Journal of Experimental Prychology
Learning, Memory, and Cognition
1994, Vol. 20, No. 5.

0278-7393/94/$3.00

OBSERVATIONS

Using Cross-Modal Lexical Decision Tasks to Investigate Sentence Processing

Janet L. Nicol, Janet Dean Fodor, and David Swinney

Recent investigations of sentence processing have used the cross-modal lexical decision task to show that the antecedent of a phonologically empty noun phrase (specifically, WH-trace) is reactivated at the trace position. G. McKoon, R. Ratcliff, and G. Ward (1994) claimed that (a) a design feature concerning the choice of related and unrelated targets is a possible confound in this work and (b) the conclusions dfawn from this previous research are therefore called into question. These claims are considered in light of both McKoon et al.'s experimental findings and results of our own experiments in which we test their linguistic materials. We argue that their results may be due to the nature of their materials. Additionally, we argue that a follow-up experiment reported by G. McKeon and R. Ratcliff (1994) used a technique that is not comparable to the cross-modal lexical decision task. It is concluded that current evidence supports the claim that structural information is used during on-line sentence processing and that the cross-modal technique is sensitive to this.

McKoon, Ratcliff, and Ward (1994) report a series of experiments designed to investigate lexical activation in relative clause sentences such as sentence 1.

1. The instructors held the skier (1) that the waitress in the lobby (2) blamed (3) for the theft.

In this sentence, the understood object of the verb *blamed* is *the skier,* even though this noun phrase does not appear in the normal object position (i.e., immediately to the right of the verb). It is widely maintained in linguistics that such structures contain a place holder, or *RN-trace*, at the position of the understood NP. This trace is linked (*coindexe*d) with the *NP* that heads the relative clause (here, *the skier*), and this NP is considered to be the antecedent of the trace. McKoon et. al (1994) probed at three points in such sentences: immediately after the head of the relative clause (the antecedent of the trace), immediately after the verb (the trace position), and immediately before the verb (a baseline position). These probe positions are shown in Example 1.

At each point, activation of *skier* was measured. This was done through semantic priming. At each probe point, a target word was presented for lexical decision. The target word was either a semantic associate of the antecedent (in this case, *snow*) or an unrelated control word. Faster response times to the semantically related target, relative to the control target, indicated that a connection had been made to the representation of the antecedent in short-term memory. We should note here that although we refer to activation and priming, we do not intend by this usage any commitment to a particular theory of lexical recognition. Although McKoon et al. (1994) did not use these terms, these terms do appear in some of the relevant literature (including Nicol & Swinney, 1989), and, for convenience, we use them here.

One of McKoon et al.'s (1994) goals was to follow up a result reported by Sharkey and Sharkey (1992), who tested for, but failed to observe, priming of semantic associates at a sentence position immediately following the prime. For this, only McKoon et al.'s first test point is relevant. We do not address that issue in this article. A second goal of their study was to replicate findings reported in Nicol and Swinney (1989)[1] which showed priming of an associate of the antecedent at the trace position. For this, McKoon et al.'s second and third test points are relevant.

In their nine experiments, McKoon et al. (1994) varied the mode of presentation (visual vs. auditory sentence presentation), and sentence complexity (simple vs. complex sentences, see below), as well as the position of the target word. Our main concern in this article is with their "complex" sentences (for reasons explained below) and with cross-modal presentation of the stimuli (auditory sentence and visual target word) because that is the method by which the prior results were

Janet L. Nicol, Department of Psychology, University of Arizona; Janet Dean Fodor, Graduate Center, City University of New York, and Haskins Laboratories, New Haven, Connecticut; David Swinney, Department of Psychology, University of California, San Diego.

The authors thank Gail McKoon and two anonymous reviewers for their helpful comments and suggestions. This work was supported by National Institute on Deafness and Other Communicative Disorders Grant DC-01409 to Janet L. Nicol and by a grant from U.S. Air Force Office of Scientific Research Grant AFOSR 91-0225 to David Swinney.

Correspondence concerning this article should be addressed to Janet L. Nicol, Department of Psychology, University of Arizona, Tucson, Arizona 85721. Electronic mail may be sent to nicol@ccit. arizona.edu.

[1] The findings of interest were actually reported in an unpublished paper by Swinney, Ford, Frauenfelder, and Bresnan (1988). For simplicity, and consistency with McKoon et al., we refer to Nicol and Swinney, 1989.

obtained, and it permits relatively natural and uninterrupted sentence processing. (For background on this method, see Swinney, 1979.) Therefore our comments focus on McKoon et al.'s Experiments 8 and 9. However, their other experiments do not contradict the points we wish to make, and we refer to them as appropriate.

Results that indicate antecedent reactivation by traces are relatively new, and theoretical morals have been drawn from them, so it is certainly in order to check their replicability. McKoon et al.'s (1994) results challenge previous' reports of evidence of antecedent reactivation by WH-trace. McKoon et al. raised the possibility that the occurrence of priming is due to the selection of the unrelated target-the control word that serves as baseline for establishing priming of the related target-rather than mediated by the linguistic structure of the sentence (specifically, the antecedent-trace relation). McKoon et al. clearly suggest that what has been construed as priming in previously published experiments is quite possibly an artifact of an unreliable method of selecting unrelated targets. The two methods they compared in their experiments are what we call the *matched-targets* design and the *switched-targets* design. In the former design, which has been used most commonly (though not exclusively) in the past, unrelated items were matched in length and frequency with related targets, but were not associates of the antecedent. (In some past studies, such as some of those reported by Nicol and Swinney, 1989, targets were matched on response time, as well as length and frequency.) In the switched-targets design, unrelated items were the same words as related items, but re-paired with sentences such that there was no association between the target word and the antecedent in the sentence. McKoon et al. suggested that the latter design is "optimal." The fact that they observed no related-unrelated response time difference with these targets, but they did find such a difference with what they regard as the inferior matched-targets design, leads them to conclude that this technique may not be a useful means by which to examine the processing of filler-gap constructions.

