Journal of Memory and Language **43,** 447–475 (2000)
doi:10.1006/jmla.2000.2708, available online at http://www.idealibrary.com on IDEAL®

Ambiguity Resolution in Sentence Processing: Evidence against Frequency-Based Accounts

Martin J. Pickering

Department of Psychology, Human Communication Research Centre, University of Glasgow, Glasgow, Scotland

Matthew J. Traxler

Florida State University

and

Matthew W. Crocker

Department of Computational Linguistics, University of Saarbrücken, Saarbrücken, Germany

Three eye-tracking experiments investigated two frequency-based processing accounts: the serial lexical-guidance account, in which people adopt the analysis compatible with the most likely subcategorization of a verb; and the serial-likelihood account, in which people adopt the analysis that they would regard as the most likely analysis, given the information available at the point of ambiguity. The results demonstrate that neither of these accounts explains readers' performance. Instead people preferred to attach noun phrases as arguments of verbs even when such analyses were unlikely to be correct. We suggest that these results fit well with a model in which the processor initially favors informative analyses.

Key Words: parsing; frequency; eye-tracking; syntax; ambiguity; garden path.

This article addresses the question of how the processor decides on its initial strategy for syntactic ambiguity resolution. At a point of ambiguity, more than one analysis is possible. An effective strategy might be to adopt the analysis that has most frequently turned out to be correct in the past. Assuming that the world stays the same in most respects, the analysis that has most frequently been correct in the past should provide a good estimate of which analysis is most likely to be correct again. Hence, by adopting this analysis, the processor should make fewer errors than if it chose any other analysis.

In this article, we contrast this frequency-based approach with alternative strategies in which the processor does not attempt to adopt the analysis that has most frequently turned out to be correct. Many current theories of parsing pay no attention to frequency in determining initial choice of analysis, but instead employ a different strategy entirely, such as selecting the syntactically simplest analysis in some well-defined sense (e.g., Frazier, 1979, 1987). Other accounts, however, do pay close attention to information about frequency (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell, Tanenhaus, & Kello, 1993). One approach is fairly directly based on frequency, with initial parsing preferences being determined by prior exposure (Mitchell, Cuetos, Corley, & Brysba-

The order of the first two authors is arbitrary. We thank Keith Edwards, Steven Frisson, Roger van Gompel, and members of the Sentence Processing Group, Universities of Glasgow and Edinburgh. The research was supported by ESRC research Grant No. R000234542 and a British Academy Postdoctoral Fellowship, both awarded to Martin Pickering, and an ESRC Research Fellowship, awarded to Matthew Crocker.

Address correspondence and reprint requests to Martin Pickering now at Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9LW United Kingdom. Fax: +44 131 650 3461. E-mail: Martin.Pickering@ed.ac.uk.

0749-596X/00 $35.00

ert, 1995). We leave consideration of specific current theories to the General Discussion.

Our purpose in this introduction is to motivate three experiments that test a particular class of account in which initial choice of analysis is determined by judgments about frequency. It turns out to be impossible to test all possible frequency-based accounts in a single set of experiments. Hence, we focus our attention on serial accounts, in which the processor considers only one analysis at a time. The other major concern is the granularity at which frequency is assessed. We therefore consider a range of serial frequency-based accounts that differ with respect to grain size. The experiments were designed to test two specific accounts that differ greatly in this respect: a *serial lexical-guidance* account and what we term a *serial-likelihood* account. We then review relevant experimental research and outline our experiments. The General Discussion considers current accounts in light of these results and sketches an alternative account in which the initial choice of analysis is based on what we term its informativity.

FREQUENCY, LEXICAL GUIDANCE, AND LIKELIHOOD

Serial, frequency-based accounts of syntactic ambiguity resolution assume that the processor selects the syntactic analysis that has most frequently been employed in the past, on the assumption that this analysis is most likely to be correct again. Such accounts therefore attempt to maximize the probability of making the right decision and minimize the need for subsequent reanalysis. We assume that the likelihood of an analysis is its probability as estimated from frequencies observed in a person's prior linguistic experience. However, this begs a very large question: frequency of what?

Let us concentrate on syntactic ambiguities that are a result of a verb's having more than one possible subcategorization. For example, the verb *realized* can take a noun-phrase object (the *object analysis*), as in *The athlete realized her potential*, or a sentential complement (the *sentential-complement analysis*), as in *The athlete realized her potential might make her a world-class sprinter*. However, *realized* takes a sentential complement much more often than an object. This can be established by counting the frequency of the different subcategorizations of *realized* within a relevant corpus. Alternatively, it can be assessed by asking participants to construct sentences containing *realized* (see pretests below). If people adopt the analysis that is compatible with the most frequent subcategorization, then they should adopt the sentential-complement analysis whenever they encounter *realized*. We discuss this *serial lexical-guidance* account in more detail below.

However, there are other serial frequency-based accounts. The serial lexical-guidance account pays attention to the verb *realized*, but ignores all other aspects of the context. For example, it is possible that the relative frequency of the object and sentential-complement analyses is different after *The athlete realized* . . . than after *realized*. There are very clear cases where the nature of the subject noun affects the likelihood of different subcategorizations. For instance, *rolled* is likely to be transitive after *The man rolled . . .*, but intransitive after *The ball rolled* . . . . This demonstrates the importance of grain size to frequency-based models (see Gibson & Schütze, 1999; Mitchell et al., 1995). Notice, however, that the subcategorization frequencies for *realized* may not be affected so much by the nature of the subject. If so, grain size will not greatly affect the predictions of frequency-based models for this verb. But it is still the case that any well-formed frequency-based model needs to specify grain size.

There are infinitely many grain sizes (based on just the verb, verb plus subject noun, verb plus some representation of discourse context, etc.). The processor could count frequencies at any of these levels. Moreover, it could pay attention to additional levels of linguistic description (e.g., the animacy of the subject) or to factors such as extralinguistic context or idiosyncrasies of the speaker or writer. However, any relatively fine-grained frequency-based model faces a very serious difficulty, known as the "sparse data problem" (see, e.g., Charniak, 1997). People will have encountered *realized* vast numbers of times, and *The athlete realized* a few times. But most complex contexts (e.g., *The athlete who holds the pole-vault record realized*) will never have been encountered before. Indeed, the relevant context need not be restricted to the target sentence. So, a frequency-based account cannot plausibly be based on the frequency of the whole preceding context.

As there are infinitely many possible grain sizes, we cannot test all accounts. Instead, we consider two extremes: one based solely on the subcategorization preferences of the verb (serial lexical guidance) and one based on preferred completions for sentence fragments (serial likelihood).

SERIAL LEXICAL GUIDANCE

The serial lexical-guidance account assumes that the parser determines its choice of analysis by paying attention to subcategorization information alone. This is the lexical-guidance account proposed by Ford, Bresnan, and Kaplan (1982), and suggested by Fodor (1978) in the context of the processing of unbounded dependencies. According to Ford et al., if *realized* is most commonly employed on the sentential-complement analysis, then the parser will adopt this analysis in preference to the object analysis during initial processing. Of course, it is more likely that the actual sentence will employ the sentential-complement analysis than the object analysis. Thus the model appears to be driven by likelihood, but solely with respect to properties of the verb. For instance, the verb *rolled* might be preferentially transitive, but after *the ball rolled* it is preferentially intransitive. Ford et al.'s parser would ignore *the ball* and would base its decision on *rolled* alone.

SERIAL LIKELIHOOD

On the serial-likelihood account, comprehenders pay attention to all available information in estimating the likelihood of possible syntactic analyses for a particular fragment. People could, in principle, be asked to judge whether the sentential complement or the object analysis was more likely to be correct after *The athlete realized* . . . (e.g., they could place bets on the outcome). In practice, this is impossible, because many people are unaware of the linguistic assumptions underlying the analyses. An alternative is to have people complete sentence fragments and to estimate the likelihood of the different analyses on the basis of the number of times that different syntactic analyses occur in the completions. The serial-likelihood account predicts that the parser will initially adopt the analysis that the people produced most often when completing the sentence fragment. (It is obviously essential that producers and comprehenders are randomly selected form the same population.)

It is not straightforward to determine the sources of information or *constraints* that people draw upon in making these completions (or, indeed, how they are integrated). Presumably, these sources of information include verb subcategorization preferences, other aspects of the linguistic context (e.g., characteristics of the subject), and extralinguistic factors. An itemization of these sources of information would be necessary for a complete account of production. To test the serial-likelihood account, however, all we need to know is the result of this process of integration. The use of (written) completions to determine off-line preferences is found in work on constraint-based theories of processing (Garnsey, Pearlmutter, Myers, & Lotowsky, 1997; Trueswell et al., 1993), as well as work critical of such approaches (Clifton, Kennison, & Albrecht, 1997; Liversedge, Pickering, Branigan, & Van Gompel, 1998; Traxler, Pickering, & Clifton, 1998). Although written completions possibly involve slight biases (e.g., participants may prefer not to write too much or may be primed by previous productions) and clearly involve an element of comprehension, they probably approximate to what speakers naturally do if they produce a sentence that begins with a particular fragment.

A model that is closely related to serial likelihood is the Linguistic Tuning hypothesis (Mitchell et al., 1995), in which parsing decisions are determined by the frequency with which the alternative analyses are used in the language (see also Brysbaert & Mitchell, 1996; Cuetos, Mitchell, & Corley, 1996; Mitchell & Cuetos, 1991). For instance, *the daughter of the* *colonel with the limp* is ambiguous between two analyses and interpretations (cf. Cuetos & Mitchell, 1988). On the one hand, *with the limp* can be attached to *the daughter of the colonel* so that it is the daughter who has the limp; on the other hand, it can be attached to *the colonel* so that it is the colonel who has the limp. These analyses differ in syntactic structure but not subcategorization. Hence, Mitchell et al. propose that listeners select the most frequent syntactic analysis. Their account is thus more general than Ford et al.'s (1982) account. However, they do not commit to any particular account of how the frequency of different analyses is assessed. The finer the grain-size employed in making this assessment, the more similar are predictions of the Linguistic Tuning hypothesis to the serial-likelihood account.

The serial-likelihood account predicts that people's initial preferences in parsing correspond to their production preferences, as measured by completions. In order to investigate this account, we employed the simple case of a minimal context consisting of a subject and a verb (e.g., *The young athlete realized*) in isolation from any discourse context. We do not know precisely which factors are relevant to determining the likelihood of a particular completion, but a complete specification of the relevant factors is not necessary to test the serial-likelihood account. Because the account predicts that the initial stages of analysis will match off-line preferences, all we need to know is off-line preferences and on-line parsing preferences. The prediction of the account is simply that these two sets of preferences are the same.

In conclusion, we can straightforwardly test two serial frequency-based accounts. One is the serial lexical-guidance account proposed by Ford et al. (1982), in which comprehenders initially adopt the analysis compatible with the most likely subcategorization for the verb. The other is the serial-likelihood account, in which comprehenders initially adopt the analysis that would be the most likely completion for the fragment. If serial lexical guidance and serial likelihood make the same predictions, it is very unlikely that the predictions of other frequency-based accounts (e.g., using different grain sizes) will be different.

