. Twenty-five pounds later, he realized his wife did love him after all."
Inference: "The heir had lost weight."

III. Medium-plausible inferences
From the line: "Now he worried that she had been after his money all along."
Inference: "The heir had not worried about her motives before marriage."

plausible, and three moderately plausible. The verb-based inferences tended to be low in frequency of generation, but high in plausibility. The so-called "high-plausible" statements were high in frequency of generation as well as plausibility. "Medium-plausible" statements were moderate in both frequency of generation and rated plausibility; however, medium-plausible statements were still quite plausible. All types of probes were modified to make sure that their referents were clear and so that the mean number of letters per sentence was constant over inference type (high, medium, or verb-based). No two statements were

TABLE 2

AN EXAMPLE FROM EACH LEVEL OF PRIOR EXPOSURE

I. Not presented
Read in Story: "Anyway, real marital strife lay elsewhere. His wife had never revealed before marriage that she was an intellectual, that she read books."
Test: "The heir did not like the fact that she read books."

II. Primed
Read in story: same as in the not-presented case.
Priming question: "The heir was delighted that she joined the book of the month club." (false)
Test: "The heir did not like the fact that she read books."

III. Presented
Read in story: "Anyway, real marital strife lay elsewhere. His wife had never revealed before marriage that she was an intellectual, that she read books. The heir did not like the fact that she read books."
Test: "The heir did not like the fact that she read books."

selected from the same point in a story, i.e., only one of the three statements at a specified point could be selected. (See Table 1 for an example of each type.)

B. Priming questions and falses. Priming questions (primes) were designed to force the subject to think about information relevant to a particular probe and perhaps to infer that statement. One priming question for each critical probe was written by the investigator. Before a prime was selected for use, three informants had to indicate that the priming question seemed to have the desired effect of making them infer the relevant statement. An example of a prime is given in Table 2. The priming questions were both true and false.

In the immediate-primed condition, the priming question and the probe to be primed were presented successively during the story. Half the time, the priming question was followed by the critical probe. When the primed inference was tested at a delay, the priming question was still asked during the story but it was followed by a filler question. The filler question was always false so that subjects could not assume that when two questions were asked in a row during the story, the second question would be true. The fillers did not prime any information that would be queried.

The probes were randomly assigned to prior exposure and delay conditions for each subject. Stories were yoked into pairs because there were 18 conditions ($3 \times 3 \times 2$) in the design, but only nine inference probes per story. Each subject read all stories.

Subjects. Forty-four students from the Human Performance Center, paid subject pool were recruited. They were paid $2.50 for approximately 1/2 hour of work. Groups of one to four subjects were run independently by computer.

Results

Figure 1 plots latencies in milliseconds for correct judgments and the error rates for the plausibility statements. The error rate for falses in both the immediate and delay conditions was approximately ten percent.

An analysis of variance was performed on subjects' means for each condition. The error term used was always the interaction of subjects with the effect(s) of interest. The standard error of the means in Figure 1, (based on the overall interaction of subjects with conditions) was less than 59 msec. The differences in reaction time due to delay was significant, $F(1,43) = 140.1$, $p < .001$, with the immediate condition faster. Prior exposure had a significant effect as well, $F(2,86) = 174.5$, $p < .001$, presented always fastest and not-presented always slowest.

In this experiment, variability due to materials (specific questions used) can be considered as part of the variability due to subjects because probes were randomly assigned to the various delay-by-prior exposure combinations for each subject. Effects of delay and prior exposure are not vulnerable to the statistical criticisms made by Clark (1973). On the other hand, test probes could not be randomly assigned to levels of inference type since this factor was defined by the statements themselves. Therefore, results that depended upon the inference effect were tested with the conservative min F' prescribed by Clark. The effect of inference type was still significant F' $(2,100) = 3.4$, $p < .05$, with high-plausible statements being faster than the other two.

