Immediate Biases in Parsing: Discourse Effects or Experimental Artifacts?

Don C. Mitchell and Martin M. B. Corley

Questions concerning relatively immediate determinants of syntactic analysis can be tackled by comparing the reading times within crucial regions of subject- and object-relative clauses. Using such a comparison, Altmann, Garnham, and Henstra (1994) have presented data which they interpreted as evidence for a discourse-driven account of parsing. This article reexamines this evidence and highlights a number of potential artifacts in the data. An experiment was conducted to demonstrate that at least some of these artifacts may have introduced distortions which were large enough to be of practical concern. It is concluded that the findings of the Altmann et al. experiment failed to provide unambiguous support for the discourse-driven model of parsing and that, overall, structure-driven models still offer a better account of the data.

There is a vigorous ongoing debate about the role of discourse factors in sentence processing. Although all current models of parsing accept that discourse plays an important role before the end of the sentence, they differ markedly in the way they construe the scheduling of these effects. Some accounts maintain that discourse factors guide initial parsing commitments and that they exert their influence before any other factors have a chance to come into play (e.g., Altmann, Garnham, & Dennis, 1992; Altmann & Steedman, 1988; Crain & Steedman, 1985). Others claim that discourse information is only brought to bear after an initial phase of analysis guided by other (nondiscourse) conflict-resolution principles (e.g., Clifton & Ferreira, 1989; Ferreira & Clifton, 1986; Mitchell, Corley, & Garnham, 1992; Rayner, Garrod, & Perfetti, 1993).[1] These two classes of theory are referred to as *immediate discourse* models and *delayed discourse* models respectively. This article focuses on the question of whether there is a point at which early nondiscourse commitments can be observed or whether the initial structural analysis of ambiguous material is governed directly by discourse alone.

Mitchell et al. (1992) argued that many, if not all, putative demonstrations of immediate discourse effects can be questioned on the grounds that the effects of context have been examined at points that are delayed two or more words from the onset of the ambiguity. Given a suitably rapid thematic-repair facility, a delay of even two or three words would provide ample opportunity to implement a discourse-based revision of the initial (discourse-independent) analysis. It follows that delayed testing procedures cannot be used to distinguish between delayed and immediate discourse parsing models. Given this observation, Mitchell et al. argued that a truly persuasive demonstration of immediate discourse effects should introduce the bias test at the earliest feasible point; that is, immediately after the word introducing the ambiguity. If the bias is tested further into the ambiguous region, the results will inevitably be subject to the criticism that they reflect the influence of delayed rather than immediate discourse effects.

Crain and Steedman (1985) argued for the immediate effect of discourse on the basis of materials such as:

(1) The politician told the woman that . . .

They observed that the word "that" introduced an ambiguity between a relative clause and a sentential complement, and presented evidence of the effect on the resolution of the ambiguity by prior discourse (an experiment later refined by Altmann, 1988). However, in both studies, the tests of subjects' syntactic commitments were introduced several words into the ambiguous region, and thus the results did not in fact differentiate between immediate and delayed discourse effects.

Mitchell et al. (1992) overcame this problem by adapting the materials used in the earlier experiment so that the resolution of the ambiguity was introduced immediately after the word "that." This was achieved by continuing the sentence with the words "had been" in the central condition. Here, the absence of an explicit noun phrase signaled conclusively that the that-clause must be a relative rather than a complement. It follows that if subjects are committed to a complement interpretation of "that" at this point, they should show an increased reading latency when this word is followed immediately by an auxiliary (compared with a control condition in which noun phrase (NP) is inserted before the auxiliary phrase "had . . ."). For the purposes of presentation, the faster processing of "had been" in "that he had been" than in "that had been" was referred to as the *object-relative advantage,* and when this occurred it was taken as evidence that the subjects had been garden-pathed after expecting a complement. To abbreviate fairly radically, Mitchell et al. demonstrated that

---

[1] This article will not consider *constraint-based* accounts which maintain that discourse and nondiscourse effects operate in a concurrent fashion (e.g., Spivey-Knowlton, Trueswell, & Tanenhaus, 1993). Given appropriate weightings, such accounts could approximate either of the models discussed here.

