Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Memory and Language 52 (2005) 144–169
Journal of Memory and Language
www.elsevier.com/locate/jml

Online syntactic storage costs in sentence comprehension

Evan Chen*, Edward Gibson*, Florian Wolf

Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received 15 June 2004; revision received 6 October 2004

Abstract

This paper presents three self-paced, word-by-word reading experiments that test for the existence of on-line syntactic storage/expectation costs in English. To investigate this issue, we compared reading times for sentence regions in which syntactic expectation costs varied, keeping other factors constant. Experiment 1 manipulated the number of verbs needed to form a grammatical sentence. It was observed that in the critical region, people read the condition in which zero verbs were predicted fastest, followed by the conditions in which one verb was predicted, with the condition in which two verbs were predicted slowest. Experiments 2 and 3 investigated whether incomplete filler-gap dependencies also incur storage costs. It was observed that people read the critical region in which a wh-filler is pending slower than if no such wh-filler is pending. The results of all three experiments demonstrate the role of online syntactic storage costs in sentence comprehension.

Keywords: Sentence comprehension; Sentence complexity; Sentence processing

It is well known that nested (or center-embedded) syntactic structures are harder to understand than their right- or left-branching counterparts (Chomsky & Miller, 1963; Yngve, 1960). For example, the right-branching English structure in (1a) is easier to understand than the nested structure in (1b), and the left-branching Japanese structure in (2a) is easier to understand than its nested version in (2b) (Gibson, 1998):

(1) a. Mary met the senator who attacked the reporter who ignored the president.
b. # The reporter who the senator who Mary met attacked ignored the president.

(2) a. [[*imooto-ga naita to*] *bebiisitaa-ga itta to*] *obasan-wa omotteiru*.
sister-nom cried that babysitter-nom said that aunt-top think
"My aunt thinks that the babysitter said that my sister cried"
b. # *obasan-wa* [*bebiisitaa-ga* [*imooto-ga naita to*] *itta to*] *omotteiru*.]
aunt-top babysitter-nom sister-nom cried that said that think
"My aunt thinks that the babysitter said that my sister cried"

The difficulty of understanding nested structures occurs despite the fact that each nested sentence has the same propositional content and lexical items as its right- or left-branching counterpart. One proposed cause for the complexity of nested structures is that

* Corresponding authors. Fax: +1 617 258 8654 (E. Gibson).
E-mail addresses: echen2@mit.edu (E. Chen), egibson@mit.edu (E. Gibson).

0749-596X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2004.10.001

they may require more syntactic memory or storage space to process than left- or right-branching structures (Abney & Johnson, 1991; Chomsky & Miller, 1963; Gibson, 1991, 1998; Lewis, 1996; Stabler, 1994). This amounts to the claim that the human parser is partially top-down in nature, with expectations about what kinds of syntactic categories are coming in the input. In the earliest of the syntactic storage accounts, Yngve (1960) and Chomsky and Miller (1963) proposed that syntactic storage at a particular parse state is quantified in terms of the number of partially processed phrase structure rules like $S \Rightarrow NP\ VP$. Wanner and Maratsos (1978) proposed that there may also be storage cost associated with incomplete filler-gap dependencies, which may not be represented within a single phrase structure rule. Other proposals for units of syntactic storage include incomplete clauses (Kimball, 1973), incomplete thematic role assignments (Gibson, 1991; Hakuta, 1981), incomplete syntactic dependencies (Gibson, 1998), incomplete case-assignments (Lewis, 1996; Stabler, 1994), and predicted syntactic heads (Gibson, 1998).

