Parallel Processes During Question Answering?

Scott P. Robertson and Kimberly C. Weber

Psychology Department
Rutgers University
Busch Campus
New Brunswick, NJ 08903

Abstract

Question answering involves several processes: representation of the question concept, identification of the question type, memory search and/or inference generation, and output. Researchers tend to view these processes as stages and have developed primarily serial models of question-answering. Word-by-word reading times of questions, however, suggest that some processing is done in parallel. Questions were read more slowly but answered quicker when the question type was apparent from the first question word (the usual English construction of a question) when compared to cases when the question word came last. Serial models can not explain such data easily. It is argued that the processes associated with a particular question type are active during processing of the question concept and that they can direct memory search during question parsing. Some parallel models of question answering consistent with the data are discussed.

Introduction

Question answering is a process that requires parsing, directed memory search, sometimes inference generation, and finally the formulation of an answer. The processes involved in many of these steps have been studied extensively using methods from both artificial intelligence (Lehnert, 1978) and cognitive psychology (Singer, 1984, 1986; Graesser, Robertson, & Anderson, 1981). Not much is known, however, about the time course of these steps and which components may be realized in parallel. In this study the hypothesis that question parsing and memory search are integrated, parallel processes is assessed.

Lehnert (1978) suggested that the process of question answering involves representation of the *question concept* and categorization of the *question type*. Once these steps are completed, information about the question type can be used to initiate appropriate memory search or inference procedures while the question concept serves as the indexing mechanism to the memory representation. For example, questions 1 and 2 below both contain the question concept *Pete drove to the store*, but they are two different question types.

(1) Did Pete drive to the store?
(2) Why did Pete drive to the store?

Question 1, a *verification* question, may be answered simply by finding the question concept in memory. The simplest procedure associated with verification specifies that if the question concept is present in memory then the answer to a verification question is "true," otherwise it is "false" (Singer, 1984). More complex models of verification, the use of plausibility criteria for example, have also been proposed (Reder, 1982). In any case, matching processes require explicit representation of the question concept in order to begin.

Question 2, a *goal* or *causal antecedent* question, requires complicated processing both to determine the question type and to retrieve an answer. The question concept must first be generated, but the next step for such questions depends on the structure of information associated with the concept. For example, if the subject of the question is an animate actor then the question is a *goal* question and memory retrieval will focus on the actor's goals. In such a case, if an explicit goal is associated with the concept, e.g. "Pete needed bread," then that goal would be a candidate for an answer; otherwise a more generic goal or preceeding state, e.g. "Pete was hungry" or "Pete was out of food," might be generated by inference as possible answers. If the subject of the question is not animate, e.g. "Why did the tree fall," then the question is a *causal anecedent* question and memory retrieval and inference processes will not involve goals but will focus on preceding, causally relevant states.

In several studies Singer (1984, 1986) has examined both verification questions and "wh- questions" (who, what, when, and where) and used a stage model to explain question-answering times. The model assumes, like Lehnert, that conceptual representation of the question occurs first, followed by question categorization, followed by memory search, finally terminating in an output stage. In Singer's studies the time to read and answer questions in which the question word was the first word of the question sentence was measured. Singer's serial model predicted the overall time taken to answer the questions, but did not reveal processes that might be occuring during reading.

To compare serial versus parallel possibilities for question answering, consider the wording of questions 3 and 4 below.

(3) Why did Pete drive to the store?
(4) Pete drove to the store, why?

Most models of question answering assume that a conceptual representation of the question is built from the sentence before other processing begins. For both cases 3 and 4 we might designate such a representation with embedded propositions as follows:

(REASON, ? (DRIVE, ACTOR:Pete, FROM:unspec, TO:Store))

If a concept is found in memory that matches the above pattern, then the portion that replaces the "?" will serve as an answer to the question. Note that it is necessary to generate the complete pattern in order to initiate matching processes. On this view the only reason to expect a difference in the time course of reading and answering between questions 3 and 4 would be if different processes were involved in generating the question concept. If question answering doesn't begin until the concept is built, then it should proceed in the same way in both cases.

An alternative view is that some retrieval processes can begin for question 3 before they do for question 4. On this view the question word, *why* in the example, triggers activation of rules (or processes) specific to the question type that can apply while the rest of the question is being parsed. This parallel process model suggests that question 3 could be answered more quickly than question 4, although extra processing occurring during reading of the main part of question 3 might slow the reading rate relative to the same parts of question 4.