such an effect manifested itself irrespective of contextual support. It turned up when the test sentence appeared alone (Experiment 1), when it appeared in the context of a paragraph designed to support the complement interpretation (Experiment 2), and, most crucially, it was equally strong in the context of discourse material designed to favor the relative reading of the that-clause (Experiment 2). On the basis of these findings, Mitchell et al. argued that whereas there was predictable evidence of context effects later in the sentence, the subjects' commitments at the early testing point were entirely unaffected by the prior discourse. They concluded that these results favored models in which the effects of context were delayed rather than accounts which maintained that the initial analysis is based on the immediate effects of discourse information.

Altmann, Garnham, and Henstra (1994) have subsequently questioned these conclusions. They accepted that disputes on the causes of initial bias can only be resolved by experiments based on the earliest feasible tests of syntactic commitment. However, they raised a number of objections to the Mitchell et al. (1992) test materials and, as a result, questioned the evidence against discourse-driven models in general. Using revised materials and eye-tracking procedures, Altmann et al. presented a new set of results which they interpreted as evidence that discourse context does, after all, influence the initial interpretation of that-clauses. The article ends with the authors reaffirming their commitment to a particular variant of the immediate discourse model. In this article, we question some of their arguments, highlight several important artifacts in their study, and argue that there are no firm grounds for modifying our earlier conclusions.

Altmann et al.'s (1994) main objection to the argument against the immediate use of discourse information centered on the effectiveness of the contexts used in Experiment 2 of the Mitchell et al. (1992) study. They suggested that these contexts failed to manipulate the available antecedents for the test sentence and were in fact all complement supporting, despite the fact that they were carefully constructed to conform to the guidelines set out in the earlier work of Altmann (1988) and Crain and Steedman (1985). In line with this earlier work, Mitchell et al.'s relative-supporting contexts made two potential antecedents available for the test sentence. But if, as Altmann et al. (1994) suggested, certain artifacts conspired to increase the salience of one, making all of the context used by Mitchell et al. complement-supporting in fact if not in intent, then it follows that the failure to eliminate the object-relative advantage in two-antecedent contexts cannot legitimately be taken as evidence against immediate discourse effects.

To be clear about our position, we should state at the outset that we do not accept that our two-antecedent contexts were uniformly biased in favor of the complement reading of the that-clause (and we presented arguments in support of this view in the original article). However, it seems unproductive to dwell on the intricacies of the earlier materials since it is impossible to quantify their relative effectiveness. A more relevant point is that we acknowledge the force of Altmann et al.'s (1994) broader observation. Delayed discourse models stipulate that there should always be an initial point at which structural choices should hold sway, independently of even the

most effective discourse manipulations. It follows that if Altmann et al. (1994) have constructed materials which genuinely succeed in showing immediate context effects, the new data could present problems for a delayed discourse model, irrespective of the quality of previous materials. Therefore, the following paragraphs focus on Altmann et al.'s claim to have eliminated the garden-path effect in relative clauses presented in relative-supporting contexts.

To make this discussion completely clear it is important to re-examine the details of Mitchell et al.'s (1992) case against context effects. This depends crucially upon the evidence for garden-path effects in subject-relative sentences. In a subject-paced reading task Mitchell et al. showed that displays of the two words "had been" take 150–300 ms less time to read when they appear in object-relative sentences like Example 2a than in subject-relative sentences like Example 2b.

(2a) The headmaster told the boy that he/had been/watching to go and wait outside his room.

(2b) The headmaster told the boy that/had been/standing there to go and wait outside his room.

This was taken as evidence that subjects systematically interpret the that-clause as a complement—with the result that they are forced to revise their analysis when they encounter the "had been" display in Example 2b. The fact that this occurred in relative-supporting contexts led us to conclude that the initial parsing choice was determined by something other than discourse context. Notice that this case pivots entirely on accepting the object-relative advantage as a garden-path effect precipitated by the mistaken initial choice of the complement analysis.