Consider the predicted syntactic head hypothesis of storage costs with respect to the contrast in (1). The most complex state in processing the nested structure (1b) in terms of syntactic storage occurs at the point of processing the most embedded subject noun phrase (NP) "Mary." Five syntactic heads are required at this point in order to form a grammatical sentence: (1) three verbs for the three subject NPs: (a) the NP "the reporter" requires its verb "ignored"; (b) the NP "the senator" requires its verb "attacked"; and (c) the NP "Mary" requires its verb "met"; and (2) two empty NP positions associated with the two filler-gap dependencies: (a) the first instance of the relative clause (RC) pronoun "who" which is eventually linked with the object position of the verb "attacked"; and (b) the second instance of the RC pronoun "who" which is eventually linked with the object position of the verb "met." In contrast, the storage requirements for the right-branching structure in (1a) consist of at most one predicted head throughout its parse. For example, at the point of processing the determiner "the," following "met," only a noun is required. The right-branching structure (1a) is therefore less complex than the nested version (1b) in terms of storage costs.

An alternative to the top-down, expectation-driven hypothesis is that the parser is purely bottom-up, with the consequence that the parser does not keep track of partially processed rules or syntactic expectations. This hypothesis predicts that the difficulty in processing nested structures is not due to syntactic storage, but is perhaps due to something else, such as integrating the words together syntactically and semantically (Gibson, 1998, 2000; Gordon, Hendrick, & Johnson, 2001). Gibson (1998, 2000) proposes that both of these mechanisms—storage and integration—are at play in normal sentence comprehension. First, it is proposed that there is a cost associated with keeping track of expected syntactic elements. Second, it is proposed that there is a cost associated with connecting an incoming word to the structure that has been built thus far, including matching earlier expectations. Integration cost has been demonstrated to be sensitive to distance, such that longer distance integrations give rise to greater complexity in terms of reading times (Gibson, 1998, 2000; Grodner & Gibson, in press).

Much evidence has been provided for the integration cost hypothesis (e.g., Gibson, 1998; Gordon et al., 2001; Grodner & Gibson, in press; King & Just, 1991; Warren & Gibson, 2002), but there is less evidence for the top-down storage cost hypothesis. The present experiments are designed to test for the existence of on-line storage costs in English, independent of other factors. Two of the most general syntactic storage hypotheses are the incomplete dependency hypothesis and the predicted syntactic head hypothesis. These two hypotheses generally make the same predictions with respect to the examples that we will discuss here, so for simplicity we will reserve our discussion to one, the predicted-head hypothesis. We return to the issue of differentiating between storage cost theories in the general discussion.

There is little previous research which tests for the presence of on-line storage costs in English. Although the intuitive contrast between sentences like (1a) and (1b) has been verified using off-line measures such as intuitive acceptability (e.g., Gibson, 1998; Gibson & Thomas, 1997; Miller & Isard, 1964; Stolz, 1967), this evidence does not address on-line sentence processing directly. Furthermore, there are at least two other factors in addition to storage costs that are involved in the comparison between (1a) and (1b). One such confounding factor is integration cost. As mentioned above, Gibson (1998) has hypothesized that another reason that nested structures are more complex than non-nested structures is that nested structures always necessitate longer distance integrations between syntactic heads. This factor may therefore be partially responsible for the contrast in (1). Another factor relevant to comparing (1a) with (1b) involves potential processing differences between modifying the subject of a verb and modifying an object of a verb. Gibson, Desmet, Grodner, Watson, and Ko (in press) observed that RCs modifying subjects are processed faster than RCs modifying objects (cf., Holmes, 1973, for related results). Gibson et al. attributed the observed reading time differences to differences in the information flow status of restrictive RCs together with the position of an NP in a sentence. Restrictive

RCs are generally used to refer to objects that are background information in the current discourse. In addition, background information is read faster when it modifies a subject rather than an object. Thus, restrictive RCs are read faster in subject position, where they are compatible with being background information, than in object position, where they are less compatible with being background information. To avoid this issue when investigating on-line storage costs, the critical region clause to be compared should be in the same syntactic position.

Gibson et al. (in press) also began to investigate the existence of on-line syntactic storage costs in English. In a self-paced reading study, Gibson et al. circumvented the potential confounds described above by comparing sentences like (3a) to the same sentences embedded within the sentential complement (SC) of a noun like "fact," as in (3b):

(3) a. **The reporter who the senator attacked on Tuesday ignored the president**.
b. The fact [that **the reporter who the senator attacked on Tuesday ignored the president**] bothered the editor.