EXPERIMENTAL EVIDENCE ON INITIAL CHOICE OF ANALYSIS

Many experimental studies have asked whether some particular source of information is used immediately in the process of syntactic ambiguity resolution or whether its use is delayed until a later stage of parsing (see, e.g., Traxler & Pickering, 1996a, for discussion). We can identify three major issues: whether semantic plausibility information, discourse context information, and detailed lexical information affect initial parsing decisions. Many early studies supported delayed use of semantic plausibility information (e.g., Ferreira & Clifton, 1986; Rayner et al., 1983). However, more recent studies have found very rapid use of semantic information in syntactic disambiguation (e.g., Boland, Tanenhaus, Garnsey, & Carlson, 1995; MacDonald, 1994; Trueswell, Tanenhaus, & Garnsey, 1994). Similarly, some studies suggest that discourse context can affect initial parsing decisions (e.g., Altmann, Garnham, & Dennis, 1992; Altmann, Garnham, & Henstra, 1994; Altmann & Steedman, 1988; Britt, 1994; Britt, Perfetti, Garrod, & Rayner, 1992; Crain & Steedman, 1985), whereas other studies suggest that it only affects reanalysis (e.g., Ferreira & Clifton, 1986; Mitchell, Corley, & Garnham, 1992; Rayner, Garrod, & Perfetti, 1992). Clearly, semantic and discourse effects can occur much more rapidly than early studies suggested, but the question remains whether these effects reflect initial parsing decisions or very rapid revision.

Our main concern is with the use of detailed lexical information. Frazier and Rayner (1982) found that readers misanalyzed sentences during reading and argued that they did this because they followed syntactic parsing strategies like Minimal Attachment and Late Closure that ignore such detailed lexical information. Such accounts have been called lexical-filtering proposals (e.g., Mitchell, 1989) and are consistent with the "Garden Path" model (Frazier, 1979, 1987). It was also soon discovered that verb biases affected processing difficulty in such sentences (Holmes, 1987; Mitchell & Holmes, 1985). According to lexical-filtering proposals, effects of verb bias must reflect reanalysis (Ferreira & Henderson, 1990; Mitchell, 1987, 1989; Rayner & Frazier, 1987). Again, however, there is evidence that lexical information can be used extremely rapidly, and the claim has been made that it can be used during initial processing (Garnsey et al., 1997; Trueswell et al., 1993). We now assess the current status of this controversy, making particular reference to complement-clause ambiguities.

Ferreira and Henderson (1990) monitored participants' eye movements while they read sentences like *He forgot Pam needed a ride home with him,* where *forgot* is preferentially used on the object analysis (e.g., *He forgot Pam.*), and sentences like *He wished Pam needed a ride home with him,* where *wished* is preferentially used on the sentential-complement analysis, as well as control sentences containing the complementizer *that* before *Pam.* At the disambiguating verb *needed,* readers' first fixations were longer when the test sentence lacked a complementizer than when a complementizer was present, and this effect was uniform across sentences containing object-biased verbs (*O sentences*) and sentences containing sentential-complement-biased verbs (*C sentences*), as evidenced by the lack of an interaction of verb bias and complementizer presence. Total times produced similar results. Finding greater first fixation and total reading times in the ambiguous sentences than in the unambiguous sentences for both object-biased and sentential-complement-biased verbs led Ferreira and Henderson to conclude that the object analysis had been computed and evaluated regardless of verb bias.

In contrast, Trueswell et al. (1993) found very different results using verbs whose biases had been carefully normed. First-pass times showed that readers had difficulty at disambiguation with O sentences but not difficulty with C sentences. This is compatible with lexical guidance. The alternative explanation is that readers initially adopted the object analysis for all verbs and rapidly reanalyzed the C sentences. In accord with this, Trueswell et al. found some increased reading time for the C sentences before disambiguation. But they argued against this explanation, and instead claimed that this effect might be because the sentential-complement analysis is more complex than the object analysis. However, Garnsey et al. (1997) found that the object analysis was much more plausible in Trueswell et al.'s O sentences than in their C sentences. This could explain Trueswell et al.'s findings and suggests that their items were imperfectly controlled in a potentially important respect.

Garnsey et al. (1997) conducted a further study on complement-clause ambiguities, using sentential-complement-biased, object-biased, and equibiased (i.e., balanced) verbs. They also manipulated the plausibility of the object analysis. O sentences caused processing difficulty at disambiguation in comparison to unambiguous control sentences with complementizers, but C sentences did not. Their manipulation of plausibility produced rather less clear results, but there was some evidence that O sentences with plausible object analyses were easier to process than O sentences with implausible object analyses. There was no evidence of similar effects in the C sentences. These results provide some evidence for lexical guidance, but there are many reasons to question this conclusion. As this result relates directly to the experiments reported below, we return to Garnsey et al. in the General Discussion.

EXPERIMENTS

Below we report three eye-tracking experiments that tested the two frequency-based accounts discussed above. The three experiments are designed to allow our conclusions to hold over different kinds of sentences and presentation conditions. Experiments 1 and 2 employ complement-clause ambiguities, whereas Experiment 3 employs locally ambiguous sentences involving preposed subordinate clauses. Experiments 1 and 3 involve sentences presented in isolation, whereas Experiment 2 presents the sentences from Experiment 1 in short discourse contexts. In all cases, we manipulated plausibility in such a way that plausibility could only have an effect if the parser initially favored an analysis that neither the serial lexical-guidance nor serial-likelihood accounts would predict it would. More informally, we manipulated the plausibility of the less frequent analysis while controlling the frequency of the more frequent analysis. If this plausibility manipulation had an effect, then readers must have been considering the less frequent analysis.

EXPERIMENT 1

Experiment 1 employed sentences like (1a) and (1b):

(1a). The young athlete realized her potential one day might make her a world-class sprinter.
(1b). The young athlete realized her exercises one day might make her a world-class sprinter.

Two relevant analyses are available after *The young athlete realized:* the object analysis (as in, e.g., *The young athlete realized her potential*) and the sentential-complement analysis [as in (1a) and (1b)]. However, both sentences are disambiguated by the word *might;* following *might,* only the sentential-complement analysis is possible. Pretests showed that the sentential-complement analysis was the most common analysis for *realized,* that it was the most common analysis following *The young athlete realized,* and that it was the most common analysis following *The young athlete realized her.* Hence, both the serial lexical-guidance and the serial-likelihood accounts predict that the processor will initially adopt the sentential-complement analysis. The experiment contained no transitive sentences, so that readers would not change subcategorization preferences as a result of the experimental session.

We manipulated the plausibility of the object analysis, while holding the plausibility of the sentential-complement analysis constant. Thus, the sentence *The young athlete realized her potential* is plausible, whereas the sentence *The young athlete realized her exercises* is implausible (or semantically anomalous). In contrast, the complete sentences (1a) and (1b) were equally plausible. Hence any plausibility effect must have been due to readers adopting the object analysis.

We predicted that if readers adopted the object analysis, they would detect that (1b) was less plausible than (1a) before the word *might,* when it becomes clear that the object analysis cannot be correct. This disruption should manifest itself either as increased reading time for (1b) versus (1a) or as more regressive eye movements in (1b) versus (1a), before *might.* Such a finding would be compatible with Pickering and Traxler (1998), who showed comparable effects with complement-clause ambiguities that were not biased toward the sentential-complement analysis, and with Traxler and Pickering (1996b), who employed unbounded-dependency constructions. Notice that this method does not require the use of unambiguous control sentences to show misanalysis. Such sentences may differ from the experimental sentences in irrelevant ways.

We might also expect to find effects after disambiguation if readers misanalyze. In a series of experiments, Pickering and Traxler (1998) found that readers experienced *more* difficulty after disambiguation in sentences whose misanalysis was plausible than sentences whose misanalysis was implausible. Readers appear reluctant to give up a plausible initial analysis. But if the initial analysis is implausible, readers commit less strongly to the initial analysis and may even abandon it before reaching disambiguation. If participants follow one or other of the frequency-based accounts discussed above, they adopt the sentential-complement analysis initially and hence never experience any processing difficulty from reanalysis. But if they do not adopt a frequency-based processing strategy, they may have more difficulty with (1a) than (1b) after reading *might.* Hence, there may be a *crossover* between two parts of the sentence: During the ambiguous region (more precisely, *potential/exercises one day*), (1b) should be harder than (1a); after disambiguation, (i.e., from *might* onward) (1a) may be harder than (1b).

The main predictions concern initial processing. However, if the parser follows a frequency-based strategy, and given that the correct sentential-complement analyses of both (1a) and (1b) are plausible, there is no reason why the object analyses should ever be considered.

Hence, frequency-based accounts predict no effects of the plausibility of the object analysis during later stages of processing.

Method

Participants. Forty native British English speakers with normal or corrected-to-normal vision from the University of Glasgow were paid to participate in the eye-tracking phase of the study. Some had participated in previous eye-tracking experiments.

Items. The items comprised 16 pairs of sentences like (1) above (see Appendix). Both (1a) and (1b) are plausible, but (1a) has a plausible object analysis, whereas (1b) has an implausible object analysis. One version of each sentence was assigned randomly to one of two experimental lists. Thus, an individual participant saw one version of each test sentence. The critical words (*potential* and *exercises*) were the same length and were matched for frequency [plausible objects: 86 per million in the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993); implausible objects: 71 per million; $F < 1$].

Verb norming. First, we generated plausible verb–object combinations for 15 verbs that could take either sentential complements or objects. To determine the preferences for each verb, we conducted a series of pretests. We were most concerned that the verbs we selected preferred reduced-sentential-complement (i.e., without *that*) continuations over object continuations. We focused on reduced sentential complements because they contain the relevant ambiguity, whereas unreduced sentential complements do not and because it is most conservative to use verbs whose reduced-sentential-complement frequency is higher than their object frequency. In the first pretest, we presented the verbs in isolation to 20 participants and asked them to write a sentence containing each verb. The verbs were included in a randomized list along with 92 other verbs. In the second pretest, we attached subject noun phrases to each verb and asked a further 29 participants to write a continuation for each subject–verb combination. Table 1 presents the results of these first two pretests. We selected verbs that (i) produced at least as many reduced-sentential-complement as object sentences in the first pretest and (ii) produced a 2:1 or greater ratio of reduced-sentential-complement to object continuations in the second pretest. Note that we employed all verbs that met our criteria (and it is unlikely that there are many more such verbs in British English). In a final pretest, we presented the beginnings of the test sentences through the article or possessive pronoun that preceded the matrix verb. We asked 20 further participants to write continuations for each fragment. At the point where the continuation began, only object or reduced-sentential-comple-

TABLE 1

Verb Preferences for Experiments 1 and 2

	Response			
	S0	NP	S-	Other
Results of first verb-preference test				
Verb				
admitted	7	7	1	5
decided	2	1	2	15
hinted	2	1	6	11
implied	8	2	8	2
pretended	6	0	4	10
realized	7	5	5	3
Results of second verb-preference test				
Stem				
The young athlete realized . . .	16	0	12	1
The doctor admitted . . .	9	4	16	0
The teacher hinted . . .	3	1	14	11
The bank admitted . . .	8	3	18	0
The investor realized . . .	11	3	14	1
The witness admitted . . .	16	0	9	4
The judge decided . . .	5	0	4	20
The slacker pretended . . .	3	0	6	20
The butler admitted . . .	10	2	17	0
The lecturer hinted . . .	4	1	15	9
The instructor implied . . .	9	0	19	1
The murderer admitted . . .	10	3	16	0
The tutor hinted . . .	5	2	15	7
The jury decided . . .	10	0	10	8
The young man realized . . .	16	2	10	1
The pub owner admitted . . .	8	2	18	1

Note. Numbers correspond to actual numbers of responses in each category. S0, reduced-sentential-complement response; NP, object response; S-, unreduced-sentential-complement response.

