Journal of Experimental Psychology:
Learning, Memory, and Cognition
1994, Vol. 20, No. 5, 1219-1228

Copyright 1994 by the American Psychological Association, Inc.
0278-7393/94/$3.00

Testing Theories of Language Processing: An Empirical Investigation of the On-Line Lexical Decision Task

Gail McKoon, Roger Ratcliff, and Gregory Ward

On-line lexical decision has been used to test major theoretical hypotheses about language comprehension. Contrary to several current models, A. Sharkey and N. Sharkey (1992) found that a word in a sentence did not give facilitation to an immediately following, highly associated test item. In this article it is shown that such facilitation can be obtained. Other theories have proposed that syntactic processes supply antecedents for implicit anaphors. In using a test item that was an associate of the antecedent of the anaphor, the authors were unable to replicate previous findings of facilitation at but not before the site of the anaphor. Across 9 experiments, obtaining facilitation depended on the choice of control condition. This dependency raises questions about previous on-line lexical decision results that have been used to support the immediacy of syntactic processing.

Theories of language comprehension vary widely in their goals. Some attempt to explain the moment-by-moment processes that construct meaning as one individual word is read after another (e.g., Kintsch, 1988). Others attempt to explain the processes that organize words into syntactic structures that show the roles played by the individual words (Fodor, 1993; Frazier, 1987; Frazier & Rayner, 1982; Nicol & Swinney, 1989; Swinney & Osterhout, 1990; Rayner & Morris, 1991). Still others are concerned with inferences that might integrate the pieces of a text into a wholistic representation in memory (e.g., Glenberg, Meyer, & Lindem, 1987; McKoon & Ratcliff, 1992a). Efforts to test all of these theories share a major problem: finding empirical procedures that allow investigation of the processes or structures of theoretical interest. In this article, we report the results of several experiments designed to analyze one empirical procedure that has frequently been used: on-line lexical decision.

In on-line lexical decision experiments, the words of a text are presented to subjects one word at a time, either visually or auditorily. At some point in the text, a test string of letters is presented visually. The subject is asked to decide, as quickly and accurately as possible, whether the string of letters is a word. Reaction time and accuracy are recorded.

The on-line lexical decision technique has been used to investigate comprehension of both word meanings and syntactic structures. One of the first uses of this technique was by Swinney (1979), whose aim was to examine the processing of ambiguous words. In his experiments, subjects listened to sentences such as "The man was not surprised when he found several spiders, roaches, and other bugs in the corner of his room," which contains the ambiguous word *bugs*. While listening, the subjects watched a fixation point on a CRT screen. Immediately after the ambiguous word, a test word replaced the visual fixation point. The lexical decision response for the test word was facilitated if it matched either of the meanings of the ambiguous word; for example, following *bugs*, responses were facilitated for both *spy* and *ant*.

More recently, on-line lexical decision has been used to test the claims of general theories of meaning comprehension. Kintsch (1988; see also Ratcliff & McKoon, 1988) has proposed that meaning is constructed from the words of a text by processes that first activate the associates of individual words and then integrate the activated concepts into a representation of the meaning of the whole text. When words are read, all of their associates—even those that will turn out to be irrelevant to the meaning of the text—are activated (with varying degrees of strength). Then, through a repeated recycling of activation, concepts that are associated to other activated concepts are strengthened, whereas concepts that are not associated to other activated concepts are weakened. Once this cyclic integration process stabilizes, the result is a representation of the meaning of the text.

It is fundamental to Kintsch's (1988) theory (and others such as Dosher & Rosedale, 1989; Ratcliff & McKoon, 1988) that relations among words be immediately available during reading. For example, if a sentence contains the word *boy*, the relation between *boy* and *girl* should be immediately available. Sharkey and Sharkey (1992) tested whether this was the case with on-line lexical decision. The words of sentences were presented visually, at a rate of 200 ms per word. When a test string was presented, it replaced the next word of the text so that the interval between onset of the word preceding the test and onset of the test was 200 ms. Sharkey and Sharkey used test words that were strong associates of words in the text and found that responses were not facilitated. In other words, when *girl* was tested 200-ms after *boy* was presented, Sharkey and Sharkey found no facilitation of the response to *girl*. If this result were supported with further empirical evidence, it

Gail McKoon and Roger Ratcliff, Psychology Department, Northwestern University; Gregory Ward, Department of Linguistics, Northwestern University.

This research was supported by National Institute of Deafness and Other Communicative Disorders Grant R01-DC01240 and U.S. Air Force Office of Scientific Research Grant 90-0246 (jointly funded by the National Science Foundation) awarded to Gail McKoon and by National Institute of Mental Health Grants HD MH44640 and MH00871 awarded to Roger Ratcliff.

Correspondence concerning the article should be addressed to Gail McKoon, Psychology Department, Northwestern University, Evanston, Illinois 60208.

would be problematic for any theory postulating the immediate availability of well-known relations among words. However, in the experiments reported in this article, we find, contrary to Sharkey and Sharkey, that relations among words do support immediate facilitation in on-line lexical decision.

From most theoretical viewpoints, our result is not surprising. That is, it is not surprising that the explicit mention of a word should lead to facilitation of associates of the word. A more controversial claim is that the implicit mention of a concept can also lead to facilitation of associates. Consider, for example, the following sentence: "The instructors held the skier that the waitress in the lobby blamed for the theft." Complete understanding of this sentence requires knowing that the person who was blamed was the skier, not the waitress or an instructor. Current psycholinguistic theories (Fodor, 1993; Nicol & Swinney, 1989; Swinney & Osterhout, 1990) have claimed that this knowledge is computed by syntactic processes. These processes compute a syntactic structure for the sentence, and in the computed structure of the sentence above, there is a "trace" following the verb *blamed.* This trace is an implicit anaphor for the object of *blamed,* and the only syntactically possible antecedent for the anaphor is *skier,* to which the anaphor should be syntactically bound. Thus, syntactic processing should associate the "gap" after *blamed* with its antecedent *skier.*

Several researchers (Fodor, 1993; Nicol & Swinney, 1989; Swinney & Osterhout, 1990) have tested syntactic gap filling with on-line lexical decision. They have hypothesized that the gap-filling process results in "activation" of the antecedent word at the gap site. For example, in the "skier" sentence, *skier* would be hypothesized to be activated immediately after the verb *blamed.* This activation, in turn, is hypothesized to lead to activation of associates of the antecedent word (e.g., *snow* as an associate of *skier*).

