JOURNAL OF MEMORY AND LANGUAGE **38,** 459–484 (1998)
ARTICLE NO. ML972562

Late Closure in Context

Gerry T. M. Altmann

University of York, United Kingdom

Kathy Y. van Nice and Alan Garnham

Laboratory of Experimental Psychology, University of Sussex, United Kingdom

and

Judith-Ann Henstra

University of York, United Kingdom

The debate surrounding the use of extra-sentential context to inform early parsing decisions has focused primarily on the class of syntactic ambiguity to which Minimal Attachment (Frazier, 1979) applies. The present paper extends the debate to Late Closure (Frazier, 1979). We argue that Crain and Steedman's (1985) Principle of Parsimony predicts a specific circumstance in which referential context should override the tendency for late closure or right association (Kimball, 1973). The first of four eye-movement reading time studies failed to confirm the prediction. However, when similar materials were embedded in a context which explicitly directed attention toward the appropriate predicate (i.e., the predicate associated with the high attachment), we found evidence of contextual override. We argue that the data fit well with the constraint-satisfaction view of sentence processing (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994), although we offer an augmentation of the account based on the *predictive* activation of forthcoming structure (cf. Elman, 1990).

The evidence both for and against the hypothesis that extra-sentential contextual information can influence the initial decisions of the syntactic processor has related primarily to the class of ambiguity to which Frazier's (1979) Minimal Attachment applies (e.g., Altmann, 1988; Altmann, Garnham, & Dennis, 1992; Altmann, Garnham, & Henstra, 1994; Altmann & Steedman, 1988; Britt, 1994; Britt, Perfetti, Garrod, & Rayner, 1992; Clifton & Ferreira, 1989; Ferreira & Clifton, 1986; Rayner, Garrod, & Perfetti, 1992; Spivey-Knowlton & Tanenhaus, 1994; Steedman & Altmann, 1989; Trueswell & Tanenhaus, 1991). The present paper explores the hypothesis that extra-sentential information can also override preferences predicted by Late Closure (Frazier, 1979). The particular example of a closure preference which we shall consider here is exemplified in

(1) She'll implement the plan she proposed tomorrow, they hope.

Frazier's (1979) statement of Late Closure strategy requires that readers attach incoming material to the phrase currently being processed—hence attachment of ''tomorrow'' to

The work reported here was supported by grants from the Joint Council Initiative on Cognitive Science (SPG8920151—Experiment 1) and from the Economic and Social Research Council (R000234973—Experiments 2 onward). The authors gratefully acknowledge many useful comments from Keith Rayner and the reviewers of earlier versions of this paper. The experimental items used in the experiments reported here can be obtained from http://www.york.ac.uk/~gtma1/closure.html.

Address reprint requests to Dr. Gerry T. M. Altmann, Department of Psychology, University of York, Heslington, York YO15DD, UK. E-mail: g.altmann@psych.york.ac.uk.

0749-596X/98 $25.00

''she proposed,'' with a consequent garden path due to the temporal mismatch between the adverb and the verb to which it initially attaches. Although the universality of the preference for late closure has been questioned in languages other than English (e.g., Cuetos & Mitchell, 1988), we do not know of any demonstrations using structures of the kind in (1) which demonstrate anything other than a late closure preference in the absence of any extra-sentential context. This paper explores the inviolability, or otherwise, of this preference: We shall first apply the principles of presupposition satisfaction developed by Crain and Steedman (1985) and Altmann and Steedman (1988) to the resolution of the closure ambiguity in (1) above. We shall argue that their principle of Parsimony predicts a situation in which extra-sentential context should override the preference seen in (1). This is tested in Experiment 1. However, within the constraint-satisfaction approach to sentence processing (cf. MacDonald, Pearlmutter, and Seidenberg, 1994; Trueswell & Tanenhaus, 1994), the satisfaction of presuppositions is just one of a number of constraints which, through interaction with a range of other constraints that must also be satisfied, can in principle influence the initial parsing process. In Experiments 2 and beyond we shall explore a further contextual constraint which may also influence the initial resolution of the closure ambiguity. According to MacDonald et al. (1994), a recency preference (i.e., low right attachment or late closure) arises because the argument structure associated with the higher attachment is less active, due to temporal decay, than the argument structure associated with the lower attachment (see MacDonald et al., 1994, for discussion). When the adverbial phrase is encountered it attaches to whichever argument structure is most active and hence most available. We shall explore a constraint which is predicted, in MacDonald et al.'s terms, to prevent the argument structure associated with the higher attachment from decaying. We shall conclude that although there does exist a structural basis for (at least certain) recency preferences, it is possible to override this basis in certain circumstances. In the General Discussion, we consider the theoretical devices which might best account for the manner in which different kinds of extra-sentential information interact with local factors to constrain the (initial) interpretation of subsequent structures.

Altmann and Steedman's (1988) principle of Referential Support suggests that if a definite noun phrase fails to identify a unique referent, subsequent material should be interpreted as a modifier—thereby providing the additional material necessary for successful reference (cf. Altmann et al., 1992). But what should happen if a definite noun phrase *does* successfully refer to a prior discourse referent? Why would the subsequent material *not* be treated as a modifier? In the case of a *simple* noun phrase, the interpretation of subsequent material as a modifier would violate the presuppositions associated with the use of the modifier. But the case of *complex* noun phrases is somewhat different. Consider (1) again. If the simple noun phrase ''the plan'' was referentially unsuccessful, but the complex noun phrase ''the plan she proposed'' *did* successfully pick out a unique discourse referent, Crain and Steedman's Principle of Parsimony could be construed as predicting that ''tomorrow'' will associate *high* with ''implement.''

If the that-less relative clause ''she proposed . . .'' is interpreted as restrictive, and ''tomorrow'' is incorporated within this relative clause, it must then be interpreted as providing given information (cf. Clark & Haviland, 1977) and as providing a relevant contrast. But in a context in which ''the plan she proposed'' is referentially successful, and in which there is no mention of *when* the proposing occurred, low attachment of the adverbial ''tomorrow'' to ''she proposed'' would be in violation of the presuppositions supported by the context. On the other hand, high attachment of the adverbial, to ''implement,'' could be interpreted as providing *new* information. According to the Principle of Parsimony, the

high attachment should therefore be preferred. Even if ''tomorrow'' were interpreted as providing new (and referentially redundant) information, so long as we assume that the processor abides by something like Grice's Maxims of Quantity—be informative, avoid redundant information—then we might suppose that the information conveyed by an apparently redundant modifier does in fact require additional presuppositions to be satisfied (and see Carreiras & Clifton, 1993, and Gilboy, Sopena, Clifton, & Frazier, 1995, who claim that parsing decisions *are* governed by such maxims). Consequently, and because these presuppositions are not supported by the context, the Principle of Parsimony would again predict the high attachment of ''tomorrow'' in (1) above. Interestingly, in the one case which is to date empirically known—the ''null context'' case (where there is no extrasentential context)—it is unclear what predictions this principle would make; much hinges on whether the incorporation of the adverbial phrase into the relative clause, in the null context, violates more presuppositions than would be violated if it were not incorporated into that clause.

EXPERIMENT 1

In Experiment 1, we explored the predictions that follow from the Principle of Parsimony using materials such as (2) and (3). In (2), the high attachment of ''yesterday'' to ''brush'' in the target sentence is anomalous, owing to the tense mismatch, whereas the low attachment is not. Conversely, in (3) the low attachment of ''tomorrow'' to ''washed'' in the target sentence is anomalous, whereas the high attachment to ''brush'' is not.

(2) *Low attachment supporting context*

Tom's got two young dogs and they like playing in the fields.

Tom washed one of the dogs yesterday but the other one last week.

Low attachment target

He'll brush the dog he washed yesterday to make its fur shine again.

(3) *High attachment supporting context*

Tom's got two young dogs and they like playing in the fields.

Tom washed one of the dogs but did not want to bother with the other dog.

High attachment target

He'll brush the dog he washed tomorrow to make its fur shine again.

High and low attachment sentences were either presented in the null context (i.e., no preceding text) or preceded by what were designed to be their respectively felicitous contexts: The low attachment supporting context in (2) introduces two dogs, both of which were washed. Consequently, the expression ''the dog he washed'' in the target sentence fails to identify a unique dog. According to the Principle of Referential Support, this will force the processor to interpret the incoming item (''yesterday'') as providing further restrictive information, and hence ''yesterday'' will be incorporated into the relative clause allowing a unique dog to be identified. Late Closure and Right Association make the same prediction about attachment, but for different, purely structural, reasons. The high attachment supporting context in (3) also introduces two dogs, but only one of them was washed. Consequently, the expression ''the dog he washed'' in the target sentence *is* referentially successful. According to our hypothesis, early closure of the modifier should ensue, leading to high attachment of ''tomorrow'' to the main verb (''brush'').

We recorded subjects' eye movements as they read each target sentence (cf. Altmann et al., 1992; Frazier & Rayner, 1982; Rayner, Carlson, & Frazier, 1983; Rayner, Sereno, Morris, Schmauder, & Clifton, 1989). Because our concern is to show early effects of context (given that all sides of the debate agree on the existence of late effects), we are mainly concerned with the first pass reading time and first pass regression data. However, we shall also report total reading times, for reasons which will become apparent below. We predict, according to our interpretation of the

Principle of Parsimony, that whereas we should observe the standard closure preference in the null context, we should eliminate the preference in the felicitous context conditions (that is, we expect an interaction between context and target).

Method

Subjects. Forty-two members of the University of Sussex were paid £4 an hour to participate in the experiment.

Apparatus. Subjects' eye movements were recorded using an infrared limbus eye-tracking system (Optoelectronic Developments type 54), with the horizontal signals being sampled every 5 ms. A complete record of viewing location, fixation duration, and fixation sequence was stored for later analysis. Although viewing was binocular, signals were recorded from one eye only (generally the left eye). The distance from the eye to the screen was approximately 60 cm, giving a resolution of one character.