Reading Behavior	Question position	
	Question-first	Question-last
Before	+ #### ##### ## ### #####	#### ##### ## ### ##### +
After	Why? Pete drove to the store.	Pete drove to the store. Why?

Table 1.

Appearance of the screen before and after the subject reads a question in the two question-position conditions.

In this study subjects read concepts from brief paragraphs and answered questions about them. In one condition the question type was known as the concept was being read, as it would be in question 3 above. In a second condition the question type was not known until after the concept had been read, as in question 4 above. Reading times for the sentences expressing the question concepts, the words indicating the question types, and the answer times were collected. An advantage in answering time when the question type was known in advance would be taken as evidence for parallel processing during reading of the question concept.

Method

Subjects.
Twenty-two undergraduate students participated for course credit. Each subject served in all conditions of the experiment.

Materials.
Twenty five- to eight-line stories were written. All stories were about a character who performed several actions. An action was chosen to be queried from each story. Each action statement chosen for a query had the same syntactic form: NOUN1 + VERB + PREPosition + DETerminer + NOUN2.

Design and Procedure.
Subjects read the twenty stories in sequence and answered a question after each story. Each story was presented as a whole on a computer screen and subjects spent as long as they wished reading the story. Subjects indicated that they had read a story by pressing a response key. When the key was pressed the story disappeared from the screen and was replaced by a *question frame*. The question frame consisted of two lines, one containing a plus-sign where the question word would appear and the second containing number-signs where words expressing the question concept would appear as a sentence. In some cases the plus sign was above the sentence line indicating that the question word would appear first. In other cases it was below the sentence line indicating that the question word would appear last. Table 1 shows examples of the question frame as it appeared in the question-first and question-last conditions for a *why* query of the sentence "Pete drove to the store."

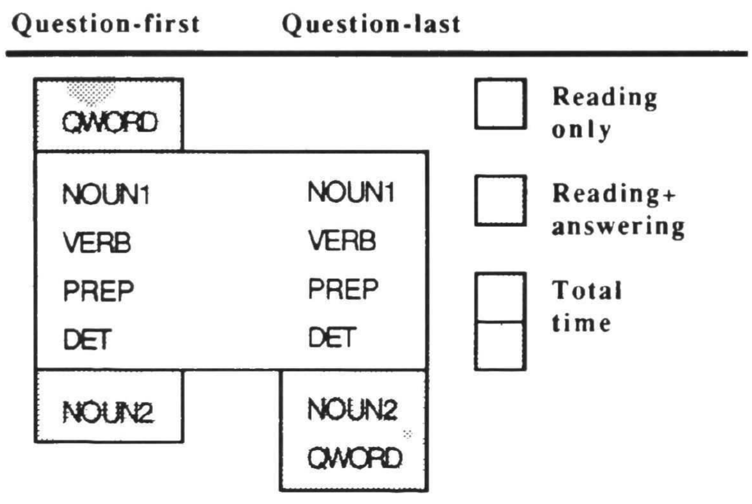

Figure 1.
A comparison of word positions and dependent measures in the two question-position conditions.

Subjects pressed the same response key repeatedly to reveal the question word (which replaced the plus sign) and each part of the question concept (words replacing the number signs). The reading time for each of these words was recorded. Subjects received special instructions for the last word (in some cases the last noun of the concept and in others a question word). They were instructed to think of the answer and press the response key as soon as the answer "came to mind." (A similar procedure was used succesfully by Reiser, Black, & Abelson, 1985, in studies of autobiographical memory.) Thus the reading time for the last word includes answer retrieval time. Subjects then wrote the answer on a sheet of paper and moved on to the next story at their own pace. All subjects received 20 practice trials to get used to the procedure.

Three types of questions were asked, *verification*, *goal*, and *time*. The question type was indicated by the presentation of the words *True?*, *Why?*, or *When?* in the question frame. Each subject answered all three types of questions (6 questions of each type) randomly intermixed. All of the 6 critical verification questions were affirmative, but two extra stories (not analysed) and several practice stories were followed by negative verification questions so that affirmation of this question type was not predictable.

There were two conditions related to the presentation position of the question words. In half of the trials the question word appeared before the concept sentence and in the other half it appeared after the concept sentence. Question type and question position were completely crossed and all subjects received stories in all combinations of these two factors. Stories were rotated through the question-type and question-position conditions across subjects. Presentation order of the stories, and hence the conditions, was random for each subject.