Altmann et al. (1994) essentially accepted without question the garden-path interpretation of the object-relative advantage and set out to use the same logic to test whether discourse effects occur with their own newly devised set of context materials. Their crucial finding was that first pass reading times for the early portions of object- and subject-relatives were indistinguishable when the test sentences appeared in the context of felicitous paragraphs (23.6 vs. 23.3 ms per character in the critical subordinate verb region). In the null context the corresponding difference was highly reliable (29.2 vs. 35.2 ms). Altmann et al. argued that the object-relative advantage, though duly corroborated in the null-context condition, was completely eliminated in the felicitous-context condition, which, they claimed, was evidence that suitably strong supportive contexts could eliminate the preference shown for a complement in the null condition. Altmann et al. interpreted this finding to be general support for the referential or immediate discourse account of initial parsing choice.

At first sight the evidence provides fairly convincing support for the immediate discourse approach. However, on closer examination it becomes clear that the findings are by no means conclusive, and in fact we make a case in the following paragraphs that they may even be more directly compatible with a delayed-discourse model. A major problem with Altmann et al.'s (1994) conclusions is that despite acknowledging a number of methodological problems, they make no systematic attempt to rule out certain entirely plausible alternatives to the garden-path account which underpins the account they offer of

their data. Mitchell et al. (1992, pp. 83–84) had considered (and eliminated) several alternative accounts of their own findings (e.g., accounts based on superficial lexical processing effects, trace detection and coindexing effects, processing catch-up, and priming, or antecedent distance effects). However, the counter arguments deployed in this analysis cannot be applied to the Altmann et al. findings as the design in the two studies differed in several important respects. In fact, the object-relative advantage in the Altmann et al. study differs from the measure used by Mitchell et al. Of all the changes made by Altmann et al., two were particularly critical: abandoning the earlier use of physically identical material and changing the design in such a way that test sentences were presented in null and supportive but never in misleading contexts.

Altmann et al.'s (1994) changes included the use of materials such as those presented in Examples 3a and 3b.

(3a) He told the woman that he'd been waiting for that they were both very lucky.

(3b) He told the woman that had been waiting for him that they were both very lucky.

Notice that in Example 3a the auxiliary "had" is replaced by " 'd". If the reading time for "that he'd been" in Example 3a is different from that for "that had been" in Example 3b, this could be the result of numerous influences other than syntactic garden-path effects. For instance, "he'd" might take longer to read than "had" because it comprises two morphemes instead of one. Altmann et al. (1994) also point out that differences might occur because "he'd" is ambiguous ("he would" or "he had") or because work has to be done to unpack the contracted form. In view of these potential low-level differences it is not at all clear whether Altmann et al.'s materials can be used to identify the presence or absence of local garden-path effects. If the contraction has the effect of reducing processing time, then this will tend to amplify any apparent garden-path effect and possibly introduce a difference where no syntactic influence is actually present. On the other hand, if the contraction increases processing time, this will have the effect of reducing, canceling, or even reversing the behavioral effects of garden-pathing. In either case the object-relative advantage becomes uninterpretable.

The other critical difference between the experiments lies in the way that context was manipulated. Altmann et al. (1994) compared the effects of felicitous contexts not against those of neutral or infelicitous paragraphs but against null contexts which were known to encourage different processing strategies. Clearly there is no way of experimentally matching the linguistic content of null and real contexts. It follows that comparisons across these conditions are not in any way controlled for extraneous factors that could potentially influence processing (e.g., lexical priming of various kinds or syntactic priming as demonstrated by Bock & Loebell, 1990). For example, all of the contexts used by Altmann et al. included the word pairs "had been" but not "he'd been." Presumably, this could have facilitated the processing of object-relative sentences like Example 3a more than subject-relative sentences like Example 3b. Since corresponding priming effects could not play a role in the null context condition, differential effects of this kind were left uncontrolled in any

comparisons between these conditions. The Altmann et al. experiment corroborated previous work that had consistently shown that processing was slowed down in null contexts (see Altmann & Steedman, 1988, p. 288; Mitchell et al., 1992, pp. 81–82). This strategic difference might in part reflect the fact that the benefits of priming effects of the kind just outlined do not have facilitatory effects in the null context condition. In other words, the failure to pair contexts with their mismatched test sentences (in the manner initially recommended by Altmann, 1988) means that it is impossible to tell what aspect of the context may be responsible for any observed changes in processing strategy. In the Altmann et al. study the context effects were just as likely to have been produced by priming phenomena as by discourse-induced garden path effects. Together with the contraction effects just considered, phenomena of this kind could well have obscured the garden-path effects which were crucial to the interpretation of the data.