Materials. Thirty-six experimental passages such as the ones in (2) and (3) were presented to subjects randomly intermixed with 56 filler passages. The only differences between the high and low attachment targets were located on the adverbial region. In half the items, the first verb was in the future tense and the second was in the past tense, whereas in the other half of the items, this tense structure was reversed. This ensured that future and past adverbials appeared equally often in both high and low attachment targets—there was thus no confound between tense and attachment. In 15 of the passages, the target adverbial was ''yesterday'' or ''tomorrow''; the remaining adverbials were selected from ''next/last week/year'' (14 passages), ''soon/recently'' (four passages), ''shortly/recently'' (two passages), and ''later/earlier'' (one passage). In 32 of the passages, each adverbial occurred in each attachment condition, such that across the conditions the adverbial region was exactly matched in number of characters. In three passages there was a one-character difference and in one passage a two-character difference. Overall, the mean number of characters in the adverbial region was: high attachment, 8.20; low attachment, 8.36. The fillers were constructed so that they resembled the experimental items to different degrees.

Design. There were four versions of each passage (2 targets $\times$ 2 contexts; the null context or the felicitous context—we did not cross context with target, so we never preceded, for example, the low attachment target with the high attachment supporting context). The design was a fully factorial repeated measures design incorporating a Latin Square. Four stimuli sets were constructed with each item represented in each set in just one of its versions. Each subject was thus exposed to all items and to all experimental conditions, but never saw more than one version of any individual item. The null context and felicitous context conditions were intermixed.

Procedure. The same procedure was adopted as in Altmann et al. (1992). Subjects' head movements were minimized using a head restraint and a bite bar. A brief practice session was included to ensure subjects' comfort and familiarity with the procedure. The eye-tracking system was calibrated prior to approximately 45% of the trials. Calibrations took place prior to all the experimental items and approximately a quarter of the fillers. This procedure lasted 30 s. In the felicitous context conditions, subjects pressed a button to view the next sentence in the context, and there was no observable delay between the button press and the presentation of the subsequent sentence. Subjects were unaware which trials would contain contexts and which would not; they were told that some trials would consist of single sentences followed by a single question. Each sentence was replaced by an aligning prompt prior to the presentation of the following sentence (or question, following the target sentence). All the experimental and filler items were followed by a simple yes/no question that was included to encourage subjects to pay attention to the target. An additional question was included in the felicitous context conditions which could only be an-

swered correctly if subjects had read the context (e.g., ''Did Tom wash two dogs?''). Subjects responded to the questions by pressing one of two response buttons in front of them. Subjects were free to come off the bite bar and pause at any stage in the experiment, although they were encouraged to do so only between trials.

Scoring regions. The target sentences were divided into regions for the purposes of scoring as follows:

. . . / the dog / he washed / yesterday /
np rel adv

to make its fur / shine again.
adv + 1 adv +2

Results

Results are reported in terms of first pass reading times per character in a region; the probability of a first pass regression—a regression during the first pass reading of a region to a point to the left of that region; and total reading times per character in a region. See Altmann et al. (1992) and Rayner et al. (1989) for further details of these measures. The per character adjustment was not used here to adjust for length differences across conditions (they were negligible), but simply for compatibility with other studies. The probability of fixating on any given region during the first pass was 0.91 and for fixating on the critical adverbial region during the first pass, 0.96. In this and all subsequent experiments, there were no systematic differences in the probability of fixation across any of the experimental conditions, and we consequently treated those occasions in which a region was not fixated as noise (that is, those trials were included in the analysis).

Separate 2-way analyses of variance (2 levels of target $\times$ 2 levels of context) were performed on the data from each region. Although we are primarily interested in the eye-movement measures from the adverbial region and beyond (see Table 1), we report as appropriate any effects observed in the earlier regions.

First pass reading times. In the two regions prior to the adverbial region, there was only a main effect of context; reading times to the two targets were slower in the null context than in the felicitous contexts—region 1 (np): $F_1(1,41) = 23.5$, $p = 0.0001$; $F_2(1,35) = 30.9$, $p = 0.0001$ and region 2 (rel): $F_1(1,41) = 16.4$, $p = 0.0002$; $F_2(1,35) = 22.6$, $p = 0.0001$. First pass reading times in the adverbial region are reported in Table 1. In this region a significant effect of context was found ($F_1(1,41) = 11.8$, $p < 0.002$; $F_2(1,35) = 16.43$, $p = 0.0003$). In addition, high attachment targets were read more slowly than low attachment targets ($F_1(1,41) = 10.5$, $p < 0.003$; $F_2(1,35) = 7.70$, $p < 0.01$). Planned comparisons revealed that although the high attachment target was read more slowly than the low attachment target out of context ($F_1(1,41) = 7.7$, $p < 0.01$; $F_2(1,35) = 5.2$, $p < 0.03$), there was no difference between the two targets when presented in their felicitous contexts ($F_1(1,41) = 1.6$, $p > 0.2$; $F_2(1,35) = 1.1$, $p > 0.3$). However, the interaction failed to even approach significance. No other significant effects, or interactions, were observed in any region, except for a marginally significant effect of context in the final region ($F_1(1,41) = 4.79$, $p < 0.04$; $F_2(1,35) = 4.02$, $p = 0.053$).

First pass regressions. See Table 1. In the two regions prior to the adverbial region, no main effects or interactions were observed. In the adverbial region, there were main effects of context ($F_1(1,41) = 12.5$, $p = 0.001$; $F_2(1,35) = 12.90$, $p = 0.001$) and target ($F_1(1,41) = 19.2$, $p = 0.0001$; $F_2(1,35) = 28.02$, $p = 0.0001$), with more first pass regressions out of this region from high attachment targets than from low attachment targets. Planned comparisons revealed that this effect held both in the null and felicitous contexts (null context: $F_1(1,41) = 19.6$, $p = 0.0001$; $F_2(1,35) = 23.1$, $p = 0.0001$; referential context: $F_1(1,41) = 4.1$, $p < 0.05$; $F_2(1,35) = 4.8$, $p < 0.04$). There was also a main effect

TABLE 1

Experiment 1: First Pass Reading Times per Character, Probability of Making a First Pass Regression out of Each Region, and Total Pass Reading Times per Character, for High and Low Attachment Targets in Each of Two Contexts: the Null Context or the Felicitous Context. Standard Errors (by Subjects) Are Shown in Parentheses

	Null context		Felicitous context	
Region	High	Low	High	Low
First pass reading times				
adv	36.2 (1.9)	31.5 (1.5)	30.2 (1.5)	28.1 (1.4)
adv +1	35.4 (2.3)	34.1 (1.6)	34.8 (2.2)	34.1 (1.9)
adv +2	32.5 (2.5)	33.8 (2.5)	30.6 (2.3)	29.8 (2.2)
Probability of regression				
adv	21.6 (0.03)	9.1 (0.02)	11.3 (0.03)	5.6 (0.02)
adv +1	20.6 (0.03)	11.1 (0.02)	18.7 (0.03)	12.3 (0.02)
adv +2	66.3 (0.04)	56.3 (0.04)	57.9 (0.04)	50.4 (0.04)
Total pass reading times				
adv	69.8 (4.9)	42.8 (2.5)	49.0 (3.5)	36.0 (2.4)
adv +1	68.9 (5.2)	48.7 (2.1)	57.3 (3.5)	46.4 (2.1)
adv +2	49.6 (4.5)	42.6 (2.9)	38.7 (2.8)	37.8 (3.1)

of target in the following region ($F_1(1,41) = 12.3$, $p < 0.002$; $F_2(1,35) = 17.01$, $p = 0.0002$). In the final region (adv +2), there was a main effect of target ($F_1(1,41) = 9.06$, $p < 0.005$; $F_2(1,35) = 10.05$, $p = 0.003$) and a main effect of context ($F_1(1,41) = 8.14$, $p < 0.007$; $F_2(1,35) = 4.41$, $p < 0.05$). No further effects, nor any interactions, were observed.

Total reading times. See Table 1. In the adverbial region a significant effect of context was found ($F_1(1,41) = 23.5$, $p = 0.0001$; $F_2(1,35) = 23.3$, $p = 0.0001$). In addition, high attachment targets were read more slowly than low attachment targets ($F_1(1,41) = 63.2$, $p = 0.0001$; $F_2(1,35) = 31.9$, $p = 0.0001$), and there was a significant interaction between context and target ($F_1(1,41) = 6.9$, $p < 0.02$; $F_2(1,35) = 7.7$, $p < 0.01$). This interaction was due to a greater difference between the two targets in the null context condition, although planned comparisons confirmed that both in the null context and the felicitous referential context the high attachment targets were read more slowly than the low attachment targets (null context: $F_1(1,41) = 51.1$, $p = 0.0001$; $F_2(1,35) = 58.7$, $p = 0.0001$; referential context: $F_1(1,41) = 11.8$, $p < 0.002$; $F_2(1,35) = 13.9$, $p < 0.001$). The same pattern persisted into the following region (although at this region, the interaction between context and target was nonsignificant). This pattern was also observed in regions 1 and 2 (np and rel), reflecting the increased times spent reinspecting these regions following regressive eye movements from the later regions in the sentence.

DISCUSSION

We had predicted that there should have been an interaction between context and type of target—the difference in the adverbial region found in the null context should have been completely eliminated in the felicitous context conditions. No hint of such an interaction was found. And although there was no different in first pass reading times at the adverbial region between the two targets when presented in their felicitous contexts, there *were* more first pass regressions from this region in the high attachment case, and the same pattern was found in the total pass reading times (which could have been expected to be sensitive to any late-occurring effects of context), with longer time spent reading (and re-

reading) the high attachment adverbial than the low attachment adverbial even in the referential context. There is thus little evidence for an early effect of context—the provision of high-attachment supporting contexts did not prevent the late closure of the relative clause. For the moment, it appears that the Principle of Parsimony (as we have interpreted it here and which motivated the construction of those high-attachment supporting contexts) is inconsistent with the data.

Within the constraint-satisfaction view of sentence processing, it is not inevitable that a single constraint will succeed in overturning the effects of others (some of which may possibly be structurally based), and we shall return to this point in the General Discussion when we shall consider the ''fate'' of the Principle of Parsimony. But if this view of sentence processing suggests that contextual override may not always obtain, can it suggest occasions when it *will* obtain?

According to MacDonald et al. (1994), the recency preference arises because the mental representations corresponding to each verb's argument structure are differentially active. Thus, in example (1) reprinted below, the argument structure corresponding to ''implement'' will become activated when that verb is encountered, as will the argument structure corresponding to ''propose'' when ''propose'' is encountered. However, because the activation of such structures decays with time, the earlier argument structure will be considerably less active than the later one when the adverb is encountered. The incorporation of the adverb into the structure corresponding to ''implement'' would require, at the very least, that this structure be sufficiently active.

