2. RESPONSE LATENCIES TO EXPLICIT AND IMPLICIT STATEMENTS AS A FUNCTION OF THE DELAY BETWEEN READING AND TEST[3]

The most important result of Exp. I and II is the question-delay effect: Explicit-version questions were answered faster than implicit-version questions when the questions were asked immediately but not when they were delayed by 15 min. Exp. III is an attempt to replicate and extend this finding.

To explain the question-delay effect, we hypothesized a surface structure memory for prose. This memory was short-term because it was lost before 20 min, but is it similar to the traditional short-term memory which is lost before 30 sec? To answer this question, a delay condition of 30 sec was included in the design with 0- and 20-min delays.

The fact that explicit-version and implicit-version questions were equally time consuming after 20 min was interpreted as evidence that both types of questions were answered from the same long-term memory representation. If the explicit and implicit information is indeed stored in the same memory representation then, with longer delays, explicit-version and implicit-version questions should continue to require equal amounts of time to answer. To test this prediction, the design included a 48-hour delay condition.

Experiment III was intended to extend the question-delay effect to new materials as well as to new time intervals. The paragraphs of Exp. I and II were all relatively short. If the assumed surface structure memory shared the limited-capacity characteristics of traditional short-term memory, then the immediate reaction time difference between explicit and implicit questions might disappear with longer materials. Experiment III paragraphs were therefore much longer than the paragraphs of the first two experiments.

The paragraphs of Exp. I and II were all similar to each other in style and construction. These paragraphs, labeled descriptive, formed listings of statements, each statement describing the topic. The sentences were related to each other mainly by conjunctions, so that sentence order was relatively unimportant. Experiment III included paragraphs with a different structure, which presented arguments in support

[3]By G. McKoon and J. Keenan.

of an idea or an explanation of a process. The sentences of these argumentative paragraphs were very strictly ordered by causation and implication, both with respect to each other and to the topic. Given the same number of surface structure words and the same number of base structure propositions, the more complex relationships of the argumentative paragraphs should require more encoding time than the simpler descriptive paragraphs. Will this increased amount of processing erase the surface structure memory and with it the question-delay effect, or is this effect a general characteristic of prose comprehension?

Method

Subjects. Forty undergraduates fulfilled a course requirement by participating in two one-hour sessions.

Design. Each subject read eight paragraphs, four descriptive and four argumentative. Concerning each of these paragraphs, the subject answered four questions: one true-explicit, one true-implicit, one false-explicit, and one false-implicit. The questions corresponding to a particular paragraph were either presented immediately after the subject read the paragraph, or after a delay of 30 sec, 20 min, or 48 hr. The dependent variables were response time per question and number of errors per question in a $2 \times 2 \times 2 \times 4$ completely within-subject factorial design. The factors were paragraph type (descriptive or argumentative), question representation (explicit or implicit), question truth (true or false), and paragraph-question interval (0, 30 sec, 20 min, or 48 hr). The false questions were included in the design for counterbalancing purposes only; no hypotheses about the process of answering false questions were proposed.

Materials. Unlike the first two experiments, the experimental paragraphs were chosen from literary sources, rather than constructed by the authors. The eight paragraphs discussed topics unfamiliar to the subjects. The four descriptive paragraphs were chosen from *American Heritage* and *Saturday Review*; these paragraphs described P. T. Barnum's midgets, a district of Athens, an early Arapahoe reservation, and the history of the piano. The number of base structure propositions in the descriptive paragraphs varied from 56 to 59 and the number of surface structure words varied from 116 to 155. The argumentative paragraphs, which represented base structure propositions in a more complex style than the descriptive paragraphs, were drawn from the writings of Bertrand Russell and from a scientific text. These paragraphs offered an argument that history cannot be viewed as a science, an argument that mathematics and logic are historically distinct disciplines, an explanation of Aristotle's claim that man is rational, and an explanation of a biochemical process. Number of underlying propositions varied from 55 to 60 and number of words from 147 to 160.

The paragraphs of this experiment differed from the paragraphs of the first and second experiments in one especially important respect: The explicit-implicit variable applied to questions and not to paragraphs. An explicit question concerned propositions explicitly stated in the surface structure while an implicit question concerned propositions of the base structure not explicitly stated in the surface

structure. Both kinds of questions were obtained from the same version of each paragraph, so that the explicit questions concerned different propositions than the implicit questions. Thus in this experiment, a particular question must be always explicit or always implicit, whereas in the first two experiments the same question could be either explicit or implicit depending on which version the subject read. Note that every paragraph contained some implicit propositions; there were no paragraphs in which all base propositions were realized in the surface structure.