It should be noted that there was a significant two-way interaction of prior exposure with delay, $F(2,86) = 35.8$, $p < .001$. The presented condi-

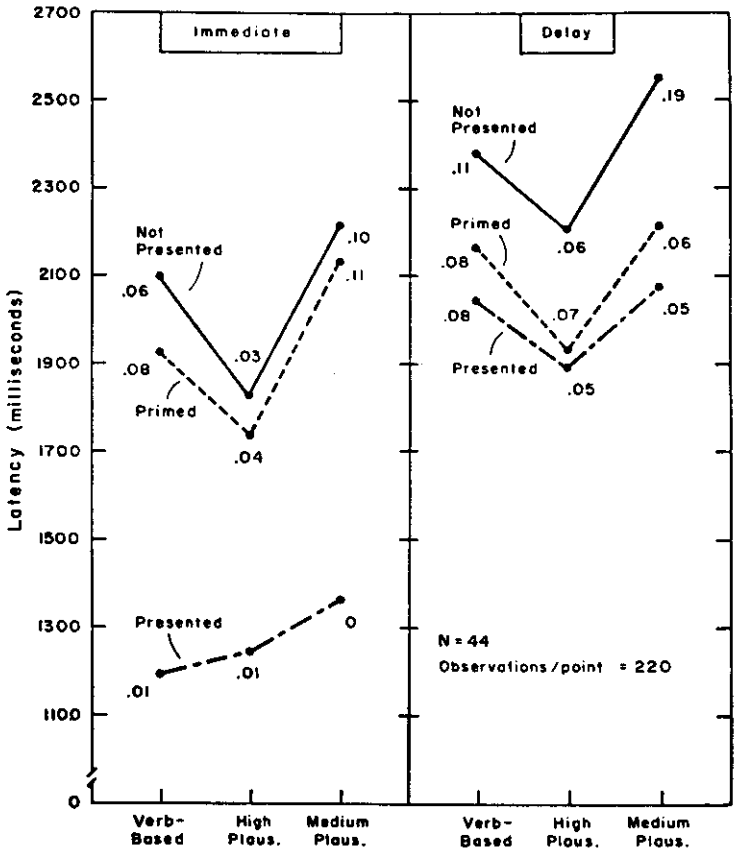

FIG. 1. Mean latencies for correct plausibility judgments (and error rates) in Experiment 1 as a function of inference type, prior exposure, and delay.

tion initially was much faster than the other two, but at a delay, the presented statements took about as long as the primed. The fact that the presented condition was initially much faster is not very surprising given that in the immediate condition the subject was shown the same sentence twice in a row. However, the interaction of prior exposure with delay is not solely attributable to a big slowdown in the presented statements. The difference between primed and not presented grew with delay, $t(132) = 2.3, p < .05$. Inference type, on the other hand, did not interact with prior exposure or delay.

The conclusion that inference type affected time to answer questions even when the statements were actually presented is derived from the significant effect of inference type and the lack of interaction between inference type and prior exposure. This conclusion was supported by a special contrast (high plausible minus the average of verb-based and

medium plausible) that showed that there is a significant effect of interence type in the delayed-presented condition, $t(172) = 2.32$, $p < .05$.

Discussion

I found the pattern of data somewhat surprising and unintuitive. I had expected the main effects of inference type, of delay, and of prior exposure; however, I also expected inference type to interact with the last two, which did not occur. For example, it is surprising that the presented condition was significantly different from a flat line and showed the same effect of plausibility as the other prior exposure conditions.

One might wonder whether the effect of inference type is attributable to some factor other than ease of deciding plausibility, e.g., reading time, or some other encoding process.[2] An experiment was conducted using the same materials, in which subjects were only required to read the critical probe sentences out of context. (After reading each unrelated statement, the subject pushed a button to signal that it had been understood. After each statement was read, a question was asked about it to insure that the subject had really processed the sentence.) The times to read the probes out of context did not show the plausibility effect typically obtained. Thus it is unlikely that the inference effect is due to reading times, per se (see Reder, 1976, Chapter V, for a fuller discussion).

A model will be described later that accounts for the obtained main effects. It also accounts for the greater advantage of primed statements over not-presented statements in the delay condition. Since the results are somewhat counterintuitive and the proposed model is post hoc, prudence demanded that a second study be conducted to insure that the results would replicate. Further, the model developed to account for the obtained data makes a prediction which can be tested in a modification of the original experiment. The derived prediction to be tested was that the advantage of priming would continue to grow with longer delays. Experiment II is quite similar to Experiment I except that it has a third level of delay, such that some questions are asked 48 hr after the story is read.