(1) She'll implement the plan she proposed tomorrow, they hope.

If, for the sake of argument, we equate the activation of a representation of something with ''attention'' to that something, we can ask what kind of contextual cue might focus attention on (that is, differentially activate) one structure (associated with one particular verb) rather than another. One way in which attention can be explicitly directed in this way is by preceding the critical sentence by a question, as in (4) below:

(4) When will Fiona implement the plan she proposed?
She'll implement the plan she proposed next week, of course.

In this example, the question sets up an expectation during the processing of ''. . . the plan she proposed'' that an adverbial phrase *relating to the first predicate* will follow (we return to the mechanism by which such expectations might be manifest in the General Discussion). If the processor adheres rigidly to a strategy such as Late Closure, the low attachment of ''next week'' to ''proposed'' should be attempted first, leading to a tense mismatch between the two. The purpose of Experiment 2 is to establish which attachment is preferred in materials such as (4). Experiment 2 uses direct question–answer pairs such as (4) above, whereas Experiments 3 and 4 embed such pairings in dialogues in which the question is asked only indirectly (as in ''The committee wondered when Fiona would implement the plan she proposed'').

EXPERIMENT 2

In Experiment 1 we did not attempt to induce a garden path in the low attachment targets; the most we could have expected was for a difference in reading times or first pass regressions in the null context conditions to disappear in the felicitous context conditions—requiring, in effect, a null result in the context conditions. In Experiments 2 to 4, we did cross contexts with targets, allowing us to explore the extent to which we could induce a garden path in the low attachment targets.

Experiments 2A and 2B used the same target sentences; in Experiment 2B they were preceded by contexts such as that shown in (4), whereas in Experiment 2A they were presented in the null context.

Experiment 2A: Method

Subjects. Forty members of the University of Sussex were paid £4 an hour to participate in the experiment.

Apparatus. The same apparatus was used as in Experiment 1.

Materials. The 40 target sentences to be used in Experiment 2B were randomly intermixed with 60 filler passages. There were two versions of each target sentence; one as in (4) above and the other with the adverbial phrase changed to refer to the opposite time period (so ''next week'' became ''last week,'' ''tomorrow'' became ''yesterday,'' and so on). Nineteen of the target sentences used the single-word adverbials ''yesterday'' and ''tomorrow,'' and 21 used two-word adverbials such as ''next year'' and ''last year.'' In this and subsequent experiments, adverbs within each target pair were matched for length (± 1 character), and log frequency (Francis & Kucera, 1982). Overall, the numbers of characters in the adverbial regions were: high attachment, 8.75; low attachment, 9.05. These were the only differences between the high and low attachment targets. The limited range of adverbial phrases arose simply due to the practicalities of finding phrases which could be used with the passages we had devised and will not be discussed further. As with Experiment 1, in half the items, the first verb was in the future tense and the second was in the past tense, whereas in the other half of the items, this tense structure was reversed (ensuring also that each adverbial occurred in both low and high attachment targets). The only difference between the target sentences in Experiments 2A and 2B is that in 2B they started with a pronoun (cf. 5 and 6 below) whereas here that first pronoun was changed to a full noun phrase (from ''She'' to ''Fiona''). The fillers were constructed so that they resembled the experimental items to different degrees. Twenty-four of the fillers were items from an unrelated experiment. The remaining fillers each resembled both experimental item types, half with and half without final appositives.

Design. There were two versions of each target sentence. The design was a fully factorial repeated measures design incorporating a Latin Square. Two stimuli sets were constructed with each item represented in each set in just one of its versions. Each subject was thus exposed to all items and to both experimental conditions (high attachment target vs low attachment target), but never saw more than one version of any individual item.

Procedure. The procedure was similar to that in the null context conditions of Experiment 1. The eye-tracking system was calibrated prior to every second trial. Calibrations took place prior to 32.5% of the experimental items and 61.7% of the fillers.

Results

Separate 1-way analyses of variance (2 levels of target) were performed on the eye-movement data from the two regions of interest: the adverbial region and the following, final, region. The probability of fixating on any given region during the first pass was 0.92 and of fixating on the critical adverbial region during the first pass, 0.91. There were no systematic differences by condition and consequently we treated those occasions in which a region was not fixated as noise.

First pass reading times per character. In the region of the adverbial, the high attachment target was read more slowly than the low attachment target (35.4 vs 32.6 ms/char; $F_1(1,39) = 9.2$, $p < 0.005$; $F_2(1,39) = 5.8$, $p < 0.03$), as predicted by Late Closure/Right Association. In the following, final, region, there was no significant difference between the two targets—high attachment target, 29.9 ms/char; low attachment target, 31.3 ms/char (both $F < 1$).

First pass regressions. There were more first pass regressions out of the adverbial region of the high attachment targets than out of the corresponding regions in the low attachment targets (27.3 vs 23.0% ($F_1(1,39) = 4.4$, $p < 0.05$; $F_2(1,39) = 4.2$, $p < 0.05$). Again, this is predicted by any account that maintains that the more recent (low) attachment is pre-

ferred. There was no such difference in the final region (41.4 vs 37%; $F_1(1,39) = 2.8$, $p > 0.1$; $F_2(1,39) = 2.5$, $p > 0.1$).

Total pass reading times per character. Total reading times for the adverbial were longer in the high attachment targets than in the low attachment targets (55.7 vs 43.4 ms/char; $F_1(1,39) = 29.1$, $p < 0.0001$; $F_2(1,39) = 29.9$, $p < 0.0001$). In the following, final, region, there was no significant difference between the two targets—high attachment target, 38.2 ms/char; low attachment target, 36.0 ms/char; $F_1(1,39) = 1.8$, $p > 0.1$; $F_2(1,39) = 1.4$, $p > 0.2$).

Experiment 2A demonstrates a clear advantage for the low attachment adverbial in the absence of any explicit prior context. In Experiment 2B we explore eye movements during the reading of these same target sentences, but preceded by direct questions of the kind shown in (4) above.

Experiment 2B

Each context question in Experiment 2B was followed by one of the two possible target sentences. Subjects were asked to judge whether the target sentence provided an answer to the preceding question (in the following example, ''yes'' or ''no'' refer to the responses that are appropriate given the judgement task):

(5) When will Fiona implement the plan she proposed? (*high attachment supporting*)
She'll implement the plan she proposed next week, of course. (*high attachment, ''yes''*)
She'll implement the plan she proposed last week, of course. (*low attachment, ''no''*)

In these cases, high/low attachment and response are perfectly confounded; high attachment targets are accompanied always by ''yes'' judgments, and low attachment targets by ''no'' judgments. Any increase in reading times, or first pass regressions, to the adverbial in the late closure case may simply be an artefact of this confound. Consequently, we included the following conditions also:

(6) Which of the plans she proposed will Fiona implement? (*low attachment supporting*)
She'll implement the plan she proposed last week, of course. (*low attachment, ''yes''*)
She'll implement the plan she proposed next week, of course. (*high attachment, ''no''*)

In these cases, the low attachment target *does* answer the context question.

We again monitored eye movements and collected data on judgments, judgment times, and reading times. If the context questions do indeed cause the processor to make the contextually appropriate attachment of the adverbial in the first pass, we would expect an interaction between question type and target type, but no main effect of target type (unless one attachment is more complex than another independently of any ambiguity resolution issue). If low attachment always obtained, we would expect a main effect of target and no interaction. In addition, because of the complexity of the low attachment ''which'' question (it is structurally more complex and both it and its answer presuppose different sets of plans and consequently more discourse entities than are presupposed by the ''when'' question), we also anticipate a main effect of question type (i.e., when vs which), with reading times to the targets following the which questions being longer.

Experiment 2B: Method

Subjects. Forty members of the University of Sussex were paid £4 an hour to participate in the experiment.

Apparatus. The same apparatus was used as in Experiment 1.

Materials. Forty experimental passages such as the ones in (5) and (6) were presented to subjects randomly intermixed with 60 filler passages. The fillers were constructed so that they resembled the experimental items to dif-

ferent degrees. In particular, they were constructed to use different question words (''what,'' ''where,'' and ''how'') and to ensure that subjects could not make the judgment (whether the target answered the question) on the basis of simply skipping directly to the clause-final adverbial phrase.

All the items consisted of a sentence and an appositive phrase separated by a comma. The final phrase in the experimental items was always two words, such as ''I'm sure,'' ''of course,'' and ''for certain.'' Final phrases in fillers varied from one to five words in length. This appositive phrase was included so as to ensure that the critical adverbial phrase was not the final phrase of the sentence.

Design. There were four versions of each passage (2 targets × 2 context questions). The design was a fully factorial repeated measures design incorporating a Latin Square. Four stimuli sets were constructed with each item represented in each set in just one of its versions. Each subject was thus exposed to all items and to all experimental conditions, but never saw more than one version of any individual item.

Procedure. The procedure was similar to that of Experiment 1, taking into account that there were no null context conditions and that the subjects' task was changed to reflect the question–answer nature of the stimuli. Subjects were told that each question would be followed by a sentence which either did or did not supply an answer to that question. They were instructed to press the ''yes'' button to view the question and then again to view the following sentence. As soon as they had read and understood that sentence, they pressed the ''yes'' button again, at which point a prompt appeared asking them to judge whether the sentence had answered the question. Eye movements during the reading of the target sentence (the potential answer to the context question) were collected, as were judgment times (between the button press after reading the target and the button press that reflected the judgment), and judgments.

The eye-tracking system was calibrated prior to every second trial. Calibrations took place prior to 30% of the experimental items and 63% of the fillers.

Scoring regions. The target sentences were divided into regions for the purposes of scoring as follows:

She'll implement	the plan	she proposed	next week,	for sure.
1	2	3	adv	final

Results

Four subjects (fortuitously, one from each of the four stimulus sets) made correct judgements on less than 70% of trials, and these subjects were excluded from all the analyses reported below. Separate analyses with these subjects included were carried out but the patterns were the same as those reported below.