We endorse McKoon et al.'s (1994) campaign to improve the design of experimental stimuli, and we agree with them that certain considerations might lead one to believe that switched targets provide less risk of poorly equated experimental and control items than do matched targets. However, this design is not without problems (see the Implications for Future Research section). Further, the antecedent reactivation studies are designed to protect them against potential problems that could arise from unequated target items. Hence, the negative conclusions that McKoon et al. wish to draw from their experimental results with the matched-target design are seriously misleading: In our article, we note several points on which we believe that McKoon et al.'s reasoning seems to be mistaken. Because the relationships among these points are complex and somewhat unexpected, we advance the argument step by step.

First we note that, with respect to evidence of reactivation of the antecedent of the trace at the trace site (Probe Point 3), there is no contrast befween McKoon et al.'s (1994) matched-target experiments and their switched-target experiments. Proof of reactivation requires proof of differential priming: Related-unrelated target differences would have to be greater at the trace position than at a nontrace position. However, McKoon et al.'s results were uniform across the baseline and trace positions in both the matched-target and the switched-target experiments. Thus, no evidence has been presented that the method of selecting unrelated targets can influence outcomes concerning antecedent reactivation.

The only effect of the choice of unrelated targets was on the amount of apparent priming observed at each position. With matched targets, there was a significant related-unrelated difference in both baseline and trace positions; with switched targets, there was none. This discrepancy must be accounted for. We suggest that it was due to poor choice of matched unrelated targets: The matched control words were not truly matched, and the inherent difference between related target words and unrelated target words would give the appearance of priming.

What then remains to be explained is McKoon et al.'s (1994) failure to find evidence of WH-trace antecedent reactivation (with either type of target selection). We suggest that this is due to the nature of their sentence materials and perhaps also to some temporal parameters of response in this task that have not previously been attended to.

Finally we consider what lessons can be learned from all of this for future experimentation in this area. We hold that cross-modal tasks (lexical decision and naming) can be informative, though they must be carefully designed and implemented.

Priming, Antecedent Reactivation, and Baselines

Relevance of Baseline Data

Priming, by definition, is the facilitation of a response to an item relative to the baseline response time for that same item. To measure priming, we therefore need to know the baseline response time for each of the related items. There are several Mays to establish this. The obvious way would be to test the same item in two contexts that are identical except for the presence of a trace. This would be a comparison across sentences. Though worth trying, such a comparison would be at the mercy of unknown variables that may affect sentence processing difficulty. There is no such thing as two contexts that are exactly identical except for the presence or absence of a trace. Whether a trace is present is necessarily correlated with other aspects of sentence structure, which might in themselves affect response time. Sometimes materials can be devised in which these differences are minimized (see below), but usually it is deemed safer to keep the sentence material constant and compare different target words instead.

A target word can be compared with a word that has the same baseline response time as the target but is not the antecedent of the trace and hence is not facilitated in this context. There are two ways to identify a word that can be used for this purpose. One is to establish response times for a variety of words in some (normal) context that has nothing to do with traces or their antecedents, and on this basis to pick a control for each related item that is identical (or close) to it in response time. The other approach, not quite so accurate, is to match related and unrelated words on the basis of length and frequency, which are believed to be the major determinants of response time in a lexical decision task. The switched-targets design that McKoon et al. (1994) advocate can be seen as an instance of the former. This design uses the very same words to guarantee that there is no inherent difference in their baseline lexical decision response times. If a related-unrelated difference is obtained using this design, it could not be due to inaccuracy in matching control words with target words.

Suppose now that priming has reliably been established at the position of the trace. This is not sufficient to show that processing of the trace has reactivated its antecedent. Priming might occur because the related target is associated generally with the meaning of the sentence or because the original activation of the antecedent has not completely declined by the time the trace is processed. To guard against these possibilities, it must be shown that the related-unrelated difference is specific to the trace site, that is, it is absent at other sentence points, and particularly between the overt antecedent phrase and the trace. At the least, it must be shown that the related-unrelated difference at the trace position is significantly greater than any such difference at a position between the antecedent and the trace. Note that this safeguard is equally essential whether the targets are switched or matched. In both cases, the sign of antecedent reactivation is a significant interaction between target type and sentence position, where the trace position is compared with a postantecedent but pretrace position such as McKoon et al.'s Position 2.

Note now that this necessary precaution of comparing priming at the trace position with priming at a baseline position eliminates the potential disadvantage of matched targets relative to switched targets. Comparison of Positions 2 and 3 (with provisos as noted below) checks for absence of residual activation before reactivation and also for accuracy of the match between related and unrelated targets. There is no danger that one control would undercut the other. If the original activation of the antecedent had not declined by Position 2, then related targets would produce faster response times than unrelated targets. This response time difference would reduce the possibility of a significant interaction; between sentence position (Probe Point 2 vs. 3) and target` type (related vs. unrelated), which is required to support claims of antecedent reactivation. To summarize, both switched-target designs and matched-target designs require a comparison of Position 3 with Position 2. A related-unrelated difference at Position 3 alone is not interpretable in terms of antecedent reactivation. Also, this comparison simultaneously disarms any inherent response time difference between related and unrelated targets.

McKoon et al.'s (1994) Findings

In the three experiments with switched targets in which Positions 2 and 3 were both tested, McKoon et al. (1994) found no significant difference between related and unrelated targets at any sentence position.[2] It can be concluded that there was no priming of the related target at the trace position, and thus no sign of reactivation of the antecedent. With matched targets, by contrast, there was a significant related-unrelated difference at Position 3 in three out of four experiments: Related targets had shorter response times than unrelated targets. However, there was also a significant difference in the same direction at Position 2 in all four of these experiments. There was no significant interaction between sentence position and target type, except in the one aberrant experiment (with simple sentences and all visual presentation) that showed a significantly greater advantage for the related targets at the baseline position than at the trace position, an unexpected outcome for which there is no obvious explanation. With the matched targets, then, just as with the switched targets, there is no evidence of reactivation of the antecedent by the trace. The significant effect at Position 3 might be the result of reactivation, but the results at Position 2 raise the possibility that it is due instead to the original activation of the antecedent that primes the related target throughout the sentence or to an inherent advantage of the related targets over the unrelated, regardless of the sentence context.

To summarize, in terms of antecedent reactivation, the matched-target experiments gave results indistinguishable from those of the switched-target experiments.