To examine the syntactic gap-filling process, Nicol and Swinney (1989) used sentences like the "skier" sentence. Sentences were presented to subjects auditorily, and lexical decision test items were presented visually. Test items were chosen so as to measure the availability of potential fillers at two sites: immediately after the verb in the relative clause (the gap site) and immediately before the verb. Nicol and Swinney's results were consistent with the gap-filling hypothesis. After the verb, but not before it, the lexical decision for an associate of the syntactically determined antecedent of the wh-trace was facilitated. Lexical decisions for associates of other nouns in the sentence were not facilitated. So, for the "skier" sentence, *snow* would be facilitated when tested after the verb, but *restaurant* would not be. The overall pattern of results—facilitation for an associate of the syntactically determined antecedent, and only this antecedent, and facilitation for this antecedent after but not before the verb—suggests that the intended filler does in fact become available at the gap site.

The research reported in this article was originally planned to extend the findings of Nicol and Swinney (1989) to other linguistic phenomena. However, we found that we could not replicate the original Nicol and Swinney results. This failure led us to explore the on-line lexical decision paradigm, and Experiments 1–9 report the results of our efforts.

Much theoretical weight has been placed on data collected with the on-line lexical decision procedure. Sharkey and Sharkey's (1992) result from on-line lexical decision stands virtually alone as data contradicting major models designed to account for relations among the meanings of words (Anderson, 1983; Dosher & Rosedale, 1989; Kintsch, 1988; Ratcliff & McKoon, 1988). These models accommodate large ranges of other kinds of data.

Similarly, the results of Nicol and Swinney (1989), Swinney and Osterhout (1990), and Fodor (1993) have been applied to important and controversial hypotheses about syntactic processing. First, facilitation of an associate of the correct antecedent at its gap site would indicate that some kind of syntactic processing is engaged early in sentence processing. Second, it has been claimed that this processing proceeds independently of other kinds of information: Swinney and Osterhout found facilitation at a gap site for the correct antecedent, even when it was much less plausible than other nouns in the sentence. For example, in the sentence "Everyone watched the enormous heavyweight boxer that the small 12-year-old boy on the corner had beaten so brutally," real-world knowledge would suggest *the boy* as the object of *beaten.* Yet facilitation was obtained only for the syntactically correct object *boxer* (Swinney & Osterhout, 1990). This result was offered in support of the highly influential notion of modularity proposed by Fodor (1983). According to this notion, syntactic processing proceeds independently of other kinds of information such as semantics or pragmatics. Third, on-line lexical decision results have formed part of the database used to distinguish among different linguistic theories (cf. Fodor, 1993). Facilitation in lexical decision has been found for the kinds of traces postulated in some linguistic theories, but not for the kinds of traces postulated by other linguistic theories. Fourth, Fodor (1993) has used the difference in patterns of results between on-line lexical decision and other tasks as part of the support for a distinction between two levels of linguistic information: phonetic form and surface structure. Finally, Chomsky (1990) pointed to the significance of gap-filling results as a reason that linguists should take the empirical research of psychologists into account in their theorizing.

All of these claims are under debate and none of the debates has been resolved. It is not our intention to present a detailed review of these theoretical positions or to contribute to the theoretical debates except indirectly through evaluation of the lexical decision procedure and results. However, this evaluation should serve to promote increased methodological concern in the design of future experiments.

Experiments 1–5

As previously mentioned, we originally designed our experiments to replicate and extend results from earlier experiments described by Nicol and Swinney (1989). Therefore, our procedures and materials were modeled on theirs. Experiments 1–5 are summarized in Table 1.

We used two sets of sentences, both of which consisted of sentences with object–gap relative clauses. One set, which we labeled complex, is exemplified by the "skier" sentence: "Two instructors held the skier that the waitress in the lobby blamed for the theft." The sentences of this set were designed to have

the same syntactic structures as those used by Nicol and Swinney (1989), with a WH-trace after the verb of the relative clause. The second set of sentences was constructed to provide some generality of results across sentence types. These sentences were simplified versions of the complex sentences, formed by simplifying the noun phrases and eliminating the prepositional phrase in the relative clause. For example, the simplified version of the "skier" sentence was "Somebody held the skier that Doctor Hillcroft blamed for the theft." The simple sentences had a gap in the same (postobject) position as the complex sentences and contained the same verbs in the relative clauses as the complex sentences with the same antecedents for the wh-traces that followed the verbs. Another example of a pair of sentences are the following sentences: "The nun hated the ballerina that the senator from the north nominated for the council," and "John hated the ballerina that an old friend nominated for the council." Each sentence had one test word, an associate of the antecedent of the wh-trace (e.g., *snow* for the antecedent *skier,* and *dance* for the antecedent *ballerina*).

In Experiments 1 and 2, sentences were presented visually, one word at a time on a PC screen. In Experiments 3–5, sentences were presented auditorily. In all the experiments, the lexical decision test items were presented visually.

Across the experiments, three different test positions were used (see Table 2). A test word in the first test position was presented immediately after the antecedent of the WH-trace (immediately after *skier* in the example sentence). In the second test position, the test word immediately preceded the verb in the relative clause. In this test position, the test word always followed the object of the prepositional phrase in the complex sentences, and it always followed the subject noun of the relative clause for the simple sentences. In the third position, the test word immediately followed the verb of the relative clause (this was the gap position).