EXPERIMENT II

Method

Procedure and Design

The procedure differed from Experiment I in the following ways: Before a subject started reading a story, a title for it would appear on the screen. The subject would press a button to have the first sentence appear after reading the title. The probes now had to be divided over

[2] The only precaution in material construction to provide comparability in reading ease was to insure that average length of sentence be constant across inference types. Other constraints in material construction were not feasible since there were so many other constraints in selecting among the elaborations. Had the items been generated by the experimenter rather than by subjects, more standardization would have been possible, but at the expense of the frequency measure.

three delay periods rather than two so there were one-third fewer at each delay.[3] Because all priming questions were asked during the story, two-thirds of the time the priming questions were followed by filler falses and only one third of the time by true probes (in Experiment I, there were equal numbers). After all stories were read, subjects were told to come back 48 hr later. They were not told why they were to come back.

In the second session, subjects were told that they were to make the same type of judgments that they had made previously with respect to the same stories. Each set of questions was preceded by the relevant story's title which they had read earlier. They were told not to push the button that advanced from the title to the questions until they remembered the story referenced. The practice story was tested first.

The only difference in design from Experiment I was an added level of delay, yielding a 3 × 3 × 3 factorial. Because each story still had only nine rated inferences, stories were yoked into triples rather than pairs as had been done in Experiment I. Only 9 of the 10 stories could be used to make up the triples. This meant that there were only three independent story-triples with one observation per story-triple in each condition. With 42 subjects, this yielded 126 observations per point. As before, inference probes were randomly assigned to conditions for each subject.

Subjects. Forty-two subjects from the Human Performance Center paid subject pool participated in the experiment. The experiment took approximately 20 min in Session I and 10 min in Session II (the 48-hr delay). Subjects were paid $4.00 for their services.

Results

As before, only the mean latencies for correct responses to the relevant conditions were analyzed. These data and the error rates are displayed graphically in Fig. 2. The error term used was always the interaction of subjects with the effects of interest. The standard error (based on the overall interaction of subjects with conditions) was 106 msec and consequently the data points are somewhat less reliable than in Experiment I. The graph corresponds to the one for Experiment I except that there are now three panels for the three levels of delay.

The difference in reaction time (RT) due to delay was significant, $F(2,82) = 61.9, p < .001$, as was the effect due to inference type, $F(2,82) = 20.5, p < .001$, and due to prior exposure, $F(2,82) = 7.6, p < .001$. As before, there were no significant interactions with inference type. The two-way interaction of prior exposure with delay was significant, $F(4,164) = 11.0\ p < .001$.

As expected by the model to be described below, the difference between primed and not-presented items grew with delay. The contrast for an increasing trend (contrasting the difference in the immediate condition with the difference at a 48-hr delay) was significant, $t(328) = 2.3, p < .025$. The effect of the priming manipulation was attenuated, however. In the immediate condition, primed items were actually slower than not-presented items, although not significantly, $t(328) = 1.3, p > .05$. At the 2- to 3 min delay, there was no difference between primed and presented conditions. At the third delay the primed inferences were faster than the

[3] This figure varies slightly because assignment of statements to conditions was random within each story-triple for each subject.