TABLE 2

Reduced-Sentential-Complement and Object Completions for Sentence Stems Used in Experiments 1 and 2

Stem	Response	
	S0	NP
The young athlete realized her . . .	16	4
The doctor admitted the . . .	16	4
The teacher hinted the . . .	13	7
The bank admitted the . . .	13	7
The investor realized a . . .	19	1
The witness admitted the . . .	20	0
The judge decided the . . .	16	4
The slacker pretended the . . .	17	3
The butler admitted the . . .	14	6
The lecturer hinted the . . .	9	11
The instructor implied the . . .	13	7
The murderer admitted his . . .	18	2
The tutor hinted the . . .	19	1
The jury decided the . . .	18	2
The young man realized his . . .	16	4
The pub owner admitted the . . .	10	10

Note. Numbers correspond to actual numbers of responses in each category. S0, reduced-sentential-complement completion; NP, object completion.

ment continuations were ever employed. Table 2 shows that including the material immediately following the matrix verb and immediately preceding the head noun of the subject of the embedded sentence did not change participants' responses.

Plausibility norming. To assess the plausibility of the postverbal noun phrases as direct objects for the matrix verbs, 20 participants assigned ratings to sentences like (2):

(2a). The young athlete realized her potential.
(2b). The young athlete realized her exercises.

We instructed participants to assign a number from 0 to 7 that reflected how much sense the sentence made. We eliminated stimuli from our set of test items if the plausible version (e.g., 2a) received mean plausibility rating below 5 or if the implausible version (e.g., 2b) received mean plausibility rating above 2.

To ensure that the items in the plausible-object and implausible-object conditions were equally plausible on the correct sentential-complement analysis, a further group of 20 participants assigned plausibility ratings on the same 0-to-7 scale. Sentences with a plausible object analysis (e.g., 1a) received a mean rating of 6.0, whereas sentences with an implausible object analysis (e.g., 1b) received a mean rating of 5.8. This very small difference was significant by participants [$F1(1,17) = 4.41$, $p < .05$, $MS_e = .076$], though not by items [$F2(1,15) < 1$]. We return to this below.

Procedure. An SRI Dual-Purkinje Generation 5.5 eye-tracker monitored participants' eye movements. The tracker has angular resolution of 10° arc. The tracker monitored only the right eye's gaze location. A PC displayed items on a VDU 70 cm from participants' eyes. The VDU displayed four characters per degree of visual angle. The tracker monitored participants' gaze location every millisecond and the software sampled the tracker's output to establish the sequence of eye fixations and their start and finish times.

Before the experiment started, participants read an explanation of eye tracking and a set of instructions. The instructions told them to read at their normal rate and comprehend the texts as well as they could. The experimenter then seated the participant at the eye-tracker and used bite bars and forehead restraints to minimize head movements. Next, participants completed a calibration procedure. Before each trial, a small "+" symbol appeared near the upper-left-hand corner of the screen. Immediately after participants fixated the "+" symbol, the computer displayed an item, with the first character of the text replacing the "+" on the screen. The "+" symbol also served as an automatic calibration check, as the computer did not display the item until it detected stable fixation on the "+" symbol. If participants did not rapidly fixate the "+" symbol, the experimenter recalibrated the eye-tracker. When participants finished reading each item, they pressed a key, and the computer either displayed a comprehension question (e.g., *Did the young athlete realise something?*), on about half of the trials, balanced across conditions, or proceeded to the next trial. Half of these questions had "yes" answers, half had "no" answers. Participants responded to the questions by pressing a button and did not receive feedback on their answers. After participants completed each sixth of the experiment, the experimenter recalibrated the equipment. Thus, the eye-tracker was calibrated a minimum of six times during the experiment and often more.

The computer displayed each experimental list in a fixed random order together with 86 other sentences. Twenty-six of these made up Experiment 3, described below. The other 60 sentences were of varied grammatical type (e.g., *The young boy wished for a brand new bicycle on his eighth birthday*.). However, these sentences never included noun-phrase objects, so that any evidence for readers' adopting the object analysis could not be due to prior exposure to related sentences.

Regions. We report analyses on four critical regions. The *noun* region comprised the head noun of the embedded sentence subject [e.g., *potential* in (1a)]. The *postnoun* region comprised the material between the noun region and the verb region (e.g., *one day*). The *verb* region comprised the first verb of the embedded sentence (e.g., *might*). This verb could be a main or an auxiliary verb and was chosen because it constituted the first point at which it became clear that the object analysis could not be correct. The *postverb* region comprised the material from the end of the verb region up to the line break (e.g., *make her*). Both the noun and verb regions were always a single word. All regions included the character space immediately before the first word in the region.

Fixations. We first determined which line of text participants were reading. This involved some judgment because the fixation point could be between two lines of text. However, in the vast majority of cases, determining which line was fixated caused no difficulty whatsoever. An automatic procedure then pooled short contiguous fixations. The procedure incorporated fixations of less than 80 ms into larger fixations within one character space and then deleted fixations of less than 40 ms that fell within three character spaces of any other fixation. Following Rayner and Pollatsek (1989), we presume that readers do not extract much information during such brief fixations.

Measures. We report three standard measures of processing during reading. *First-pass time* sums all fixation times starting with the reader's first fixation inside the region until the reader's gaze leaves the region, either to the left or to the right. For regions consisting of a single word, this corresponds to Rayner and Duffy's (1986) gaze duration measure. *Total time* sums all fixation times within the region. *First-pass regressions* include all leftward eye movements that cross the region's left boundary and which immediately follow a first-pass fixation in the region. Hence there can only be one first-pass regression from a region per trial. First-pass regressions are reported as raw score means (e.g., if a participant encounters eight stimuli per condition, a mean of 2 = 25% of trials involved a first-pass regression). We also report two less standard measures of early processing (which are, nevertheless, related to measures first employed by Rayner and Duffy). The reason is that first-pass times are sometimes truncated by readers entering a region and rapidly regressing, which can occur when they encounter processing difficulty (Brysbaert & Mitchell, 1996; Hemforth, Konieczny, Scheepers, & Strube, 1994; Liversedge, Pickering, & Traxler, 1996; Traxler, Bybee, & Pickering, 1997; cf. Rayner & Pollatsek, 1989). Thus, two pieces of text, one that causes processing difficulty and one that does not, can have similar first-pass times. In the first case, readers rapidly exit to the left; in the second case, readers rapidly exit to the right. *Right-bounded time* is the sum of all fixations within a region before the eye fixates any region to the right of the region. *Regression-path time* includes all of the fixations within a region and all subsequent fixations on prior regions, until the eye crosses the region's right boundary.

Note that if the eye fixates a point beyond the end of a region before landing in the region for the first time, then the analysis software returns a value of 0 ms for first-pass time, first-pass regressions, right-bounded time, and regression-path time. All such 0 ms values were excluded from the analyses reported below. This

TABLE 3

Experiment 1 Means for First-Pass Regressions, First-Pass Time, Right-Bounded Time, Regression-Path Time, and Total Time Data by Region and Condition

		Region			
Measure	Object analysis	Noun	Postnoun	Verb	Postverb
First-pass regressions	Plausible	0.78	1.05	1.48	2.45
	Implausible	0.83	1.15	1.30	2.60
First-pass time (ms)	Plausible	255	301	279	353
	Implausible	261	323	270	377
Right-bounded time (ms)	Plausible	265	318	310	391
	Implausible	279	363	287	412
Regression-path time (ms)	Plausible	310	353	362	471
	Implausible	341	442	357	479
Total time (ms)	Plausible	367	443	363	471
	Implausible	430	514	339	506

criterion eliminated 11.2% of the data. Further, we removed trials on which the eye-tracker lost track of the participant's gaze location before analyzing the data. Specifically, we removed trials on which two consecutive regions received no first-pass fixation. This criterion eliminated a further 3.9% of the data. Analyses that treated such regions as having 0 ms reading times produced nearly identical results.

Results and Discussion

Table 3 presents mean first-pass regressions, first-pass time, right-bounded time, regression-path time, and total time for Experiment 1. Table 4 presents statistical analyses for these data.

We first subjected the data from Experiment 1 to separate 2 (Plausibility: plausible vs implausible object analysis) × 2 (Region: postnoun vs verb region) ANOVAS with participants and items as random factors (see the first block of Table 4). If participants constructed the object interpretation for the sentences in Experiment 1, then the manipulation of the plausibility of the object interpretation should have produced longer reading times (or more regressions) for sentences with implausible object analyses than sentences with plausible object analyses. Adopting a plausible object analysis should have caused participants to have greater difficulty processing disambiguating material, and so sentences with plausible object analyses should have produced longer reading times (or more regressions) following the point of syntactic disambiguation than sentences with implausible object analyses. This would constitute the crossover effect discussed above. The reliable interaction of plausibility in the right-bounded time, regression-path time, and total time measures demonstrates that participants did regularly construct the object analysis.

We next analyzed the data from each scoring region separately. Mean total time on the noun region was longer for sentences with implausible object analyses than for sentences with plausible object analyses (see the second block of Table 4), but the other measures revealed no reliable differences between conditions. The postnoun region produced reliable differences between condition means in the right-bounded, regression-path, and total-time measures (see the third block of Table 4). Thus, sentences with implausible object analyses caused participants greater difficulty than sentences with plausible object analyses before readers encountered syntactically disambiguating material. Analyses on the data from the verb and postverb regions produced reliable differences between sentences with plausible and implausible object analyses only in the verb region, and there only in the right-bounded time measure (see the fourth and fifth blocks of Table 4). This finding

TABLE 4

Experiment 1 ANOVAs for First-Pass Regressions, First-Pass Time, Right-Bounded Time, Regression-Path Time, and Total Time Data

	Within participants and items			
Source	*F*1		*F*2	
Plausibility × Region (postnoun vs verb regions)				
First-pass regressions	<1	(1.49)	<1	(2.92)
First-pass time	2.38	(6839)	5.07*	(1205)
Right-bounded time	7.41**	(7878)	16.46***	(1178)
Regression-path time	5.62*	(19770)	12.37**	(3181)
Total time	8.62**	(11297)	34.71***	(1134)
Plausibility (noun region)				
First-pass regressions	<1	(0.77)	<1	(1.62)
First-pass time	<1	(1449)	<1	(1768)
Right-bounded time	1.49	(2443)	<1	(2290)
Regression-path time	2.09	(8885)	1.81	(4159)
Total time	10.88**	(7277)	5.44*	(5825)
Plausibility (postnoun region)				
First-pass regressions	<1	(1.12)	<1	(3.26)
First-pass time	1.41	(6564)	2.47	(1423)
Right-bounded time	5.86*	(6909)	5.93*	(1782)
Regression-path time	9.51**	(16884)	13.40**	(3850)
Total time	7.66**	(13319)	11.45**	(3406)
Plausibility (verb region)				
First-pass regressions	<1	(1.33)	<1	(1.53)
First-pass time	1.86	(3805)	4.00	(658)
Right-bounded time	5.43*	(3638)	7.14*	(1242)
Regression-path time	<1	(10891)	<1	(3945)
Total time	3.17	(4696)	2.16	(3192)
Plausibility (postverb region)				
First-pass regressions	<1	(1.73)	<1	(2.13)
First-pass time	2.47	(1423)	<1	(4856)
Right-bounded time	1.10	(7708)	<1	(6668)
Regression-path time	<1	(20116)	<1	(14260)
Total time	2.13	(11161)	1.26	(5371)

Note. Values within parentheses represent mean-squared errors. Degrees of freedom are (1, 39) for participants and (1, 15) for items.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

suggests that the bulk of processing necessary for recovery from misanalysis was accomplished via a combination of refixations on previously encountered material and longer fixation on the verb itself. This is consistent with the general absence of effects in the other measures on the verb region.