Table 1
Results of Experiments 1–5: Response Times and Error Rates

Sentence	Test Position 1 RT (ms)	Test Position 1 Error (%)	Test Position 2 RT (ms)	Test Position 2 Error (%)	Test Position 3 RT (ms)	Test Position 3 Error (%)
Visual						
Simple: Exp. 1						
Associate	775	2	—	—	—	—
Control	798	3	—	—	—	—
Simple: Exp. 2						
Associate	—	—	972	2	926	3
Control	—	—	977	3	938	6
Auditory						
Simple: Exp. 3						
Associate	730	1	738	1	—	—
Control	759	1	770	0	—	—
Simple: Exp. 4						
Associate	—	—	776	3	760	2
Control	—	—	770	2	765	2
Complex: Exp. 5						
Associate	—	—	760	1	774	1
Control	—	—	771	3	762	1

Note. Dashes indicate data not applicable for test position. RT = response time; Exp. = experiment.

Table 2
Examples of Sentences With Test Words and Test Positions

Sentence type	Example
Complex	The instructor held the skier$_1$ that the waitress in the lobby$_2$ blamed$_3$ for the theft.
Simple	Somebody held the skier$_1$ that Doctor Hillcroft$_2$ blamed$_3$ for the theft.

Note. Associate test word is *snow.*

A critical feature of Experiments 1–5 was the choice of a baseline against which to measure facilitation for the associate of the antecedent of the WH-trace. For example, if *snow* were tested in Position 1, immediately after *skier,* then we might expect to see facilitation of the response time to *snow.* The question is though, facilitation with respect to what control test word? We chose as a control test word the associate of the antecedent from some other sentence. For example, the associate test word for the "skier" sentence was *snow,* and the control test word might have been *dance.* Thus, the same words were used as test items in the two conditions: the associated condition, in which a sentence was tested with the test word associated to the antecedent for the wh-trace, and the control condition. The only difference was that in the control condition a sentence was tested with the associate of some other sentence. This choice for control test words has several design advantages: First, it controls for any characteristics of the individual test words that might affect lexical decision response times or accuracy rates. For example, the frequencies in English of the control test words are exactly the same as the frequencies of the associate test words because they are the same words. Second, the mean response times for associated test words represent means across exactly the same words as the mean response times for the control test words, again because they are exactly the same words. Third, any interactions between test words and test positions are controlled. Some possible test words might be facilitated or inhibited because they somehow "fit" or failed to fit the test positions in ways other than those under study. For example, an inanimate test word might show inhibition in a test position immediately following a verb because most of the verbs in our sentences take animate objects. Once more, using the same test words in both conditions controls for this potential problem.

Method

Materials. The set of complex sentences contained 28 sentences of the following form: noun phrase, verb, noun phrase, *that,* noun phrase, prepositional phrase, verb, adjunct phrase. These sentences averaged 15 words in length. Each complex sentence was changed into a simple sentence by simplifying the first and third noun phrases and deleting the prepositional phrase. The simple sentences averaged 12 words in length. The second noun phrase and the verb of the relative clause were the same in both the simple and the complex versions. The test word for each sentence was an associate of the noun in the second noun phrase (which was the antecedent of the WH-trace following the relative clause verb). The complete set of antecedents and their associated test words consisted of the following: skier–snow, journalist–news, ballerina–dance, architect–building, locksmith–key, gardener–flowers, secretary–typing, convict–prisoner, boy–girl, photographer–

camera, woman–lady, millionaire–rich, sculptor–statue, victim–injury, writer–novel, duchess–duke, poet–verse, gangster–mob, soldier–army, cowboy–Indian, baker–bread, doctor–nurse, junkie–drugs, comedian–laugh, jockey–horse, zoologist–animals, cobbler–shoes, musician–song. The complete set of complex sentences is shown in the Appendix. The simple sentences were used in Experiments 1–4 and the complex sentences in Experiment 5.

There were also 48 filler sentences, averaging 14 words in length. Each of the filler sentences had one test item; 14 of these were words and 34 were nonwords. The test positions for these items were scattered randomly through the sentences so that subjects could not anticipate which word in a sentence would be followed by a test item.

Visual presentation procedure. Sentences and test items were presented on a PC screen, with responses collected from the keyboard of the PC. Stimulus presentation and response recording were controlled by a real-time computer system.

In Experiments 1 and 2, the sentences and test items were presented visually. The experiments began with a practice list of 30 lexical decision test items (without any sentences) to familiarize subjects with the response keys. The 28 experimental sentences and the 48 filler sentences were then presented in random order, with the random order changed after each second subject. Displayed on the PC screen, each sentence began with an instruction to press the space bar on the keyboard to initiate a sentence. The words of a sentence were presented one at a time, with all letters in lowercase, except for the first letters of the first words of sentences and the first letters of proper nouns. Each word was displayed for 170 ms plus 17 ms multiplied by the number of letters in the word; the word was then erased from the screen, and the next word was displayed. Each word was displayed at the same location on the PC screen. Test items were displayed five spaces to the right of the location for words of the sentences, and test items were marked with two trailing asterisks. There was no extra time between a word of a sentence and the test item that immediateley followed it, so the stimulus onset asynchrony between the word of the sentence and the test item was 170 ms plus 17 ms multiplied by the number of letters in the sentence word. Test items were displayed in lowercase. A test item remained on the screen until subjects made a response, the question mark and slash key for "word" and the z key for "nonword." The test word was then erased, and the words of the sentence continued after a 170-ms pause. Subjects were instructed to respond quickly and accurately to the test items. To encourage the subjects to read the sentences, they were occasionally given a recall test: After 8 randomly chosen sentences, subjects were asked to write down the last sentence they had read. One test item proved problematic with visual presentation: We presented *indian* (used as the associate of *cowboy*) without the first letter capitalized and, probably as a consequence, it showed slow responses overall, so it was deleted from the analyses of results.

Auditory presentation procedure. In Experiments 3–5, the sentences were presented auditorily through headphones, and the test items were presented visually on a PC screen. The sentences were recorded by a male speaker at a natural speaking rate. Test positions for a sentence were located by examining an amplitude–time plot of the sentence; a test position following a word of the sentence was defined as the point of lowest activity between that word and the next word. If there was no single point at which activity was lowest, then the test position was located at the end of the range of lowest activity farthest from the preceding word, but never overlapping the next word.

