

## Sentential Context and On-Line Lexical Decision

Gail McKoon and Roger Ratcliff

J. L. Nicol and D. Swinney (1989) reported that lexical decision response times to a test word that was related to the implicit object of a verb were faster, when tested immediately after the verb, than response times to a control test word. They concluded from this result that the relation between the implicit object and the verb was understood during comprehension. In G. McKoon, R. Ratcliff, and G. Ward (1994), another interpretation was suggested, that the faster lexical decision response times were due to the better semantic-pragmatic fit of the related test words to the sentential contexts, and experimental data were presented to support this interpretation. In response, J. L. Nicol, J. D. Fodor, and D. Swinney (1994) pointed out some possible problems with these experiments. The experiment presented in this article undermines their arguments. By using sentences in which there was no implicit object, exactly the same pattern of results was found as originally reported by J. L. Nicol and D. Swinney, thereby implicating some factor other than syntactic processing of an implicit object as an explanation of their finding.

McKoon, Ratcliff, and Ward (1994) and Nicol, Fodor, and Swinney (1994) discussed various problems that can arise when lexical decision probe items are inserted into sentence comprehension processes. In the course of these discussions, a reader might find the authors using such words as *syntax*, *design*, *subjects*, *results*, and *confounding*. These words fit the general semantic and pragmatic context of the articles. However, it is most likely a reader would not find the word *petunia*—it does not fit.

In McKoon et al. (1994), we suggested that the semantic and pragmatic (and also syntactic) fit of a probe word to the context in which it was tested could affect how quickly and accurately subjects respond to that word. This suggestion was prompted by the results of testing two sets of words in the same sentence contexts, one set of words intuitively fitting better to the contexts of the sentences than the other. Responses for the better fitting words were consistently faster than responses for the other words, even though the two sets of words were equated for length and frequency of appearance in English (Kucera & Francis, 1967). The advantage for the better fitting words was quite large, about twice as large as the difference between response times for the two sets of words when they were tested outside the sentence contexts (in a list of single test items; Nicol et al., 1994).

In McKoon et al. (1994), the better fitting test words had faster response times at both of the two test positions in a

sentence. This suggests that there was a good fit between a word and its context sentence as a whole, a fit that did not vary between the earlier test position and the later test position. In this article, we extend our argument and provide data to show that the fit can also vary with test position. These new data should of course not be surprising: At later test positions in a sentence more and different information may be available to make a word more or less consistent with the sentence than it might have been at earlier test positions.

Consider, for example, a sentence that begins *The baby*. Many probe words might fit the *baby* context about equally well, including the probes *bench* and *fruit*—babies can be sat on benches or look at benches, and they can eat fruit. So *bench* and *fruit* might have about equal response times if they were tested immediately after *The baby*. If the probe position were moved to later in the sentence, after *The baby threw*, then the context for the probes would change. Babies can throw fruit but they cannot throw benches; the hypothesis was that response time to the probe *bench* should suffer accordingly. This hypothesis follows from the same reasoning that was presented in our earlier article (McKoon et al., 1994), in which we suggested that subtle interactions among the semantic, syntactic, and pragmatic contexts of a sentence and test position could affect response times for test words; *marshmallow*, we suggested, would not fit well into a test position that immediately followed verbs such as *assault*, *drive*, or *bribe*.

We used materials that were adapted from materials previously used in experiments reported by Nicol and Swinney (1989, provided by Nicol et al., 1994, Appendix). An example of one of their sentences is shown in Sentence 1 below. In this sentence, the syntactic object of the verb *accused* is *the boy*. Nicol and Swinney hypothesized that the language processing system understands what the object of the verb is and understands it immediately when the verb is comprehended. The consequences of this processing can be detected empirically, they proposed, by fast response times immediately after the verb to a word highly related to the object. Specifically, for Sentence 1, the result of understanding that the object of the verb *accused* is *boy* would be to speed response times for the test word *girl* from Test Position 1, where *the boy* is not involved

---

Gail McKoon and Roger Ratcliff, Psychology Department, Northwestern University.