Note that the very small difference in plausibility ratings of the experimental sentences on the sentential-complement analysis cannot have caused the experimental results. The difference between the two conditions emerged before participants had encountered the verb *might* (i.e., before they had reached the verb region). It is hard, therefore, to see how any plausibility effect based on the plausibility of the complete sentence could have caused the effects to emerge so early. Additionally, such an account could not explain the evidence for difficulty upon encountering the verb in the sentences with plausible versus implausible object analyses.

We conclude that participants regularly adopted the object analysis for the experimental materials. This finding is incompatible with both serial lexical guidance and serial likelihood. The pattern of results in Experiment 1 was consistent, however, with previous findings for the same construction using a similar plausibility manipulation but with experimental sentences that were generally biased toward the sentential-complement analysis (Pickering & Traxler, 1998). As in Pickering and Traxler, we found evidence for the crossover pattern: more difficulty for sentences with implausible object analyses before disambiguation, and some evidence for more difficulty for sentences with plausible object analyses after disambiguation.

EXPERIMENT 2

Experiment 2 tested whether the results of Experiment 1 replicated when participants read passages rather than isolated sentences. Our prediction was that the same pattern of results should occur and therefore that participants would not follow serial lexical guidance or serial likelihood in extended texts any more than in isolated sentences.

Method

Participants. Twenty new participants from the same population as Experiment 1 were paid to participate in the eye-tracking phase of the study.

Items. The stimuli were the test sentences from Experiment 1 embedded within short passages. The target sentences were preceded by a title and a single sentence and followed by a final sentence (see Appendix).

> THE YOUNG SPRINTER
> The girl did a series of strenuous exercises every day before she reached her full potential. The young athlete realized her potential/exercises one day might make her a world-class sprinter. She dreamt about winning a gold medal in the Olympic Games.

Both versions of the noun that was manipulated between conditions (i.e., *potential* and *exercises*) appeared in the first sentence of the passage, in different orders for different items. The sentence with the plausible object analysis is given before the slash; the noun used in the sentence with the implausible object analysis is given after the slash. In other respects, the method was identical to Experiment 1.

Procedure. Participants were presented with 48 paragraphs in all, 16 experimental paragraphs from the current experiment and 32 paragraphs from an unrelated experiment investigating instrument inferences. In other respects, the procedure was identical to Experiment 1. We eliminated 8.0% of the data due to participants skipping adjacent regions. Dropping data from any region that was not fixated eliminated a further 13.5% of the data. As in Experiment 1, analyses including these data produced nearly identical results.

Results and Discussion

Table 5 presents mean first-pass regressions, first-pass time, right-bounded time, regression-path time, and total time for Experiment 2. Table 6 presents statistical analyses for these data.

As in Experiment 1, we subjected the data to separate 2 (Plausibility: plausible vs implausible object analysis) × 2 (Region: postnoun vs verb region) ANOVAs (see the first block of Table 6). If participants constructed the object interpretation for the sentences in Experiment 1, then the manipulation of the plausibility of the object interpretation should have produced longer reading times (or more regressions) for sentences with implausible object analyses than sentences with plausible object analyses. Adopting a plausible object analysis should cause participants to have greater difficulty processing disambiguating material, and so sentences with plausible object analyses should have produced longer reading times or more regressions following the point of syntactic disambiguation than sentences with implausible object analyses. The reliable interaction of plausibility in the total-time measure demonstrates

TABLE 5

Experiment 2 Means for First-Pass Regressions, First-Pass Time, Right-Bounded Time, Regression-Path Time, and Total Time Data by Region and Condition

		Region			
Measure	Object analysis	Noun	Postnoun	Verb	Postverb
First-pass regressions	Plausible	1.65	0.75	0.90	2.15
	Implausible	1.30	0.65	1.00	1.75
First-pass time (ms)	Plausible	252	306	241	354
	Implausible	253	346	245	353
Right-bounded time (ms)	Plausible	269	347	263	434
	Implausible	263	391	264	418
Regression-path time (ms)	Plausible	339	393	303	583
	Implausible	332	453	319	563
Total time (ms)	Plausible	333	472	369	515
	Implausible	365	527	338	501

that participants regularly constructed the object analysis.

We next analyzed the data from each scoring region separately. The postnoun region produced reliable differences between sentences with plausible and implausible object analyses in the first-pass and right-bounded time measures (see the third block of Table 6). The participants analysis of the regression-path time measure also produced a statistically significant difference, but the items analysis did not.

In this experiment, the strongest evidence that readers adopted the object analysis comes from elevated reading times for sentences with implausible object analyses in the postnoun region in the first-pass and right-bounded time measures. Although an interaction of plausibility and region indicated that the magnitude and direction of the plausibility effect in the postnoun and verb regions differed, the verb region in isolation did not produce reliable plausibility effects.

In conclusion, Experiment 2 provides additional support for the conclusions of Experiment 1 and allows the conclusions to be generalized to the reading of more naturalistic texts.

EXPERIMENT 3

Experiments 1 and 2 showed that people do not follow either the serial lexical-guidance or the serial-likelihood account in reading complement-clause ambiguities. In order to determine whether these conclusions held more generally, Experiment 3 investigated the processing of another type of sentence using the same method. We employed subordinate-clause ambiguities like (3):

(3a). While the pilot was flying the plane that had arrived stood over by the fence.
(3b). While the pilot was flying the horse that had arrived stood over by the fence.

In these sentences, the ambiguity concerns whether *the plane* in (3a) and *the horse* in (3b) is the object of flying or the subject of *stood.* The latter analysis turns out to be correct, and *was flying* turns out to be intransitive. However, experimental evidence shows that people often encounter difficulty processing such sentences, which suggests that they often initially consider the possibility that *the plane* or *the horse* is the object of *was flying* (Clifton, 1993; Ferreira & Henderson, 1991; Frazier, 1979; Frazier & Rayner, 1982; Mitchell, 1987; Pickering & Traxler, 1998; Warner & Glass, 1987).

Our question is whether readers still consider this possibility when the verb in the subordinate clause (here, *was flying*) is more likely to be intransitive than transitive and when the whole context makes the same prediction. According to either frequency-based account, readers should ignore this object analysis and should

TABLE 6

Experiment 2 ANOVAs for First-Pass Regressions, First-Pass Time, Right-Bounded Time, Regression-Path Time, and Total Time Data

Source	Within participants and items			
	*F*1		*F*2	
Plausibility × Region (postnoun vs verb)				
First-pass regressions	<1	(0.83)	<1	(0.92)
First-pass time	3.69	(1805)	3.66	(1697)
Right-bounded time	3.12	(2910)	3.45	(2449)
Regression-path time	<1	(10135)	1.25	(8286)
Total time	6.95*	(5361)	7.76*	(3947)
Plausibility (noun region)				
First-pass regressions	1.15	(1.07)	1.09	(0.72)
First-pass time	<1	(1344)	<1	(892)
Right-bounded time	<1	(2046)	<1	(1208)
Regression-path time	<1	(5133)	<1	(2012)
Total time	1.42	(7242)	1.48	(4164)
Plausibility (postnoun region)				
First-pass regressions	<1	(0.57)	<1	(0.93)
First-pass time	6.21*	(2667)	4.54*	(3310)
Right-bounded time	5.76*	(3277)	4.49*	(3694)
Regression-path time	4.91*	(7247)	3.38	(9280)
Total time	2.25	(13426)	2.32	(10009)
Plausibility (verb region)				
First-pass regressions	<1	(0.63)	<1	(0.79)
First-pass time	<1	(691)	<1	(406)
Right-bounded time	<1	(2059)	<1	(632)
Regression-path time	<1	(8914)	<1	(3246)
Total time	1.90	(5161)	1.95	(4628)
Plausibility (postverb region)				
First-pass regressions	1.57	(0.83)	<1	(0.92)
First-pass time	<1	(4977)	<1	(3660)
Right-bounded time	<1	(6625)	<1	(3557)
Regression-path time	<1	(21365)	<1	(8286)
Total time	<1	(14048)	<1	(8812)

Note. Values within parentheses represent mean-squared errors. Degrees of freedom are (1, 19) for participants and (1, 15) for items.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

immediately assume that *the plane* or *the horse* is the subject of the main clause. The processor should therefore not be affected by the implausibility of flying a horse. But if readers do not follow either of these accounts and treat *the plane* and *the horse* as the object of *was flying*, then we predict a crossover pattern: difficulty with (3b) compared with (3a) during the ambiguous region (more specifically, *plane/horse that had arrived*), and difficulty with (3a) compared with (3b) after disambiguation (i.e., from *stood* onward).

Method

Participants. Twenty new participants from the same population as Experiment 1 were paid

TABLE 7

Intransitive and Noun-Phrase Object Completions for Sentences from Experiment 3

Verb	Response: Intransitive	Response: NP
Completions for verbs presented in isolation		
chanted	10	10
cheered	15	5
lectured	14	6
marched	18	2
preached	14	6
sailed	14	6
sang	17	3
smoked	10	10
walked	12	8
was flying	17	3
was running	14	6
was swimming	18	2

Stem	Response: Intransitive	Response: NP	Response: Other
Completions for verbs presented with their stems			
As the monk chanted . . .	14	0	0
While the supporters cheered . . .	10	3	1
While the pilot was flying . . .	12	2	0
As the professor lectured . . .	14	0	0
As the sergeant marched . . .	13	1	0
As the minister preached . . .	12	2	0
As the old man sailed . . .	10	4	0
As the choir sang . . .	12	1	1
When the man smoked . . .	10	3	1
While the boy was swimming . . .	14	0	0
While the man walked . . .	11	3	0
As the pickets chanted . . .	14	0	0
Because the boss lectured . . .	14	0	0
When the kidnapper marched . . .	14	0	0
While the captain sailed . . .	10	4	0

to participate in the eye-tracking phase of the study.

Items. The items consisted of 26 sentences like (3) above (see Appendix). Both (3a) and (3b) are plausible, but (3a) has a plausible object analysis, whereas (3b) has an implausible object analysis.

Verb norming. We wanted to select intransitive-preference verbs. We completed a two-stage selection process as before. Table 7 presents the results of the two norming tasks. In the first norming phase, participants wrote sentences for 107 verbs. Fifteen were the sentential-complement verbs tested for Experiments 1 and 2. The rest were of varying types. This task uncovered 12 verbs that produced intransitive responses as often as or more often than object continuations. We then wrote stimuli containing these 12 verbs. We presented the beginnings of the stimuli, up to and including the verb and asked another group of 14 participants to complete each stem. We selected 26 stems that produced a ratio of intransitive to object completions of 2:1 or greater.