The one respect in which the two methods differ in their outcomes is the across-the-board advantage of related targets in the matched-target experiments. It is unclear that this is of any importance, given that the methods agree on the absence of antecedent reactivation. However, McKoon et al. (1994) emphasized this difference, taking it as a sign of a general unpredictability of outcomes with matched targets. Such concern though would be groundless if there were a straightforward explanation for the matched-target results. We turn to this now.

Why Did Matched and Switched Targets Differ?

We noted above the possibility that the baseline Position 2 was too close to the original occurrence of the antecedent phrase. The experiments reported in Nicol and Swinney (1989) guarded against this by the inclusion of a prepositional phrase following (and modifying) the subject of the relative clause. Its purpose was to provide time for the activation of the antecedent noun to decline before Positions 2 and 3. McKoon et al. (1994) omitted this "padding" in their simple sentences so that only two or three words intervened between the relative pronoun (which is what signals that the antecedent noun is the antecedent of a trace) and Position 2. It is possible, then, that Position 2 was in fact too close for comfort to the antecedent in these simple sentences. McKoon et al.'s switched-target experiments afford two comparisons between experiments with simple sentences and experiments with complex sentences tested under similar conditions at Position 2.. In one comparison (Experiments 4 vs. 5), neither sentence type showed a related-unrelated difference; in the other (Experiments 3 vs. 5), the simple did and the complex did not. Obviously, no firm conclusion can be drawn, though the latter comparison suggests that McKoon et al.'s simple sentences may fall just at the

[2] Although only the experiments that test Positions 2 and 3 are relevant to findings reported in Nicol and Swinney (1989), we note here that McKoon et al. did find significant facilitation of rcsponse times to associates at Probe Position 1.

Table I
Mean RTs (in Milliseconds) and Error Rates for McKoon et al.'s (1994) Word Targets (Experiment 1)

Target type	RT (M)	Error (%)
Related	631	4.6
Unrelated	656	6.7
Unrelated-related	25	2.1

Note. RT = response times.

margin of safety with respect to decline of initial activation before the baseline position.

For the complex sentences, however, this seems a less likely explanation. These sentences did include a padding phrase comparable to that in the Nicol and Swinney (1989) sentences. The only hard data on complex sentences tested with switched targets comes from McKoon et al.'s (1994) Experiment 5, and as noted, this showed no related-unrelated difference at Position 2, in contrast to the two experiments (7 and 9), which tested complex sentences with matched targets. Again, the evidence is scant, but the best conclusion seems to be that the padding was indeed sufficient. We should look elsewhere for an explanation of the positive effect that the matched targets show at the baseline position.

A second obvious source for the response time difference between related and matched unrelated targets would be an a priori difference in baseline lexical decision time for these targets. Because the McKoon et al. (1994) experiments did not speak to this, we tested its plausibility by means of a lexical decision task on their related and unrelated items in isolation to see whether the latter were inherently harder to respond to than the former.

Experiment 1: Lexical Decision to Target Items

Method

Subjects. Forty-three University of Arizona undergraduates participated in this experiment for course credit.

Materials and procedure. We used the 56 targets used by McKoon et al. (1994). Following McKoon et al. we presented the pair "Indian-warren" to subjects, but removed it from data analysis. (The same pattern of results obtained with and without this pair.) These 56 words were randomly interspersed among 56 pronounceable, orthographically legal nonwords and were preceded by 8 practice items. Order of presentation was random, with a different random order for each subject.

Again following McKoon et al. (1994) each target appeared on the computer screen until the subject responded or for 1,800 ms. We instructed subjects to make a lexical decision about each item and to respond as quickly as possible by pressing an appropriately labeled button on a button box. The button press stopped a clock that had been started with the appearance of the item. Subjects pressed a foot pedal to advance to the next item. Stimulus presentation and data storage were controlled by the DMASTR system.[3] Items were presented to subjects on a Princeton Ultrasync color monitor controlled by a 286 PC and appeared in normal IBM text font.

Results

Before analysis, we treated the data in the following fashion. Data from two subjects who responded with a greater than 40% error rate were discarded. All response times greater than 1500 ms were excluded from analysis. Only correct responses were included in the analyses.

The mean response times to the two sets of words are shown in Table 1. Note that the terms *related* and *unrelated* are used to classify the target sets as they were used in the McKoon et al. (1994) experiments.

We conducted two analyses of variance (ANOVAs): one with subjects (F_1) and one with items (F_2) as the random variable. Mean response times to the related targets were significantly faster than response times to unrelated targets: $F_1(1, 40) = 13.0$, $p < .001$, $MS_e = 885$; $F_2(1, 26) = 4.38$, $p < .05$, $MS_e = 2{,}127$. Error rates in the two conditions mirror the response time pattern, there were more erroneous responses to the unrelated targets than to the related targets: F1(l, 40) = 5.02, $p < .05$, $MS_e = 19$; $F_2(1, 26) = 3.5$, $p < .08$, $MS_e = 18$. These data indicate that the two sets of targets, though matched for length and frequency, are inherently imbalanced; length and frequency are not perfect predictors of lexical decision time. Thus, the significant related-unrelated differences that McKoon et al. observed at all sentence positions for these targets could be due to this inherent bias rather than to any kind of priming.

Is the 25-ms difference between target types in isolation sufficient to account for the differences observed in the cross-modal sentence processing situation? Certainly, the 25 ms is not as great as the differences in McKoon et al.'s (1994) matched-target experiments, which ranged from 28 ms to 66 ms. However, it is a common observation that response time differences are larger when response times are longer. The response times for the targets in conjunction with sentences in the McKoon et al. experiments are approximately 170 ms longer than response times for the same words in isolation in our Experiment 1, so the response time difference in their experiments would be expected to be correspondingly larger. Just how much larger, we cannot estimate. Clearly, though, McKoon et al. would need to establish that their response time differences go beyond the reach of this explanation to make the case that the matched-target design gives unaccountable results. No reason has been given to believe that properly matched targets will show spurious priming. As we have seen, even spurious priming due to mismatched targets will not be mistakable for antecedent reactivation.

Could Matched Targets Somehow Produce Spurious Selective Priming?