In (3b), the dependency between the subject NP "the fact" and its verb "bothered" is pending when people process the embedded sentence "the reporter who the senator attacked on Tuesday ignored the president," leading to a predicted verb over this region. Consistent with the hypothesis that keeping track of this predicted syntactic element consumes processing resources, Gibson et al. found slower reading times during the processing of the embedded sentence in (3b) as compared with the same region in (3a). However, there is a complication with the materials in Gibson et al.'s design. The target region is interpreted relative to the noun "fact" (as its complement) in the high storage condition (3b), whereas the same region is not semantically dependent on any previous linguistic material in (3a). It could be that interpreting the embedded clause relative to the preceding thematic context is what makes people read this region slower.[1]

[1] In addition, the target region is three words further along in (3b) than in (3a). If later regions are read more slowly, then this is an additional confound in interpreting Gibson et al.'s results. However, Gibson et al. found that later regions are not generally read more slowly than earlier regions. Thus, word position is an unlikely explanation of the effects that they observed.

Several other experiments by Gibson and colleagues indirectly investigated the existence of on-line storage costs in English by comparing RC structures and verbal SC structures like those in (4) (Gibson & Warren, 2004; Grodner, Gibson, & Tunstall, 2002; Warren & Gibson, 2002):

(4) a. Relative Clause (RC) Structure:
The witness [who **the evidence that was examined by the lawyer** implicated ___] was lying.
b. Verbal Sentential Complement (SC) Structure:
The witness thought [that **the evidence that was examined by the lawyer** implicated his next door neighbor].

During the processing of the embedded subject NP region in bold, the structures, and hence the integrations, are the same. Two extra syntactic heads are predicted during the processing of the bold region in the RC structure (4a): (1) a verb associated with the matrix subject NP "the witness"; and (2) a wh-trace associated with the RC pronoun "who." Thus, the storage cost hypothesis predicts longer reading times in the critical bold region in sentences like (4a) when compared to (4b), and this is the pattern of data that was observed. Gibson and Warren (2004) and Warren and Gibson (2002) presented similar comparisons with similar results. Although these results are suggestive, the materials in these experiments were not specifically designed to address the storage cost hypothesis. Consequently, the regions prior to the target region were not always controlled. For example, the NPs preceding the critical region in Grodner et al. (2002) items as exemplified in (4) were not always the same across the high and low load conditions, although they are in the example provided.

The primary goal of the present experiments is to test for the existence of on-line storage costs in controlled English materials using self-paced reading. A secondary goal is to attempt to distinguish among types of syntactic storage hypotheses, if the presence of storage costs can be measured. According to one class of hypotheses, storage cost may only be measurable in complex structures, with multiple interfering expectations (Lewis, 1996). For example, according to Kimball's (1973) principle of two sentences, the processor has extreme difficulty in comprehending sentences that require more than two expected verbs to come. Similarly, Lewis (1996) proposes that processing difficulty may result from multiple open expectations of the same type, such as subject–verb relationships (specifier of IPs in X-bar notation, Chomsky, 1986), or filler–gap relationships (specifier of CPs in X-bar notation). It is an open question whether storage costs are measurable only in these highly interfering circum-

stances (as in the materials in (3) and (4) above), which may exacerbate the effects, or whether they are present even without such interference.

Another open question is which kinds of predicted syntactic elements (if any) are associated with storage cost. All previous work that showed evidence of online syntactic storage cost compared conditions in which the high-storage conditions involved more predicted verbs than the low-storage conditions. It is therefore possible that storage cost is restricted to predicted verbs (i.e., partially processed clauses). Experiments 2 and 3 test this prediction by examining another kind of incomplete dependency, the expectation of a wh-trace inside a relative clause (cf., Wanner & Maratsos, 1978).