Plausibility norming. We constructed two versions of each target sentence. In one version, the subject of the complement clause made a plausible object of the preceding verb [e.g., (3a)]. In the other version, the subject of the complement clause made an implausible object of the preceding verb [e.g., (3b)]. In order to determine this, raters assigned numbers between 0 and 7 to sentences like (4) according to how much sense they made.

(4a). The pilot was flying the plane.
(4b). The pilot was flying the horse.

We eliminated items as in Experiment 1.

To ensure that there were no plausibility differences on the correct reading, a further group of 29 participants assigned plausibility ratings to the complete experimental sentences [i.e., to (3a) and (3b)]. Items in the plausible condition received mean ratings of 5.3, whereas items in the implausible condition received mean ratings of 4.5. This fairly small difference produced a significant value in the participants analysis [$F1(1,28) = 72.04$, $p < .0001$, $MS_e = 0.245$] that did not quite attain significance in the items analysis [$F2(1,25) = 3.14$, $p < .10$, $MS_e = 2.78$]. We return to this below. In other respects, the method was identical to Experiment 1.

Procedure. Experiment 3 was run with Experiment 1, and hence the procedure was identical. We eliminated 4.4% of the data due to participants skipping adjacent regions. Dropping data from any region that was not fixated on the first pass eliminated a further 12% of the

TABLE 8

Experiment 3 Means for First-Pass Regressions, First-Pass Time, Right-Bounded Time, Regression-Path Time, and Total Time Data by Region and Condition

		Region			
Measure	Object analysis	Noun	Postnoun	Verb	Postverb
First-pass regressions	Plausible	1.50	1.13	3.70	4.30
	Implausible	1.58	2.60	3.00	3.90
First-pass time (ms)	Plausible	291	597	347	306
	Implausible	305	606	326	275
Right-bounded time (ms)	Plausible	307	651	404	374
	Implausible	322	748	378	302
Regression-path time (ms)	Plausible	362	696	530	604
	Implausible	394	873	485	410
Total time (ms)	Plausible	418	1041	541	441
	Implausible	467	1054	489	358

data. An in the other experiments, analyses including these data produced nearly identical results.

Results and Discussion

Table 8 presents mean first-pass regressions, first-pass time, right-bounded time, regression-path time, and total time for Experiment 3. Table 9 presents statistical analyses for these data.

We again subjected the data to separate 2 (Plausibility: plausible vs implausible object analysis) × 2 (Region: postnoun vs verb region) ANOVAs (see the first block of Table 9). If participants constructed the object analysis for the sentences in Experiment 3, then the manipulation of the plausibility of the object interpretation should have produced longer reading times or more regressions for sentences with implausible object analyses than sentences with plausible object analyses. Adopting a plausible object analysis should have caused participants to have greater difficulty processing disambiguating material, and so sentences with plausible object analyses should have produced longer reading times or more regressions following the point of syntactic disambiguation than sentences with implausible object analyses. Hence, we predicted a "crossover" between regions before and after disambiguation, as discussed above. The reliable interaction of plausibility in the first-pass regressions, right-bounded time, regression-path time, and total time measures demonstrates that participants regularly constructed the object analysis.

We next analyzed the data from each scoring region separately. Mean regression-path time and total time on the noun region was longer for sentences with implausible object analyses than for sentences with plausible object analyses (see the second block of Table 9). Analyses of the right-bounded measure produced a statistically reliable difference in the items analysis and a nearly reliable difference in the participants analysis. Analyses of the postnoun region revealed statistically reliable differences in three of the four measures of early processing (see the third block of Table 9). The first-pass and total time measures did not reveal any differences between sentences with plausible and implausible object analyses. The plausibility effect observed in the noun and postnoun regions reversed itself in the verb and postverb regions (see the fourth and fifth blocks of Table 9). Analyses of the data from the verb region produced reliable differences in the first-pass and total-time measures, with an additional reliable difference in the items analysis of the right-bounded measure. The postverb region produced reliable differences on the right-bounded time, regression-path time, and total-time measures. Analyses of the first-pass time data re-

TABLE 9

Experiment 3 ANOVAs for First-Pass Regressions, First-Past Time, Right-Bounded Time, Regression-Path Time, and Total Time Data

Source	Within participants and items			
	*F*1		*F*2	
Plausibility × Region (postnoun vs verb)				
First-pass regressions	16.12***	(3.07)	13.76***	(6.04)
First-pass time	1.43	(6643)	3.72	(1982)
Right-bounded time	22.93***	(6646)	26.21***	(4052)
Regression-path time	24.40***	(20252)	15.94***	(19860)
Total time	4.52*	(9408)	6.58*	(6281)
Plausibility (noun region)				
First-pass regressions	<1	(1.24)	<1	(1.88)
First-pass time	2.41	(1663)	3.59	(1181)
Right-bounded time	3.09	(1645)	4.36*	(1382)
Regression-path time	4.71*	(4368)	5.64*	(3105)
Total time	6.90**	(6998)	4.22*	(10184)
Plausibility (postnoun region)				
First-pass regressions	17.68***	(3.38)	26.75***	(2.76)
First-pass time	<1	(9738)	<1	(2638)
Right-bounded time	36.83***	(5163)	16.23***	(7922)
Regression-path time	65.87***	(9564)	14.14***	(30495)
Total time	<1	(25077)	<1	(26892)
Plausibility (verb region)				
First-pass regressions	3.38	(3.33)	3.21	(5.76)
First-pass time	4.21*	(2202)	8.01*	(1058)
Right-bounded time	3.54	(3797)	5.82*	(1796)
Regression-path time	1.87	(21706)	2.19	(8833)
Total time	8.97**	(6099)	4.25*	(6676)
Plausibility (postverb region)				
First-pass regressions	1.14	(2.14)	<1	(3.45)
First-pass time	3.53	(5338)	7.98**	(2064)
Right-bounded time	9.66**	(10846)	10.30**	(5418)
Regression-path time	10.32**	(73437)	10.78**	(32230)
Total time	9.69**	(14384)	12.69**	(8103)

Note. Values within parentheses represent mean-squared errors. Degrees of freedom are (1, 39) for participants and (1, 25) for items.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

vealed a reliable effect in the items analysis, but not in the participants analysis.

Just as in Experiment 1, the small difference in plausibility ratings of the sentences on the intransitive analysis cannot have caused the experimental results. The clearest reason for concluding that is that the sentences with plausible object analyses, which also had slightly more plausible intransitive analyses than the sentences with implausible object analyses, were harder to process than the sentences with implausible object analyses after disambiguation. This effect is predicted if participants initially adopted the object analysis (Pickering & Traxler, 1998), but cannot be explained if participants initially adopted the intransitive analysis.

An additional reason for concluding that the small plausibility difference cannot have been the source of the experimental results is that the difference between the two conditions emerged before participants had reached the main verb (e.g., *stood*). We conclude that participants initially adopted the object analysis in Experiment 3. This is again inconsistent with the serial lexical-guidance and serial-likelihood accounts.

GENERAL DISCUSSION

Experiments 1–3 tested the predictions of two serial frequency-based models: a lexical-guidance account in which the parser initially adopts the analysis compatible with the most frequent subcategorization frame of the verb (Ford et al., 1982), and a more general serial-likelihood account in which the parser adopts the analysis that readers would produce as their sentence completion given prior context. Neither of these accounts is compatible with the results. Experiments 1 and 2 showed that participants adopted the object analysis in complement-clause ambiguities, even when both prior context and subcategorization frequencies supported the sentential-complement analysis. Experiment 3 showed that they adopted the transitive analysis of subordinate-clause ambiguities, even when both prior context and subcategorization frequencies supported the intransitive analysis.

In more detail, we constructed sentences that were locally ambiguous: A noun phrase could either serve as the object of a preceding verb or as the subject of a following complement clause. Pretests demonstrated that the object analysis was the less frequent analysis for the verb, and the less frequent continuation, given the prior context. Pretests also demonstrated that the two versions of the experimental items differed in plausibility on the less frequent object analysis. The eye-tracking experiments showed that participants were affected by the plausibility manipulation. Most importantly, participants found processing of the ambiguous region more difficult if the object analysis was implausible than if it was plausible. Additionally, the results demonstrated that the conditions exhibited a crossover between regions before and regions after the point of disambiguation (Pickering & Traxler, 1998). Experiments 1 and 3 showed effects with two different sentence types; Experiment 2 showed effects when the sentences from Experiment 1 were presented in discourse context. Hence the results indicate that readers adopted the object analysis, even though both frequency-based accounts predicted that they would have adopted the alternative analysis.

Contrast with Garnsey et al. (1997)

Our results appear to contradict those of Garnsey et al. (1997). The crucial difference is that we find plausibility effects with sentential-complement-biased verbs (i.e., C-biased verbs), both during the critical ambiguous region and after disambiguation, and they do not. We identify three possible reasons for this. First, our experiments had much greater power. We employed 8 items per cell (13 per cell in Experiment 3, using a second sentence type) rather than 4 per cell; and Garnsey et al.'s results depended on interactions between experimental sentences and control sentences with the complementizer *that*, whereas our did not. Second, their ambiguous region was shorter than ours. Our effects are clearest in the postnoun region, which they did not have. In our laboratory, we have repeatedly found plausibility effects that emerge immediately after a critical head noun (e.g., Pickering & Traxler, 1998; Traxler et al., 1998).

Finally, Garnsey et al.'s (1997) conclusions depend to a very large extent on the comparison between the C-biased and O-biased verbs. Essentially, they find some evidence for early plausibility effects with the O-biased verbs, but none with the C-biased verbs. But there is reason to believe that effects with C-biased verbs would be smaller than effects with O-biased verbs (e.g., Mitchell & Holmes, 1985; Trueswell et al., 1993), perhaps reflecting differences in difficulty of reanalysis. Hence, they are faced with the problem of distinguishing a smaller effect from no effect. No direct comparisons are made between C- and O-biased verbs. Moreover, their evidence for plausibility effects with O-biased verbs is weak. The critical interaction only approached significance weakly on the first pass ($.05 < p < .1$ by participants; $p > .1$ by items) and was nonsignificant on total time. In the self-paced-reading replication, the effect actually *reversed* (though nonsignificantly), and there was a marginal reversed effect with the third set of verbs, the equibiased verbs. At disambiguation, there was a marginal interaction for the O-biased verbs between plausibility and complementizer presence on total time only. In general, the effects of plausibility were weak with the O-biased verbs. Given their short ambiguous region and small number of items, it is not surprising that Garnsey et al. (1997) did not detect plausibility effects for C-biased verbs.

In conclusion, we were looking for a very specific effect; whereas they searched for a much more complex pattern of effects, which resulted from the manipulation of type of verb, presence versus absence of complementizer, and reading method. It is therefore not surprising that we found an effect that their technique failed to discover; whereas, in contrast, their experiments provided very important evidence about other aspects of processing. Hence we can be convinced that readers do not always adopt the most likely analysis at a point of local ambiguity. We should not, however, that Garnsey et al. (1997) also reject a serial lexical-guidance account. They found that some of the effects in their C sentences correlated with the difference in the plausibility of the object analysis between test sentences. This suggests that readers evaluated the object analysis when the verb preferred a sentential complement, *contra* the serial lexical-guidance account.