Judgments. The percentages of correct judgments made in each of the four conditions are given in Table 2. Two-way ANOVAs (2 targets × 2 questions) revealed significant main effects of question type ($F_1(1,35) = 21.4$, $p = 0.0001$; $F_2(1,39) = 27.4$, $p = 0.0001$) and target type ($F_1(1,35) = 12.2$, $p < 0.002$; $F_2(1,39) = 11.7$, $p = 0.002$) with an interaction between the two ($F_1(1,35) = 4.9$, $p < 0.04$; $F_2(1,39) = 6.0$, $p < 0.02$). There were thus fewer correct responses to ''which'' questions (78% correct) than to ''when'' questions (88%). The interaction is due to the fact that subjects made more correct ''no'' responses than correct ''yes'' responses to ''which'' questions (which we had expected given the complexity of the relationship between the ''which'' questions and their answers).

Judgment times. The average times for correct judgments in each of the four conditions are given in Table 2. Two-way ANOVAs (2 targets × 2 questions) revealed a significant main effect of question type ($F_1(1,35) = 6.1$, $p < 0.02$; $F_2(1,39) = 8.3$, $p < 0.007$)—it took longer to make the appropriate judgments

TABLE 2

Experiment 2B: Percentage of Correct Judgments, and Mean Reaction Times (ms) to Make Correct Judgments, for Each of the Two Question Types (When vs Which) and Each of the Two Target Types (High vs Low). Standard Errors (by Subjects) Are Shown in Parentheses

	When question		Which question	
	High	Low	High	Low
Correct response	yes	no	no	yes
Percentage correct	88.6 (1.7)	88.1 (2.1)	85.0 (2.4)	71.7 (3.1)
Judgment time	783 (57)	929 (74)	931 (81)	975 (83)

following ''which'' questions (953 ms) than following ''when'' questions (833 ms)—a main effect of target type ($F_1(1,35) = 6.6$, $p < 0.02$; $F_2(1,39) = 7.8$, $p < 0.009$), but no interaction between the two ($F_1(1,35) = 2.1$, $p > 0.16$; $F_2(1,39) = 2.0$, $p > 0.16$).

Eye-movement data. See Table 3. Separate 2-way analyses of variance (2 targets × 2 questions) were performed on the eye-movement data from each region. All analyses were performed on just those trials on which subjects responded correctly in the judgment task (overall, subjects responded correctly on 83.35% of trials). The probability of fixating on any given region was 0.92 and for fixating on the critical adverbial region, 0.97. There were no systematic differences by condition and consequently we treated those occasions in which a region was not fixated as noise.

First pass reading times per character. In region 1 there were no effects of question type or target type and no interaction between the two. In region 2, there was a main effect of question type only—the noun phrase was read faster after ''when'' question than after the ''which'' question ($F_1(1,35) = 17.3$, $p = 0.0002$; $F_2(1,39) = 8.7$, $p < 0.006$). This effect persisted, marginally, into the third region (the relative clause); $F_1(1,35) = 4.0$, $p < 0.06$; $F_2(1,39) = 8.9$, $p < 0.005$. The data for the adverbial region and the final region are given in Table 3. In the adverbial region, there was a significant interaction between question type and target ($F_1(1,35) = 7.7$, $p < 0.009$;

TABLE 3

Experiment 2B: First Pass per Character Reading Times, Probability of Making a First Pass Regression out of each Region, and Total Pass Per Character Reading Times, for Each of the Two Target Types (High vs Low) after Each of the Two Question Types (When vs Which). Standard Errors (by Subjects) Are Shown in Parentheses

	When question		Which question	
Region	High	Low	High	Low
First pass reading times				
adv	36.9 (1.9)	38.5 (2.1)	39.6 (2.6)	35.0 (2.1)
final	41.0 (3.3)	43.1 (4.4)	39.0 (3.9)	42.5 (3.4)
Probability of regression				
adv	0.20 (0.03)	0.28 (0.04)	0.35 (0.05)	0.21 (0.04)
final	0.32 (0.04)	0.52 (0.04)	0.43 (0.04)	0.37 (0.04)
Total pass reading times				
adv	49.9 (3.0)	66.4 (4.4)	69.3 (5.6)	50.8 (3.8)
final	52.7 (5.2)	62.1 (6.2)	56.7 (5.9)	54.6 (4.7)

$F_2(1,39) = 4.1$, $p < 0.05$), but no main effects of question type (both $F < 1$) or target ($F_1(1,35) = 1.3$, $p > 0.26$; $F_2 < 1$). Planned comparisons revealed that there were no differences in reading times to the two targets following ''when'' questions ($F_1(1,35) = 1.1$, $p > 0.3$; $F_2 < 1$), but that following ''which'' questions, high attachment targets were read more slowly than low attachment targets, although this just missed significance on the by-items analysis ($F_1(1,35) = 8.3$, $p < 0.007$; $F_2(1,39) = 3.7$, $p = 0.06$). In the following, final, region, there were no significant effects.

First pass regressions. In regions 2 and 3 there were main effects of question type only (with more first pass regressions after ''which'' questions—region 2, $F_1(1,35) = 8.6$, $p = 0.006$; $F_2(1,39) = 6.7$, $p < 0.02$ and region 3, $F_1(1,35) = 9.9$, $p < 0.004$; $F_2(1,39) = 6.2$, $p < 0.02$). The first pass regressions out of the adverbial and final regions are given in Table 3. In the adverbial region there was no main effect of target type ($F_1(1,35) = 1.9$, $p > 0.17$; $F_2 < 1$), an effect by items only of question type ($F_1(1,35) = 1.8$, $p > 0.19$; $F_2(1,39) = 4.8$, $p < 0.04$), but an interaction between the two ($F_1(1,35) = 21.1$, $p = 0.0001$; $F_2(1,39) = 16.8$, $p = 0.0002$). Planned comparisons revealed that there were more regressions from the low attachment target than from the high attachment target following the ''when'' questions ($F_1(1,35) = 4.8$, $p < 0.04$; $F_2(1,39) = 4.8$, $p < 0.04$), but more regressions from the high attachment target than from the low attachment target following the ''which'' questions ($F_1(1,35) = 18.5$, $p = 0.0001$; $F_2(1,39) = 13.1$, $p = 0.0009$). In the final region, there was no effect of question type (both $F < 1$), a main effect of target type ($F_1(1,35) = 5.2$, $p < 0.03$; $F_2(1,39) = 5.0$, $p < 0.04$), and a significant interaction between the two ($F_1(1,35) = 19.2$, $p = 0.0001$; $F_2(1,39) = 15.4$, $p = 0.0003$)—following the ''when'' question, there were more regressions from the low attachment target than from the high attachment target (52 vs 32%), ($F_1(1,35) = 22.8$, $p = 0.0001$; $F_2(1,39) = 19.0$, $p = 0.0001$), but no difference between the targets following the ''which'' question (43 vs 37%): ($F_1(1,35) = 2.0$, $p > 0.16$; $F_2(1,39) = 1.4$, $p > 0.24$).

Regression-contingent analyses were also performed on the data from the adverbial region, the region of most interest. Specifically, we inspected first pass reading times on those trials when the eyes left the region for the first time with a rightward saccade (i.e., on trials without a first pass regression)—see Altmann (1994), Altmann et al. (1992), and Rayner & Sereno (1994) for the rationale and discussion concerning such analyses. The patterns of first pass reading times were unchanged.

Total pass reading times per character. See Table 3. At region 4, there was no effect of target type (both $F < 1$), no effect of question type (both $F < 1$), but an interaction between question and target ($F_1(1,35) = 42.7$, $p = 0.0001$; $F_2(1,39) = 31.3$, $p = 0.0001$). Planned comparisons revealed that there were longer times for the low attachment target than for the high attachment target following the ''when'' question ($F_1(1,35) = 19.1$, $p = 0.0001$; $F_2(1,39) = 17.1$, $p = 0.0001$), but longer times for the high attachment target than for the low attachment target following the ''which'' questions ($F_1(1,35) = 23.7$, $p = 0.0001$; $F_2(1,39) = 14.3$, $p = 0.0005$). In the following, final, region, there was no effect of target type ($F_1(1,35) = 2.4$, $p > 0.13$; $F_2 < 1$), no effect of question type (both $F < 1$), but an interaction between the two that was significant on the by-items analysis only ($F_1(1,35) = 2.8$, $p > 0.10$; $F_2(1,39) = 4.7$, $p < 0.04$). Planned comparisons revealed that there were marginally longer times for the low attachment target than for the high attachment target following the ''when'' question ($F_1(1,35) = 3.7$, $p = 0.06$; $F_2(1,39) = 4.3$, $p < 0.05$), but no difference following the ''which'' question (both $F < 1$).

DISCUSSION

In Experiment 2A, reading times and the probability of making a regressive eye movement away from the adverbial were determined by whether the target required a high

or low attachment. The one pervasive pattern in all the data from Experiment 2B, on the other hand, is that there was an interaction between context and target exactly as predicted on the assumption that the question context would ensure that the appropriate attachment decision would be made. In effect, we have found evidence of contextual override of Late Closure.

Although we have shown significant interactions between context and target, we found only limited evidence of any processing difficulty on the low attachment adverbial when preceded by the high attachment supporting context (the ''when'' questions). We found in this case that there were no differences at the adverbial region between first pass reading times across the two versions of the target, although there was a significant difference in the first pass regression data (and these two patterns were quite distinct from those observed in Experiment 2A and in Experiment 1). In the late closure supporting context (the ''which'' questions) we did find consistent evidence across the first pass measures for differences between the targets. To what extent does this asymmetry in the first pass reading times pose a problem for the hypothesis that contextual override obtained?

In part, this asymmetry may be explained by the fact that the position in the target sentences at which it can be determined that the context question has *not* been answered changes as a function of which question was asked. In the ''when'' contexts, it is only *after* the comma in the low attachment targets that it can be determined that the question has not been answered (in principle, a second, high attaching, adverbial could follow the first, low attaching, one); in the ''which'' contexts, this can be determined *before* the comma. This difference is reflected in the pattern of first pass regressions. For example, in the ''when'' contexts, there were significantly more regressive eye movements away from the final region (that is, after the comma) of the low attachment targets (requiring a ''no'' response) than the same region of the high attachment targets (with probabilities of 0.52 and 0.32, respectively). In the ''which'' contexts, on the other hand, there was no difference between the probabilities of making a regressive eye movement away from this region (0.37 and 0.43, respectively). Not surprisingly, given the explicit judgment that subjects were required to make, it would appear that subjects were sensitive to when such a judgment could be made. It follows that our data in Experiment 2B are in part an artefact of the question-and-answer judgment task which was used in that study and that the first pass reading time data may themselves reflect aspects of that task which would not be expected to occur during normal reading for comprehension. The fact that we observe reading time differences on the adverbial only when subjects are able unequivocally to reject the ''no'' response (that is, in the ''which'' conditions) is compatible with the suggestion that our data are in part confounded by the task we have adopted. With this in mind, Experiment 3 abandoned the judgment task and instead used *indirect* questions embedded within a discourse context (cf. Hanna, Spivey-Knowlton, & Tanenhaus, 1996; Liversedge, Pickering, Branigan, & van Gompel, 1998).