If imperfect matching of targets can sometimes create an appearance of priming, as outlined above, could it sometimes do so at the trace position but not at the baseline position? If so, this would be a more serious defect because mismatched targets could then produce spurious demonstrations of antecedent reactivation. In their General Discussion, McKoon et al. (1994) raised the prospect that using a matched-targets design could create spurious effects of this kind, that is, differential

[3] DMASTR Display Software was developed by Ken Forster and others, at Monash University and has recently been modified by Jonathan Forster at the University of Arizona.

facilitation at the trace position. This is not a result that they observed in any of their experiments. Nevertheless, they sketch a possible mechanism for this disturbing possibility on the basis of findings such as those of Wright and Garrett (1984), who reported that responses to a target word are facilitated if it is syntactically congruent with the preceding sentence context. Consider how this syntactic priming effect would work (n sentences such as Sentence 1. How could the related target be differentially favored at the trace position? Either the unrelated target would have to be more syntactically congruent at Test Point 2 than at Test Point 3, the related target would have to be more congruent at Point 3 than at Point 2, or both. However, this is feasible only if two quite unlikely coincidences hold. First, the target words would need to be imbalanced in such a way that the unrelated targets were predominantly verbs (which would constitute a syntactically congruent continuation of the sentence at Point 2, but not at Point 3), the related targets were predominantly nouns or adjectives (which would fit better at Point 3 than at Point 2), or both. Examination of the targets used in Nicol and Swinney (1989; see our Appendix) indicates that this is not the case: Related-unrelated target pairs were matched for syntactic category. Second, for differential syntactic priming to occur, subjects would have to be able to "splice" the visually presented target into a precise syntactic position in the auditorily presented sentence. It is important to note that in the Wright and Garrett (1984) study (and others since, such as Gorrell, 1989), the sentence was presented visually, and it was stopped when the probe word appeared. However, in Nicol and Swinney and other studies using cross-modal presentation, the sentence does not stop when the target word appears; it continues, with words that can be integrated only if the visual target is not integrated. We know of no evidence that subjects do attempt to incorporate visual target words into precise syntactic positions in auditorily presented sentences (or indeed whether they are capable of doing so), and in our experience the paraphrases that subjects give show no such intrusions of targets into the sentences.: Thus, to the best of our knowledge, no concrete cause for concern about the question of spurious differential priming has been reported. As noted, certainly, McKoon et al.'s own data do not exhibit any such effect. With only one exception (unexplained and in the opposite direction), their data are fully consistent across test positions.

One final note on the possibility that selective priming at trace positions in previous studies might be artifactual: Not all of the studies that have found antecedent reactivation effects are susceptible even in principle to this counter explanation. For example, experiments in Nicol (1988) and the experiments reported in Fodbr (1993), cited by McKoon et al., examined reactivation'of;the antecedent of a pronoun or reflexive. The experiment used sentences such as Sentence 2 (# indicates the probe point).

2. The boxer told the skier that the doctor for the team was sure to blame himself/him # for the injury.

Note that this permits a well-controlled across-sentence comparison. The context up to the critical word is the same, so target words presented at that point could not differ in syntactic congruence. Furthermore, the related target for the reflexive version of the sentence serves in effect as the unrelated target for the pronoun version, and vice versa, so that selection or misselection of targets could not be responsible for the significant interaction between targets and sentence versions (pronoun-reflexive). Materials can be devised that allow nonovert elements such as WH-trace to be studied in similar fashion. Though sentencq with and without traces cannot be exactly identical in other respects, they can be made similar in lexical and syntactic structure in the neighborhood of the trace. For instance, relative clauses (with trace) can be compared with complement clauses (without trace), as in Sentences 3 and 4.

3. The receptionist knew the doctor that the journalist had telephoned (trace) after the meeting.
4. The receptionist and the doctor knew that the journalist had telephoned after the meeting.

With sufficient ingenuity in creating sentential materials, it may be possible to use more across-sentence designs in future, to complement the across-target designs. If the two continue to give comparable results, then this would increase confidence that what has been taken to be antecedent reactivation is not an artifact of any particular method of study.

Nonreplication of the WH-Trace Antecedent Reactivation Effect

We have observed so far that McKoon et al.'s (1994) data do not show that the matched-targets method gives different results from the switched-targets method with respect to antecedent reactivation at a trace position relative to a baseline. Because of the mismatching of their matched targets, McKoon et al.'s data do not even support the claim that the matched-targets design, properly implemented, can create spurious priming effects, either across the board or at selective positions. Hence, no reason has been given for believing that the positive results of previous experiments with the matched design are untrustworthy. We ourselves have used both switched and matched designs and have found WH-trace antecedent reactivation with both. The earlier published studies used matched targets; more recent studies have used switched . targets (Nicol & Pickering, 1993; Nicol, 1993; Nicol, Conway, & Denham, 1993). Results with the two kinds of targets have been similar.

However, the McKoon et al. (1994) experiments-with both matched and switched targets-did fail to show any WH-trace antecedent reactivation effect. In none of their experiments did McKoon et al. observe a significantly greater related-unrelated difference at the trace position than at the baseline position. Such a consistent failure to replicate the WH-trace antecedent reactivation effect reported in Nicol and Swinney, 1989, (also Hickok, 1993; Nicol & Pickering, 1993; Swinney & Osterhout, 1990; Zurif, Swinney, Prather, Solomon, & Bushell, 1993; and the studies cited above) is disturbing. It is potentially damaging to the interesting theoretical conclusions that have been drawn from the apparent existence of the antecedent reactivation phenomenon. In this section, we consider some possible reasons for the absence of a WH-trace effect in these experiments.

General Experimental Variables

With only one exception, McKoon et al.'s (1994) data can be summarized by the generalization that within each of the eight experiments that tested more than one sentence position, the results were similar across the positions tested: A related-unrelated difference was found at both positions or at neither. (The one exception is Experiment 6, in which there was a significant advantage for related targets at the baseline, an advantage for unrelated targets at the trace position, and a significant interaction between positions. McKoon et al. themselves remark that there is no obvious explanation for this.) For instance, even with switched targets, results for Position 2 were variable across experiments but in each case were similar to those at the other position tested (whether Position 1 or Position 3). This across-the-board pattern of results raises the possibility that the primary determinant was not the experimental variable of interest, but some environmental variable such as the level of background noise against which the sentence was heard, the quality of the audiovisual stimulus, or both, the motivation of the subjects, the clarity of the task instructions given by the experimenter, and so forth. Though general factors of this kind might possibly swamp out any differential effect between Sentence Positions 2 and 3, they are unlikely to produce differential effects. If so, there would obviously be no general cause for concern about the replicability of the antecedent reactivation effect raised by these particular results.