Experiment 1

Experiment 1 was designed to test whether predicted verbal heads are associated with incremental storage costs depending on the number of predicted verbs. To test for effects of storage cost, we used materials containing SC-biased verbs (e.g., "suspected" and "knew") and their nominalizations (e.g., "suspicion" and "knowledge") in order to manipulate the number of predicted verbs across a critical embedded clause. The materials were constructed in a 2×2 design, crossing the syntactic category of the first SC-taking word (verb or noun) with the category of the second SC-taking word (verb or noun). An example of the four conditions that were tested is given in (5):

(5) a. Zero predicted verbs (Verb 1/Verb 2):
The detective suspected that the thief knew that **the guard protected the jewels** and so he reported immediately to the museum curator.
b. One late predicted verb (Verb 1/Noun 2):
The detective suspected that the knowledge that **the guard protected the jewels** came from an insider.
c. One early predicted verb (Noun 1/Verb 2):
The suspicion that the thief knew that **the guard protected the jewels** worried the museum curator.
d. Two predicted verbs (Noun 1/Noun 2):
The suspicion that the knowledge that **the guard protected the jewels** came from an insider worried the museum curator.

The critical region in this design consists of the embedded clause "the guard protected the jewels," in bold. Because this clause has the same structure in all conditions, integration costs are identical across the four conditions. Furthermore, this clause is in a similar syntactic position in each condition, the complement of a related verb or noun. Thus, any observed differences in reading difficulty during this region are not attributable to integration cost differences or syntactic position differences.[2]

For the zero predicted verbs condition in (5a), the critical material "the guard protected the jewels" is embedded as the SC of the verb "knew" which is itself part of a clause embedded as the SC of the matrix verb "suspected." Because both verbs "knew" and "suspected" are encountered immediately after their respective subject nouns, no additional verbs are predicted after the critical embedded clause. For the one-late predicted verb condition in (5b), the verb "knew" is nominalized to "knowledge" with the result that the critical clause is a SC of the noun "knowledge." The change to the embedded subject NP "the knowledge" results in the requirement for an additional verb during the processing of the critical region. Similarly, for the one-early predicted verb condition in (5c), the matrix verb "suspected" is nominalized to "suspicion" and the embedded SC "the thief knew that the guard protected the jewels" is an argument of the matrix subject NP "the suspicion." Once again, a verb is required after the critical region. Finally, for the two predicted verbs condition in (5d), both the verbs "suspected" and "knew" are nominalized and two verbs are required following the critical region. Thus, if storage costs are proportional to the number of incomplete verb dependencies or predicted verbs held in memory, then the zero predicted verbs condition should be read fastest of the four conditions in the critical region. The one-late and one-early predicted verb conditions should be read

[2] The positions of the words in the critical embedded clause are not identical across the four conditions, because of the inclusion of a subject noun in the verbal versions of the items (e.g., "the detective suspected that ..." vs. "the suspicion that ..." in (5)). This small difference between the noun/verb versions is not likely to be confounded with the predictions of storage costs in the critical region. For this to be so, it would need to be the case that people read later word positions more slowly. This is generally not so in English, as demonstrated for example in Gibson et al. (in press) using similar materials. Second, another version of Experiment 1 was also run in which the subjects of the verbs in the verbal conditions were also included as genitive subjects in the nominal conditions (e.g., "the detective suspected that ..." vs. "the detective's suspicion that ..." in items like (5)). The same numerical patterns were observed during the critical region in this variant of the current experiment as here, with many of the numerical comparisons also significantly different. We do not present this experiment because not all of the numerical trends in the critical region were significant and because the reading times in the earlier nominal regions, which included the genitive NPs, were much slower than in the current experiment, possibly reflecting the unusualness of repeated genitive subjects of NPs.

more slowly, and the two predicted verbs condition should be read the slowest of the four.

On-line syntactic storage cost hypotheses also predict reading time differences in other regions of the target sentences. First, there is an extra predicted verb in the one-early condition relative to the zero predicted verbs condition during the second clause "the thief knew that." Second, there is an extra predicted verb in the two predicted verbs condition relative to the one-late condition during the region "the knowledge that." Third, there is an extra predicted verb during the region "came from an insider" (following the critical region) for the two predicted verb condition relative to the one-late condition. Reading times are predicted to be longer for each of the conditions with more predicted verbs.