Implications of These Results for Other Models

The results rule out two possible kinds of serial, frequency-based accounts: serial lexical guidance and serial likelihood. They might, in theory, be compatible with other serial frequency-based accounts that make their decisions at some different fineness of grain from either of the accounts that we have tested. In the absence of any clear accounts, we can only claim that it is unlikely that such accounts would make different predictions from both of the accounts we have tested. Hence, we can rule out Ford et al.'s (1982) lexical guidance account and suggest that the Linguistic Tuning account (Mitchell et al., 1995) is likely to be incompatible with our data. In what follows, we discuss other kinds of models. Two of these classes are currently influential; the third is a new alternative.

Parallel, frequency-based accounts. By a parallel frequency-based account, we mean any account in which different analyses can be considered concurrently and are ranked on the basis of their frequency. The most explicit of these is Jurafsky's (1996) model, which estimates likelihood using a combination of top-down (grammatical) and bottom-up (lexical preference) information and considers a bounded number of ranked analyses in parallel. While it is difficult to make specific predictions for his model with respect to our experimental materials, we presume it would initially rank the sentential-complement or the intransitive analysis higher than the object analysis and should thus predict no processing delays as a result of plausibility for the object reading. The effects we observe are therefore not directly predicted or explained by the parallel likelihood parsing mechanism he proposes.

However, the best known parallel accounts are constraint-based accounts, in which parsing preferences are determined by the simultaneous interaction of multiple constraints (Garnsey et al., 1997; MacDonald, 1994; MacDonald et al., 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Trueswell, 1996; Trueswell et al., 1993, 1994; cf. Taraban & McClelland, 1988; Tyler & Marslen-Wilson, 1977). These constraints relate to any properties of the encountered sentence that may influence its continuation, including subcategorization preferences, other syntactic cues, the meaning of the fragment, the nature of the discourse context, and prosody or punctuation. The relative activations of different analyses (i.e., their rankings) change dynamically as new constraints become available.

Because the parser pays attention to multiple constraints, such accounts have more in common with fine-grained than coarse-grained accounts. This is evidenced by the fact that data about these constraints is typically estimated from corpus counts or from production tasks where participants complete sentences from the point of ambiguity onward. For instance, a few participants might complete *The young athlete realized* . . . using an object (e.g., *her potential*), but more using a sentential complement (e.g., *her potential was enormous*); and most might complete *The young athlete realized her* using a sentential complement (e.g., *her potential was enormous*). From this, such models would assume that the parser would prefer the sentential-complement analysis to the object analysis (e.g., Garnsey et al., 1997; Trueswell et al., 1993) and would foreground it by *her* at the latest.

There are two ways in which our experiments do not directly test constraint-based accounts. First, it is not clear that we employ the correct fineness-of-grain to test the models. Clearly, constraint-based accounts are more general than the serial lexical-guidance account in that subcategorization preferences are not the only factor used to determine preferences. They do appear to have more in common with the serial-likelihood account, if only because proponents of such theories have employed sentence-completion data to determine their predictions.

However, some constraint-based models have been given particular connectionist implementations which are not entirely compatible with the assumption that rankings should be determined by completion preferences. Thus, Juliano and Tanenhaus (1994) assumed a network consisting of an input layer representing lexical items, a hidden unit layer, and a layer representing syntactic categories. As Mitchell et al. (1995) pointed out, lexical information must be stored in the set of weights between the input and the hidden layers. During training, every input leads to adjustment of all other weights via back propagation. Thus, the weights between the hidden units and the syntactic category units will necessarily reflect frequencies of broader categories than the individual lexical items—so-called neighborhood effects. The consequence is that subcategorization preferences will be affected by subcategorization frequencies for classes of verbs, not just individual verbs. For instance, the processor might note that more verbs are transitive than intransitive, say, and prefer a transitive analysis for a low-frequency intransitive-preference verb.

Given, however, that both of our pretests (construction of a sentence containing a verb like *realized* and completion of a fragment like *The athlete realized* . . .) make the same prediction about the favored analysis, it is highly unlikely that any current constraint-based model would predict the opposite. It is therefore reasonable to conclude that the object analysis should be disfavored in both constructions considered in this article.

The second way in which our experiments do not directly test constraint-based accounts is that we tested the adequacy of serial frequency-based accounts, not parallel ones. Parallel models to predict the activation of the object analysis. Assuming the above discussion, they predict that the sentential-complement analysis (or intransitive analysis in Experiment 3) would be activated most strongly (i.e., foregrounded) and the object analysis would be less strongly activated (i.e., backgrounded).

Hence, constraint-based accounts would explain the plausibility effects of the object analysis that occur in all three experiments as stemming from the backgrounded analysis. The later effects (after syntactic disambiguation) are explained by claiming that the object analysis becomes more strongly activated if it is plausible, and hence reanalysis becomes more difficult; whereas it becomes less strongly activated if it is implausible, and hence reanalysis becomes easier. Notice that this account assumes that backgrounded analyses are semantically interpreted with respect to general knowledge. This is a much stronger claim than the claim that backgrounded analyses are computed.

This claim is probably impossible to disprove in general (a more powerful parallel account can always capture the same data as a more parsimonious serial account). However, it is far from clear that it is compatible with the assumptions of current constraint-based models. The reason is that different analyses compete against each other for activation, and more competition occurs if the two analyses reach a similar level of activation than if one is much more activated than the other. According to Spivey-Knowlton and Sedivy (1995), "processing delays are a manifestation of direct competition between opposing alternatives . . . . Near equal activation levels of the two alternatives will result in lengthy competition, hence greatly slowed reading time at that point of ambiguity" (p. 260). The sentential-complement (or intransitive) analysis receives most activation on the basis of its frequency. If the object analysis is plausible, then it will receive considerable activation from this source. Hence, the two analyses should complete, resulting in processing difficulty. But if the object analysis is implausible, neither frequency nor plausibility will provide it with activation. Thus, only the sentential-complement (or intransitive) analysis should be strongly activated, and no competition should ensue. The striking conclusion is that current constraint-based models predict the opposite pattern during processing of the ambiguous region from that observed in the experiments reported here.

In other words, a parallel account in which activation levels are determined by frequency of analyses or probability of completions might be compatible with our data. But we do not believe that current constraint-based accounts could be compatible, because of their reliance on competition as the mechanism of syntactic disambiguation. In contrast, parallel likelihood models such as Jurafsky (1996) rank analyses according to their prior probability. Thus the sentential-complement (or intransitive) analysis will be ranked higher than the object analysis for the items in our experiments. In the case where the noun phrase following the verb is implausible as a direct object, such a model should therefore predict no processing difficulty during reading (and no crossover pattern), since the object analysis is disfavored and its implausibility means there is no reason to change this ranking. Thus the object analysis will never be foregrounded. This is inconsistent with the findings we present here, where the implausibility of the object does lead to an increased complexity, suggesting that the object analysis is initially foregrounded.

Restricted serial accounts. Our results are compatible with a number of *restricted* serial accounts in which initial decisions are based on some information sources alone. Most current accounts use grammar-based principles to make initial parsing decisions (e.g., Abney, 1989; Crocker, 1996; Ferreira & Henderson, 1990; Frazier, 1979, 1987; Gorrell, 1995; Kimball, 1973; Mitchell, 1987, 1989; Pritchett, 1992; Sturt & Crocker, 1996). In all of these models, the parser initially adopts the object analysis rather than the sentential-complement analysis in Experiments 1 and 2, and the object analysis rather than the intransitive analysis in Experiment 3.

We illustrate this with respect to the "Garden Path" model, proposed by Frazier (1979, 1987). After *The young athlete realized her*, the parser follows the principle of minimal attachment and adopts the object analysis, essentially because it requires the postulation of fewer nodes in a phrase structure tree than the sentential-complement analysis. Hence, it predicts effects of plausibility after *potential* versus *exercises.* After *While the pilot was flying the,* the parser follows the principle of late closure and adopts the object analysis, because it prefers to incorporate the new noun phrase into the subordinate clause rather than begin a new clause. Hence, it predicts effects of plausibility after *plane* versus *horse.* In both cases, the object analysis is infrequent; but the parser's initial decision pays no attention to frequency (indeed, it ignores subcategorization preferences entirely). Other restricted, serial accounts predict that the parser initially adopts the object analysis in both types of sentence and are therefore also compatible with our findings.

Frazier (1979) argued that the parser obeys minimal attachment because the analysis compatible with minimal attachment requires fewer computational steps than alternatives. The different analyses compete in a race, and the minimal attachment analysis wins. This account suggests that minimal attachment exists because of the limitations of the parser; it cannot access other sources of information during initial processing, even if those sources of information in fact meant that the parser obtained an analysis that would prove more useful ultimately.

Informativity. Our data suggest that people do not base their initial parsing decisions on frequency-based estimates of likelihood for an analysis alone. The parser does not simply select the analysis compatible with the most frequent subcategorization frame of the verb; nor does it select the analysis that the reader would produce on the basis of all available contextual information.

The data are, as noted above, consistent with some restricted serial accounts. However, the findings of Trueswell et al. (1993) and Garnsey et al. (1997) indicate that frequency information is employed during early stages of processing. Our results point against strict likelihood models, but they do not rule out other kinds of probabilistic accounts. One way of reconciling our results with those of Trueswell et al. and Garnsey et al. is to claim that frequency plays a role during initial processing, but its impact is not sufficiently strong that the processor simply always favors the most frequent analysis. Indeed, a serial account that uses frequency as the dominant decision mechanism is not supported by our findings.

We now sketch an alternative account in which the parser does pay attention to frequency information but is not exclusively guided by it. Instead, we suggest that the parser may be guided by a principle that we call *informativity* (Chater, Crocker, & Pickering, 1998; Crocker, Pickering, & Chater, 1999).

Let us assume that the parser is designed to operate as accurately and effectively as possible so as to allow the reader to reach the correct analysis for the sentence as quickly as possible. This is of course only an assumption, but similar assumptions have been exploited in various areas of cognitive psychology and beyond (e.g., Anderson, 1991; Legge, Klitz, & Tjan, 1997; Oaksford & Chater, 1994; cf. Marr, 1982). If so, it would seem strange to ignore potentially useful information, such as frequency. Indeed, it might appear sensible to be guided entirely by frequency in order to reduce the need for reanalysis to a minimum.

But in fact there are good reasons to assume that such a parser might allow other factors to affect its choice of analysis. The vast majority of local ambiguities cause little processing difficulty (e.g., the complement-clause ambiguity: see Pritchett, 1992; Sturt, Pickering, & Crocker, 1999). In many cases, the processor must initially favor the wrong analysis, without causing any marked difficulty. So the parser need not simply be guided by a desire to reduce reanalysis as much as possible.