Experiment

Rather than leaving these objections as hypothetical problems, we offer a simple empirical demonstration that at least one of the effects (the contraction effect) was large enough to confound the interpretation of any object-relative advantage obtained with such materials. In order to set aside syntactic garden-pathing effects, we used sentences that were not subject to the complement/relative ambiguity as in Examples 4a, 4b, and 4c.

(4a) He'd cut the pie in two and shared it with his friends.

(4b) He had cut the pie in two and shared it with his friends.

(4c) He would cut the pie in two and share it with his friends.

If the contracted form "he'd" causes processing difficulties, then the display "He'd cut" should take longer to read than both "He had cut" and "He would cut" (when corrected for display length). Evidence of this kind would indicate that the artifact has to be taken seriously.

Method

Subjects. There were 20 subjects, all of whom were students at the University of Exeter. All spoke English as a first or only language. Subjects were paid £2 ($3) to participate in the study. None had taken part in other psycholinguistic experiments.

Apparatus. A BBC Microcomputer was used to display the materials and record response times and answers to questions; the EXMORE experimentation package (Mitchell & Barchan, 1984) and a standard CRT display were used.

Materials and design. The materials consisted of 25 sets of five sentences generated from the form "Pro Aux Part ... and V ..." (Pro = pronoun, Aux = auxiliary, Part = participle, V = tensed verb), where Aux was either the contracted form " 'd" or one of the full forms "had" or "would," and V was either in the present or past tense. The participles were chosen so that the infinitive and past forms were identical ("he would come"/"he had come"). The verbs used were: come, run, bet, burst, bid, read, wed, let, cut, hit, put, split, set, case, rid, become, hurt, wet, shed, slit, spread, reset, cost, thrust, and shut. Four grammatical sentences were obtained from each participle form; a further test sentence was generated using a different segmentation (see Table 1 for an example material set).

Table 1

Viewing Times (in Milliseconds per Character) for Test Displays in Each Experimental Sentence

Experimental sentence	Viewing times
(a) **He'd cut**/the pie in two/and share/it with his friends.	133.2
(b) **He'd cut**/the pie in two/and shared/it with his friends.	134.5
(c) **He would cut**/the pie in two/and share/it with his friends.	95.2
(d) **He had cut**/the pie in two/and shared/it with his friends.	117.4
(e) He/**had cut**/the pie in two/and shared/it with his friends.	100.9

Note. Test displays are set in boldface.

The materials were sorted into five files with each set represented once in each file, and each sentence type appearing five times in each file. Thus, in each file, there were 5 sentences with the word "would" explicitly, 10 with the word "had" (with two different segmentations), and 5 each with " 'd" resolving as "would" and "had" respectively. A set of 35 "filler" sentences was added to each file of 25 sentences. In order to counteract any possible biases in the experimental materials, 5 of these filler sentences contained an explicit "would" (equalizing the proportion of "had" and "would" trials) and 10 consisted of sentences containing various other uses of apostrophes. These, and the remaining 20 sentences, were of mixed construction and complexity. Of the 60 sentences in each file, 12 (20%) were followed by a yes–no question which concentrated on some aspect of the content of the sentence. These questions were evenly distributed across material types and answers.

Procedure. Each subject underwent a practice session that consisted of 5 sentences specially created for that purpose before the experiment commenced. This session was identical to the experimental session in all respects other than randomization and content. The instructions appeared on the computer screen before each session and informed the subjects that they would have to read a number of sentences split into chunks, followed in some cases by yes–no questions. They were advised to locate the Y and N keys and to rest their fingers on the space-bar, which was used to summon up successive displays. No indication was given of the purpose of the experiment and no subject guessed correctly before debriefing. The experimenter sat with the subjects during the practice session and answered any procedural queries. For the experimental session, one of the five experimental files was randomly picked by the computer, the only constraint being that no more than two of the 25 experimental sentences would ever appear in sequence.