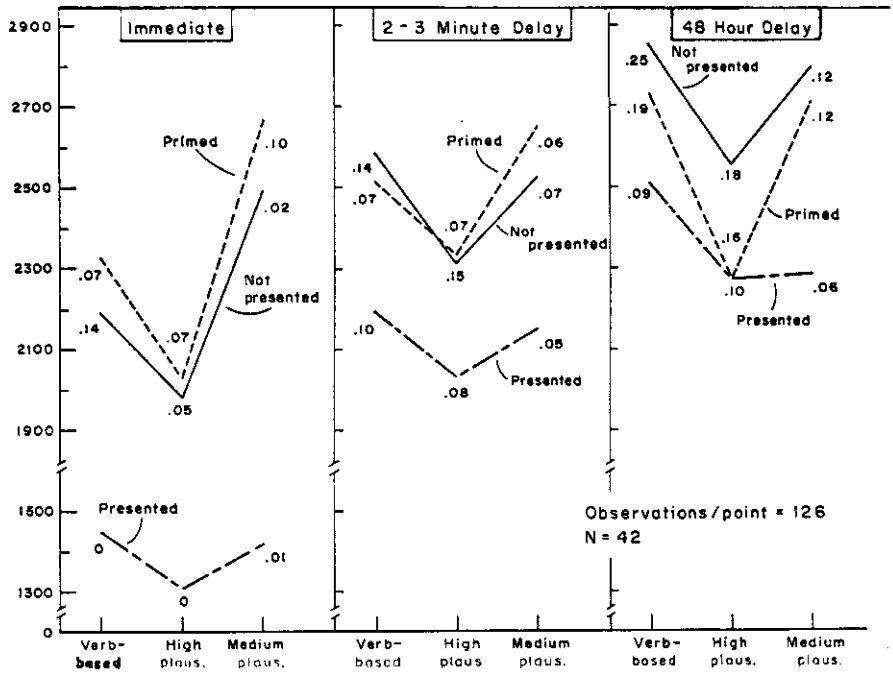

FIG. 2. Mean latencies for correct plausibility judgments (and error rates) in Experiment II as a function of inference type, prior exposure, and delay.

not-presented ones, $t(328) = 1.93$, $p < .05$, one tailed. This difference at a 2-day delay is smaller than the difference at the shorter delay in Experiment I. An explanation for the attenuated effect will be given later.

Discussion

A number of results from these experiments should be highlighted. Inference type did not interact with prior exposure, yet both factors had an effect on latencies. Delay also affected latencies. Since statements in the not-presented condition took longer to judge than the primed statements, it is not reasonable to conclude that subjects infer nothing, that they store only the input sentences. Priming must have caused more inferences or elaborations to be made. It is also *not* reasonable to conclude that statements probed were always inferred during reading and then retrieved when needed. If they were, there would have been no effect of prior exposure and probably no effect of inference type.

One might want to argue that differences in inference type reflect differences in time to retrieve the specific proposition, that it takes longer to retrieve a medium-plausible statement from memory than a highly plausible one. This notion does not seem viable for several reasons: First, in the not-presented condition RTs are so much slower, yet the effect of inference type is the same as in the primed and presented conditions; second,

the effect of inference type is as large in the immediate condition as at a delay. That means, to account for the inference-type effect as a difference in retrieval time, one would have to posit that the difference in accessibility of high-plausible versus medium-plausible statements is the same immediately as at a delay. This seems untenable because there is relatively little trouble in finding information immediately, as evidenced by the overall faster RTs at the short delay. When there is little trouble in finding information, the difference in retrieval time for highly plausible statements compared with moderately plausible statements should be considerably less. However, if anything, the effect of inference type is greater in the not-presented immediate condition than in the not-presented delay condition.

The conclusion that some inferences are made during comprehension is consistent with previous findings which suggest that we store more than the input (e.g., Bransford & Johnson, 1973; Sulin & Dooling, 1974). However, one might have thought that subjects simply retrieve the appropriate inference, if it is stored, to judge if a statement is true of a passage. It may be that readers are more likely to store highly plausible inferences during reading, but if so, they do not "take advantage" of this, by retrieving it, at test. One might have thought that at least presented "inferences" would be retrieved at test, yet these statements also show the effect of inference type, i.e., the presented condition lines in Figs. 1 and 2 are not flat.[4]

The fact that even presented statements showed the inference effect might lead one to speculate that subjects are not storing the presented inferences. This possibility has been ruled out by a recognition memory experiment using basically the same materials (Reder, in preparation). The task involved subjects reading stories with some plausible inferences inserted in the text. At test, the story contained not only the "old" plausible inferences, but some "new" inferences as well. All inference statements were now underlined and subjects had to discriminate the previously presented "inferences" from the new ones. The assignment of which statements were to be "old" and which were to be "new" inferences was counterbalanced across subjects. Subjects would rate from +3 to −3 how confident they were that the underlined statement had been presented initially, +3 indicating certainty that the sentence had been presented before, −3 indicating certainty that it was new. The mean rating for old sentences was 1.75; the mean rating for new was −1.89. Not only is this difference large in magnitude, it is also highly significant—all 20 subjects showed the difference. It therefore seems clear that subjects do store the presented sentences.

[4] The presented lines are not significantly flatter, but the curves do seem slightly attenuated. This can be explained by a slightly greater chance of finding an exact match during search in this condition. Finding an exact match obviates the judgment stage (see discussion of the two-stage model), but this is not thought to happen often.