Results

Reading times were collected for the question words and each word expressing the question concept. Here the overall reading and answering time, the time to read the first four words of each sentence (combined and separately), and the time to read the question word plus the last word of each sentence are reported. These dependent measures are illustrated in Figure 1. As Figure 1 shows, the question-word *(QWORD)* reading time includes answering time in the question-last condition but not in the question-first condition and the *NOUN2*

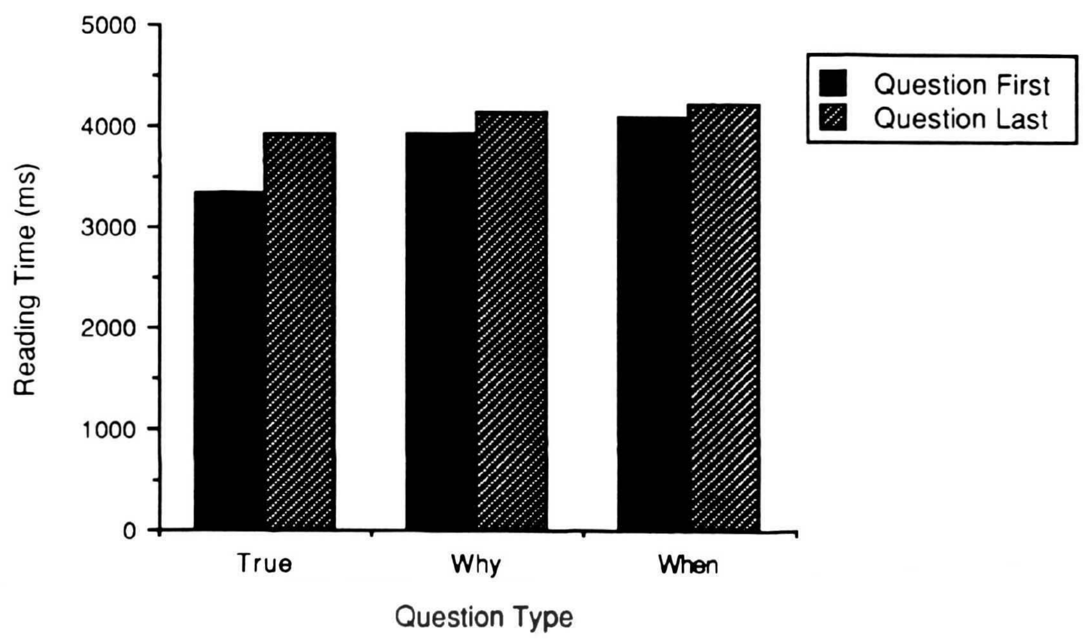

Figure 2. Total reading and answering time.

reading time includes answering time in the question-first condition but not in the question-last condition. By reporting *QWORD+NOUN2* in all conditions we are reporting the answering time combined with the (presumably constant) time to read each word separately.

Each subject provided three times in each of the six conditions. All reading times above 8000ms were eliminated. These times usually occurred when subjects forgot to press the reponse key before writing an answer. The mean times for as many good observations as a subject provided in each condition were used. In only one case was data missing because all three times from a subject in a particular condition were above the threshold. In this case the overall mean of that condition was used as a replacement value.

Figure 2 shows the overall mean reading and answering time for all words in the two question positions and across the three question types. The total time was shorter when the question word came first than when it came last, 3780ms versus 4090ms repectively, $F(1,21)=9.96$, $p<.01$, MSe=317,940. Also, there was a difference among the total times across question types, with the verification question being fastest, 3631ms, the why-question next, 4029ms, and the when question longest, 4145ms, $F(2,42)=3.42$, $p<.05$, MSe=937,618. There was no interaction between these factors.

Figure 3 shows the mean time to read *NOUN1+VERB+PREP+DET* combined in the two question positions and across the three question types. This time does not include answering time. The mean time to read this material was longer when the question was known, 1456ms versus 1355ms for the question-first and question-last conditions respectively, $F(1,21)=10.95$, $p<.01$, MSe=30,478. A question-type effect was not present for these words nor was an interaction.