The experiments began with the same 30 lexical decision practice items as for the visual presentation experiments. The 28 experimental sentences and the 48 filler sentences were then presented in random order, the same random order for each subject. A row of plus signs was displayed on the PC screen as a fixation point at all times, except when a test item was presented. The sentences were presented one after another with about a 2-s pause between each sentence. At the test position for a sentence, the plus signs were replaced by the test item, which remained on the screen either until the subject responded or until 1,800 ms had elapsed. Auditory presentation of the sentence continued during the interval that the test item remained on the screen. Subjects were instructed to respond quickly and accurately to the test item, pressing the question mark and slash key for a word and the z key for a nonword. As in the visual experiments, they were asked to recall in writing 8 randomly chosen sentences.

Subjects and design. In each experiment, there were 32 subjects participating for credit in an introductory psychology class at Northwestern University.

For the first experiment, there was one test position, immediately following the second noun of the sentence (which was the antecedent of the wh-trace), Position 1 in Table 2. There were two experimental conditions: The test word for a sentence was either the associate of the second noun of the sentence (the associated condition) or the associate of the second noun of some other sentence (the control condition). These two conditions were combined with groups of subjects and groups of sentences in a Latin square design.

Experiments 2–5 all had the same design, each using two test positions. In Experiment 3, these positions were immediately after the second noun (Test Position 1, as in Experiment 1) and immediately before the verb of the relative clause (Test Position 2, see Table 2). In Experiments 2, 4, and 5, the second and third positions (immediately before and after the verb of the relative clause) were used. In each case, there were four experimental conditions: the two test positions crossed with the two test word conditions (associated and control). The four conditions were combined with groups of subjects and groups of sentences in a Latin square design.

When a sentence was tested in the control condition, the test word was the associate of the antecedent of another of the 28 experimental sentences. Which other sentence was chosen randomly (without replacement), with the randomization changed after every second subject. No test item was presented to a subject more than once.

Results

We excluded slow outlier response times (times longer than 1,500 ms) from the analyses; these made up about 1.5% of the data in each experiment. Means of correct responses were calculated for each subject and each test item in each condition, and means of these means are shown in Table 1. Analyses of variance (ANOVAs) were performed on the means, with both subjects, F_1, and items, F_2, as random variables ($p < .05$).

The pattern of results is presented in Table 1. First, when a test word immediately followed its associate in a sentence (Test Position 1), response time was facilitated. This was true both in Experiment 1 with visual presentation and in Experiment 3 with auditory presentation. This finding stands in clear contrast to Sharkey and Sharkey's (1992) failure to find facilitation in a similar experiment.

Second, at the gap position (Position 3) following the verb, where there was hypothesized to be a wh-trace to serve as an anaphor, there was little evidence of facilitation. In these experiments, implicit mention of the antecedent through its anaphor did not serve to significantly facilitate responses for the associate of the antecedent.

The only test position at which results were somewhat equivocal was at Test Position 2, immediately before the verb of the relative clause. In Experiment 3, the associate of the antecedent was facilitated, but this was not the case in

Experiments 2, 4, and 5. We cannot offer any reason for this discrepancy.

ANOVAs confirmed the conclusions just stated. For the first test position, there was significant facilitation of response times in Experiment 1, $F_1(1, 31) = 5.55$ and $F_2(1, 26) = 4.03$. In Experiment 3, there was significant facilitation at both the first and second test positions, $F_1(1, 31) = 7.33$ and $F_2(1, 24) = 6.78$. A planned test confirmed facilitation at the first test position, $F_1(1, 28) = 4.37$ and $F_2(1, 24) = 5.36$.

There were no significant effects on response times of any other variables in any of the experiments ($Fs < 2.7$), except that in Experiment 2, responses were significantly faster in Test Position 3 than in Test Position 2 in the analysis of the subject means, $F_1(1, 31) = 4.51$ and $F_2(1, 26) = 3.31$. There were no significant differences among error rates ($Fs < 2.7$).

The standard errors of the response time means in the five experiments were 7.3 ms, 22.3 ms, 12.8 ms, 13.0 ms, and 11.6 ms, respectively. Response times and error rates for filler test items are shown in Table 3.

We performed an additional analysis on the data from Test Positions 2 and 3 to investigate the possibility that the failure to obtain a difference between the associated and control conditions at the second and third test positions was due to spuriously fast responses in the control condition. Fast responses could arise in the control condition if the test words in that condition happened, by random assignment, to be associated (against our intentions) either to the antecedent of the implicit anaphor or to other words in the sentences with which they were tested. To eliminate this possible explanation of the results, we eliminated from the analyses all of the test words that were associated with any words in any sentences other than their own sentence. We eliminated all the test words that were associated in any way we could think of, by even quite weak associations, a total of 16 test words (which eliminated data about equally across the four counterbalancing groups of items). For example, we eliminated the test word *girl* because it might be associated to words from other sentences than its own, such words as *secretary* or *woman*. If such associations had speeded responses in the control condition, then eliminating these test words should lead to slower responses in the control condition than in the associated condition, but this did not happen. Responses in the two conditions were still virtually identical, differing by no more than 5 ms.

Table 3
Response Times and Error Rates for Filler Test Items

	Words		Nonwords	
Experiment	RT (ms)	Error (%)	RT (ms)	Error (%)
1	826	8	818	6
2	999	8	1,051	5
3	829	3	899	4
4	813	6	852	5
5	779	3	850	5
6	905	7	952	6
7	908	6	990	7
8	823	4	868	5
9	793	2	865	5

Note. RT = response time.

Discussion

The results obtained in Experiments 1–5 contradict previous findings. Contrary to Sharkey and Sharkey (1992), we found that a word in a sentence facilitated response time on an immediately following test of an associated word. Our result, unlike Sharkey and Sharkey's, is consistent with current models of the processing of relations among words. Models that postulate spreading activation processes predict that presentation of a word will facilitate subsequent decisions on other words related to it (Anderson, 1983; Kintsch, 1988). Models that postulate compound cue kinds of retrieval mechanisms similarly predict that relations among related words will be quickly available to facilitate decisions (Dosher & Rosedale, 1989; Ratcliff & McKoon, 1988).