A TWO-STAGE MODEL FOR PLAUSIBILITY JUDGMENTS

The Additive Factor Method promoted by Sternberg (1969) was applied to the reaction time data in order to infer the staging of the cognitive processes involved in the plausibility judgment task. Since inference type did not interact with prior exposure, yet both had main effects, they are assumed to affect different stages of processing. Since delay and prior exposure did interact, they are assumed to affect the same stage. The model that seems most parsimonious and viable consists of first, a retrieval stage in which information relevant to making a judgment is retrieved, and second, a plausibility judgment stage in which the plausibility of the probe is computed from the retrieved information. The retrieval stage is affected by prior exposure and delay; the plausibility judgment stage is affected by inference type. Subjects retrieve information that will be relevant to making a judgment rather than trying to retrieve the exact statement queried. Since many questions asked of subjects are not about statements they are likely to have stored, and since they are only asked to decide if the statement is plausible, not if it was presented, it is reasonable that they would behave in this fashion.[5]

The notion of computing plausibility rather than retrieving verbatim statements is similar to the reconstructive position most often associated with Bartlett (1932). In remembering, one pieces together fragments of many memories to induce what else must be true. One's world knowledge is employed in "putting the pieces together," i.e., figuring out what is probably true. One difference between my position and Bartlett's is that he believed that a subject only reconstructs if a proposition is not stored in memory. I believe that even though the statement may be in memory, the subject still tries to reconstruct rather than search for an isolated proposition; the latter strategy would be too inefficient. Note, this view predicts that subjects should be faster to make plausibility judgments than to make presented versus not-presented judgments.

Prior exposure is assumed to affect retrieval by manipulating how much relevant information is available from which to sample. The duration of the retrieval stage depends on the density of relevant to irrelevant propositions associated with the concepts in the probe. It is assumed that propositions are represented in a network structure, where nodes correspond to concepts, and arcs are relational connections between concepts (cf. Anderson, 1976; Collins & Loftus, 1975). When the test question is presented, it is first parsed. The parse causes the concepts of the question to be activated. Search for relevant information is carried on in parallel from all concepts stated in the probe. Since, for this task, no one particular proposition need be found, any subset of the relevant propositions can

[5] If, in searching for relevant information, a subject finds the exact proposition, he may decide to skip the second stage and respond plausible on the basis of the match. The data indicate that this happens infrequently.

satisfy the criteria of the retrieval stage. Thus, the more relevant propositions there are, the faster a given number of them can be found in a group of relevant and irrelevant ones. These relevant propositions most likely include some that were not explicitly presented in the story; priming the statement or presenting it encourages the subject to attend to a specific part of the story and allows him or her to generate more relevant elaborations. Retrieval of the requisite number of propositions will be faster in these two conditions because there are more relevant propositions.

With delay, subjects get slower at making judgments because they lose some of their relevant (and irrelevant) story elaborations. The approximate number of preexperimental propositions attached to the concepts of a probe stays constant while the amount of useful information for the retrieval stage grows smaller with delay. Therefore, the advantage of additional elaborations is thought to increase with delay: The fewer the number of relevant propositions remaining, the more valuable each remaining one becomes. (Presented statements are much slower at a delay than in the immediate condition because they no longer have the advantage of the same statement presented twice in succession.) A more detailed, quantified model showing how the increased advantage is predicted is given in Reder (1976). A discussion of the plausibility judgment stage is also given there.

FURTHER ANALYSES

This simple model with relatively few assumptions can account for the basic pattern of data from the two experiments. However, there remains a discrepancy between Experiments I and II concerning the attenuation of the effect of priming. One explanation for the attenuation that was supported by a postexperimental test posits that subjects are elaborating less in the priming condition in Experiment II. Perhaps the manipulation intended to focus attention on the to-be-queried information was less effective because there was less motivation to attend to the primed material. In Experiment II, the proportion of questions followed by a second relevant question (as opposed to a filler false or no question at all) was less than in Experiment I. In other words, in Experiment I, subjects probably learned faster that it was worthwhile to attend to the priming questions.