This research was supported by National Institute of Deafness and Other Communicative Disorders Grant R01-DC01240, U.S. Air Force Office of Scientific Research Grant 90-0246 (jointly funded by National Science Foundation [NSF]), and NSF Grant SBR-9221940 to Gail McKoon and by National Institute of Mental Health Grants HD-MH44640 and MH00871 to Roger Ratcliff. We thank Chuck Clifton, Gary Dell, Steve Greene, and Janet Nicol for useful comments on an earlier version of this article.

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

in processing, to Test Position 2, where it has been understood as the object. Such a speedup from Position 1 to Position 2 should be observed for *girl* but not for the control test word *body*. The data reported by Nicol and Swinney matched this predicted pattern.

The police stopped the boy that the crowd at

the party 1 accused 2 of the crime. (1)  
           girl          girl  
           body         body

In this article, we show evidence for an alternative explanation for the pattern of results reported by Nicol and Swinney (1989), an explanation in terms of the semantic, pragmatic, and syntactic fit of the test words to the test positions. The reason a related test word (e.g., *girl*) has faster response times in one test position than another is that it fits the context in one test position better than the other, in ways that do not apply to the control test word. To test our explanation, we designed our experiment to be like the experiment reported by Nicol and Swinney (1989) by using sentences from that experiment and the same test words for the sentences, except, for each sentence, we moved the initial part of the sentence that supplied the potentially understood object of the verb to the end of the sentence, past the test positions. Sentence 1 above became Sentence 2 below. If, for our new modified sentences, we find that response times for the related word speedup from Test Position 1 to Test Position 2, whereas response times for the control word do not, then it cannot be due to processing of an implicit object because there was none.

The crowd at the party 1 accused 2 the boy. (2)  
                           girl          girl  
                           body         body

## Method

### Subjects

Twenty-four subjects participated in the experiment to receive credit for an introductory psychology class.

### Materials

For the experimental sentences, we modified 36 of the 47 sentences from the appendix of Nicol et al. (1994; the sentences originally used for the experiment reported by Nicol and Swinney, 1989); these were the first 36 sentences except for 3 that we replaced with others of their sentences to eliminate repetitions of content words between sentences. For each sentence, we used the same test words as given by Nicol et al. We used only 36 sentences to make the experimental design comparable to the designs used by McKoon et al. (1994) and to keep the experimental session less than 30 min in length. We modified each sentence as previously described in the introduction. (We did not use these materials for the experiments in McKoon et al. because they were not available to us at that time.) Some other examples of modified sentences are the following: "The baby in the carriage threw the apple," test words *fruit* and *bench*; "The maid at the inn poured the drink," test words *wine* and *rain*; "The clown at the show caught the thief," test words *crook* and *drone*; "The actor from the studio needed the cloak," test words *robe* and *goat*. In each example, the first test word was the one related to the object of the verb in the original Nicol and Swinney sentence. However, in our versions of the materials, the

object had not yet occurred before the test positions, so it could not be responsible for any obtained results. Instead, our alternative proposal is plausible: In each example, the first test word seems to fit the second test position better than the second test word does; babies are more likely to throw fruit than benches.

For filler sentences, we used the same 48 sentences as were used in the experiments described in McKoon et al. (1994).

### Procedure

The sentences and test words were both presented visually on the screen of a PC, with the same procedure as described in McKoon et al. (1994). Each word of a sentence was displayed for 170 ms plus 17 ms multiplied by the number of letters in the word; then the word was erased and the next word of the sentence presented in the same location. A test word was presented five spaces to the right of the location for words of the sentences, and it was marked with two trailing asterisks. There was no extra time between the offset of a word of a sentence and a test word that followed it. After a subject responded to a test word, any remaining words of the sentence were displayed. To encourage the subjects to read the sentences, they were occasionally asked to recall the sentence they had just read; there were eight such recall tests. For further details see McKoon et al.

### Design

There were four conditions in the experiment: A test word was presented either immediately before the verb or immediately after the verb, and either the first test word (the "related" one in the Nicol and Swinney, 1989, experiments) or the second "control" test word was presented. The four conditions were crossed with groups of subjects (6 per group) and sets of sentences (9 per set) in a Latin square design. We randomized the order of presentation of sentences, with the randomization changed after every second subject.