Each trial was preceded by the words "Press space-bar." This press triggered the first display of the trial, with subsequent presses triggering each subsequent display. If the material was to be followed by a question, the word "Question" was displayed for 2 s, allowing the subjects time to find the Y and N keys to make a response. Records were kept of the key pressed and latency of the response (in milliseconds) for each display. Subjects completed the experimental session alone and were informed by the computer when the session had finished.

Results and Implications

The mean reading times per character for the five different aux + participle conditions are shown in Table 1.[2] The

apostrophe was treated as a character for the purpose of calculating per character latencies. These data were entered into repeated measures analyses of variances (ANOVAs) with form of aux + participle display as a fixed effect and subjects (or materials) as random effects. Overall, the results showed that reading time per character was markedly affected by the form of display, $F_1(4, 76) = 8.72$, $MS_e = 745.1$, $p < 0.001$; $F_2(4, 96) = 12.33$, $MS_e = 659.2$, $p < 0.001$.

Planned comparisons were used to examine the form effect in more detail. A comparison between the reading latencies for "He'd + participle" displays in the first two form conditions revealed no statistical difference between them, $F_1(1, 19) = 0.02$; $F_2(1, 24) = 0.03$; a result which was to be expected given that the two sets of materials were identical at this point. A comparison of the latencies for these two conditions against the two alternative expanded forms ("He had + participle" and "He would + participle") confirmed the prediction that the contracted forms would take longer to process than the full forms, $F_1(1, 19) = 30.56$, $MS_e = 496.5$, $p < 0.001$; $F_2(1, 24) = 26.86$, $MS_e = 706.4$, $p < 0.001$.[3] Presumably the reason for this is either that the contracted forms have to be unpacked in the course of interpreting the material or that they are ambiguous with respect to the "had/would" expansion. Either type of effect would have to be controlled for in any informative comparison between reading times for contracted and uncontracted displays.

In the Altmann et al. (1994) study there was no explicit pronoun in the uncontracted condition, and it is conceivable that the inclusion of the word "he" (or "she") could have reduced the reading time (per character) over the region considered. However, a comparison between conditions (d) and (e) indicated that the latency was, if anything, increased rather than reduced when the pronoun was included in the display, $F_1(1, 19) = 4.04$, $MS_e = 534.4$, ns; $F_2(1, 24) = 6.49$, $MS_e = 671.8$, $p < 0.05$. More importantly, the contraction effect was still clearly evident when the "he'd + participle" display in conditions (a) and (b) was compared to the "had + participle" display excluding the pronoun in condition (e), $F_1(1, 19) = 14.69$, $MS_e = 983.2$, $p < 0.005$; $F_2(1, 24) = 29.97$, $MS_e = 602.6$, $p < 0.001$.

---

[2] Various other measures could have been employed in the data analysis (e.g., reading latencies uncorrected for display size or data adjusted by means of linear regression techniques). However, the central purpose of the study was to evaluate the method and conclusions drawn by Altmann et al. (1994), and so for the sake of comparability we have opted to follow their approach and use reading time per character as the main measure in our analyses.

[3] Any examination of the contraction effect involves making comparisons between displays of different sizes. Since latencies increase with display size (e.g., Mitchell & Green, 1978), little can be learned by comparing raw reading latencies. For the record, size variations were sufficient to ensure that the mean (full) latency for the contracted forms (a) and (b) (1,217 ms) was not significantly different from that for the expanded forms (1,273 ms), $F_1(1, 19) = 1.92$, $MS_e = 33,325$; $F_2(1, 24) = 0.93$, $MS_e = 86,538$. As already indicated, the size-corrected measure almost certainly provides a more relevant basis for reassessing the Altmann et al. (1994) findings.