This explanation suggests that subjects in Experiment I initially should not realize the utility of expending extra attention on primed information any more than the subjects in Experiment II. After a few stories, however, subjects in Experiment I should be focusing more attention on prime-related sentences since they kept being asked, during and immediately after the story, questions relating to the priming question. With this in mind, the data from the 10 stories in Experiment I were divided in half. The data from the first five paragraphs that subjects read (call these the

first half) should show an attenuated priming effect similar to the pattern in Experiment II. (Subjects initially should make more elaborations in the primed conditions than other conditions, but not a lot more.)

The results were as predicted. The average difference between primed and not presented in the first half of the experiment was 80.5 msec (−28 msec in the immediate condition—primed slower than not presented—and 189 msec., primed now faster in the delayed condition); in the second half of the experiment, the difference was 279 msec (169 msec, primed faster than not presented immediately, and 389 msec at a delay.)

CONCLUSION

It is not the case that inferences are made during comprehension in order to avoid making them when the material is tested. Nonetheless, many inferences and elaborations are generated while reading. A subset of these and of the input propositions are later retrieved in order to reconstruct the propositions and to compute whether a statement seems true. This is not to say that subjects could not display good memory for the exact input statements if forced to do so. I reported data showing that subjects can discriminate presented "inferences" from implicit ones.

The reason subjects do not hunt for the exact proposition is that it would be less efficient to do so. Even if every plausible probe to be judged was stored in memory, a search strategy would often fail. Half of the probes are implausible. Further, there are many other propositions in memory that would have to be sifted through to find the exact proposition. I suspect that many elaborations that are generated have subpropositions in common, i.e., are partially redundant. A related inference would be a hindrance in a task of searching for the *exact proposition* [like a Collins and Quillian-(1969) spurious connection]. The strategy used by subjects to compute plausibility takes advantage of the related inferences. The more relevant elaborations there are, the faster a subset can be selected for the judgment stage (no particular one needs to be used).

Finally, I want to mention briefly the usefulness of using a plausibility judgment task in reaction time experiments. If the task had asked subjects to decide if a statement had been *presented*, subjects would have been forced to answer the questions by using a fact retrieval process. This would have obscured the fact that other strategies are often used in everyday judgments, and we would not have recognized the role that elaborations can play in memory.

APPENDIX

Example Story (written by L. Reder) Used in Experiments I and II

The heir to a large hamburger chain was in trouble.
He had married a lovely young woman who had seemed to love him.
Now he worried that she had been after his money all along.

He sensed that she was not attracted to him.
Perhaps he consumed too much beer and french fries.
No, he couldn't give up the fries.
Not only were they delicious, he got them for free!
Anyway, real marital strife lay elsewhere.
His wife had never revealed before marriage that she read books.
Sometimes she used words that were many syllables long.
The proud husband decided that she was showing off.
At least, he thought, she stayed at home.
It is not too late, he resolved.
The heir decided to join Weight Watchers.
Twenty-five pounds later, the heir realized his wife did love him after all.
He vowed never to eat another french fry.
He also told his father that he wanted no part of his greasy food fortune.
The wife of the ex-heir smiled as they went jogging into the sunset.
Tonight she would teach him to read.

REFERENCES

Anderson, J. R. *Language, memory, and thought*. Hillsdale, NJ: Lawrence Erlbaum, 1976.

Anderson, J. R., & Bower, G. H. *Human associative memory*. Washington, D.C.: Winston, 1973.

Bartlett, F. C., *Remembering: A study in experimental and social psychology*. London: Cambridge Univ. Press, 1932.

Bransford, J. D., & Johnson, M. K. Considerations of some problems of comprehension. In W. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973.

Clark, H. H. The language-as-fixed-effect-fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 335–359.

Collins, A. M., & Loftus, E. F. A spreading-activation theory of semantic processing. *Psychological Review*, 1975, **82**, 407–428.

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 240–247.

Kintsch, W. *The Representation of meaning in memory*. Hillsdale: Lawrence Erlbaum, 1974.

Reder, L. M. *The role of elaborations in the processing of prose*. Unpublished doctoral dissertation, Univ. of Michigan, 1976.

Sternberg, S. Memory scanning: Mental process revealed by reaction time experiments. *American Scientist*, 1969, **57**, 421–457.

Sulin, R. A., & Dooling, D. J. Intrusion of a thematic idea in retention of prose. *Journal of Experimental Psychology*, 1974, **103**, 255–262.