Four true-false questions were constructed for each paragraph. Explicit questions were copied directly from a sentence or clause of the paragraph, preserving meaning and word usage. An explicit question obtained in this manner was falsified by negation or by the substitution of some incorrect noun from the paragraph. The types of inferences represented in the implicit questions varied as in Exp. I and II, except that none were Level-1 inferences; no attempt was made at categorization, as opposed to Exp. I. In every case, the inference was required for understanding of the paragraph, but not explicitly stated in the surface structure. For example, one descriptive paragraph stated that "when the Arapahoe first settled on a reservation they knew little of the white men's culture" and that "stairsteps, built to take people up to a house on top of another house, amused them." The unstated but necessary inference is that the Arapahoe had never used stairsteps, and the resulting false-implicit question is "Arapahoes had used stairsteps before white men came." The questions of Exp. III were not limited to one proposition but contained 3 or 4 propositions. The questions varied from 7 to 11 words.

When the paragraph-question interval was 30 sec or 20 min, the interval was filled with a sentence unscrambling task. This task was chosen because it involves processing of linguistic information and thus should prevent rehearsal of the experimental paragraphs. One hundred sentences of 9 to 12 words were constructed and the word order of each sentence was randomized. The sentences did not deal with topics discussed in the experimental paragraphs. The subjects attempted to unscramble the words and form a sentence, completing as many such sentences as possible during the intervals.

Procedure. Except for introductory instructions read by the experimenter, all experimental materials were presented on the screen of a Uniscope keyboard/display unit, 100 model, controlled by a Sigma 3 computer. Slightly below and in front of this screen was a response box displaying three buttons labeled "true," "false," and "finish."

Subjects were tested for two one-hour sessions separated by 48 hr. The order of presentation of the experimental materials during these sessions was determined in the following manner for each subject. First each paragraph was assigned to a paragraph-question interval, one argumentative and one descriptive paragraph to each interval. The subject read the two paragraphs assigned to the 48-hr interval during the first session and answered the corresponding questions during the second session. One of the 30-sec interval paragraphs, one of the 20-min interval paragraphs, and one of the 0 interval paragraphs were also presented with their questions on the first day.

In the second session, the subject read the remaining three paragraphs, followed by questions, and answered questions for the two 48-hr interval paragraphs. The order of the 5 events on each day was randomized. When the paragraph-question interval was 20 min or 30 sec, the interval was filled by the sentence unscrambling task. Five minutes of the unscrambling task followed reading of the 48-hr interval paragraphs on the first day, and preceded answering of the corresponding questions on the second day. The assignment of paragraphs to intervals and subjects was completely counterbalanced. All possible combinations of paragraphs were chosen equally often.

The experimenter began the first session by reading instructions that described the several experimental tasks: reading paragraphs, unscrambling sentences, and answering questions. The subject was asked to read each paragraph carefully for understanding and to press the finish button when he reached the end. The finish button caused the reading time to be recorded and the next task to be presented on the screen. For unscrambling sentences, the instructions emphasized speed. The subject was required to write his solution sentences and then press the finish button; the next scrambled sentence then appeared, unless the sentence-unscrambling interval was over. The subject answered the true-false questions "quickly but accurately" with one index finger on the true button, the other index finger on the false button. When one of these buttons was pressed, response time was recorded and the next question was presented. Each experimental task was preceded by a short instruction which the subject read from the screen and which told him what to do next; for example, "Now unscramble the following sentence" or "Now read the following paragraph." The instruction for questions included one or two words (e.g., "piano paragraph questions") to indicate about which paragraph the subject was to be interrogated. The subject was nevere informed about the length of the delay interval.

The experimental materials of the first session were preceded by practice materials that included each kind of task. No practice materials preceded the experimental materials in the second session, but the instructions were reviewed.