Implications for the Interpretation of the Altmann et al. (1994) Results

The immediate implication of the potential contraction artifact is that there is no way of ruling out the possibility that there is a residual tendency to interpret that-clauses as complements even when they are presented in the contexts devised by Altmann et al. (1994). Altmann et al. reported that the contracted and uncontracted first-pass reading times were essentially identical in such contexts. However, the results of the current experiment showed that such comparisons have to be treated with caution. We have presented evidence that the contracted forms take reliably longer to process than uncontracted forms (both with and without pronouns in the display). Assuming that such effects play a role in eye-tracking data as well as in self-paced reading studies, it seems reasonable to conclude that if syntactic garden-path effects had been completely eliminated in the felicitous condition of the Altmann et al. study, this experiment would then have ended up revealing a reliable contraction effect in the first-pass data. The fact that there was no such effect (0.3 ms per character in the wrong direction) strongly suggests that the difference must have been canceled by equivalent differences in the opposite direction. If these opposing effects are syntactic garden-path effects, then the overall pattern of the data implies (contrary to Altmann et al.'s claims) that the new contexts failed to eliminate the prevailing tendency to interpret that-clauses as complements. To this extent the data argue against immediate use of discourse. As we stressed earlier (Mitchell et al., 1992, p. 83), immediate discourse models maintain that the sole determinant of initial syntactic choice is the prior referential context. If it is really true that people show a complement bias even in the context of relative-supporting paragraphs, then an immediate discourse model has severe problems in explaining where this bias comes from.

It is tempting to conclude that, when properly interpreted, these data go further than rejecting immediate discourse models and actually provide further support for delayed discourse models. For example, a syntax first model such as that of Frazier (1979, 1987) would predict that, once normalized for the contraction effect, the underlying effects should occur even in relative-supporting contexts. However, the chain of inference is too weak to warrant drawing such a firm conclusion. We have no way of estimating how large the contraction effect would be in eye-tracking data (as opposed to the self-paced data presented here). Nor can we make any firm statements about the magnitude or direction of any artifactual priming effects. Because of this we cannot be confident that the effects would exactly compensate for competing syntactic effects in the manner outlined previously. The point of these observations, therefore, is merely to establish that the issue remains open, and that no one will be in a position to evaluate the effects of the Altmann et al. contexts until a number of methodological problems have been ironed out.

In light of these arguments, Altmann et al. (1994) might concede that there may be obscure garden-path effects in their felicitous context condition. They may then point to the significant Context × Sentence Type interaction, arguing that

this, rather than the alleged absence of the garden-path effect, provides the crucial evidence against delayed discourse models. However, there are numerous problems with this fall back position. First, there is some question about the reliability of this effect; the interaction fails to reach significance when the region of analysis is changed marginally to exclude the word "that" (Altmann et al., 1994, p 213). Second, the interaction may be partly attributable to differential priming effects in the (nonequivalent) felicitous contexts. Effects of this kind cannot occur in the null context condition and would therefore tend to show up in the form of an interaction. Third, the interaction (such as it is) may simply be a statistical artifact. In the null condition reading time was slower overall and so the standard deviation (and hence most normal dependent effects) would be expected to be magnified in this condition even if they were functionally equivalent to corresponding effects in the context condition. In certain circumstances range or variance artifacts of this kind can be handled by carrying out the analyses on log-transformed data. It is not clear that the interaction would survive such a transformation. Also, most importantly, in the absence of "crossed" contexts, an interaction, even if genuine, only provides circumstantial support for immediate discourse models such as the referential support model. This is because even if one were able to dismiss all of the potential problems with Altmann et al.'s contexts, it would still not be clear which aspect of the felicitous contexts was causing the hypothetical garden-path effect to disappear. At the very least it is essential to demonstrate that with appropriate (i.e., in this case infelicitous) contexts the same manipulation can induce a garden-path where none was previously observed. In short, like the evidence with the object-relative advantage itself, Altmann et al.'s interaction fails to provide unequivocal support for the immediate discourse model.