EXPERIMENT 3

Thirty-two of the target sentences used in Experiment 2 were used in this experiment with one minor modification that is described below. Each target sentence was preceded by one of two contexts, corresponding to the indirect versions of the ''which'' and ''when'' questions used in the last experiment.

(7) *When context* (*high attachment supporting*)
Last week Fiona presented a new funding plan to her church committee. The other committee members wonder when Fiona will implement the plan she proposed.

Which context (*low attachment supporting*)
Last week Fiona presented yet another funding plan to her church committee. The other committee members wonder which of the plans that Fiona's proposed she'll implement.

These targets are identical to the ones used in Experiment 2 except that the final appositive (''they hope'') was not restricted to just two words and explicitly referred to whoever ''wondered when'' or ''wondered which''—in Experiment 2, the appositive was neutral (e.g., ''of course''):

(8) She'll implement the plan she proposed next week, they hope. (*high attachment target*)
She'll implement the plan she proposed last week, they hope. (*low attachment target*)

We again monitored eye movements as subjects read each target sentence. We anticipated an interaction between context and target, with the high attachment targets taking longer to read than the low attachment targets in the ''which'' contexts, but with the low attachment targets taking longer in the ''when'' contexts. If initial attachment decisions are based purely on structural considerations such as Late Closure or Right Association, we expect a main effect of target (high attachments taking longer than low attachments), but no interaction between target and context.

Method

Subjects. Sixty-four members of the University of Sussex were paid £4 an hour to participate in the experiment.

Apparatus. The same apparatus was used as in Experiments 1 and 2.

Materials. See Appendix. Thirty-two experimental passages such as the ones in (7) and (8) were presented to subjects randomly intermixed with 64 filler passages. Fourteen of the experimental passages used the single-word adverbials ''yesterday'' and ''tomorrow,'' and 18 used two-word adverbials such as ''next year'' and ''last year.'' As with the previous two experiments, in half the items, the first verb was in the future tense and the second was in the past tense, whereas in the other half of the items, this tense structure was reversed. Overall, the numbers of characters in the adverbial region were: high attachment, 9.00 and low attachment, 8.94. These were the only differences between the high and low attachment targets. The fillers were constructed so that they resembled the experimental items to different degrees. Thirty-two of the fillers were items from an unrelated experiment. The remaining 32 fillers were a cross between the two interleaved item types (for example, context passages which resembled the unrelated experiment structurally, but ended with when/which questions in the final lines).

As in Experiment 2, all the target items consisted of a sentence and a short appositive phrase separated by a comma.

Design. There were four versions of each passage (2 targets $\times$ 2 contexts). The design was a fully factorial repeated measures design incorporating a Latin Square. Four stimuli sets were constructed with each item represented in each set in just one of its versions. Each subject was thus exposed to all items and to all experimental conditions, but never saw more than one version of any individual item.

Procedure. A similar procedure was adopted as in Experiment 1, taking into account that there were no null context conditions. Eye movements were collected during the reading of the target sentence.

The eye-tracking system was calibrated prior to every second trial. Calibrations took place prior to 53.1% of the experimental items and 48.4% of the fillers.

Scoring regions. The target sentences were divided into regions for the purposes of scoring as follows (and all regions were exactly matched for number of characters across conditions):

She'll implement | the plan |
1 2

she proposed | next week, | they hope.
3 adv final

Results

Separate 2-way analyses of variance (2 targets × 2 contexts) were performed on the eye movement data from each region. The probability of fixating on any given region during the first pass was 0.93 and of fixating on the critical adverbial region during the first pass, 0.94. There were no systematic differences by condition and consequently we treated those occasions in which a region was not fixated as noise.

First pass reading times per character. In region 1 there were no effects of context or target type and no interaction between the two. In region 2, there was a main effect of context only—the noun phrase was read faster after the ''when'' context than after the ''which'' context ($F_1(1,63) = 11.1$, $p < 0.002$; $F_2(1,31) = 16.5$, $p = 0.0003$). This effect persisted into the third region (the relative clause); $F_1(1,63) = 5.5$, $p < 0.03$; $F_2(1,31) = 6.9$, $p < 0.02$. The data for the adverbial and final regions are given in Table 4. In the adverbial region, there was a significant effect of target type ($F_1(1,63) = 18.4$, $p = 0.0001$; $F_2(1,31) = 16.5$, $p = 0.0003$), an effect of context ($F_1(1,63) = 5.9$, $p < 0.02$; $F_2(1,31) = 6.8$, $p < 0.02$), and a marginally significant interaction between context and target ($F_1(1,63) = 3.7$, $p = 0.06$; $F_2(1,31) = 8.0$, $p = 0.008$). Planned comparisons revealed that there was a significant difference in reading times following ''which'' contexts, with high attachment targets being read more slowly than low attachment targets ($F_1(1,63) = 19.3$, $p = 0.0001$; $F_2(1,31) = 41.3$, $p = 0.0001$). Following ''when'' contexts, there was no difference between the two targets on the by-subjects analysis, although according to the by-items analysis high attachment targets were read more slowly than low attachment targets ($F_1(1,63) = 2.8$, $p = 0.1$; $F_2(1,31) = 5.9$, $p = 0.02$). In the following, final, region, there were no significant effects.

First pass regressions. In region 2 the only significant effect was an interaction between context and target, although it was only marginally significant on the by-subjects analysis: $F_1(1,63) = 3.6$, $p < 0.07$; $F_2(1,31) = 5.4$, $p < 0.03$. This interaction persisted into the third region, although now it only approached significance on the by-items analysis ($F_1(1,63) = 5.2$, $p < 0.03$; $F_2(1,31) = 3.3$, $p < 0.08$). In addition, there was also a main effect at this region of context ($F_1(1,63) = 4.7$, $p < 0.04$; $F_2(1,31) = 5.6$, $p < 0.03$), with more regressions out of the region following ''which'' contexts (11%) than following ''when'' contexts (8%). The first pass regressions out of the adverbial and final regions are given in Table 4. In the adverbial region there was a main effect of target type ($F_1(1,63) = 11.6$, $p < 0.002$; $F_2(1,31) = 8.3$, $p < 0.008$), no effect of context (both $F < 1$), but an interaction between the two ($F_1(1,63) = 9.3$, $p < 0.004$; $F_2(1,31) = 10.7$, $p < 0.003$). Planned comparisons revealed that there were more regressions from the high attachment targets than from the low attachment targets following the ''which'' contexts ($F_1(1,63) = 20.1$, $p = 0.0001$; $F_2(1,31) = 23.2$, $p = 0.0001$) but no difference in the ''when'' context (both $F < 1$). In the final region, a main effect of target type approached significance only in the by-subjects analysis ($F_1(1,63) = 3.4$, $p = 0.07$; $F_2(1,31) = 2.1$, $p > 0.1$). There was no effect of context (both $F < 1$), but a significant interaction between the two ($F_1(1,63) = 27.7$, $p < 0.0001$; $F_2(1,31) = 36.5$, $p < 0.0001$)—following the ''when'' context, there were more regressions from the low attachment target than from the high attachment target (61 vs 44%), ($F_1(1,63) = 21.3$, $p = 0.0001$; $F_2(1,31) = 28.0$, $p = 0.0001$), but more regressions from high attachment targets than from low attachment targets following the ''which'' context (57 vs 47%), ($F_1(1,63) = 8.0$, $p < 0.007$; $F_2(1,31) = 10.6$, $p < 0.003$).

Total pass reading times per character. See Table 4. At the adverbial region there was a

TABLE 4

Experiment 3: First Pass per Character Reading Times, Probability of Making a First Pass Regression out of Each Region, and Total Pass Per Character Reading Times, for Each of the Two Target Types (High vs Low) after Each of the Two Contexts (When vs Which). Standard Errors (by Subjects) Are Shown in Parentheses

	When Context		Which context	
Region	High	Low	High	Low
First pass reading times				
adv	32.2 (1.5)	29.8 (1.4)	31.7 (1.5)	25.2 (0.8)
final	34.8 (1.8)	37.4 (2.5)	35.2 (2.2)	35.2 (2.1)
Probability of regression				
adv	0.20 (0.02)	0.19 (0.02)	0.24 (0.03)	0.14 (0.02)
final	0.44 (0.03)	0.61 (0.03)	0.57 (0.03)	0.47 (0.04)
Total pass reading times				
adv	48.3 (2.6)	53.4 (2.7)	62.7 (3.5)	37.1 (1.6)
final	42.3 (2.3)	53.6 (3.9)	52.5 (3.6)	42.2 (2.7)

significant effect of target type ($F_1(1,63) = 32.2$, $p < 0.0001$; $F_2(1,31) = 26.3$, $p < 0.0001$), no effect of context (both $F < 1$), but a significant interaction between context and target ($F_1(1,63) = 82.3$, $p < 0.0001$; $F_2(1,31) = 132.7$, $p < 0.0001$). Planned comparisons revealed that there were longer times for the low attachment targets than for the high attachment targets following the ''when'' contexts ($F_1(1,63) = 4.5$, $p < 0.04$; $F_2(1,31) = 6.9$, $p < 0.02$) but longer times for the high attachment targets following the ''which'' contexts ($F_1(1,63) = 114.8$, $p = 0.0001$; $F_2(1,31) = 186.5$, $p = 0.0001$). In the following, final, region, there were no main effects of target or context (all $F < 1$), although there was a significant interaction between the two ($F_1(1,63) = 34.2$, $p < 0.0001$; $F_2(1,31) = 126.3$, $p < 0.0001$). Planned comparisons revealed that there were longer times for the low attachment targets than for the high attachment targets following the ''when'' contexts ($F_1(1,63) = 18.8$, $p = 0.0001$; $F_2(1,31) = 69.8$, $p = 0.0001$), but longer times for the high attachment targets following the ''which'' contexts ($F_1(1,63) = 15.5$, $p = 0.0002$; $F_2(1,31) = 56.8$, $p = 0.0001$).