In contrast, some reanalyses are clearly harder than others (Sturt et al., 1999). One aspect of this is that reanalysis is in general likely to be easier if the parser has been misled for a short than a long time. Experimental evidence in general supports this, though length is not the only determinant of difficulty (Ferreira & Henderson, 1991; Frazier & Rayner, 1982; Sturt et al., 1999; Warner & Glass, 1987). This suggests that an efficient parser need not always adopt the most likely analysis immediately: It may be better to favor a somewhat less likely analysis that can be abandoned quickly and straightforwardly over a somewhat more likely analysis that cannot be abandoned without great processing difficulty.

On this basis, we propose that it may be preferable to favor what we call a *testable* analysis. Informally, a testable analysis is one for which the parser is likely to receive good evidence quickly (e.g., on the next constituent) about whether the analysis is likely to be correct or not. Consider (1a) and (1b):

(1a). The young athlete realized her potential one day might make her a world-class sprinter.
(1b). The young athlete realized her exercises one day might make her a world-class sprinter.

If the parser adopts the object analysis after *The young athlete realized* or *The young athlete realized her,* it is likely to encounter a noun that provides good evidence about whether it was correct or not. If it encounters *potential* [as in (3a)], *dreams,* or *ambitions,* then the object analysis is very likely (e.g., because *realized her potential* would have occurred frequently on the object analysis). In this case, the parser will retain the object analysis. If it encounters *exercises* [as in (3b)], *house,* or in fact almost any other word, then the object analysis is effectively impossible. In this case, the parser is able to abandon the object analysis and adopt the sentential-complement analysis with confidence. In either case, the parser immediately encounters fairly reliable information about the object analysis. Hence the object analysis is highly testable.

But if the parser adopts the sentential-complement analysis, then the noun that it encounters will not provide good evidence about whether the analysis is correct or not. Almost any noun is possible after *The young athlete realized her* on the sentential-complement analysis. So it would not know whether to retain the analysis or abandon it after encountering *potential* or *exercises*, and it would have no reason to treat *potential* and *exercises* differently. Hence this analysis is not highly testable. The parser would presumably retain the sentential-complement analysis, and would be right more often than not, because this analysis is most likely. But if the object analysis were correct, it would have no evidence for this until eventual syntactic disambiguation, and so, by choosing the sentential-complement analysis, it would be led a long way up the garden path. Similar arguments hold for the intransitive analysis in Experiment 3.

For other ambiguities, the advantages of choosing a testable analysis may be more substantial. Consider the following hypothetical case. The parser has to choose between Analysis A, correct 40% of the time, and Analysis B, correct 60% of the time. Analysis A is completely testable, so that the parser knows by the next noun (say) whether Analysis A is correct or not. Analysis B is untestable, so that no further information becomes available to indicate whether the right choice has been made or not (as Analysis A is not computed and interpreted). If the parser selects Analysis B, it will be right 60% of the time and will never reanalyze. If it selects Analysis A, it will be right 100% of the time: 40% of the time immediately, 60% of the time after rapid revision. Assuming reanalysis is straightfoward in this case, adopting Analysis A initially is clearly the more efficient strategy.

We do not propose that the parser should ignore likelihood entirely. Instead, we hypothesize that it pays attention to both likelihood and testability. On this assumption, the favored analysis is the most *informative* analysis: one that is both fairly likely and fairly testable. We use the term "informative" because the most informative analysis denotes the analysis that the following material will provide most evidence or information about whether it is likely to be correct.[1] We therefore argue that the sentence processor will have a preference to build testable analyses over nontestable ones, except where the testable analysis is highly unlikely. The result will be a greater number of easy misanalyses (induced by less probable but more testable analyses) and a smaller number of difficult misanalyses (induced by more probable but less testable analyses). The result will be that the ultimately correct analysis will usually be obtained quickly, either initially or after rapid reanalysis. However, it would not be optimal to favor testable analyses without any regard to frequency, and some balance between the two influences should be struck.

We refer to the principle which balances likelihood with testability as *the principle of informativity:*

5. For all s_i in S, select s_i where $I(s_i)$ is maximized, where $I(s_i) = f(P(s_i), T(s_i))$.

In other words, at a point of local ambiguity, the processor should favor the syntactic analysis (s_i) which maximizes its informativity (I). This is a function of the prior probability (P) of an analysis and its testability (T). Crocker et al. (1999) present a particular characterization of this function (for background, see Chater et al., 1998). The prior probability of an analysis is simply its likelihood at that point in the sentence. Crocker et al. argue that testability be instantiated by a measure called *specificity* (S), which reflects the extent to which a particular analysis constrains (i.e., predicts) what words

[1] The term *informative* should not be confused with the notion of information as developed in Information Theory (Shannon, 1948). However, to the extent that informativity characterizes the constraints that a (partial) analysis places on its continuation, it can be seen as a measure of association not dissimilar from the notion of mutual information.

will be encountered next and can be estimated from a corpus. The fewer words that are permitted, the more specific the analysis, and hence the more informative the words (i.e., the evidence) will be in evaluating whether the foregrounded analysis is correct. They then present a rational analysis, which derives the following function for I:

6. $I(s_i) = P(s_i) \times S(s_i)$.

In other words, they claim that specificity and prior probability should be multiplied to achieve the optimal function.

This model makes a range of clear predictions based on prior probability and specificity. Strictly speaking, we can only make these predictions for the processing of an individual sentence after we have determined both the priors and the specificity for that sentence. There are a number of ways to approximate these values. In the case of the ambiguities considered in this article we suggest that the priors and specificity will be principally determined by properties of the verb. We can obtain estimates of priors for verbs (say) by determining how often the verb is used in one construction or another, ideally with reference to the person who is processing the sentence. More rough approximations can be obtained from corpora or from completion studies. Obtaining estimates of specificity values is somewhat harder because of the need to estimate the range of words that can occur on a particular analysis. In other words, for priors were merely need word occurrences on a particular analysis, whereas for specificity we need word co-occurrences. This rapidly leads to a sparse data problem, even for large corpora. From corpora or completion studies, we can obtain a sample of sentences using a particular verb on a particular analysis and see how much variability there is in the words employed.

However, the predictions for a considerable proportion of locally ambiguous sentences can be made on the basis of a single reasonable assumption. This is that the specificity between a pair of words or constituents that share a direct dependency will be much greater than the specificity between words or constituents that do not share such a dependency. For example, in verb-initial languages, the specificity that a verb places on one of its postverbal arguments will normally be relatively high (e.g., because the verb places selectional restrictions on the argument). In contrast, the specificity that a verb places on a postverbal phrase that is not one of its arguments will be relatively low, since these two elements will normally place few restrictions on each other.

Such a situation occurs in both of the local ambiguities considered in this article. For instance, *realized* places far more constraints on the following noun phrase on the object analysis than it does on the sentential-complement analysis [see (1)]; and *was flying* places far more constraints on the following noun phrase on the object analysis than it does on the intransitive analysis [see (3)]. This difference in specificity is likely to be very considerable and will outweigh the difference in priors. Thus, the object analysis will be the more informative analysis in both types of sentence. Such a situation will, we suggest, occur with most sentential complement-preference and intransitive-preference verbs. However, it may be possible for the prior preference for these analyses to be so vast that the difference in testability could be outweighed (as may happen with verbs like *sneezed* that only occur transitively in very restricted circumstances and thus very infrequently; see Mitchell, 1987). But in general, there will be a strong initial preference for an analysis that allows a dependency to be formed over one that does not.

The primary purpose of the informativity model is to provide an abstract characterization of how the parser should behave. A separate question is how such behavior is realized by the human language-processing mechanism. The strongest hypothesis would be that the parser attempts a faithful implementation of informativity and actually computes the informativity values in a word-by-word manner. It is equally possible, however, that the parser uses heuristic strategies which approximate to informativity, thereby yielding similar behavior for the majority of the utterances people encounter. For example, the general predictions of specificity might be realized as simple heuristics (e.g., prefer argument attachment to modifier attachment). A further possibility is a hybrid account in which people combine probabilistic estimates of likelihood with heuristic strategies like argument attachment. Such a hybrid approach could exploit information stored at the lexical level (e.g., concerning the relative frequency of possible subcategorization frames for individual verbs and possibly the specificity of these alternatives). Storing more complex probabilistic information, for instance associated with partial syntactic structures, would be much more complex, as it would entail computing frequencies for structures that are presumably constructed during processing and are not stored.

Our experiments tested the predictions of the serial-lexical-guidance and serial-likelihood accounts and showed that neither was compatible with the results of three eye-tracking experiments. Instead, the processor appeared to favor less frequent analyses over more frequent ones. Such evidence would traditionally be used to support restricted, two-stage parsing models. In this final section, we have presented an alternative account of these findings. We suggest that the processor obeys the principle of informativity, under which initial selection of analyses is dependent both on how frequent an analysis is and how testable it is.

APPENDIX

The noun before the shill (/) was used in the version of the item containing the plausible object analysis, whereas the noun after the shill was used in the version of the item containing the implausible object analysis. Note that some nouns occur twice, but repeats never occurred in an item list. Regions are indicated by "|" marks. The final "|" in each item also indicates the line break.

Experiment 1 items

1. The young athlete realized her | potential/exercises | one day | might | make her | a world-class sprinter.
2. The doctor admitted the | patient/surgery | one day | would | be forgotten | by the staff.
3. The teacher hinted the | solution/appendix | sometimes | could | be found in | the back of the book.
4. The bank admitted the | mistake/weather | sooner or later | would | cause a lot | of trouble.
5. The investor realized a | profit/banker | apparently | would | save the | company from going broke.
6. The witness admitted the | truth/sound | in time | would | make the victim | angry.
7. The judge decided the | issue/crook | in time | would | confuse the | jurors.
8. The slacker pretended the | illness/surgery | somehow | caused | his | absence.
9. The butler admitted the | visitor/candles | apparently | distressed | the | duchess tremendously.
10. The lecturer hinted the | solution/building | clearly | would | confuse the | architecture students.
11. The instructor implied the | answer/doctor | sooner or later | would | open up | new lines of research.
12. The murderer admitted his | guilt/knife | one day | was | uncovered by | the police.
13. The tutor hinted the | answer/spider | clearly | would | amaze the | students.
14. The jury decided the | verdict/sheriff | apparently | would | satisfy people's | demand for justice.
15. The young man realized his | goals/shoes | somehow | would | be far out of | reach.
16. The pub owner admitted the | punters/matches | sometimes | bored | him | senseless.

Experiment 2 items

THE YOUNG EXECUTIVE
The young man thought about his current goals as he packed his shoes in the luggage compartment of the bus. The young man realized his goals/shoes somehow would be far out of reach. But he struggled on anyway.

THE YOUNG SPRINTER
The girl did a series of strenuous exercises every day before she reached her full potential. The young athlete realized her potential/exercises one day might make her a world-class sprinter. She dreamt about winning a gold medal in the Olympic Games.

AT THE MERCHANT BANK
The banker studied the company's books in detail trying to figure out how it could make a profit. The investor realized a profit/banker apparently would save the company from going broke. He felt certain that the company's fortunes would take a turn for the better.

AT THE HOSPITAL
The wise old doctor wondered whether the experimental surgery would help the new patient. The doctor admitted the patient/surgery one day would be forgotten by the staff. They had all been working extremely hard.

ON TRIAL FOR BLACKMAIL
The witness thought she knew the truth and the sound coming from the tape player convinced her even more. The witness admitted the truth/sound in time would make the victim angry. The crime had shocked the entire community.