It should be stressed that all of the criticisms raised here relate to the changes introduced in the Altmann et al. (1994) study, and that for this reason they do not apply to the earlier work. Mitchell et al. (1992) did not use contracted forms; they did not make any direct comparisons between null and paragraph contexts, and when they did compare contexts they attempted to minimize priming differences by restricting the changes to minimal noun phrase replacements (incidentally following the elegant procedures pioneered by Altmann, 1988; Altmann & Steedman, 1988; and Crain & Steedman, 1985). The evidence from the simple experiment reported here suggests that precautions of this kind are vital if one wants to draw conclusions about discourse context effects.

In order to demonstrate that their new contexts exert early guiding effects in parsing, it is vital for Altmann and his colleagues to conduct a study in which effects other than garden-path effects can be ruled out convincingly. Such a study would ideally incorporate control conditions to evaluate trace coindexing accounts and crossed-contexts (Altmann's own proposal) to minimize artifacts of lexical and syntactic priming. It would also be advisable to avoid the complications introduced by making comparisons between contracted and uncontracted forms. An experiment of this kind, with strong contexts could go a long way toward resolving the question of how initial syntactic choices are made.

Summary and Conclusion

Altmann et al. (1994) criticized the Mitchell et al. (1992) study on the grounds that the relative-supporting contexts may have been too weak to exert any measurable influence on the parsing process. They made various changes designed to strengthen the contextual materials and conducted an experiment to determine whether this discourse information plays an early role in selecting syntactic analyses in a subsequent test sentence. In practice, they provided no convincing evidence that their contexts are indeed stronger than the paragraphs used in the earlier study. Patterns of data which they attributed to referential effects could equally well be explained in terms of various artifacts that have nothing to do with discourse. Because of these artifacts, and particularly the contraction effect demonstrated here, it is not possible to determine whether these contexts (improved or otherwise) are capable of eliminating or even reducing garden-path effects which arise in certain sentences when people initially opt for the complement reading of a that-clause. The study by Altmann et al. therefore fails to resolve the issue that it sets out to test. The Mitchell et al. study was not subject to the artifacts outlined here, and the subsequent experiment provided no compelling justification for qualifying our original claim; namely that there is no evidence that initial commitments are moderated by discourse context.

References

Altmann, G. T. M. (1988). Ambiguity, parsing strategies, and computational models. *Language and Cognitive Processes, 3,* 73–97.

Altmann, G. T. M., Garnham, A., & Dennis, Y. I. L. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language, 31,* 685–712.

Altmann, G. T. M., Garnham, A., & Henstra, J. A. (1994). Effects of syntax in human sentence parsing: Evidence against a structure-based proposal mechanism. *Journal of Experimental Psychology: Language, Memory, and Cognition, 20,* 209–216.

Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30,* 191–238.

Bock, J. K., & Loebell, H. (1990). Framing sentences. *Cognition, 35,* 1–39.

Clifton, C., Jr., & Ferreira, F. (1989). Ambiguity in context. *Language and Cognitive Processes, 4,* 77–103.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context in the psychological syntax processor. In D. R. Dowty, L. Kartunnen, & A. M. Zwicky (Eds.), *Natural language parsing* (pp. 320–358). Cambridge, England: Cambridge University Press.

Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25,* 348–368.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies.* Bloomington, IN: Indiana University Linguistics Club.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 601–681). Hillsdale, NJ: Erlbaum.

Mitchell, D. C., & Barchan, J. (1984). *EXMORE: Exeter module for on-line reading experiments.* Exeter, Devon, UK: University of Exeter.

Mitchell, D. C., Corley, M. M. B., & Garnham, A. (1992). Effects of context in human sentence parsing: Evidence against a discourse-based proposal mechanism. *Journal of Experimental Psychology: Language, Memory, and Cognition, 18,* 69–88.

Mitchell, D. C., & Green, D. W. (1978). The effect of context and content on immediate processes in reading. *Quarterly Journal of Experimental Psychology, 30,* 609–636.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior, 22,* 358–374.

Rayner, K., Garrod, S., & Perfetti, C. A. (1993). Discourse influences during parsing are delayed. *Cognition, 45,* 109–139.

Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology, 47,* 276–309.