Figure 4 shows the mean time per word for *NOUN1*, *VERB*, *PREP*, and *DET* separately in the two question-order conditions combined over question types. A word by question-position by question-type ANOVA showed no main effect of question type and no interactions of question-type with other factors. The apparent interaction between question position and word was significant $F(3,63)=3.84$, $p<.01$, MSe=2014. Individual comparisons revealed the suprising result that the question position effect was present for *NOUN1*, $F(1,21)=7.28$,

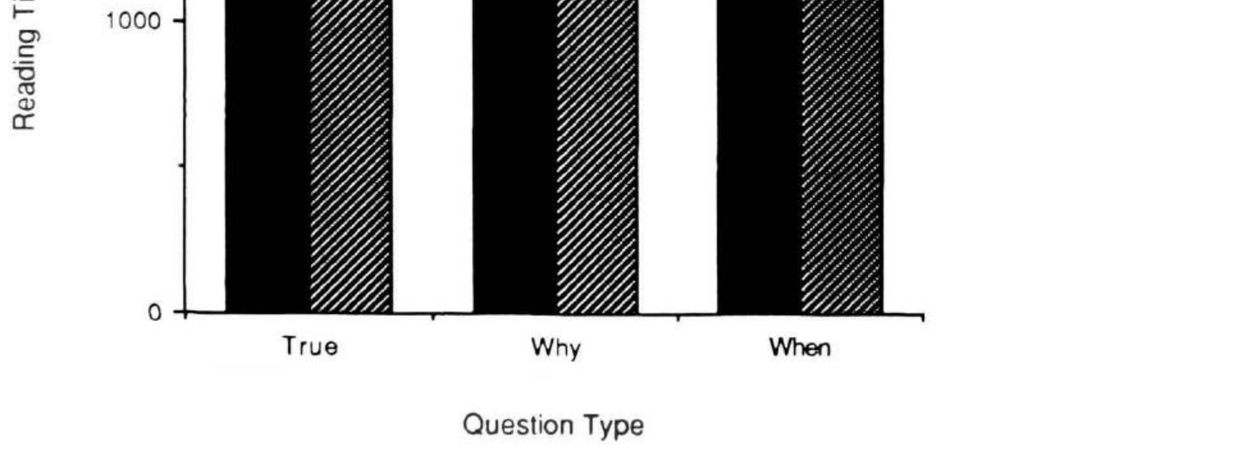

Figure 3. Reading time for NOUN1+VERB+PREP+DET.

$p<.01$, MSe=4489; *PREP*, $F(1,21)=5.96$, $p<.05$, MSe=3323; and *DET*, $F(1,21)=16.32$, $p<.01$, MSe=3651; but not for *VERB*. The means for each significant question-position effect were 367ms versus 336ms for *NOUN1*, 363ms versus 339ms for *PREP*, and 395ms versus 353ms for *DET*, in the question-first and question-last conditions respectively. There were no effects of question type and no interactions between question position and question type for individual words.

Figure 5 shows the mean times to read the question word plus the last noun *(QWORD+NOUN2)* in the two question positions and across the three question types. This time includes answering time. In this comparison the mean time was faster when the question was known, 2452ms versus 2769ms for the question-first and question-last conditions respectively, $F(1,21)=9.36$, $p<.01$, MSe=449,676. There was also a main effect of question type, with the verification question being fastest, 2163ms, the why question next, 2829ms, and the when question longest, 2840ms, $F(2,42)=4.48$, $p<.05$, MSe=1,316,768. There was no interaction between the two factors.

Discussion

The main hypothesis that the time to read and answer a question would be faster when the question type was known was confirmed. A reasonable explanation of this effect is that processes related to answer generation can be executed during parsing of the information in the question concept.

Separate analyses of times that included question-answering time versus those that did not showed that knowledge of the question actually increased reading times of words in the body of the question concept. The advantage in reading time when the question type was known came on (or after) the last word and more than made up for the earlier slower reading times. This strengthens the view of parallel processes since it suggests that resources involved in question processing and concept representation compete during reading to slow overall time but that by the end of the question the answer is much closer if the question was known.