Although both the one-late and the one-early conditions require the maintenance of one predicted verb over the critical region, the expectation is initiated at different points in the two sentences. If the amount of time that an incomplete dependency is held in memory increases the load, then reading times in the critical region of the one-early condition should be longer than in the one-late condition. Such an hypothesis was made in Gibson's (1998) syntactic prediction locality theory (SPLT) in order to partially explain distance effects in nested materials. However, it was also noted by Gibson that the integration component of sentence comprehension would account for distance effects, with the result that the locality-based storage hypothesis may be unnecessary in such a processing system. Indeed, results reported in Gibson et al. (in press) failed to provide evidence for a syntactic storage cost metric that increases over distance (using a different syntactic structure manipulation). This evidence provided some of the motivation for Gibson's (1998, 2000) more recent hypothesis, the dependency locality theory (DLT), whose storage component is not assumed to be locality based. Because of the lack of distance-based storage effects in the earlier experiments, we do not anticipate such a difference in reading times between the one-late and one-early conditions for the critical region.

Method

Participants

Seventy-four participants from the MIT and Boston University communities were paid for their involvement. All were native speakers of English and were naïve as to the purposes of the study.

Materials and design

Forty sets of sentence were constructed, each with the four conditions as exemplified in (5). The first region in each item consisted of either an SC-biased subject–verb combination consisting of a subject NP and an SC-biased verb (e.g., "the detective suspected"), or the nominalized form of the SC-biased verb (e.g., "the suspicion"). The second region consisted of the beginning of the first embedded SC, which began with the word "that" and was followed by a region similar in structure to the first region: either an SC-biased subject–verb combination consisting of a subject NP and an SC-biased verb (e.g., "the thief knew"), or the nominalized form of the SC-biased verb (e.g., "the knowledge"). The third region consisted of the most embedded SC (e.g., "that the guard protected the jewels"), which made up the critical region. Finally, various materials, which diverged for the four conditions, completed the sentences grammatically and plausibly. Because a clause following an SC-taking noun is also temporarily ambiguous with an RC, we chose SC-taking nouns that that were lexically biased to take SCs (72% on average across all nouns) according to the norms in Kennison (2000). Because of this bias, participants would probably not follow a potential RC interpretation. We investigate the potential effect of this ambiguity in more depth in Experiment 2.

The target sentences were split into four lists balancing all factors in a Latin-Square design. Each list was combined with 60 fillers of various types. Appendix A provides a complete list of the stimuli along with the SC-bias of each SC-noun in the items. The stimuli were pseudo-randomized separately for each participant so that at least one filler item intervened between two targets.

Procedure

The task was self-paced, word-by-word reading using a moving window display (Just, Carpenter, & Woolley, 1982). The software used to run the experiment was Linger by Doug Rohde. Each trial began with a series of dashes marking the length and position of the words in the sentences, printed approximately a third of the way down the screen. A single line displayed up to 100 characters. Participants pressed the spacebar to reveal each word of the sentence. As each new word appeared, the preceding word returned to dashes. The amount of reading time (RT) the participant spent on each word was recorded as the time between key-presses. After the final word of each item was a comprehension task in which the participant was presented with a fill-in-the-blank statement, which asked about information contained in the preceding sentence. There were four choices provided as possible answers. Example comprehension questions for the items in (5) are provided in (6):

(6) a. The detective suspected that the thief knew that the guard protected the jewels and so he reported immediately to the museum curator.
Question: The _____ reported to the museum curator.
Choices: detective, witness, guard, policeman;
Answer: detective.

b. The detective suspected that the knowledge that the guard protected the jewels came from an insider.
Question: The _____ may have come from an insider.
Choices: suspicion, knowledge, protection, jewels;
Answer: knowledge.

c. The suspicion that the thief knew that the guard protected the jewels worried the museum curator.
Question: The suspicion worried the _____.
Choices: detective, thief, museum guard, museum curator; Answer: museum curator.

d. The suspicion that the knowledge that the guard protected the jewels came from an insider worried the museum curator.
Question: The _____ worried the museum curator.
Choices: suspicion, knowledge, insider, guard;
Answer: suspicion.