## Results

We calculated mean response times and error rates for each subject and item in each condition (eliminating responses longer than 2,000 ms, about 2% of responses), and differences were tested with analyses of variance (ANOVAs),  $F_1$  for subjects' effects and  $F_2$  for items' effects ( $p < .05$ ). Mean response time for filler test words was 952 ms (8% errors). Mean response time for filler nonwords was 1,002 ms (7% errors).

If Nicol and Swinney (1989) were correct in their interpretation of their results, if facilitation for a related test word in the test position following a verb were due to understanding what was the object of the verb, then there should have been no facilitation in our experiment because no potential object was presented before the test positions. However, we did find facilitation: The data show the interaction that would have been predicted if there had been a prior object. Because there was no such object, the interaction must have come about for some other reason.

The mean response times for the first test word (the related word in the experiment reported by Nicol and Swinney, 1989) were 969 ms (5% errors) before the verb and 876 ms (2% errors) after the verb, a speedup of 93 ms. The corresponding means for the other test word, the control word, were 921 ms (5% errors) before the verb and 924 ms (8% errors) after the verb. We had expected that response times for the first test

words would be faster than response times for the second test words when tested after the verb because we thought they were better fits to the context at that point. We had not anticipated that the first test words would be slower when tested before the verb, but it is possible that their fit was worse at that point. Obviously, much more research will be needed before we have a firm grasp of what variables affect the fit of different words at different test points.

The critical interaction was significant,  $F_1(1, 23) = 7.00$  and  $F_2(1, 32) = 4.18$ . There was a marginally significant main effect such that response times in the first test position were slower than response times in the second test position,  $F_1(1, 23) = 4.24$  and  $F_2(1, 32) = 5.86$ . The main effect of test word was not significant,  $F_s < 1.0$ . The standard error of the response time means was 18 ms. In error rates, the interaction between test position and test word was significant in the subjects analysis,  $F_1(1, 23) = 5.37$ , but not the items analysis,  $F_2(1, 32) = 3.33$ . No other error rate effects were significant. The standard error of the error rates was 0.02.

Nicol et al. (1994) suggested that the experiments described in McKoon et al. (1994) failed to show interactions of the kind found here because of problems with "general experimental variables" that might have tended to "swamp out" interactions (although these variables only swamped out some interactions, not others; see McKoon et al.). Nicol et al.'s suggestion is unlikely to be correct in light of the current results, obtained with almost all general experimental variables held constant with the McKoon et al. experiments.

Nicol et al. (1994) also suggested that McKoon et al.'s (1994) experiments were problematic because the sentences used in the experiments were difficult to understand (although it was not specified exactly how difficulty of understanding might override the modular syntactic processing assumed by their view of sentence comprehension). However, their comment does not apply here because the sentences used were essentially the same as those used in the experiments reported by Nicol and Swinney (1989).

The results reported here show the same interaction pattern as would be predicted by Nicol and Swinney (1989) and Nicol et al. (1994) for their versions of the sentences, the versions in which a potential object for a verb was presented in advance of the verb. Therefore, it would be desirable to compare directly the interaction found here with the one reported by Nicol and Swinney. This is not possible, however, because baseline response times, error rates, and statistical analyses were not included in Nicol and Swinney's overview of the relevant data.

It is noteworthy that both our experiment and the experiment reported by Nicol and Swinney (1989) obtained interactions, and that the interactions were of similar form, even though we used a procedure in which the words of the sentences and the test words were presented visually, whereas in the Nicol and Swinney experiment, a procedure was used in which the words of the sentences were presented auditorally and the test words visually. It might be thought possible that subjects would adopt different strategies for the two procedures, but the similarity of results makes this seem unlikely. The two procedures also gave the same patterns of results in our previous experiments (McKoon et al., 1994) and we know of no other, contradictory, evidence that the two procedures

give different results in experiments like these. Thus, the postulation of two different strategies appears to be theoretically uneconomical in the light of current evidence.