The effects of question type have been observed before and were not suprising (Singer, 1884, 1986). While

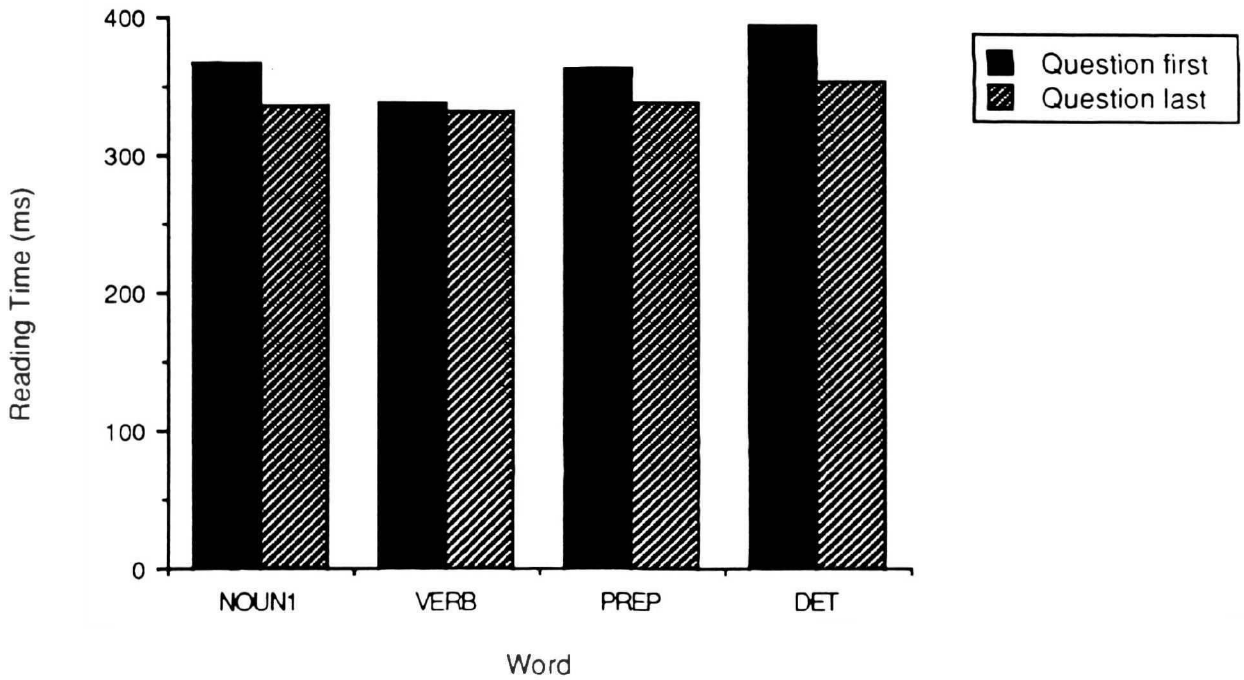

Figure 4. Reading time by word over all question types.

the mechanism behind this effect has never been specified, it is clearly related to the different operations that are required for answering different questions. For example, there are more potential answers to *goal* or *causal antecedent* questions than there are to *verification* questions, and so either memory retrieval time or output preparation time should be longer for the former types of questions. The fact that question-type effects appear only in the answering times, and not in the time for reading the main part of the sentence, may suggest that memory retrieval directed by processes specific to the question type does not occur in parallel with sentence comprehension. This would mean that processes common to all questions, preparation of question-answering processes for example, are what are occuring during comprehension. This phenomenon deserves further study.

These results rule out extreme serial models of question answering, e.g. Singer's (1981, 1984, 1986) VAIL models, retrieval mechanisms described by Graesser et al. (1981), or a strict interpretation of Lehnert's (1978) model. Such models hold that separate stages are involved in building the question concept, accessing the question-answering procedure, activating the question concept, and finally applying the procedure to the activated knowldge. The results may be explained, however, by different versions of parallel models.

In one parallel scenario the question word triggers proceeses that access the appropriate question-answering procedures. These procedures are assembled during reading, held in working memory during parsing of the question concept, and are then ready for application when the question concept is finally built. This model would explain why the question-position effect appears at the beginning of the sentence and why it disappears on the second word (once ready the question-answering procedures are not applied until later) but not why the effect *returns* toward the end of the sentence. Once question-answering procedures are present in working memory their effect on reading time should stay constant.

In a second parallel scenario question-answering procedures are accessed at the start of the question and begin to apply whenever there is enough information to do so. This model is consistent with the reappearance of the question-position effect later in the sentence (due to different retrieval processes). It is puzzling, however, under this model why the verb does not trigger parallel search processes since it clearly identifies the unique concept being queried. We are now repeating the experiment using different syntactic forms of the question