DISCUSSION

The results from this study are somewhat equivocal; although the low attachment targets engendered faster reading times and fewer regressions following the ''which'' contexts, the high attachment targets did not engender faster reading times or fewer regressions following the ''when'' contexts (and although there was no difference in the first pass regression measure, low attachment targets engendered marginally faster reading times in these contexts). The first pass regression data thus suggest the same asymmetry as was observed in Experiment 2B—in the ''which'' contexts, the low attachment was easier than the high attachment, but in the ''when'' contexts, there was no difference (although there was a hint of the high attachment being easier than the low attachment in the final frame of the targets). Consequently, we cannot blame the question-and-answer judgment task as employed in Experiment 2B for this asymmetry. It reflects a more general fact about these context/target pairings. Of course, it is conceivable that even in conditions where subjects do not have to make an explicit judgment, as in the case of this last experiment, they may nonetheless make some implicit judgment that is susceptible to the same constraint with respect to when it can be determined that the target does not answer the question. However, on the basis of the results from Experiment 4 below, we do not believe this to be the case.

There is, in fact, one further reason for ex-

pecting an asymmetry between the ''when'' and ''which'' questions. Consider again how the MacDonald et al. (1994) account of recency effects applies to the materials used in Experiments 2 and 3 (reproduced below):

(8) She'll implement the plan she proposed next week, they hope. (*high attachment target*)
She'll implement the plan she proposed last week, they hope. (*low attachment target*)

When the adverbial is encountered following the ''when'' question, the argument structure associated with ''implement'' will be activated, and hence accessible, due to the contextual input (according to our earlier hypothesis). However, the argument structure associated with ''proposed'' will *also* be active; it was introduced, after all, on the immediately preceding word. It is unclear whether one should expect one structure to be more active than the other, but if the two structures were equally accessible, both the high and low attachments would be equally possible, and there would be no tense mismatch associated with the low attachment target in the ''when'' context. When the adverbial is encountered following the ''which'' question, the argument structure associated with ''proposed'' will be the more active because the ''which'' question does not prevent the decay of the argument structure associated with ''implement.'' Consequently, an attempt would therefore be made to incorporate the adverbial into the most activated structure and there should be evidence with the high attachment target of a processing difficulty (due to the tense mismatch being noticed). This is exactly the pattern we observed in Experiments 2B and 3.

As a partial test of this account, we conducted one further study in which we carried out a manipulation which we predicted would allow the argument structure associated with ''proposed'' to decay, leaving the argument structure associated with the higher verb ''implement'' (still active on the basis of its contextual support) to become now the more strongly active of the two. We did this by interposing a prepositional phrase between the second verb and the target adverbial:

(9) She'll implement the plan she proposed to the committee next week, they hope. (*high*)
She'll implement the plan she proposed to the committee last week, they hope. (*low*)

EXPERIMENT 4

The target sentences used in Experiment 4 were the same as the ones used in the preceding experiment except that a prepositional phrase was introduced between the second verb and the adverbial as shown in (9) above, and (where necessary) the appositive phrase was shortened so as to limit the number of characters to no more than 80 (the maximum that could be displayed on one line of the screen). Each target sentence was preceded by one of two contexts. One of these corresponded to the ''when'' context from Experiment 3:

(10) *When context* (*high attachment supporting*)
Last week Fiona presented a new funding plan to her church committee. The other committee members wonder when Fiona will implement the plan she proposed.
The other was a modified version of this context in which we replaced the sequence ''wonder when'' with ''are guessing that'':

(11) *That context*
Last week Fiona presented a new funding plan to her church committee. The other committee members are guessing that Fiona will implement the plan she proposed.

The purpose of this manipulation was to provide a context that was essentially identical to the when-context but which did not focus attention on the temporal aspects of the predicate ''implement.''

We again monitored eye movements as subjects read each target sentence. We anticipated that in the ''when'' contexts the adverbial would engender longer reading times in the low attachment targets than in the high attachment ones. We anticipated also that in the ''that'' contexts, the presence of the prepositional phrase between the second verb and the adverbial would lead to a weakening of the late closure preference that we observed in the null contexts of Experiments 1 and 2A—the argument structure associated with that second verb should have decayed somewhat by the time the adverbial is encountered and, consequently, the differential between the activation level of the second argument structure and the (decayed) first argument structure would not be as great as it was in those earlier experiments which did not include the preposed prepositional phrase.

Method

Subjects. Thirty-two members of the University of Sussex were paid £4 an hour to participate in the experiment.

Apparatus. The same apparatus was used as in Experiments 1 to 3.

Materials. See Appendix. Thirty-two experimental passages consisting of a context (cf. 10) and target (cf. 9) were presented to subjects randomly intermixed with 64 filler passages. The targets were the same as those used in Experiment 3 except for the interposed prepositional phrase discussed above and consisted of a sentence and a short appositive phrase separated by a comma. The fillers were simplified versions of those used in Experiment 3, constructed so that they resembled the experimental items to different degrees.

Design. There were four versions of each passage (2 targets × 2 contexts). The design was a fully factorial repeated measures design incorporating a Latin Square. Four stimuli sets were constructed with each item represented in each set in just one of its versions. Each subject was thus exposed to all items and to all experimental conditions, but never saw more than one version of any individual item.

Procedure. See Experiment 3. Calibrations took place prior to 53.1% of the experimental items and 48.4% of the fillers.

Scoring regions. The target sentences were divided into regions for the purposes of scoring as follows (and all regions were exactly matched for number of characters across conditions):

She'll implement	the plan	she proposed
1	2	3

to the committee	next week,	they hope.
4	adv	final

Results

Separate 2-way analyses of variance (2 targets × 2 contexts) were performed on the eye movement data from each region. The probability of fixating on any given region during the first pass was 0.93 and of fixating on the critical adverbial region during the first pass, 0.93 also. There were no systematic differences by condition and consequently we treated those occasions in which a region was not fixated as noise.

First pass reading times per character. In regions 1 through to 4 (the postverbal prepositional phrase) there were no effects of context or target type and no interaction between the two. The data for the adverbial and final regions are given in Table 5. In the adverbial region, there was no effect of target type (both $F < 1$). There was an effect of context, although this just missed significance on the by-items analysis ($F_1(1,31) = 5.2$, $p < 0.03$; $F_2(1,31) = 4.0$, $p = 0.055$) and an interaction between context and target ($F_1(1,31) = 5.9$, $p = 0.02$; $F_2(1,31) = 6.7$, $p < 0.02$). Planned comparisons revealed that there was a significant difference in reading times following ''when'' contexts, with high attachment targets being read faster than low attachment targets ($F_1(1,31) = 5.4$, $p < 0.03$; $F_2(1,31) = 6.1$, $p < 0.02$). Following ''that'' contexts, there was no difference between the two targets ($F_1(1,31) = 1.2$, $p > 0.2$; $F_2(1,31) = 1.4$,

TABLE 5

Experiment 4: First Pass per Character Reading Times, Probability of Making a First Pass Regression out of Each Region, and Total Pass Per Character Reading Times, for Each of the Two Target Types (High vs Low) after Each of the Two Contexts (When vs Which). Standard Errors (by Subjects) Are Shown in Parentheses

	When context		That context	
Region	High	Low	High	Low
First pass reading times				
adv	30.1 (1.2)	33.7 (1.5)	30.1 (1.6)	28.4 (1.3)
final	32.7 (1.4)	36.1 (1.6)	37.6 (2.3)	37.9 (2.3)
Probability of regression				
adv	0.14 (0.03)	0.23 (0.02)	0.17 (0.03)	0.12 (0.02)
final	0.45 (0.05)	0.54 (0.04)	0.43 (0.04)	0.45 (0.05)
Total pass reading times				
adv	41.8 (2.5)	50.4 (3.3)	41.2 (2.1)	38.5 (2.0)
final	38.1 (2.4)	47.0 (4.0)	44.0 (3.2)	44.6 (4.3)

$p > 0.2$). In the following, final, region, there was a marginal effect of context ($F_1(1,31) = 6.6$, $p < 0.02$; $F_2(1,31) = 3.1$, $p > 0.08$), but no effect of target and no interaction between the two.

First pass regressions. In region 3 there was only a main effect of context (with more first pass regressions after ''that'' contexts (5.5%) than after ''when'' contexts (2.7%) ($F_1(1,31) = 4.7$, $p < 0.04$; $F_2(1,31) = 5.0$, $p < 0.04$). There were no other main effects or interactions in regions 2 to 4 (the prepositional phrase). The first pass regressions out of the adverbial and final regions are given in Table 5. In the adverbial region there was no effect of target type (both $F < 1$), no effect of context ($F_1(1,31) = 2.3$, $p > 0.1$; $F_2(1,31) = 2.8$, $p > 0.1$), but an interaction between the two ($F_1(1,31) = 9.7$, $p < 0.004$; $F_2(1,31) = 9.7$, $p < 0.004$). Planned comparisons revealed that there were more regressions from the low attachment targets than from the high attachment targets following the ''when'' contexts ($F_1(1,31) = 8.2$, $p < 0.008$; $F_2(1,31) = 8.2$, $p < 0.008$) but no difference in the ''that'' context ($F_1(1,31) = 2.4$, $p > 0.1$; $F_2(1,31) = 2.4$, $p > 0.1$). In the final region, a main effect of target type approached significance ($F_1(1,31) = 3.7$, $p < 0.07$; $F_2(1,31) = 3.5$, $p < 0.08$). There was a significant effect of context ($F_1(1,31) = 6.1$, $p < 0.02$; $F_2(1,31) = 4.5$, $p < 0.05$), but no interaction between the two ($F_1(1,31) = 1.0$, $p > 0.3$; $F_2(1,31) = 1.0$, $p > 0.3$).