Results

Reading Time. The mean descriptive paragraph reading time and the mean argumentative reading time were computed for each subject. A test of the average difference between these means showed that descriptive paragraphs were read faster than argumentative paragraphs, $t(39) = 4.88$, $p < .01$. The average reading time for a descriptive paragraph was 55.35 sec, and the average reading time for an argumentative paragraph was 62.90 sec. This difference cannot be accounted for by number of base structure propositions because this number was equivalent for the two types of paragraphs. The number of words for descriptive and argumentative paragraphs was equal for only three of the descriptive paragraphs (the fourth one was shorter than either the other descriptive paragraphs or the argumentative paragraphs), but reading times still differed when number of words was controlled. Thus, the finding that descriptive paragraphs are read faster than argumentative paragraphs can be attributed neither to differences in the number of propositions in the base structure of the

paragraphs nor to differences in the number of words. It must be attributed to the more complex structure of the argumentative paragraphs which required more processing time.

Errors. Table 5 shows the percentage of error responses in each condition of the experiment. The table summarizes the data for those questions for which true was the correct answer. The data from questions for which false was the correct answer will not be analyzed here. No experimental hypotheses pertained to these data, since how subjects infer that a sentence is false in the present experiment is rather beyond speculation. Furthermore, the observed pattern of results did not prove to be particularly suggestive.

For each subject the mean number of explicit question errors and the mean number of implicit question errors was computed. The difference between these means was significant, $t(39) = 5.08$, $p<.01$; subjects averaged 1.35 implicit errors (on a total of eight questions), but only .40 explicit errors (also on eight questions). A similar t test on the error means of descriptive and argumentative paragraphs showed that this variable did not significantly affect the number of errors. Analysis of variance on the error means for each question delay interval also showed no significant differences; number of errors did not increase as the interval between paragraph and questions increased.

Question reaction times. Table 5 and Fig. 1 summarize the relevant results. The reaction times are based on correct responses only. Extreme reaction times were

TABLE 5

Error Rates, Mean Response Latencies, and Standard Errors for "True" Responses in Experiment III

Interval	Argumentative paragraphs		Descriptive paragraphs	
	Explicit	Implicit	Explicit	Implicit
0:				
Per cent error:	2	17	12	15
Response time:	4.7	5.4	3.3	3.6
S.E.	.4	.4	.2	.2
30 sec:				
Per cent error:	7	25	2	10
Response time:	4.6	5.8	3.2	4.2
S.E.	.4	.4	.2	.3
20 min:				
Per cent error:	5	17	2	12
Response time:	5.1	5.3	4.3	4.2
S.E.	.4	.4	.3	.2
48 hrs:				
Per cent error:	2	25	5	12
Response time:	5.3	6.6	4.4	4.3
S.E.	.4	.4	.4	.3