## Discussion

If a subject is reading a sentence about a baby throwing something, it is intuitively not surprising that response times to the probe *fruit* are faster than response times to the probe *bench*. The experiment in this article demonstrates that this intuition is correct. The factors that contribute to the intuitive feeling that a word "fits" a test position may include semantic interactions, pragmatic interactions, syntactic interactions (cf. Levin, 1993), associations in the language (see McKoon & Ratcliff, 1992b), associations in the real world, meaningfulness values, concreteness values, and the context formed by all the sentences in the experiments. Which of these factors are important to comprehension processes is currently a matter for speculation. However, despite the speculative nature of explanations of our finding, its consequences—both negative and positive—are critical. The negative consequence is to draw attention to methodological problems in on-line text experiments. The positive consequence is to draw attention to the power of the language processing system. In the sections that follow, we elaborate on these two points.

### *The Down Side*

The results of our experiment raise the possibility that some previous results are artifactual. In any on-line experiment in which response times (or accuracy values) are compared for two different test words, there must be some control to show that the differences are due to the variable of interest, not to some general fit to the context of one test word over the other. In particular, both the current finding and those in McKoon et al. (1994) raise serious doubts about interpretations of previous results thought to show processing of implicit syntactic information (cf. Fodor, 1993; Hickok, 1993; Nicol & Pickering, 1993; Nicol & Swinney, 1989; Swinney & Osterhout, 1990).<sup>1</sup>

It must be stressed that the fit of a test item to the context in which it is tested is an issue whose importance is not limited to on-line lexical decision. In principle, there is no reason to think that the semantic, pragmatic, and syntactic fit of a word would be any less a factor for on-line recognition, on-line naming latencies, or on-line reading times obtained from keypresses or eye movement monitoring. Furthermore, many kinds of changes in sentence or discourse contexts could potentially affect the fit

<sup>1</sup> It is possible that the fit of a test word to the sentential context in which it is tested is determined by quite subtle factors. Nicol and Pickering (1993), for example, used two versions of a sentence, differing only in the conjunction *that* versus *why*: "The receptionist informed the doctor that-why the journalist had phoned about the events."

The words *that* and *why* may differ in the effects they have on the entities in the discourse model constructed by comprehension processes. *That* might more likely increase the accessibility of *doctor* in the discourse model and therefore increase the fit of a related test word (e.g., *nurse*) to the sentential context. A subtle effect similar to this was found by McKoon et al. (1993) for verbs.

of test probes. In the experiment reported here, there was a large difference in response time as the context changed from immediately before a verb to immediately after a verb; that is, the context changed by one word that was immediately proximate to the test probe. What other kinds of changes could be important is a question for further research.

Our results imply that future empirical efforts to investigate the effects of one kind of psycholinguistic variable on comprehension independently of the effects of other kinds of variables must be undertaken with great care. For the experiment reported by Nicol and Swinney (1989; and perhaps the other experiments cited above), our results suggest that a syntactic variable was not adequately separated from semantic and pragmatic variables. Guaranteeing such separation in future experiments will doubtless require a range of different control words, control materials, and control experiments. We expect that no single control will rule out all possible confounds. Neither using different sets of experimental and control test words (e.g., Nicol & Swinney, 1989) nor using the same set of test words in both experimental and control conditions (e.g., McKoon et al., 1994) will necessarily ensure success, and what combinations of control conditions will be required to address different theoretical questions remains to be worked out.

### *The Up Side*

The most important aspect of our results is the power of the language comprehension system that they demonstrate. The amount of time that intervened between the two test positions was only about 250 ms. Yet the fit of a test word to the context was changed by 93 ms on a baseline of about 922 ms, a change of 10%. This large change came about even though there were no words in the sentence context that were individually highly associated to the test word. Discussion in Nicol et al. (1994) offered the temptation to think of this result as "spurious." We think just the opposite—the result is a major clue to how language comprehension works.

Often in discussions of language processing, two theoretical viewpoints conspire to prevent researchers from thinking of the semantic-pragmatic fit of a test word to a sentence context as a critical factor in predicting empirical results. The first is a concentration on syntactic processing. For example, in Nicol et al.'s (1994) discussion of the possibility of spurious effects, they concentrated only on syntactic factors and no others. The second theoretical preconception is that the processing involved in making a decision about a test word is understood as separate from the processing of the sentence context that preceded the test word. Specifically, when a test word is presented, facilitation on the decision about that word is assumed to result from some effect of prior sentence processing on the word's representation in long-term memory. This effect is often described as *activation* or *priming*, with these terms used to embody a theoretical notion that something has happened to a word in memory before its presentation as a test word.