Total pass reading times per character. At frame 4 (the prepositional phrase) there was no effect of target ($F_1(1,31) = 3.1$, $p > 0.09$; $F_2(1,31) = 2.3$, $p > 0.1$), a marginal effect of context ($F_1(1,31) = 3.4$, $p > 0.07$; $F_2(1,31) = 4.2$, $p < 0.05$), and a significant interaction between the two ($F_1(1,31) = 12.2$, $p < 0.002$; $F_2(1,31) = 12.4$, $p < 0.002$). There were no main effects of either variable separately. The total pass reading times in the adverbial and final regions are given in Table 5. At the adverbial phrase, there was no effect of target ($F_1(1,31) = 2.9$, $p > 0.09$; $F_2(1,31) = 1.6$, $p > 0.2$), but there was an effect of context ($F_1(1,31) = 13.9$, $p < 0.0009$; $F_2(1,31) = 10.0$, $p < 0.004$) and an interaction between the two ($F_1(1,31) = 11.4$, $p = 0.002$; $F_2(1,31) = 9.8$, $p < 0.004$). Planned comparisons revealed that there were longer times for the low attachment targets than for the high attachment targets following the ''when'' contexts ($F_1(1,31) = 13.3$, $p = 0.001$; $F_2(1,31) = 11.4$, $p = 0.002$) but no difference between the two targets following the ''that'' contexts ($F_1(1,31) = 1.3$, $p > 0.2$; $F_2(1,31) = 1.1$, $p > 0.2$). In the following, final, region, there was a main effect of target ($F_1(1,31) = 5.3$, p

< 0.03; $F_2(1,31) = 8.5$, $p < 0.007$), no effect of context (both $F < 1$), but a marginally significant interaction between the two ($F_1(1,31) = 6.9$, $p < 0.02$; $F_2(1,31) = 3.4$, $p > 0.07$). Planned comparisons revealed that there were longer times for the low attachment targets than for the high attachment targets following the ''when'' contexts ($F_1(1,31) = 15.8$, $p = 0.0004$; $F_2(1,31) = 7.7$, $p < 0.01$), but no difference following the ''that'' contexts (both $F < 1$).

DISCUSSION

The data from Experiment 4 unequivocally demonstrate contextual override of the null context preference for low attachment. Following ''when'' contexts, the high attachment adverbial engendered faster reading times and fewer regressive eye movements than the low attachment adverbial. Interestingly, there was no statistical difference between these measures following the ''that'' contexts, suggesting that the interposition of the additional prepositional phrase between the second verb and the target adverbial did lessen the strength of the low attachment preference that might normally be expected in these conditions (although the means reported in Table 5 do go in the expected direction). We note, in this regard, MacDonald and Thornton's (1996) demonstration regarding the sentence pairs in (12) and (13):

(12) a. Mary likes swimming very slowly
b. John likes swimming very much

(13) a. Mary likes it when the dolphins at Sea World are swimming very slowly
b. Mary likes it when the dolphins at Sea World are swimming very much

According to MacDonald et al.'s (1994) account of recency preferences, the earlier argument structure (corresponding to ''likes'') does not decay sufficiently to make (12b) any harder to process than (12a). In (13), however, the corresponding argument structure does decay sufficiently to make (13b) harder to process than (13a). These predictions were supported in a self-paced reading task which showed that ''very much'' was read more slowly than ''very slowly'' in (13) but not in (12).

We do not claim on the basis of our data alone that we have unequivocally proven the predictions made by MacDonald et al.'s account of recency preferences (which include the late closure preference). That is not the purpose of the present paper. Instead, our purpose has been to show that attachment preferences can be mediated, and indeed overridden, by contextual factors. Experiment 4 showed exactly that.

It could be argued, however, that our data might reflect a contextual sensitivity which developed during the course of the experiment (perhaps because the ''when'' contexts were biased toward high attachment whereas the ''that'' contexts were not, themselves, biased one way or another). We do not believe this to be the case because the targets resolved equally often to the high and the low attachments and because this should, if anything, encourage subjects to ignore any contextually induced attachment bias. Evidently, they did not ignore context.

GENERAL DISCUSSION

If the recency preference can be overridden by context, by what mechanism is this override achieved? In a sentence fragment such as (14) below, the most recent argument structure at the offset of ''committee'' (and hence, according to MacDonald et al., 1994, the most representationally active) is the one corresponding to ''proposed.''

(14) She'll implement the plan she proposed to the committee . . .

But the language user must be sensitive to the fact that in answer to the question in (15) below, any of the alternatives given in (16) is possible:

(15) When will Fiona implement the plan she proposed to the committee?

(16) Next week.
She'll implement the plan next week.
She'll implement the plan she pro-

posed next week.
She'll implement the plan she proposed to the committee next week.

Once this relationship between question and answer is learned (and exactly *how* it is learned is beyond the scope of this paper), the language user will ''expect'' an answer when given a question such as that in (15) and will expect that answer to take one or other of the forms given in (16). Or rather, given that it is impossible to predict the exact form of any sentence that might follow (15), the language user will continually adjust his or her predictions on the basis of what has been encountered so far of the following sentence. Thus, there will be no expectation of an adverbial following the fragment ''She'll,'' or ''She'll implement,'' or ''She'll implement the plan she.'' Instead, as the sentence unfolds through time, the language user will only expect the adverb to occur in very specific locations. In representational terms, the language user will *predictively activate* some representation of the upcoming adverbial at each of the positions in which an adverb like ''next week'' can occur—in much the same way as, for example, Elman's (1990) recurrent network learned, in effect, to predict the syntactic categories that could occur in particular sentential positions. In effect, then, the question sets up an expectation regarding the location of the relevant information in the target sentence (cf. Cutler & Fodor, 1979), and this is manifested as the predictive activation of structures which support the subsequent integration of that information. But how does this notion of predictive activation fit within MacDonald et al.'s model of recency effects, which is based primarily on the notion that representations corresponding to *argument structures* are activated and, subsequently, decay?

We shall begin by supposing that two broad kinds of representation are activated during sentence processing; one kind corresponds to argument structures, as in the MacDonald et al. model, and the other kind corresponds to other kinds of dependency which, although not a part of the argument structure of a verb, are licensed by (and hence predictable from) the grammar—adjuncts, for example. To put the proposal in concrete terms, after the question in (15), a hearer or reader might, without any further input, activate a representation corresponding to an adverbial that could be *predicted* to occur in first position within a subsequent utterance. This representation would rapidly decrease in activation when ''She'll,'' and then ''implement'' are encountered (because neither permits the predicted adverbial to occur immediately after). At ''implement,'' a representation corresponding to the argument structure of ''implement'' will activate, although by the time ''the plan'' is encountered, this will have started to decay (simply because of the passing of time—cf. MacDonald et al. 1994). However, when ''the plan'' is encountered, a representation corresponding to a predicted (or indeed, *projected*) adverbial will again become activated, because the contextually contingent adverbial could now occur at the offset of this fragment. When ''she'' is encountered, the argument structure associated with ''implement'' (or rather, the representation corresponding to this structure) will continue to decay, and the representation corresponding to the predicted adverbial will also decline (because no adverbial can be predicted to follow ''she''). At ''proposed,'' the representation for ''implement'' continues to decay, whilst a representation associated with the argument structure of ''propose'' will become active. Also at this point, a representation corresponding to the predicted adverbial that is contingent on the contextual cue will again become active (because the adverbial could occur in the next position). However, because each argument structure licenses adjunct-like predictions, an alternative adverbial, contingent on ''propose,'' will also be predicted at this point (as licensed by the grammar—that is, by the knowledge that allows us to predict what class of word can occur where). But with the subsequent words ''to the committee,'' both the argument structure associated with ''propose,'' *and* its con-

tingent predictions, will decay. Finally, at the offset of ''She'll implement the plan she proposed to the committee,'' a representation corresponding to the contextually predicted adverb will again activate. But at this point in the sentence, there will be relatively less predictive activation of any representation corresponding to an adverbial that could be contingent on ''proposed,'' because the argument structure associated with this verb, and any consequent contingent predictions, will have decayed.

The claim that structure can be predictively activated (or projected) in the manner just outlined, and that such activation supports subsequent integration of the information corresponding to whatever structure is activated, is little different from the existing claim that representations corresponding to argument structures are activated in response to the lexical realization of verbs (or whatever other items support argument structures). An argument structure is, after all, a projection of structure, some of which is yet to be encountered within the sentence. And presumably that structure does support the subsequent integration of information at particular locations within the sentence. In this regard the distinction between the two broad kinds of representation introduced above—between argument structures and other grammatically licensed nonargument dependencies—is not as clear-cut as it may at first appear. Arguments and adjuncts are both dependencies licensed by the grammar. The difference, presumably, is in the nature of the dependency—that is, in the nature of the predictive contingency between the argument/adjunct and the item that licenses the corresponding projection. As pointed out by Frazier (1995), adjunct representations must somehow be incorporated within any theory of constraint-satisfaction if the behavioral phenomena concerning such structures is to be explained within these theories. Although we cannot offer here a complete account of such a theory, we do tentatively propose that a predictive element does play a crucial role in the processing of the structures we have explored empirically here.

One final consideration concerns the status of Crain and Steedman's (1985) Principle of Parsimony and the notion of presupposition. A context (or knowledge about what is true in the world) can be said to support the presuppositions associated with the usage of a particular construction if there exists a (statistical) interdependence between some property of that context and the target construction. If the presuppositions associated with a particular structure are indeed statistical dependencies between those structures and the contexts in which they occur, then those dependencies could be one of a range of statistically derived constraints which, according to the constraint-based approaches to sentence processing, are applied incrementally during sentence processing. And whether one chooses to refer to the interdependence as a ''presupposition,'' or as a ''contextually based constraint on activation'' is just a matter of terminology. Although the data from Experiment 1 do appear to compromise the efficacy of these presuppositional constraints with regard to the processing of closure ambiguities, we do nonetheless note that the contextual constraint which *did* succeed in overriding the recency preference did also require the interposition of an extra prepositional phrase, as in ''She'll implement the plan she proposed *to the committee* . . .'' We do not currently have data on what would occur if such target sentences were presented in the referential contexts we used in Experiment 1, just as we do not have data on what would occur if such target sentences were processed in the null context. Clearly, further research is required to establish the independent contributions of the alternative constraints which may influence the recency preference. The data we have presented thus far do strongly suggest that extra-sentential contextual information *can* override the recency preference and that this preference is most likely due to the temporal (and hence, in some respect, structural) factors suggested by MacDonald et al. (1994). Our explanation

of contextual override requires that comprehension is achieved through the satisfaction and interaction of multiple probabilistic, *and hence predictive,* constraints.