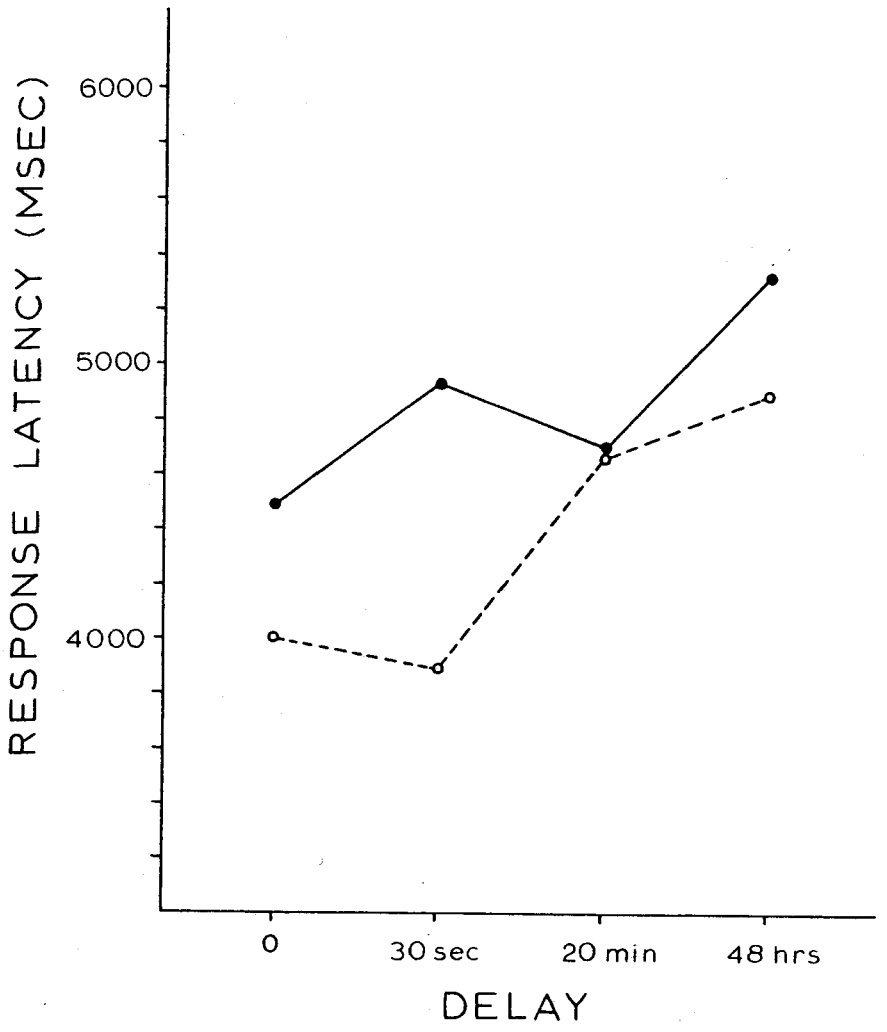

FIG. 1. Response latencies for explicit statements (broken line) and implicit statements (solid line).

discarded if they fell more than two standard deviations above a subject's mean, and a reaction time equal to the second standard deviation point was substituted. Response times lost because of error were replaced by substituting a value proportional to both the row and column mean.

Questions concerning argumentative paragraphs required more time to answer than questions concerning descriptive paragraphs, $F(1,554) = 59.75$, $p<.001$, equally for explicit and implicit questions. No interaction involving the paragraph type variable was significant. Descriptive and argumentative paragraphs involved the same number of base structure propositions and there were no differences in error rate. Hence, one can assume that the memory representations for both types of paragraphs were equally long and equally accurate. The differences in question reaction times must then be attributed to the greater structural complexity of the memory representation of the argumentative paragraphs. Apparently, access time to information in episodic memory depends not only upon the total size of the unit that contains the required information (as was suggested by the differences between long

and short paragraphs in Exp. I and II, and as will be demonstrated more conclusively in Chapter 9) but also on the complexity of the base structure.

Question response time increased as the delay between reading and the test increased, $F(3,554) = 4.78$, $p<.005$. The effect of explicit presentation of a test statement was also significant; implicit questions required more time to answer than explicit questions, $F(1,554) = 9.44$, $p<.005$. The most interesting result, however, concerns the interaction between the delay interval and the presentation mode. It was hypothesized that at delays where a memory representation of the surface features of the paragraph is still available, explicitly presented statements will be answered more rapidly than implicitly presented statements, but that this difference will disappear with longer delays, because responses on the basis of memory for surface features are no longer possible. Specifically, this implies that performance on the immediate test and after a 30-sec delay should be equivalent, and that performance after a 20-min and 48-hr delay should be equivalent, but that the superiority of explicit questions should be restricted to the 0- and 30-sec delays. These predictions were tested by means of orthogonal comparisons. Because the paragraph type variable did not interact with any of the other experimental variables the data were averaged over argumentative and descriptive paragraphs. These averages are shown in Fig. 1. The comparison between the 0-interval condition and the 30-sec delay condition yielded a nonsignificant $F(1,554) = 1.95$. The difference between the 20-min delay and the 48-hr delay also failed to reach statistical significance, $F(1,554) < 1$. However, there was a statistically significant difference between explicit and implicit test statements for the 0- and 30-sec delay conditions, $F(1,554) = 11.40, p<.01$. For the 20-min and 48-hr delay conditions, the difference between explicit and implicit response times was not statistically significant, $F(1,554) = 1.54$. Thus, the question delay effects of Exp. I and II are fully replicated.

Even though it did not affect the statistical analyses very much, there is one deviant result, as can be seen from an inspection of Table 5. The mean response latency for implicit questions from argumentative paragraphs after 48-hr delay is quite a bit longer than any of the other times that have been obtained. Is this an indication of differential decay rates for explicit and implicit information, or is it merely a consequence of selecting a few particularly difficult test statements for that experimental condition? Although there is no way of deciding this question here, one must proceed with caution: Perhaps the data are more complex than the statistical analysis makes them appear.

Discussion

Sentences must often be regarded as incomplete in their coding of the intended content. In this sense, comprehension must always involve some processing of implicit information. Some implicit information that must be processed involves syntactic deletions of only a single case within a proposition. For example, (LARGE,BOX)&(RED,BOX) are not usually realized as *the box is large and the box is red* but rather as *the large red box*, where one occurrence of box is deleted. The question that the present studies were designed to answer is whether a reader will

process implicit information when that information consists of *whole* propositions. The textual conditions under which the question was investigated were such that the processing of successive propositions depended upon the encoding of the implicit proposition.