There is one factor unique to the cross-modal design, however, which deserves closer attention. The cross-modal paradigm is valuable in that it affords access to sentence parsing processes as they are occurring. An event of interest occurs in a sentence-in the present case, an event of recognizing a trace and coindexing it with its antecedent. By assumption, that event creates a temporary state (activation of the meaning of the antecedent) that can influence the processing of the target word.[4] Consider the timing of the word processing and the sentence event processing. Typically the word is presented simultaneously with (what is believed to be) the onset of the state evoked by the sentence event. However, there is no reason to suppose that the onset of this state is instantaneous. Both common sense and results on similar phenomena (e.g., Bever & McElree, 1988; MacDonald & MacWhinney, 1990; McElree & Bever, 1989) suggest that it would increase over the course of some milliseconds and then decline again. Suppose it is n ms before it is at a detectable level, p ms before it reaches its peak, and m ms before it declines to a nondetectable level. Now consider the word processing. Minimally, this will consist of the following stages: Stage 1, visual processing; Stage 2, lexical access (and decision); and Stage 3, motor response planning and performance. Let us suppose that only Stage 2 is influenceable by the concurrent linguistic processing of the sentence; Stages 1 and 3 are in different modules and hence immune. Then any part of the linguistic phenomenon of interest that occurs during Stage 1 or Stage 3 of the lexical processing will not be registered in subjects' performance on the lexical decision task.

Previous applications of the cross-modal priming paradigm have simply taken it on trust that Stage 2, the linguistically sensitive portion of lexical decision making, will coincide with, or at least overlap with, the $m - n$ ms during which the antecedent of the trace is activated; if the experimenter so arranges things that the visual stimulus is presented synchronous with the offset of the word that precedes (early) or follows (late) the trace position. This may not be a safe assumption. The unpredictable failures of the task suggest that it is not. Whether the necessary overlap occurs will presumably depend on how long the lexical decision takes and what proportion of the response time is attributable to visual processing as opposed to lexical processing, which in turn will depend on the quality and size of the visual stimulus, possibly the exposure time of the target, and the length and frequency of the word. The relative timing of sentence and word processing may also depend on how difficult the sentence processing is in two ways: A high general processing load at that or earlier points in the sentence may have resulted in a temporary processing lag and the complexity of the specific process of interest (the steps involved in detecting and coindexing the trace) could delay onset of its measurable consequence (the activation of the meaning of the antecedent). Of course, sensory stimulus variables will be relevant for the time course of sentence processing too: A high-quality audio signal will likely be processed faster and more uniformly than a low-quality input.

What all of this suggests is that in this or any other paradigm in which the experimenter decides when and where to look for subjects' response to a stimulus, a null result may only indicate that the test point happens not to have been optimally chosen. Before rejecting the experimental hypothesis, it would therefore be appropriate to check alternative timing relations between sentence and probe. (Note that the alignment between sentence and probe presumably could not create an effect such as a related-unrelated difference in response time, though it could very well destroy one.)

Comprehension Difficulty of Sentences

A second possibility to be considered, in case of a failure to find antecedent reactivation, are complications in the sentence materials that could obscure the relation between the WH-trace and its antecedent so that the antecedent is not consistently reactivated at the trace site. Again, if this were so, it would not be a cause for general concern because the antecedent reactivation hypothesis applies only to sentences in which structure and meaning can be decoded on-line by the perceiver.

McKoon et al. (1994) note that the syntactic structure of their sentences was modeled after that of the sentences used in Nicol and Swinney (1989). However, the two sets of sentences were not identical, and on inspection, it does seem possible that the McKoon et al. materials were less comprehensible than those of Nicol and Swinney. It is not easy to pin down exactly what makes the McKoon et al. sentences difficult to understand, but our own intuition is that in many cases the constituent meanings cannot easily be fitted together into coherent propositions. Though it is certainly not impossible to

[4] Let us assume, for the sake of the present argument, that the activation is real.

make sense of them, a considerable amount of inference generation is required to do so, and this could create processing difficulty (see, for example, Crain and Steedman's [1985] discussion of effects of presuppositional complexity on processing). Note that the problem is not simply that the McKoon et al. sentences were implausible: Swinney & Osterhout, 1990, have shown that implausible sentences such as Sentence 5 show an antecedent priming effect just as plausible sentences do:

5. Everyone watched the enormous heavyweight boxer that the small 12-year-old boy on the corner had beaten so brutally.

However, such sentences are implausible in a quite specific and very different fashion than McKoon et al.'s. Determining what proposition they express is fairly easy; it is just that the proposition in question is recognized as unlikely to be true. In McKoon et al.'s sentences, the problem seems to be more that of establishing what proposition is expressed at all.

Consider, for example, the McKoon et al. (1994) Sentence 6, as compared with a Nicol and Swinney (1989) Sentence 7.

6. The optometrist aided the victim that the barber in the airport hurt in the fight.
7. The cyclist fixed the light that the motorist in a hurry had hit with his van.

Subjects reading or hearing Sentence 6 may start to run into trouble on encountering the word *victim*. This word carries rich presuppositions (someone has unfairly suffered harm, physical or otherwise, inflicted by someone else, probably deliberately, etc.), but the word occurs before any mention of an event that could supply such a background, and there is no prior discourse that satisfies these presuppositions. There is the added complication of having to construct a scenario in which an optometrist, under that description, would be aiding a victim. By the time the barber arrives, the effort to contextualize this story may take more resources than subjects have available (especially because it is only one sentence of more than 80, of which at least the 28 experimental sentences are similar in presuppositional complexity to Sentence 6). By contrast, Sentence 7 describes a fairly common and comprehensible scenario. Subjects hearing or reading this sentence need to infer that a light (probably a bicycle light) needed fixing, but the characters and the activities fit together in a natural way.