DOWN AT THE OLD BAILEY
Without being prompted, the crook raised the issue that made the old judge scramble for his law books. The judge decided the issue/crook in time would confuse the jurors. They might never uncover the truth.

ANOTHER CASE SOLVED
The sheriff hoped that the jury would announce the verdict before the crowd outside got out of hand. The jury decided the verdict/sheriff apparently would satisfy people's demand for justice. The crowd grew larger as the trial dragged on.

AVOIDING WORK
Jeff thought that the shirker had made up the whole story about his illness and surgery that he needed. The slacker pretended the illness/surgery somehow caused his absence. He had been fired from his last four jobs.

AT THE ARCHITECTURE COLLEGE
The old lecturer knew that the students could not design the building properly until they came up with the solution to the central support problem. The lecturer hinted the solution/building clearly would confuse the architecture students. They were having a very hard time mastering the subject material.

LEARNING ABOUT GENETICS
The instructor told the students that the doctor had found the answer to the most serious problem in genetics. The instructor implied the answer/doctor sooner or later would open up new lines of research. New treatments for inherited diseases would be just around the corner.

DOWN AT THE BANK
One of the accountants discovered a mistake that the bank made predicting how the weather would affect the crops. The bank admitted the mistake/weather sooner or later would cause a lot of trouble. A lot of people were likely to lose a lot of money.

IN THE PALACE
The duke insisted that every candle in the palace be lit because the visitor wanted to stay up very late. The butler admitted the visitor/candles apparently distressed the duchess tremendously. She had been under a lot of stress lately.

AT THE PUB
Every Saturday afternoon, dozens of punters would pack the pub to watch the matches on television. The pub owner admitted the punters/matches sometimes bored him senseless. He tried to distract himself with thoughts of his upcoming holidays.

DONE FOR MURDER
When the newspaper wrote that a knife had been recovered, the murderer knew that his guilt would be proven. The murderer admitted his guilt/knife one day was uncovered by the police. However, he never showed any remorse for what he had done.

IN BIOLOGY CLASS
The tutor had to tell the students all about the spider before they figured out the answer to the question about its feeding habits. The tutor hinted the answer/spider clearly would amaze the students. He always believed that students learned more when the teacher was enthusiastic.

THE MATHS TEXTBOOKS
The teacher said that the appendix appeared in one edition of our textbooks and that we might check the solution. The teacher hinted the solution/appendix sometimes could be found in the back of the book. The students learned a lot more when they were able to check their answers.

Experiment 3 items

1. While the pilot was flying the | plane/horse | that had arrived | stood | over | by the fence.
2. As the monk chanted the | prayer/priest | that the bishop knew | impressed | a | foreign visitor.
3. While the supporters cheered the | team/wind | that came in | bothered | the | coaching staff.
4. As the professor lectured the | students/sparrows | that were still | became | a | bit restless.
5. As the sergeant marched the | squad/tanks | that the man saw | attacked | a | fortified position.
6. As the minister preached the | sermon/bishop | that we heard | excited | the | congregation.
7. As the old man sailed the | boat/isle | that the tycoon | owned | appeared on | the horizon.
8. As the choir sang the | hymn/game | that the vicar liked | was played by | the organist.
9. When the man smoked the | cigar/child | that belonged to a friend | burned | a | hole in his suit.
10. While the boy was swimming the | river/horse | from the hills | raced | past | the pond.
11. While the man walked the | dog/car | that his wife bought | stopped | next | to the fire hydrant.
12. As the pickets chanted the | threats/workers | that upset the boss | created | a | big disturbance.
13. Because the boss lectured the | clerk/break | that was late | was | delayed an | entire hour.
14. When the kidnapper marched the | hostage/vehicle | that he seized | stopped at | the side of the road.
15. While the captain sailed the | ship/bird | that passed by | became | visible | alongside the reef.
16. While the girl sang the | song/lamp | that she liked | irritated | the old | woman.
17. When the girl walked the | poodle/cradle | that her sister | owned | rested in | the front hall.
18. While the protesters chanted the | slogans/workers | that we knew | annoyed | an | elderly manager.
19. While the speaker lectured the | crowd/train | that remained | prepared | to | depart.
20. As the guards marched the | convict/tractor | that broke down | received | some | attention.
21. As the helmsman sailed the | vessel/island | that the navy aided | repulsed | a | pirate attack.
22. When the tenor sang the | opera/crowd | that pleased the producer | made | the | director ecstatic.
23. As the teacher lectured the | boy/bee | that was outside entered | the | classroom.
24. As the pirate sailed the | vessel/maiden | that was shipwrecked | tossed | in a | rough sea.
25. As the soprano sang the | song/room | which we arranged | displeased | the | conductor.

26. While the athlete was running the | race/baby | that we saw produced | an | enthusiastic response.

REFERENCES

Abney, S. P. (1989). A computational model of human parsing. *Journal of Psycholinguistic Research,* **18,** 129–144.

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

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

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

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioural and Brain Sciences,* **14,** 471–517.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database* (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language,* **34,** 774–806.

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

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

Brysbaert, M., & Mitchell, D. C. (1996). Modifier attachment in sentence parsing: Evidence from Dutch. *Quarterly Journal of Experimental Psychology A,* **49,** 664–695.

Charniak, E. (1997). Statistical techniques for natural language parsing. *AI Magazine,* **18,** 33–44.

Chater, N. J. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review,* **101,** 566–581.

Chater, N. J., Crocker, M. W., & Pickering, M. J. (1998). The rational analysis of inquiry: The case of parsing. In N. J. Chater & M. R. Oaksford (Eds.), *Rational models of cognition* (pp. 441–469). Oxford: Oxford Univ. Press.

Clifton, C., Jr. (1993). Thematic roles in sentence parsing. *Canadian Journal of Experimental Psychology,* **47,** 222–246.

Clifton, C., Jr., Kennison, S. M., & Albrecht, J. E. (1997). Reading the words *her, him, his:* Implications for parsing principles based on frequency and on structure. *Journal of Memory and Language,* **36,** 276–292.

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

Crocker, M. W. (1996). *Computational psycholinguistics: An interdisciplinary approach to the study of language.* Dordrecht: Kluwer.

Crocker, M. W., Pickering, M. J., & Chater, N. J. (1999). *A rational analysis of parsing and interpretation.* Unpublished manuscript.

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.

Cuetos, F., Mitchell, D. C., & Corley, M. M. B. (1996). Parsing in different languages. In M. Carreiras, J. Garcia-Albea, & N. Sabastian-Galles, (Eds.), *Language processing in Spanish.* Hillsdale, NJ: Erlbaum.

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

Ferreira, F., & Henderson, J. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **16,** 555–568.

Ferreira, F., & Henderson, J. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language,* **30,** 725–745.

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry,* **9,** 427–473.

Ford, M., Bresnan, J. W., & Kaplan, R. M. (1982). A competence based theory of syntactic closure. In J. W. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 727–796). Cambridge, MA: MIT Press.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies.* Ph.D. thesis, University of Connecticut.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 559–586). Hillsdale, NJ: Erlbaum.

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.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language,* **37,** 58–93.

Gibson, E., & Schütze, C. T. (1999). Disambiguation preferences in noun phrase conjunction do not mirror corpus frequency. *Journal of Memory and Language,* **40,** 263–279.

Gorrell, P. (1995). *Syntax and perception.* Cambridge: Cambridge Univ. Press.

Hemforth, B., Konieczny, L., Scheepers, C., & Strube, G.

(1994). *First analysis, reanalysis, and repair.* Freiburg: Institut fur Informatik und Gesellschaft.

Holmes, V. M. (1987). Syntactic parsing: In search of the garden path. In M. Coltheart (Ed.), *Attention and performance XII.* Hillsdale, NJ: Erlbaum.

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

Juliano, C., & Tanenhaus, M. K. (1994). A constraint based lexicalist account of the subject–object attachment preference. *Journal of Psycholinguistic Research,* **23,** 459–471.

Jurafsky, D. (1996). A Probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science,* **20,** 137–194.

Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. Chips: An ideal-observer model of reading. *Psychological Review,* **104,** 524–553.

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.

Liversedge, S. P., Pickering, M. J., & Traxler, M. J. (1996). *A comparative analysis of qualitatively different eye movement measures.* Poster presented at the Ninth CUNY Sentence Processing Conference, New York.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes,* **9,** 157–201.

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

Marr, D. (1982). *Vision.* San Francisco, CA: Freeman.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language,* **38,** 283–312.

Mitchell, D. C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 601–618). Hillsdale, NJ: Erlbaum.

Mitchell, D. C. (1989). Verb guidance and lexical effects in ambiguity resolution. *Language and Cognitive Processes,* **4,** 123–154.

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

Mitchell, D. C., & Cuetos, F. (1991). The origin of parsing strategies. In C. Smith (Ed.), *Current issues in natural language processing.* Austin, TX: University of Texas.

Mitchell, D. C., Cuetos, F., Corley, M. M. B., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (non-lexical) statistical records. *Journal of Psycholinguistic Research,* **24,** 469–488.

Mitchell, D. C., & Holmes, V. M. (1985). The role of specific information about the verb in parsing sentences with local structural ambiguity. *Journal of Memory and Language,* **24,** 542–559.

Oaksford, M., & Chater, N. J. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review,* **101,** 608–631.

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **24,** 940–961.

Pritchett, B. (1992). *Grammatical competence and parsing performance.* Chicago: Univ. of Chicago Press.

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

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition,* **14,** 191–201.

Rayner, K., & Frazier, L. (1987). Parsing temporarily ambiguous complements. *Quarterly Journal of Experimental Psychology,* **39,** 657–673.

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

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading.* Englewood Cliffs, NJ: Prentice Hall.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal,* **27,** 623–656.

Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition,* **55,** 227–267.

Sturt, P., & Crocker, M. W. (1996). Monotonic syntactic processing: A cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes,* **11,** 449–494.

Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty. *Journal of Memory and Language,* **40,** 136–150.

Taraban, R., & McClelland, J. R. (1988). Constituent attachment and thematic role assignment in sentence processing: Influence of content-based expectations. *Journal of Memory and Language,* **27,** 597–632.

Traxler, M. J., Bybee, M. D., & Pickering, M. J. (1997). Influence of connectives on language comprehension: Eye-tracking evidence for incremental interpretation. *Quarterly Journal of Experimental Psychology A,* **50,** 481–497.

Traxler, M. J., & Pickering, M. J. (1996a). Case-marketing in the parsing of complement sentences: Evidence from eye movements. *Quarterly Journal of Experimental Psychology A,* **49,** 991–1004.

Traxler, M. J., & Pickering, M. J. (1996b). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, **35,** 454–475.

Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, **39,** 558–592.

Trueswell, J. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, **35,** 566–585.

Trueswell, J., Tanenhaus, M. K., & Garnsey, S. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, **33,** 285–318.

Trueswell, J., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **19,** 528–553.

Tyler, L. K., & Marslen-Wilson, W. D. (1977). The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior*, **16,** 683–692.

Warner, J., & Glass, A. L. (1987). Context and distance-to-disambiguation effects in ambiguity resolution: Evidence from grammaticality judgments of garden path sentences. *Journal of Memory and Language*, **26,** 714–738.

(Received March 29, 1999)
(Revision received December 7, 1999)