Forster (1981) long ago suggested an alternative idea about how on-line responses to test words are decided. He proposed that an obligatory part of processing a test word (the same obligatory processing as for any word read in a discourse) is

checking the meaning of the word against the context to see if it fits. A good fit facilitates response times, a bad fit inhibits them. In accord with Forster's idea, Fischler and Bloom (1979) and West and Stanovich (1986; Stanovich & West, 1983) found facilitation in response times for words that were highly likely completions of the sentences after which they were tested. Here, we have extended their results to words that are not so highly likely and not (by themselves) possible completions of their sentences.

More recently, some researchers have proposed models by which each word in a discourse interacts with the previous context (Doshier & Rosedale, 1989; MacDonald, Pearlmutter, & Seidenberg, 1993; McKoon & Ratcliff, 1992a; McKoon, Greene, & Ratcliff, 1993; Ratcliff & McKoon, 1988, 1994). The interactive processing is assumed to proceed whether the next word is part of the text or a test word. McKoon et al. (1994) described compound cue theory, by which a test word is assumed to combine with whatever information is in short-term memory to form a compound cue with which to access long-term memory. For sentences and test words like those used in the experiment here, compound cue theory offers a simple description of the result: The combination of *the baby in the carriage threw* with *bench* is not as good as the combination of *the baby in the carriage threw* with *fruit*.

When interactive processing is considered, obvious questions of representation become salient. What is it in the representations of the meanings of *the baby in the carriage threw* and *fruit* that allows them to combine into a good fit with each other? We have no complete answer to this question, but we think the experiment described here attests to its central importance on the agenda of language processing research.

### References

- Doshier, 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.
- Fischler, I., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior*, 18, 1–20.
- Fodor, J. D. (1993). Processing empty categories: A question of visibility. In G. T. M. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 351–400). Hillsdale, NJ: Erlbaum.
- Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology*, 33A, 465–495.
- 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.
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago, IL: The University of Chicago Press.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1993). *The lexical nature of syntactic ambiguity resolution*. (Cognitive Science Tech. Rep. UIUC-BI-CS-93-13 [Language Series]) Urbana: University of Illinois, Beckman Institute.
- McKoon, G., Greene, S. B., & Ratcliff, R. (1993). Discourse models, pronoun resolution, and the implicit causality of verbs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1–13.

- 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.
- McKoon, G., Ratcliff, R., & Ward, G. (1994). Testing theories of language processing: An empirical investigation of the on-line lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1219–1228.
- Nicol, J. L., Fodor, J. D., & Swinney, D. (1994). Using cross-modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20*, 1229–1238.
- 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. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, *18*, 5–20.
- 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.
- Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General*, *112*, 1–36.
- 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. Inferences and text comprehension* (pp. 17–33). San Diego, CA: Academic Press.
- West, R. F., & Stanovich, K. E. (1986). Robust effects of syntactic structure on visual word processing. *Memory and Cognition*, *14*, 95–103.

Received March 25, 1994

Revision received April 22, 1994

Accepted April 28, 1994 ■

### P&C Board Appoints Editor for New Journal: *Journal of Experimental Psychology: Applied*

In 1995, APA will begin publishing a new journal, the *Journal of Experimental Psychology: Applied*. Raymond S. Nickerson, PhD, has been appointed as editor. Starting immediately, manuscripts should be submitted to

Raymond S. Nickerson, PhD  
Editor, *JEP: Applied*  
Department of Psychology  
Tufts University  
Medford, MA 02155

The *Journal of Experimental Psychology: Applied* will publish original empirical investigations in experimental psychology that bridge practically oriented problems and psychological theory. The journal also will publish research aimed at developing and testing of models of cognitive processing or behavior in applied situations, including laboratory and field settings. Review articles will be considered for publication if they contribute significantly to important topics within applied experimental psychology.

Areas of interest include applications of perception, attention, decision making, reasoning, information processing, learning, and performance. Settings may be industrial (such as human-computer interface design), academic (such as intelligent computer-aided instruction), or consumer oriented (such as applications of text comprehension theory to the development or evaluation of product instructions).