The answer that Exp. II and III have provided to this question is unequivocally "yes." If subjects are given a paragraph in which one of the base structure propositions is not represented explicitly, this proposition will be inferred during reading and stored in memory if *(a)* the proposition is redundant in the sense that it is possible to infer it on the basis of general knowledge as well as the specific paragraph context, and *(b)* if such an inference is required in order to preserve the continuity of the text. The conclusion that subjects do infer redundant implicit propositions during reading is in agreement with results obtained by Frederiksen (1972), who studied subjects' recall of text, and in particular the kind of errors and distortions that occur during recall. On the basis of such data, Frederiksen concluded that inferences are made during reading, in addition to being formed during the reconstructive activity at the time of recall.

Error rates were higher for implicit material than for explicit material but did not increase with the delay interval in Exp. III. This result supports the interpretation of the data offered here that inferences are made during reading, rather than at the time of retrieval. Inferring a proposition during reading is more difficult than decoding one that is explicitly represented in the surface structure, but once that proposition is stored in memory it is treated like all other propositions.

The difference observed here between descriptive and argumentative paragraphs deserves some comment. These paragraphs were included because it seemed possible that memory for surface features might not be as good for the complex argumentative paragraphs as for the rather simple descriptive paragraphs. For example, surface memory for argumentative paragraphs might contain less information because argumentative encoding time was longer. Or, surface memory might contain only a listing of propositions and not the structure connecting those propositions. Such a listing would be adequate for answering descriptive explicit questions, but not for answering argumentative explicit questions. If this had been the case, the superiority of explicit questions at short delays would have been reduced or absent for the argumentative paragraphs. Instead, the question-delay effect was as great for these paragraphs as for descriptive paragraphs.

Both reading times and question response times were longer for argumentative paragraphs than for descriptive paragraphs. These results show that the argumentative paragraphs were indeed more complex semantically. If the two paragraph types had differed only in surface or syntactic complexity, then reading times would have differed as they did, but question response times would have been equal (see Chapter 5). In other words, if the two paragraph types had the same kind of memory structure, then times to access information in those structures should also be the same. But access time for argumentative paragraphs was longer than access time for descriptive paragraphs both in long-term memory and in surface memory.

How do the structures of the two paragraph types differ? They do not differ in the

size of the memory representation, because number of surface words and number of base propositions were controlled. The difference must therefore lie within the memory representations. Either the argumentative propositions themselves are more complex (e.g., more abstract) or the argumentative organization is more complex. If the organization is more complex, then any single proposition would be more difficult to locate. Thus argumentative access times would be longer than descriptive access times. This same interpretation, that the propositions of an encoded paragraph are not content-addressable, described the paragraph length effect in Exp. I and II. Other interpretations are, of course, possible; for example, argumentative paragraphs are more abstract. But however the argumentative-descriptive differences are interpreted, the important result remains: The question-delay effect applies to both types of text.

The most interesting result of the studies reported here was an unexpected one: the failure to observe the predicted results in Exp. I, and for the short delays, in Exp. III, which we interpreted to be a consequence of the fact that memory for text is a multi-level phenomenon. Specifically, we have observed two different memory representations: a relatively stable propositional representation of the meaning of the paragraph, which was the same whether or not redundant propositions were explicitly expressed in the text, and a less stable memory for surface features of the text, perhaps for the actual words used, which was available only right after reading the paragraph or after a 30-sec delay. Twenty minutes of reading other material effectively interfered with this memory for surface features and forced subjects to rely upon the propositional memory. Access time to surface-type information was faster than access time for propositional-type information. The latter was approximately equal for explicitly presented propositions tested after a delay, and implicitly presented propositions, irrespective of delay. For the data in Fig. 1, the average response time for explicitly presented material with no (or 30-sec) delay was 3.9 sec, while for the same material a response time of 4.8 sec was observed after a delay, which is quite comparable to the 4.7 and 5.0-sec response times for implicit propositions, tested either without or with delay. There is, in addition, a suggestion in Fig. 1 that response times increase as a function of delay, even if one disregards the 0- and 30-sec tests of explicitly presented material. This would mean that access time to propositional memory itself becomes longer as a function of delay. There may be various explanations for this phenomenon. For example, it might mean that as the interval between reading and test increases, the time to access the right paragraph in memory becomes longer.