We can only speculate about why we were able to demonstrate immediate facilitation and Sharkey and Sharkey (1992) were not. They used fewer subjects, and perhaps variance was higher in their experiment. This is plausible because a 45-ms effect in their experiment (due to the position in a sentence at which a test word was presented) was not significant. Also, in their experiment, lexical decision test items were distinguished from words of the sentences by color of the lettering, green versus white. Perhaps the green lettering served in some way to switch processing away from the words of the sentences.

Our results were also different from previous findings when we tested for facilitation due to an implicit presentation of an associate of a test word. Nicol and Swinney (1989) reported facilitation at the site of an implicit anaphor. In sentences with syntactic structures like our sentences, they found a pattern of facilitation at the wh-trace site following a verb but no significant facilitation before the verb. Our results show no evidence of this pattern.

We thought that the reason for our failure to find the previously reported pattern of facilitation might be our choice of control condition. As explained in the introduction, we believed that using the same pool of words in both conditions, associated and control, was an optimal experimental design. However, the control condition that has been used by Nicol and Swinney (1989), Swinney and Osterhout (1990), and Fodor (1993) was different; they used a different pool of words in the two conditions. In their designs, there were two test words for any given sentence, always the same two words. One of the words was the associate of the antecedent of the trace (e.g., the associate *snow* for the antecedent *skier*). The other word, the control, was a word unrelated to the meaning of the sentence, with the same number of letters and the same frequency in the English language as the associated word. We thought that this difference in choice of control condition between our Experiments 1–5 and previous experiments might account for the difference in results, and we tested this hypothesis in Experiments 6– 9.

Experiments 6–9

These four experiments are outlined in Table 4. We used both the simple and the complex versions of the sentences, and sentences were presented both auditorily and visually. The only difference from the comparable experiments in the first

Table 4
Results of Experiments 6–9: Response Times and Error Rates

Sentence	Test Position 2 RT (ms)	Test Position 2 Error (%)	Test Position 3 RT (ms)	Test Position 3 Error (%)
	Visual			
Simple: Exp. 6				
Associate	872	4	871	1
Control	922	6	854	6
Complex: Exp. 7				
Associate	845	5	827	3
Control	873	9	881	6
	Auditory			
Simple: Exp. 8				
Associate	760	3	753	4
Control	816	8	832	5
Complex: Exp. 9				
Associate	753	3	742	2
Control	819	9	784	7

Note. RT = response time; Exp. = experiment.

series (Experiments 1–5) was in the control condition. We choose a new pool of control words, one word for each sentence, so that the control word for a sentence had the same number of letters and approximately the same frequency in English as the associate test word (according to Kucera & Francis, 1967).

Method

Materials and procedure. The sentences and their associated test words were the same as in Experiments 1–5, and the only change was in the words used in the control condition. The procedures for the experiments were also the same as in Experiments 1–5. The antecedents with their new control words were the following: skier–uses, journalist–clay, ballerina–equal, architect–material, locksmith–add, gardener–evident, secretary–afloat, convict–symmetry, boy–trade, photographer–affect, woman–file, millionaire–camp, sculptor–morale, victim–define, writer–stone, duchess–buys, poet–marks, gangster–ads, soldier–list, cowboy–warren, baker–seeds, doctor–graph, junkie–dried, comedian–shots, jockey–doubt, zoologist–perfect, cobbler–grown, musician–dust.

Subjects and design. There were 32 subjects in each of Experiments 6 and 7, 24 subjects in Experiment 8, and 20 subjects in Experiment 9, all from the same population as in Experiments 1–5. Except for the new control words, the designs of the experiments and randomization procedures were the same as in the earlier experiments.

Results

The data were analyzed in the same manner as for Experiments 1–5, and the means are displayed in Table 4.

In Test Position 2, responses to the associate test word were faster than responses to the control test word in every one of the experiments. The same was true for Test Position 3, except in Experiment 6. For Experiments 7, 8, and 9, responses to the associate were faster than responses to the control word at Test Position 3, but this pattern reversed in Experiment 6, for no apparent reason.

ANOVAs confirmed these observations. For Experiments 7, 8, and 9, the main effect of faster responses for the associate words than for the control words was significant; for these three experiments, respectively: $F_1(1, 31) = 7.21$ and $F_2(1, 26) = 4.47$; $F_1(1, 23) = 16.01$ and $F_2(1, 27) = 24.98$; $F_1(1, 19) = 9.28$ and $F_2(1, 27) = 11.85$. Other effects on response times were not significant (all Fs < 3.23). The standard errors for the means were, respectively, 10.5 ms, 14.2 ms, and 17.8 ms. There were generally more errors on the control words than on the associate words, and this effect was sometimes significant. For the three experiments, respectively: $F_1(1, 31) = 5.74$ and $F_2(1, 26) = 2.90$; $F_1(1, 23) = 4.02$ and $F_2(1, 27) = 1.84$; $F_1(1, 19) = 6.33$ and $F_2(1, 27) = 6.20$. All other effects on error rates were not significant (Fs < 2.3). For all of Experiments 6–9, the standard errors on the error rates varied between 1.0% and 1.5%.

The pattern in Experiment 6 was different. The interaction between test word and test position was significant for response times, $F_1(1, 31) = 7.36$ and $F_2(1, 26) = 11.18$. The main effect of test word was also significant in the subjects analysis, $F_1(1, 31) = 8.80$, but not in the items analysis, $F_2(1, 26) = 2.25$. The main effect of test position was not significant (Fs < 2.05). The standard error of the response time means was 12.0 ms. There were marginally more errors on the control test words, $F_1(1, 31) = 4.14$ and $F_2(1, 26) = 3.58$. Other effects on error rates were not significant (Fs < 1.85).