Participants pressed one of four keys to respond. After an incorrect answer, the word "INCORRECT" flashed briefly on the screen. No feedback was given for correct responses. Correct answers were balanced across the four possible response positions. All parts of the sentences were probed, across items and conditions, in roughly equal proportions, with the consequence that participants had to pay attention to the content of the entire sentence in order to answer the questions. Participants were asked to read sentences at a natural rate and to be sure that they understood what they read. They were told to answer the questions as quickly and accurately as they could and to take wrong answers as an indication to read more carefully.

Before the main experiment, a short list of practice items and comprehension tasks was presented in order to familiarize the participant with the task. Each session with a participant averaged 30 min. For most participants, this experiment was followed by an unrelated experiment using the same self-pace reading procedure. Participants were given short breaks between the two experiments.

Results

Eleven participants' data were omitted from all analyses because of poor comprehension question performance (<60% accuracy overall). The statistical patterns are identical if the accuracy cutoff is down as low as 58%, allowing the inclusion of data from four more participants.

Comprehension task performance

On average, the comprehension tasks for the experimental items in Experiment 1 were answered correctly in 81% of the trials. The percentages of correct answers for each condition are presented in Table 1.

An omnibus 1-way ANOVA (0, 1-Late, 1-Early, and 2 predicted verbs) showed a significant effect ($F1(3,186) = 47.78$, $MS_{\text{within}} = 0.0110$, $p < .0001$; $F2(3,117) = 11.97$, $MS_{\text{within}} = 0.0284$, $p < .0001$). A 2×2 ANOVA (Noun or Verb in the first clause/NP region × Noun or Verb in the second clause/NP region) showed a significant main effect of the syntactic category of the first clause ($F1(1,62) = 44.47$, $MS_{\text{within}} = 0.0098$, $p < .0001$; $F2(1,39) = 7.904$, $MS_{\text{within}} = 0.0360$, $p < .008$), a significant main effect of the syntactic category of the second clause ($F1(1,62) = 45.87$, $MS_{\text{within}} = 0.0118$, $p < .0001$; $F2(1,39) = 10.94$, $MS_{\text{within}} = 0.0328$, $p < .003$), and a significant interaction ($F1(1,62) = 52.64$, $MS_{\text{within}} = 0.0114$, $p < .0001$; $F2(1,39) = 22.94$, $MS_{\text{within}} = 0.0164$, $p < .0001$). These effects were carried by the fact that response accuracies to the two predicted verbs condition were lower than to any of the other conditions in planned pair-wise comparisons licensed by the significance of the interaction (Fs > 20; ps < .0002).

Reading times

To adjust for differences in word length as well as overall differences in participants' reading rates, a regression equation predicting RTs from word length was derived for each participant, using all filler and target items (Ferreira & Clifton, 1986; see Trueswell, Tanenhaus, & Garnsey, 1994, for discussion). At each word position, the RT predicted by the participant's regression equation was subtracted from the actual measured raw RT to obtain a residual RT. We present tables of both residual and raw RTs for each experiment. To save space, we restrict the presentation of statistical tests to the residual RTs. The statistical analyses gave similar patterns of significance for raw RTs in the critical region of all three experiments

Table 1
Mean (standard error) comprehension task performance for Experiment 1 in percent correct as a function of incomplete verb dependencies by participants

Number of predicted verbs			
0	1-Late	1-Early	2
85 (1.5)	86 (1.6)	87 (1.7)	68 (1.9)