Might the simple versions of the McKoon et al. (1994) sentences be exempt from this danger? We cannot say for certain because a full listing of the simple sentence stimuli did not accompany their article. For reasons previously given, we do not think these simple sentences are appropriate for antecedent reactivation testing. However, we note also that they too may have been difficult, though for other reasons. From the examples in the text it appears that proper names, or expressions like "somebody," were substituted for some of the descriptive NPs. These simpler NPs may have been easier to incorporate into the sentential semantics than descriptive NPs, but, because their referents were totally unknown, the sentence content may also have been more difficult for subjects to remember.

To test the hypothesis that the meanings of McKoon et al.'s (1994) complex sentences were more difficult to grasp than those of the Nicol and Swinney (1989) sentences, we presented both sets of sentences to subjects in the following judgment experiment.

Experiment 2: Sentence Judgments

Method

Subjects. Sixteen subjects from the University of Arizona participated in this experiment for course credit.

Materials and procedure. McKoon et al. (1994) tested twenty-eight complex sentences, and Nicol and Swinney (1989) tested 48. In both cases, one sentence from each set was omitted from data analysis. Hence, in this experiment, 27 McKoon et al. sentences and 47 Nicol and Swinney sentences were tested. The McKoon et al. sentences are published with their article in this issue; the Nicol and Swinney sentences are in the Appendix. As can be seen, the two sets of sentences are roughly comparable in terms of syntactic structure and word length. Sentences were presented word by word in centered position on a computer screen, at a rate of 300 ms per word. This method and rate of presentation are similar to that used by McKoon ct al. in their Experiments 1, 2, 6 and 7, with all visual presentation (though our presentation rate was fixed and theirs varied with word length). At the end of the sentence, the prompt N/Y appeared, signaling the subject to make a sentence judgment. We asked subjects to read each sentence, to judge whether they had grasped the sentence well enough to be able to paraphrase what the sentence had been about, and to indicate their judgment by pressing an appropriately labeled response key. (This is the "got it" judgment task of Frazier, Clifton, and Randall, 1983). They were warned that, periodically, and independent of their sentence judgment, they would be asked to writc down the gist of the last sentence they had read. Eight sentences were probed in this way. We included this task, as in McKoon et al.'s experiments, to keep the subjects' attention; paraphrases themselves were not quantified or analyzed. The apparatus and software used tc conduct this experiment were identical to that of Experiment 1.

Results

Our primary interest was in the categorization of sentences into ones that subjects thought they had grasped and the ones they thought they had not. The mean percentage of yes responses-responses indicating that subjects had grasped the sentence-were as follows: 51.8% for the McKoon et al. (1994) sentences and 72.2% for the Nicol and Swinney (1989) sentences.

We note that the McKoon et al. (1994) Sentence 6 received a yes response 50% of the time, and the Nicol and Swinney (1989) Sentence 7 received a yes response 69% of the time, so these examples are fairly representative of the sets from which they are drawn. An ANOVA, with subjects as the random variable, revealed that subjects responded *yes* significantly more often to the Nicol and Swinney sentences than to the McKoon et al. sentences, $F(1, 15) = 24.74$, $p < .001$, $MS_e = 134$.

Any response time that exceeded the subject's mean by at least two standard deviations was replaced with the exact 2-SD value (accounting for 3% of the response time data). Response times were relatively unrevealing, though suggestive: Response times to respond yes were slightly faster to the Nicol and Swinney (1989) sentences (1,300 ms vs. 1,388 ms for the McKoon et al. sentences), and response times to respond no were slightly faster to the McKoon et al. sentences (1,871 ms vs. 1,961 ms for Nicol and Swinney sentences). Although neither difference was significant (p = .195 and p = .27, respectively), their interaction (Response x Stimuli) approached significance, $F(1, 15) = 3.97$, $p < .07$, $MS_e = 31,799$, and this tends to support the idea that subjects were more confident in their grasp of the Nicol and Swinney sentences than of the McKoon et al. sentences.

As noted previously., this difference in comprehensibility could be the causeof McKoon et al.'s (1994) failure to replicate the WH-trace effect. If the demand on memorial and interpretational'resources is too great, then subjects may lose track of the antecedent by the time they get to the position of the trace. Or there may be a failure to recognize the existence of a trace at that position; its presence is not overtly signaled in the input but must be deduced by the parsing routines. In either case, no priming of the antecedent would be predicted. (Alternatively, subjects might begin to rehearse the sentence during its presentation, in which case they might show priming for the antecedent at a number of points throughout the sentence.) More. generally, if subjects are doing something other than comprehending the sentences in a normal way, effects of linguistic structure cannot be counted on.

Implications for Future Research

We have argued here that the difference between the results of McKoon et al.'s (1994) experiments and those reported elsewhere in the literature is unlikely to be due to whether target items are matched or switched and is more likely to be due to poor matching of the matched targets, comprehension difficulty of the sentences, and possibly some experiment-specific environmental variables. The question of whether a matched- or switched-target design is optimal is still open as are other questions about how to maximize the reliability and sensitivity of the cross-modal priming paradigm.

In practical terms, switching targets is simpler than matching them. However, it does require some care to avoid inadvertently pairing a supposedly unrelated target with a sentence to which it bears some relation. It is also the case that in the switched targets paradigm, an unrelated target for one sentence is a word chosen specifically because it is highly related to another sentence. If the latter sentence were to occur just before the former, then this association might cause unwanted priming of the unrelated item, thus weakening the apparent priming of related targets. Temporal separation of the relevant sentences in presentation to subjectsis an obvious precaution here, but the safety margin is not known.

The use of switched targets raises a technical question about what constitutes an item: a sentence or a lexical target. If the items analysis is based on sentences, then the comparison will be between a related target and an unrelated target not matched with it in response time. If target words are what count as items, then the analysis compares response times to a word as it appears with two quite different sentences not matched in complexity, plausibility, and so forth. Neither alternative is optimal. Both potentially introduce greater variability than the matched-targets approach, reducing the chance of obtaining significant results. Both presuppose, without grounds, that sentence complexity at the probe position does not interact with target word frequency. For this reason, it may be useful, in a switched-targets design, to pair the target words on the basis of response time (previously determined) or length and frequency (as approximate predictors of response time). Each sentence would be tested with a related target and with an unrelated target, and the two targets would be comparable in all relevant respects except relatedness. The same two targets would appear with another sentence, but now switched, that is, the unrelated target for the first sentence would be the related target for the second sentence and vice versa.