Two aspects of the data should be pointed out. First, over the series of nine experiments, which included 17 different comparisons of associate and control response times, results were inconsistent for two of the comparisons (Test Position 2 in Experiment 2 and Test Position 3 in Experiment 6). This suggests that any results from the on-line lexical decision procedure should be replicated across experiments to ensure a high degree of confidence in the general patterns that emerge. Second, the F values for significant effects were always higher with auditory presentation of the sentences than with visual presentation. This might have come about for a variety of reasons, but it is worth bearing in mind for future research.

The conclusions from Experiments 6–9 and comparisons of their results with those of Experiments 1–5 are straightforward. In the first 5 experiments, we used the same pool of words as test words in the associated and control conditions. For these experiments, in six out of seven cases there was no facilitation at Test Positions 2 or 3. In the last four experiments, we used different pools of words as test words in the associated and control conditions. For these experiments, in seven out of eight cases there was facilitation at both of Test Positions 2 and 3. It appears that the choice of control word was critical in determining the results.

General Discussion

We designed the experiments reported here to investigate the use of on-line lexical decision tests in the study of sentence comprehension. Lexical decision test words were presented at one of several points during a sentence. In the associated condition, the test word was highly associated to one of the words in the sentence, and it was tested either immediately after the associated word in the sentence or at one of two later positions in the sentence. The results of our experiments depended on the choice of control test words; whether the

control test words were the same words as for the associated condition (simply switched to sentences for which they were not associated) or whether the control test words were different words from the associated test words. If the control words were the same as the associated words, then there was facilitation of response times for the associated words relative to the control words at the immediate test position but not at later test positions. If the control test words were different from the associated test words, then facilitation was observed at the later test positions. These two conclusions held up over 15 of the 17 comparisons afforded by the nine experiments.

The finding that an associated word is facilitated when it is tested immediately after a related word in a sentence is intuitively compelling and also not surprising from most theoretical viewpoints. It would be expected that a lexical decision test of *snow* immediately following the sentence fragment *the skier* would result in facilitation of response time to *snow,* and this is what we found. Although Sharkey and Sharkey (1992) recently failed to find immediate facilitation, their result may well be anomalous. The variance among response times in their experiment appears to have been high (as mentioned above), and their failure is inconsistent not only with the results described here but also with a considerable amount of previous research. On-line facilitation has been found with lexical decision test positions at the ends of sentences or sentence fragments (McKoon & Ratcliff, 1989a, 1989b; O'Seaghdha, 1989; Till, Mross, & Kintsch, 1988) and with on-line text experiments that use a variety of other paradigms including measurements of word by word reading times, phoneme monitoring latencies, and naming latencies (cf. Foss & Speer, 1991; McKoon & Ratcliff, 1981, 1989c; Simpson, Peterson, Casteel, & Burgess, 1989; Stanovich & West, 1981). On-line facilitation for associated test words is also consistent with on-line facilitation for the multiple meanings of ambiguous words (Onifer & Swinney, 1979; Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979). Furthermore, the finding of on-line facilitation for associated words gains considerable validation in another important way: Consistency with a wide range of different kinds of data is established by virtue of its incorporation into comprehensive theories of memory (Anderson, 1983; Kintsch, 1988; Ratcliff & McKoon, 1988). Thus, a large body of previous research argues in favor of accepting the validity of our finding of immediate facilitation.

It is important to stress the differences among the theories with which immediate facilitation is consistent. According to spreading activation theories (e.g., Anderson, 1983; Kintsch, 1988), presentation of a word in a sentence activates the concept in memory that corresponds to the word. The activation spreads to other related concepts, so that they, in turn, become activated. If one of these activated concepts is then presented as a test word for lexical decision, its response time will be facilitated because it was already activated before its presentation. In these theories, activation spreads quickly so that the response on a test word can be facilitated even if presentation of an associated word preceded it by as little as 100 ms. The main competitors for spreading activation theories are theories that assume memory retrieval is based on a compound cue mechanism (Dosher & Rosedale, 1989; Ratcliff & McKoon, 1988). In these theories, the process by which immediate facilitation occurs is very different than spreading activation. There is no anticipatory activation of the test word. Instead, words presented to the system are assumed to join together in short-term memory to form a compound cue. This cue has some degree of familiarity, where familiarity is determined by the strengths of associations between the compound in short-term memory and items in long-term memory. Familiarity is calculated by a matching process that matches the cue in short-term memory against all the items in long-term memory. The immediate facilitation observed in the experiments reported here is consistent with the compound cue view because a lexical decision for an associated test word will be facilitated by a high familiarity value for the cue made up of the test word and the immediately preceding word of the sentence. Recently, compound cue theories and spreading activation theories have been extensively tested against each other, but both still seem to be viable accounts of retrieval from long-term memory (McKoon & Ratcliff, 1992b; McNamara, 1992a, 1992b; Ratcliff & McKoon, 1994).

The implications of the immediate facilitation effect found in our experiments are quite different when viewed from the two different theoretical perspectives. For spreading activation, immediate facilitation would be taken to indicate that reading a word in a sentence makes related concepts in memory immediately available. For compound cue theories, though, immediate facilitation does not, in itself, indicate what happens during reading of the words in sentences. No conclusions can be drawn about what would happen if the test word were not presented. The facilitation in response time is a reflection only of the situation in which short-term memory contains both the word of the text and the test word. What the two kinds of theories share is the assumption that, however the facilitation comes about, it should happen quickly, within about 100 ms.

Although our finding of immediate facilitation for related text and test words is consistent with most previous work, the patterns of facilitation we obtained for tests of implicit anaphors are not. A number of researchers have reported testing for the availability of antecedents at several different kinds of gap sites (Fodor, 1993; Nicol & Swinney, 1989; Swinney & Osterhout, 1990). For sentences such as "Two instructors held the skier that the waitress in the lobby blamed for the theft," Nicol and Swinney found that response times for an associate of the antecedent for the wh-trace following the verb of the relative clause were facilitated when tested immediately after the verb but not when tested immediately before the verb; that is, *snow* would be facilitated when tested after *blamed* but not when tested before *blamed.* This pattern of facilitation after the verb but not before is the finding that has been used to argue for the reactivation of the antecedent of the WH-trace. In neither of our sets of experiments though did we find this pattern. When we chose control test words from the same pool of words as the associated test words, we did not find facilitation either before or after the verb. When we chose control test words from a different pool of words than the associated test words, we found facilitation at both test points.