APPENDIX

Experimental Items Used in Experiments 3 and 4

Each item consisted of three sentences: a lead-in sentence which introduced the entities subsequently referred to; a sentence which included the indirect ''when'' question, ''which'' question (Experiment 3) or ''that'' statement (Experiment 4); and the target sentence. The lead-in sentences, which were identical across conditions within each experiment, are not included here for reasons of space. The full items, and fillers, are available from the first author at http://www.york.ac.uk/~gtma1/closure.html. The slashes indicate the alternative options within and across the two experiments (in the order ''when'' option / ''that'' option / ''which'' option). In the final sentence, the slashes indicate the alternative attachments (in the order 'high' / 'low'); the material included within square brackets was only present in Experiment 4.

Her brother has no doubt when / that Alice will wear the jumper she washed / which of the jumpers that Alice washed she'll wear. She'll wear the jumper she washed [with special detergent] tomorrow / yesterday, he's sure.

Her dad wonders momentarily when / rather assumes that Amy built the kite she'll fly / which of the kites that Amy will fly she built herself. She built the kite she'll fly [in the competition] last / next week, he's almost positive.

Her husband doesn't need to ask when / be told that Ann will freeze the beans she picked / which of the beans that Anne's picked she'll freeze. She'll freeze the beans she picked [from the garden] tomorrow / yesterday, he's quite sure.

His supervisors can guess when / are guessing that Arthur arrested the shoplifter he'll prosecute / which of shoplifters that Arthur will prosecute he arrested himself. He arrested the shoplifter he'll prosecute [for theft] last / next week, they figure.

Her boyfriend wonders when / assumes that Carol will try the wine she bought / which of the wines that Carol's bought she'll try. She'll try the wine she bought [in a sale] next / last week, he's pretty sure.

Her boyfriend can guess when / is guessing that Cathy will burn the wood she gathered / which of the wood that Cathy has gathered she'll burn. She'll burn the wood she gathered [in the forest] next / last week, he's nearly positive.

Her boss wondered briefly when / rather assumed that Christine had chosen the applicant she'll interview / which of the applicants that Christine will interview she chose herself. She chose the applicant she'll interview [on the phone] yesterday / tomorrow, he's guessing.

His mother wonders when / hopes that Colin will marry the woman he rescued / which of the women that Colin rescued he'll marry. He'll marry the woman he rescued [from the river] next / last week, she's assuming.

The band members wonder when / are guessing that David will employ the musician he auditioned / which of the musicians that David auditioned he'll employ. He'll employ the musician he auditioned [for the band] tomorrow / yesterday, they're hoping.

Her husband wonders when / is guessing that Eileen will expose the judge she blackmailed / which of the judges that Eileen has blackmailed she'll expose. She'll expose the judge she blackmailed [in the sting] next / last month, he figures.

Her son wonders when / is guessing that Elaine wrote the speech she'll deliver / which of the speeches that Elaine will deliver she wrote herself. She wrote the speech she'll de-

liver [to her constituents] last / next week, he's pretty sure.

The customer wonders when / hopes that Graham will service the vehicle he drove / which of the vehicles that Graham had driven he'll service. He'll service the vehicle he drove [at high speeds] tomorrow / yesterday, she's pretty sure.

His mum can't help wondering when / thinking that Ian created the sculpture he'll display / which of the sculptures that Ian will display he created himself. He created the sculpture he'll display [in the gallery] last / next month, she believes.

His parents can guess when / suspect that Jack will meet the friend he phoned / which of the friends that Jack phoned he'll meet. He'll meet the friend he phoned [from a call box] tomorrow / yesterday, they're guessing.

The secretary wonders momentarily when / rather assumes that James will complete the project he started / which of the projects that James started he'll complete first. He'll complete the project he started [in his lab] next / last month, she's pretty sure.

Her supervisor wondered vaguely when / rather assumed that Jane ran the experiment she'll analyse / which of the experiments that Jane will analyse she ran already. She ran the experiment she'll analyse [for the project] yesterday / tomorrow, he figures.

The owner wonders when / is guessing that Joan last / already groomed the horse she'll ride / which of the horses that Joan will ride she'd groomed herself. She groomed the horse she'll ride [around the track] yesterday / tomorrow, he's pretty sure.

His wife wonders when / is guessing that Joe inoculated the sheep that he'll shear / which of the sheep that Joe will shear he'd inoculated. He inoculated the sheep he'll shear [with the old shears] yesterday / tomorrow, she figures.

His mother wonders when exactly / suspects that John hired the clerk he'll promote / which of the clerks that John will promote he'd hired himself. He hired the clerk he'll promote [to a regular position] last / next month, she figures.

His wife wonders when / suspects that Keith tore the shirt he'll mend / which of the shirts that Keith will mend he tore himself. He tore the shirt he'll mend [by hand] yesterday / tomorrow, she's almost positive.

His children wanted to know when / to be sure that Kenneth will ice the cake he baked / which of the cakes that Kenneth baked he'll ice. He'll ice the cake he baked [in the microwave] tomorrow / yesterday, they're almost positive.

Her parents can't help but wonder when / think that Diane will visit the man she described / which of the men that Diane's described she'll visit. She'll visit the man she described [to her family] next / last week, they suspect.

The church members wonder when / are guessing that Fiona will implement the plan she proposed / which of the plans that Fiona's proposed she'll implement. She'll implement the plan she proposed [to the committee] next / last week, they figure.

Her parents wonder when / hope that Linda will show the slides she took / which of the slides that Linda has taken she'll show. She'll show the slides she took [with her camera] next / last month, they're pretty sure.

His friends wondered when / are guessing that Malcolm organised the party he'll hold / which of the parties that Malcolm will hold he organised himself. He organised the party he'll hold [on his patio] last / next month, they're pretty sure.

Her kids asked when / hoped that Mandy caught the fish she'll prepare / which of the fish that Mandy will prepare she'd caught. She caught the fish she'll prepare [for the family] yesterday / tomorrow, they're pretty sure.

His girlfriend can't help wondering when /

assuming that Martin brewed the beer he'll serve / which of the beers that Martin will serve he brewed himself. He brewed the beer he'll serve [at the beertasting] last / next week, she's guessing.

His mother wonders when / is guessing that Mike watered the lawn he'll mow / which of the lawns that Mike will mow he watered as well. He watered the lawn he'll mow [with the hand mower] yesterday / tomorrow, she's pretty sure.

The manager wonders when / suspects that Penny had insulted the candidate she'll challenge / which of the candidates that Penny will challenge she'd insulted already. She insulted the candidate she'll challenge [in the debate] yesterday / tomorrow, he thinks.

His girlfriend wonders when / assumes that Peter will answer the letter he received / which of the letters that Peter's received he'll answer. He'll answer the letter he received [from his ex-wife] next / last week, she thinks.

His wife wondered when / hoped that Tim had moved the wardrobe he'll restore / which of the wardrobes that Tim will restore he'd moved already. He moved the wardrobe he'll restore [in his workshop] last / next week, she's pretty sure.

His wife didn't need to ask when / be told that Tom will plant the field he weeded / which of the fields that Tom weeded he'll plant first. He'll plant the field he weeded [with his hoe] next / last week, she's sure.

REFERENCES

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

Altmann, G. T. M. (1994). Regression-contingent analyses of eye movements during sentence processing: Reply to Rayner and Sereno. *Memory and Cognition,* **22,** 286–290.

Altmann, G. T. M., Garnham, A., & Dennis, Y. (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: Learning, Memory & Cognition,* **20,** 209–216.

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

Britt, M. A. (1994). The interaction of referential ambiguity and argument structure. *Journal of Memory and Language,* **33,** 251–283.

Britt, M. A., Perfetti, C. A., Garrod, S., & Rayner, K. (1992). Parsing in discourse: Context effects and their limits. *Journal of Memory and Language,* **31,** 293–314.

Carreiras, M., & Clifton, C. (1993). Relative clause interpretation preferences in Spanish and English. *Language and Speech,* **36,** 353–372.

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given–new contract. In R. O. Freedle (Eds.), *Discourse production and comprehension.* (pp. 1–40). Norwood, NJ: Ablex.

Clifton, C., & Ferreira, F. (1989). Ambiguity in context. *Language and Cognitive Processes,* **4,** SI 77–104.

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

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition,* **30,** 73–105.

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition,* **7,** 49–59.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science,* **14,** 179–211.

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

Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton-Mifflin.Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies.* Bloomington, IN: Indiana University Linguistics Club.

Frazier, L. (1995). Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research,* **24,** 437–468.

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

Gilboy, E., Sopena, J. M., Clifton, C., & Frazier, L. (1995). Argument structure and association prefer-

ences in Spanish and English complex NPs. *Cognition,* **54,** 131–167.

Hanna, J. E., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1996). Integrating discourse and local constraints in resolving lexical thematic ambiguities. In *Sixteenth annual cognitive science society meeting* (pp. 266–271). Hillsdale, NJ: Erlbaum.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition,* **2,** 15–47.

Liversedge, S. P., Pickering, M. J., Branigan, H. P., & van Gompel, R. P. G. (1998). Processing arguments and adjuncts in isolation and context: The case of by-phrase ambiguities in passives. *Journal of Experimental Psychology: Learning, Memory and Cognition,* **24,** 461–475.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review,* **101,** 676–703.

MacDonald, M. C., & Thornton, R. (1996). *Constraint-based models and modification ambiguities.* Poster presented to the 9th Annual CUNY Conference on Sentence Processing. New York.

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 Behaviour,* **22,** 358–374.

Rayner, K., Garrod, S., & Perfetti, C. A. (1992). Discourse influences during parsing are delayed. *Cognition,* **45,** 109–139.

Rayner, K., & Sereno, S. C. (1994). Regressive eye movements and sentence parsing: On the use of regression contingent analyses. *Memory and Cognition,* **22,** 281–285.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C. (1989). Eye movements and online language comprehension processes. *Language and Cognitive Processes,* **4,** SI 21–50.

Spivey-Knowlton, M., & Tanenhaus, M. (1994). Referential context and syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing.* (pp. 415–439). Hillsdale, NJ: Erlbaum.

Steedman, M. J., & Altmann, G. T. M. (1989). Ambiguity in context: A reply. *Language and Cognitive Processes,* **4,** SI 105–122.

Trueswell, J. C., & Tanenhaus, M. K. (1991). Tense, temporal context, and syntactic ambiguity resolution. *Language and Cognitive Processes,* **6,** 303–338.

Trueswell, J. C., & Tanenhaus, M. K. (1994). Towards a lexicalist framework of constraint-based syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing.* (pp. 155–179). Hillsdale, NJ: Erlbaum.

(Received June 3, 1997)
(Revision received December 15, 1997)