Additionally, some thought should be given to the timing of target word presentation relative to the position of interest in the sentence. What is the ideal timing for detecting the maximum effect of a sentence event? Clearly, there is no general answer to this question. If visual processing is easy and sentence processing is difficult, then a "zero" delay for the target might be too early. The converse is also possible, though perhaps less likely in practice, that is, if visual processing is slow and sentence processing is easy, the lexical task might miss all but the tail end of the sentence events. Because these factors can vary across subject populations, experiments and laboratories, the safest course for avoiding a spurious null result would be to test a range of target positions around the point of interest in the sentence.

Note that the cross-modal dual task paradigms are not alone in relying on alignment of the target stimulus and the sentence. Monitoring tasks, for example, also require a guess as to optimal placement of the target. All such tasks are to this extent more vulnerable than are off-line tasks, such as postsentential judgment, paraphrase, or probe recognition. However, it is hardly to be wondered at if greater sensitivity and greater vulnerability go hand in hand (as they tend to in sciences other than psycholinguistics). However, because on-line measures can be so much more informative than off-line tasks are about the process of assigning structure and meaning to sentences, it would be misguided to reject them. Rather, we should put our efforts into understanding and improving them.

As a final point, McKoon and Ratcliff (1994) follow up this article with another of their own. In this second article, McKoon and Ratcliff report the results of an experiment that used modified versions of .75% of the sentences used by Swinney et al. (reported in Nicol and Swinney, 1989; as noted above, Swinney et al.'s full set is listed in the Appendix). They altered the sentences so that they were no longer filler-gap sentences: In fact, the sentences no longer contained any mention of the filler before the probe point; hence, of the accompanying related-unrelated target words the related items were no longer, strictly speaking, related to a critical word that had appeared upstream from the probe. They used an all-visual serial presentation, in which the probe item appeared instead of the next word in the sentence (the sentence continued only once subjects had made a response). Surprisingly, they found a pattern of response times similar in one respect to that reported by Swinney et al.: faster response times for the related targets than to controls at the postverbal probe point. They concluded that this result suggests that the findings reported in Nicol and Swinney were entirely artifactual: Fast response times are obtained for any item that easily fits-syntactically, semantically, pragmatically-into the ongoing sentence.

However, there are a number of reasons why the results reported in McKoon and Ratcliff (1994) are simply not relevant to those reported in Nicol and Swinney (1989). The primary reason is that the all-visual presentation is different from the cross-modal presentation in two important respects. First, sentences are presented visually rather than auditorily. Differences in auditory versus visual presentation of stimuli have been noted to produce different effects (Darwin, Turvey & Crowder, 1972; Nicol, 1993; Sperling, 1960). Second, a target word appears, instead of a word in the sentence rather than concurrent with the words in the sentence. This is a critical difference. Such a technique is likely to encourage just what McKoon and Ratcliff were looking for, integration of the probe word into the sentence. After all, the all-visual task invokes many of the demand characteristics of a sentence completion task, which is known by experimenters in the field to be an integrative task, one in which sentential context plays a role in facilitating or inhibiting the decision response to a word appearing late in the sentence. Anyone familiar with the technique might have predicted that plausibility, goodness-of-fit in the Cloze sense, and other such related factors would enter into the results that they found with this technique. (Further, their relatively long response times [200-300 ms longer than those reported in many cross-modal experiments] suggest that subjects would have time to gauge general goodness-of-fit.) What does come as a surprise is that they did not obtain a similar finding in their first round of all-visual experiments reported in McKoon et al. (1994), although, as we have discussed above, the comprehension difficulty posed by their sentences may have eliminated such integration effects. In any case, one cannot simply take for granted that the cross-modal and all-visual presentations are comparable, and the conclusions drawn from findings obtained with one technique cannot, without due evidence, be applied to those obtained with the other. That they are not comparable is suggested by the differences in the pattern of response times reported in McKoon and Ratcliff versus those reported in Nicol and Swinney. Both found faster response times to related probes after the verb. However, McKoon and Ratcliff found much slower response times for related probes before the verb, whereas Nicol and Swinney reported almost no difference in response times before the verb. Without statistical tests, we cannot know whether these two patterns of data . differ significantly, but the difference does suggest that the two tasks might be producing different effects. Further, some experiments have used the same local sentential context for both experimental and baseline conditions (as in Nicol and Pickering [1993], in which the probe always appeared after an embedded verb); for these, it would be especially difficult to argue that `congruence effects are at play. In summary, although McKoon and Ratclifl's results suggest that the all-visual presentation-in which a probe word appears instead of a word in the sentence-could be vulnerable to congruency effects, they have not demonstrated that the cross-modal technique is equally yulnerable to such factors. In the end, we see no grounds for disbelieving the claim that syntactic knowledge guides on-line sentence processing and that the cross-modal technique, when used carefully, is sensitive to those effects.

References

Bever, T. G., & McElrec, B. (1988). Empty categories access their antecedents during comprehension. *Linguistic Inquiry*, *19*, 35-45.

Crain, S., & Steedman, M. J. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowry, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing*: *Psychological, computational and theoretical perspectives* (pp. 320-358). Cambridge, England: Cambridge University Press.

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, *3*, 255-267.

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

Frazier, L, Clifton, C., & Randall, J. (1983). Filling gaps: Decision principles and structure in sentence comprehension. *Cognition*, *13*, 187-222.

Gorrell, P. (1989). Establishing the loci of serial and parallel effects in syntactic processing. *Journal of Psycholinguistic Research*, *18*, 61-74.

Hickok, G. (1993). Parallel parsing: Evidence from reactivation 'in garden-path sentences. *Journal of Psycholinguistic Research*, *22*, 239-250.

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

MacDonald, M. C., & MacWhinney, B. (1990). Measuring inhibition and facilitation from pronouns. *Journal of Memory and Language*, *29*, 469-492.

McElree, B., & Bever, T. G. (1989). The psychological reality of linguistically defined gaps. *Journal of Psycholinguistic Research*, *1*8, 21-36.

McKoon, G., & Ratcliff, R. (1994). Sentential context and on-line lexical decision. *Joumal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, XXXX-XXXX.

McKoon, G., Ratcliff, R., & Ward, G. (1994). Testing theories of language processing: An empirical investigation on the on-line lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, XXXX-XXXX.