It is not easy to decide exactly what the relationship is between the memory for surface features of a text, as observed here, and the short-term memory identified in list-learning studies. Classical short-term memory is severely restricted in its capacity, and the type of memory that was observed here goes well beyond this capacity, even if one grants that exactly what constitutes a "chunk" of text is still a little puzzling. If one equates short-term memory with the contents of a subject's consciousness, as was suggested in Chapter 4, it is again obvious that 120-word paragraphs exceed the capacity of consciousness, though one would be hard put to state precisely what that capacity might be. Therefore, it appears that memory for

surface features of a text is something else than the familiar short-term memory, a level intermediate between it and memory for meaning. Short-term memory proper has been demonstrated in studies concerned with speech and recall in the work of Jarvella (1970). Jarvella showed that when spoken discourse was interrupted and the listener was asked for immediate recall, subjects could, essentially, recall perfectly the current syntactic clause, but that verbatim recall dropped off sharply prior to that. The memory level identified by Jarvella may be much more like the acoustic short-term memory buffer that keeps reappearing in list-learning work than the memory for surface features that we were forced to postulate here. The latter has a much larger capacity and is not acoustic or phonemic but is at a deeper, perhaps verbal-linguistic level: It was not disrupted by the slight changes in the wording between text and question that were made in Experiment I.

The alternative to a short-term memory explanation is a depth-of-processing hypothesis, as suggested by Craik and Lockhart (1972). Craik and Lockhart posit a limited-capacity central processor which may be deployed in a number of different ways. One type of processing involves deeper analysis of the stimulus. Memory is, thus, a continuum from the transient products of sensory analyses to the more stable products of semantic operations. Superimposed on this basic memory system, however, is a second way in which stimuli are retained—recirculating information at one level of processing. Whenever information is maintained at one level, the phenomenon of short-term memory appears. While limited capacity is a function of the processor itself, the number of items held depends upon the level at which the processor is operating. The deeper the level of processing, the greater its apparent capacity. Likewise, information is lost at a rate appropriate to the processing level: Deeper levels involve slower rates.

The observed phenomenon of a surface memory for prose can be easily derived from the processing assumptions of Craik and Lockhart. First of all, processing of the textual information used in our studies must proceed to deeper levels of analysis than processing of single words or word pairs in list-learning tasks. Short-term memory for prose should, therefore, exceed the capacity of short-term memory observed in list learning, since the level at which the central processor maintains the information is deeper. Furthermore, since rate of loss from short-term memory is a function of processing level, one should expect short-term textual information to be lost more slowly than the 20- to 30-sec temporal limit observed for single words. The only addition to Craik and Lockhart that is required to explain our results is that access time is a function of level. This additional assumption is necessary because reaction times to explicit propositions were significantly faster than those to implicit propositions when the surface representation was still available.

Summary

Comprehension consists in constructing a text base from whatever cues are available in a text. Usually, each proposition in the text base is explicitly cued by the text, but it happens quite frequently that a coherent text base can only be constructed if some propositions are included that are not directly represented in the surface struc-

ture of the text. The processing of such implicit propositions was studied in a series of experiments wherein subjects read paragraphs in which a given critical proposition either was or was not explicitly expressed. Immediately after reading, a test sentence based upon the critical proposition was shown to the subject who gave a true-false response. Subjects were very well able to recognize such sentences as true even when these sentences had not been explicitly a part of the paragraph, but they responded faster when they had actually just read the sentence. On the other hand, if the true-false test was delayed 20 minutes or more, reaction times to explicit and implicit sentences were equal. The latter finding was interpreted as support for the notion that whether or not critical sentences are explicitly given in a paragraph, the reader will infer the corresponding proposition and store it in memory as part of the text base that is his memory for the meaning of the paragraph. Therefore, when asked a question about it, he will answer it with equal ease whether or not the sentence had actually been read—except that if the test question is asked while he still has other information about the text (memory for the actual words used), he can answer on the basis of this auxiliary memory, speeding up his response time. Memory for surface features is, however, subject to strong interference and is quickly lost; hence, the initial superiority of explicit sentences is also lost and subjects must rely upon their propositional memory, which is, presumably, equal in the cases studied here for explicit and implicit test sentences.