Why did we fail to replicate previous results? One possible answer to this question is suggested by the dramatic effect of

the choice of control condition. We got very different patterns of facilitation with the two different control conditions. This logically opens up the possibility that with other sets of control words, other patterns of data might emerge. With another set of control words, we might have replicated exactly the pattern that has been obtained in previous experiments (e.g., Nicol & Swinney, 1989). The most serious issues raised by our results are how to choose the "right" set of control words, whether there is any one correct set, and how researchers might go about defending the choice of control words used in their experiments over some other choice.

We can only offer tentative suggestions about why the choice of control words might be so important. We know that the syntactic fit of a test word to its test position can affect response times (Clifton, Frazier, & Connine, 1984; Wright & Garrett, 1984). In Wright and Garrett's experiments, a test word either fit the syntactic context of the sentence fragment that preceded it or it did not, and lexical decisions were slowed when it did not. This suggests that there might also be a host of other reasons why different words have different response times at different test positions in a sentence, including the meaningfulness values of the words, concreteness values, likelihoods of appearing in sentences of the type used in the experiments, and so on. For example, consider the sentences used in our experiments; they almost all took the form that "some person *verbed* someone that another person *verbed.*" Some words, because of their semantics or pragmatics, just will not easily fit in such sentences. *Marshmallow* is a case in point. In a context that includes sentences about an employer confronting a secretary that an accountant fired, *marshmallow* seems out of place. Moreover, there may be subtle interactions between the syntactic and semantic contexts of a sentence and test position. To give a few examples of verbs from our sentences, we cannot *blame, suspect, bribe, nominate, appoint, drive,* or *assault* a *marshmallow,* so *marshmallow* might fit particularly badly in a test position following a verb and perhaps less badly in a test position at the end of a phrase before the verb. Again, our current state of knowledge about these issues only allows speculation. The important point is that attention must be paid to the choice of control words in future experiments. As this issue is investigated further, we may be able to understand why previously used sets of control words have given the results they did, whether the control words in an experiment should come from the same pool of words as the associated words, and what the important variables are that govern the response time for a word tested in the middle of a sentence.

In conclusion, the theoretical implications from our results can be easily outlined. First, previous research on syntactic gap filling and the suggestion from that research that syntactic processes occur early and fast are called into question. Until we understand better how control words should be chosen, it may be that the case for fast syntactically based gap-filling processes will have to be made from other paradigms (cf. Bever & McElree, 1988; Boland, Tanenhaus, & Garnsey, 1990; Foss & Speer, 1991; Frazier & Clifton, 1989; Garnsey, Tanenhaus, & Chapman, 1989; McElree & Bever, 1989; Rayner & Morris, 1991; Stowe, 1986). Second, theoretical enterprises that have depended on on-line lexical decision results (cf. Fodor, 1989, 1993; Nicol & Swinney, 1989; Swinney & Osterhout, 1990) will have to be reworked, either with new lexical decision evidence or with reliance on other kinds of empirical evidence.

References

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Bever, T., & McElree, B. (1988). Empty categories access their antecedents during comprehension. *Linguistic Inquiry, 19,* 35–43.

Boland, J. E., Tanenhaus, M. K., & Garnsey, S. M. (1990). Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language, 29,* 413–432.

Chomsky, N. (1990, July). *Address to the Cognitive Science Society.* Boston, MA.

Clifton, C., Jr., Frazier, L., & Connine, C. (1984). Lexical expectations in sentence comprehension. *Journal of Verbal Learning and Verbal Behavior, 23,* 696–708.

Dosher, B. A., & Rosedale, G. (1989). Integrated retrieval cues as a mechanism for priming in retrieval from memory. *Journal of Experimental Psychology: General, 2,* 191–211.

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Fodor, J. D. (1989). Empty categories in sentence processing. *Language and Cognitive Processes, 4,* 155–209.

Fodor, J. D. (1993). Processing empty categories: A question of visibility. In G. T. M. Altmann & R. Shillcuck (Eds.), *Cognitive models of speech processing* (pp. 351–400). Hillsdale, NJ: Erlbaum.

Foss, D., & Speer, S. (1991). Global and local context effects in sentence processing. In R. Hoffman & D. Palermo (Eds.), *Cognition and the symbolic processes: Applied and ecological perspectives.* Hillsdale, NJ: Erlbaum.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading.* Hillsdale, NJ: Erlbaum.

Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Process, 4,* 93–126.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in analysis of structurally ambiguous sentences. *Cognitive Psychology, 14,* 178–210.

Garnsey, S. M., Tanenhaus, M. K., & Chapman, R. M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research, 18,* 51–60.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language, 26,* 69–83.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95,* 163–182.

Kucera, H., & Francis, W. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

McElree, B., & Bever, T. G. (1989). The psychological reality of linguistically defined gaps. *Journal of Psycholinguistic Research, 18,* 21–35.

McKoon, G., & Ratcliff, R. (1981). The comprehension processes and memory structures involved in instrumental inference. *Journal of Verbal Learning and Verbal Behavior, 20,* 671–682.

McKoon, G., & Ratcliff, R. (1989a). Semantic association and elaborative inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 326–338.

McKoon, G., & Ratcliff, R. (1989b). Assessing the occurrence of elaborative inference with recognition: Compatibility checking vs. compound cue theory. *Journal of Memory and Language, 28,* 547–563.

McKoon, G., & Ratcliff, R. (1989c). Inferences about contextually defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1134–1146.

McKoon, G., & Ratcliff, R. (1992a). Inference during reading. *Psychological Review, 99,* 440–466.

McKoon, G., & Ratcliff, R. (1992b). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1155–1172.