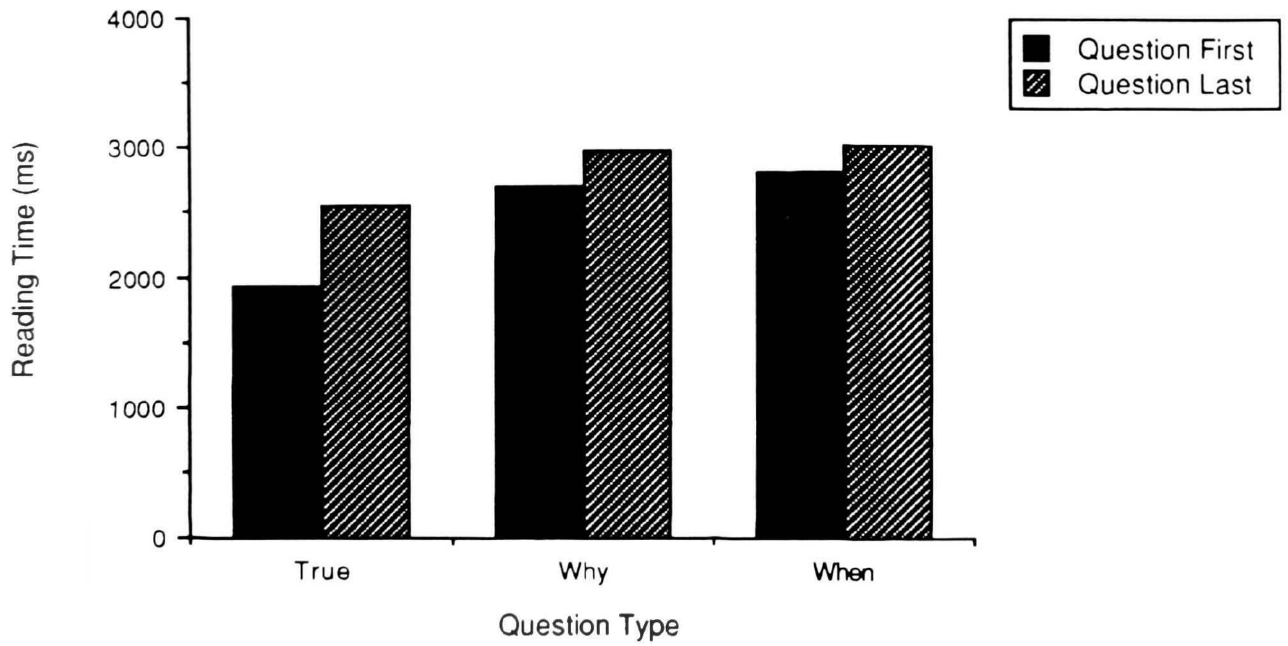

Figure 5. Reading time for QWORD+NOUN2.

concept in order to explore the intricacies of question processing during parsing.

Finally some comments should be made about symbolic versus connectionist models of this process. While the data indicate parallelism in question processing they should not be construed as evidence for or against a connectionist architecture. Wording such as "assembly of procedures for answering a why-question," a description from the point of view of a symbol-manipulation device, might easily be construed as "state of the network such that nodes related to goals and prior causal states are more likely to become active," a description from the point of view of a PDP device. It is not clear that one architecture is superior to another in explaining this data, although efforts to model the process should proceed along both lines.

To conclude it appears that question-answering processes and representation of the question concept can occur in parallel. This view conflicts with most current models of question answering. Details of a new model of these processes are now being developed. Also, further experiments to distinguish among various parallel algorithms are being conducted. Since language usually co-occurs with other cognitive processes, the ability to parallel process during language comprehension is important and the tradeoffs involved should receive careful attention from cognitive scientists.

References

Graesser, A.C., Robertson, S.P., & Anderson, P.A. (1981). Incorporating inferences in narrative representations: A study of how and why. *Cognitive Psychology, 13,* 1-26.

Lehnert, W.G. (1978). *The process of question answering.* Hillsdale, NJ: Erlbaum.

Reder, L.M. (1982). Plausibility judgements versus fact retrieval: Alternative strategies for sentence verification. *Psychological Review, 89,* 250-280.

Reiser, B.J, Black, J.B., & Abelson, R.P. (1985). Knowledge structures in the organization and retrieval of autobiographical memories. *Cognitive Psychology, 17,* 89-137.

Singer, M. (1981). Verifying the assertions and implications of language. *Journal of Verbal Learning and Verbal Behavior, 20,* 46-80.

Singer, M. (1984). Toward a model of question answering: Yes-no questions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 285-297.

Singer, M. (1986). Answering wh- questions about sentences and text. *Journal of Memory and Language, 25,* 238-254.