Nicol, J. L. (1988). *Coreference processing during sentence comprehension*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Nicol, J. L (1993, March). *Processing syntactic dependencies: Task-specific effects*. Poster presented at the Sixth Annual City University of New York Conference on Human Sentence Processing, University of Massachusetts at Amherst.

Nicol, J. L, Conway, L., & Denham, K. (1993). *Priming in visually-presented passive sentences*. Poster presented at the Sixth Annual CUNY Conference on Human Sentence Processing, University of Massachusetts at Amherst.

Nicol, J. L., & Pickering, M. J. (1993). Processing of syntactically ambiguous sentences: Evidence from semantic priming. *Journal of Psycholinguistic Research*, *22*, 207-237.

Nicol, J. L, & Swinney, D. A. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, *18*, 5-19.

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

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, *74*, 1-29.

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

Swinney, D., Ford, M., Fraugnfelder, U., & Bresnan, J. (1988). *On the temporal course of gap-filling and antecedent assignment during sentence comprehension*. Unpublished manuscript.

Swinney, D., & Osterhout, L (1990). Inference generation during auditory language comprehension. In A. Graesser & G. Bower (Eds.), *The psychology of learning and motivation* (Vol. 25, pp. 17-33). New York: Academic Press.
Wright, B., & Garrett, M. (1984). Lexical decision in sentences. *Memory & Cognition, 12*, 31-45.
Zurif, E. B., Swinney, D., Prather, P., Solomon, J., & Bushell, C. (1993). An on-line analysis of syntactic processing in Broca's and Wernicke's aphasia. *Brain & Language, 45*, 448-464.

Received October 20,1993
Revision received March 14, 1994
Accepted March 15,1994

Appendix

List of Sentence and Target Stimuli for Swinney, Ford, Frauenfelder, and Bresnan (1988)

Sentence	Related target	Unrelated target
1. The police stopped the boy that the crowd at the party accused of the crime.	girl	body
2. The old man picked up the apple that the baby in the carriage threw in the gutter.	fruit	bench
3. The priest enjoyed the drink that the maid at the inn poured for the guests.	wine	rain
4. The worker repaired the truck that the horse at our place kicked in the fender.	van	net
5. The woman smelled the rose that the nun at our store picked for the vase.	thorn	louse
6. The judge convicted the theif that the clown at the show caught with the money.	crook	drone
7. The baby ate the bread that the bird at the zoo dropped by the fence.	crust	snail
8. The diner found the ring that the chef at the club lost in the soup.	gold	seat
9. The officer found the brick that the kids from the suburbs threw at the house.	clay	film
10. The bird ate the fish that the waves from the storm washed onto the shore.	trout	spice
11. The tailor hemmed the cloak that the actor from the studio needed for the performance.	robe	goat
12. The hunters caught the deer that the hounds in the field chased into the cave.	antelope	dictator
13. Lightning struck the house that the trees on the hill blocked from the wind,	home	well
14. The teacher caught the snake that the nurse at the school saw near the barn.	reptile	booklet
15. The farmer sold the fruit that the cat at our house knocked from the tree.	pear	cord
16. The people recognized the shirt that the prisoner near the fence stole from the old woman.	pants	beans
17. The seamstress designed the suit that the prince on the horse tore during the hunt.	dress	chair
18. The cop found the car that the addict from the city sold to the garage.	jeep	dish
19. The lifeguard kicked the box that the guide from the tour left on the beach.	carton	coupon
20. The pianist played the song that the wife of the mayor heard at the wedding.	tune	coin
21. The union leader addressed the woman that the supervisor at the factory fired after the strike.	female	cousin
22. The woman cleaned up the mud that the guests at the party brought into the house.	dirt	coat
23. The doctor treated the tourist that the tiger from the zoo had bitten on the hand.	trip	ship
24. The waiters cleaned up the milk that the customer at the table spilled on the tablecloth.	cream	tooth
25. The jogger saw the accident that the ice in the street had caused yesterday morning.	crash	delay
26. The painter lit the pipe that the lawyer in Germany had sent as a Christmas present.	smoke	bread
27. The boy found the pottery that the Indians from the Badlands had made years ago.	bowl	plug
28. The scientist removed the rats that the monkey in the cage watched with great interest.	mice	doll
29. The lady removed the branches that the storm from the west had ripped from the trees.	twigs	prams
30. The student brought the book that the professor at Yale had recommended the year before.	novel	rifle
31. The old woman watched the parade that the mayor on the grandstand had planned for many weeks.	march	visit
32. The visitors cooked the food that the campers near the mountain brought as a gift.	meat	cast
33. The youth delivered the letter that the secretary in our department wrote for the boss.	mail	bond
34. The man rescued the hikers that the pilots on a test run spotted near the river.	climber	drummer
35. The audience liked the wrestler that the reporter from the country condemned for foul language.	fighter	cleaner
36. The people cheered the athlete that the water from the road splashed in the face.	runner	helmet
37. The director watched the skater that the comedian from the show suggested for the musical.	ice	cup
38. The receptionist greeted the dignitary that the newspapers from Washington had described in great detail.	famous	active
39. The butler insulted the merchant that the millonaire from Hollywood had hired on a trial basis.	business	question
40. The bellhop delivered the telegram that the manager in the office expected since early last week.	message	housing
41. The cyclist fixed the light that the motorist in a hurry had hit with his van.	bulb	lamb
42. The gourmet tasted the wine that the waiter in the white pants had poured from an old bottle.	alcohol	nursery
43. The busboy pocketed the coins that the speaker from the conference had left on the table.	wallet	peanut
44. The man caught the fish that his buddy on the bank picked up with the net.	fin	ram
45. She read the note that the plumber in the blue van left on the door.	card	disk
46. The sergeant wrote on the paper that the soldier in dress whites brought to the office.	pencil	branch
47. The therapist examined the leg that the captain of the team injured in a tennis match.	foot	boat

Antecedents are underlined; following each sentence, target pairs are shown. Note that average frequency for related and control probes across all items is 68.6 and 74.0, respectively (Kucera & Francis, 1967). However, related-control probe pairs were chosen on the basis of matching words for reaction time in isolated lexical-decision format; this a priori pretest was conducted on the same population (but different sample) of students that participated in the experiments using these probes.