McNamara, T. P. (1992a). Theories of Priming: I. Associative distance and lag. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1173–1190.

McNamara, T. P. (1992b). Priming and constraints it places on theories of memory and retrieval. *Psychological Review, 99,* 650–662.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research, 18,* 5–20.

Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition, 9,* 225–236.

O'Seaghdha, P. G. (1989). The dependence of lexical relatedness effects on syntactic connectedness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 73–87.

Rayner, K., & Morris, R. K. (1991). Comprehension processes in reading ambiguous sentences: Reflections from eye movements. In G. Simpson (Ed.), *Understanding word and sentence* (pp. 175–198). Amsterdam: North-Holland Press.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 95,* 385–408.

Ratcliff, R., & McKoon, G. (1994). Retrieving information from memory: Spreading activation theories versus compound cue theories. *Psychological Review, 101,* 177–184.

Sharkey, A., & Sharkey, N. (1992). Weak contextual constraints in text and word priming. *Journal of Memory and Language, 31,* 543–572.

Simpson, G., Peterson, R., Casteel, M., & Burgess, C. (1989). Lexical and sentence context effects in word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 88–97.

Stanovich, K. E., & West, R. F. (1981). The effect of sentence context on ongoing recognition: Tests of a two-process theory. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 658–672.

Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes, 1,* 227–245.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18,* 645–659.

Swinney, D., & Osterhout, L. (1990). Inference generation during auditory language comprehension. In A. Graesser and G. Bower (Eds.), *The psychology of learning and motivation* (Vol. 25, pp. 17–33). New York: Academic Press.

Tanenhaus, M., Leiman, J., & Seidenberg, M. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior, 18,* 427–440.

Till, R. E., Mross, E. F., & Kintsch, W. (1988). Time course of priming for associate and inference words in a discourse context. *Memory & Cognition, 16,* 283–298.

Wright, B., & Garrett, M. (1984). Lexical decision in sentences: Effects of syntactic structure. *Memory & Cognition, 12,* 31–45.

Appendix

Set of Complex Sentences Used in Experiments 1–4

Two instructors held the skier that the waitress in the lobby blamed for the theft.

The banker bribed the journalist that the cops inthe subway suspected of the break-in.

The nun hated the ballerina that the senator from the north nominated for the council.

The pilot trusted the architect that the judge in the city acquitted of the forgery.

All the tenants appreciate the locksmith that the tailor in the basement chose for the job.

Three brothers pitied the gardener that the attorney for the museum banned from the show.

The employer confronted the secretary that the accountant at the racetrack fired for gross insubordination.

The witness recognized the convict that the teller in the cafeteria accused of violent behavior.

The clown amused the boys that the actress in the mink drove to the stadium.

The hostess greeted the photographer that the swimmer with pale skin encountered at the meeting.

The janitor called the woman that the farmer in the store saved from the blaze.

The cabby contacted the millionaire that the mailman on the scooter struck on the head.

Few parents knew the sculptor that the professor of African geography appointed to the committee.

The optometrist aided the victim that the barber in the airport hurt in the fight.

The chef envied the writer that the soprano with blue eyes followed all over town.

The announcer interviewed the duchess that the painter without a passport defrauded of the treasure.

Many artists admired the poet that the priest from the mountain visited at the penitentiary.

The bride identified the gangster that the carpenter at the barbecue attacked with a knife.

The dentist treated the soldier that the athlete with a beard punched in the tavern.

The bartender criticized the cowboy that the trucker from the factory assaulted with a rifle.

The lifeguard rescued the dog that the hobo with a rock forced off a cliff.

The students cheered the doctor that the firemen in the parade applauded for tremendous bravery.

(Appendix continues on next page)

The warden released the junkie that the sailor in the desert forgave for grand larceny.
The boxer heckled the comedian that the referee with striped pants invited to the club.
The librarian comforted the jockey that the outlaw at the funeral threatened with a stick.
The butler summoned the zoologist that the sheriff with strong arms arrested for extreme cruelty.
The king punished the cobbler that the ambassador on the patio caught with the jewels.
A bee stung the musician that the usher with the radio reprimanded for public drunkenness.

Received June 29, 1992
Revision received July 2, 1993
Accepted August 20, 1993 ■

Call for Papers

Beginning in 1995, there will be a new peer-reviewed journal in the emerging interdisciplinary specialty area devoted to work and well-being. Its mission statement is as follows.

> The *Journal of Occupational Health Psychology* publishes research, theory and public policy articles in occupational health psychology (OHP), an interdisciplinary field representing a broad range of backgrounds, interests, and specializations. OHP concerns the application of psychology to improving the quality of worklife and to protecting and promoting the safety, health, and well-being of workers. The *Journal* has a threefold focus on the work environment, the individual, and the work–family interface. The *Journal* seeks scholarly articles, from both researchers and practitioners, concerning psychological factors in relationship to all aspects of occupational health. Included in this broad domain of interest are articles in which work-related psychological factors play a role in the etiology of health problems, articles examining the psychological and associated health consequences of work, and articles concerned with the use of psychological approaches to prevent or mitigate occupational health problems. Special attention is given to articles with a prevention emphasis. Manuscripts dealing with issues of contemporary relevance to the workplace, especially with regard to minority, cultural, or occupationally underrepresented groups, or topics at the interface of the family and the workplace are encouraged. Each article should represent an addition to knowledge and understanding of OHP.

Manuscripts should be prepared according to the *Publication Manual of the American Psychological Association* and should be submitted in quadruplicate to:

James Campbell Quick, Editor
Journal of Occupational Health Psychology
University of Texas at Arlington
P.O. Box 19313
Arlington, Texas 76019
Phone number: (817) 273-3514
FAX number: (817) 273-3515
E-mail Internet address: JOHP@willard.uta.edu

Express mail: 701 South West Street
Room 514
Arlington, Texas 76010

The *Journal of Occupational Health Psychology* (*JOHP*) will be published quarterly by the Educational Publishing Foundation (EPF), an imprint of the American Psychological Association devoted to the quality publication of interdisciplinary journals.