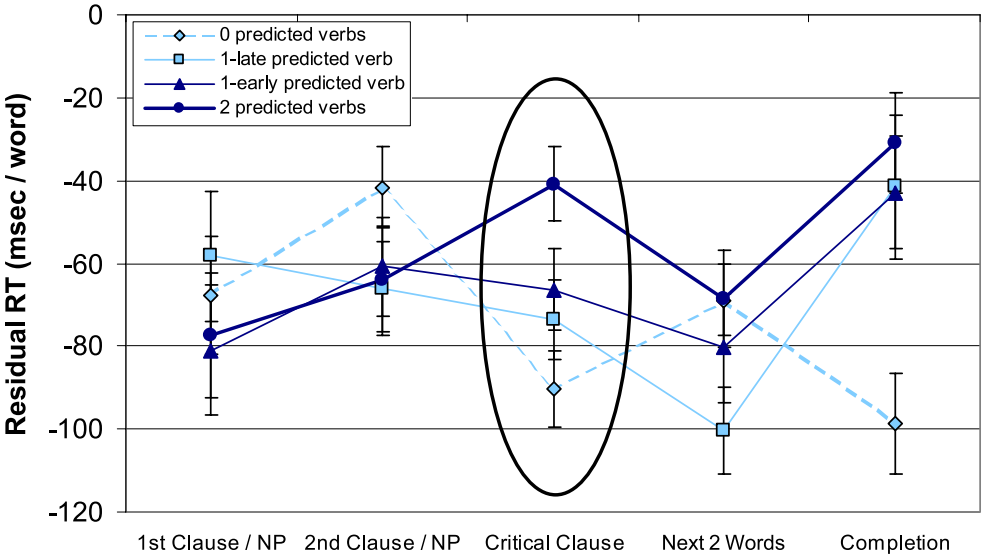

Fig. 1. Plot of mean (standard error) residual reading times (RTs) per word by region by participants in Experiment 1.

and where there are differences in significance, they are noted.[3]

Only items with correctly answered comprehension questions were analyzed. Furthermore, residual RT data points that were greater than 3 standard deviations from the mean for each region within a condition were also excluded from the analysis, affecting less than 1.4% of the data for Experiment 1 overall. Fig. 1 presents the mean residual RTs per word (ms/word) across the four conditions in this experiment. For the purpose of the analyses, we divided the materials into five regions, as shown in (7):

The first region consisted of the material up to the first SC-taking word followed by the complementizer "that." The second region consisted of the material up to the second SC-taking word followed by the complementizer "that." The third region was the critical region. The fourth region consisted of the first two words following the critical region. We analyzed only the first two words in this region because some items ended on the third word for some conditions and not others. Including the sentence-final word would have resulted in end-of-sentence effects confounding any other possible effects. The fifth region (which was not analyzed because of large item differences between conditions in this region) consisted of the remaining of material in each item. Table 2 presents the mean RTs per word for these regions, in both residual and raw forms.

(7)	The detective suspected that	the thief knew that	the guard protected the jewels	and so	⋯
	The suspicion that	the knowledge that		came from	⋯
				worried the	⋯
	Clause 1 / NP1	Clause 2 / NP2	Critical clause	two words	Completion

Reading times in the first two regions

In the first clause/NP region, a 2 × 2 ANOVA (Noun 1/Verb 1 × Noun 2/Verb 2) revealed a main effect of the syntactic category of the first SC-taking word (Noun 1/Verb 1; e.g., "suspicion"/"suspected") by participants only ($F1(1,62) = 6.753$, $MS_{\text{within}} = 2441$, $p < .02$; $F2(1,39) = 1.973$, $MS_{\text{within}} = 5397$, $p = .168$) in which

[3] As noted by a reviewer, readers in this experiment were reading the materials at a rate of between 400 ms/word and 500 ms/word, which is slower than in typical single-sentence reading experiments. The materials in the current experiment are quite complicated, as are the materials that served as filler items, which may account for the relatively slow reading times. Support for this hypothesis comes from analyses of other experiments conducted in our laboratory on simpler materials, consisting of at most two clauses. In such experiments, participants typically read in the range of 250–400 ms/word.

Table 2
Mean residual RTs (ms/word) by participants as a function of predicted verbs, for the five main regions (raw RTs in parentheses)

	Number of predicted verbs			
	0	1-Late	1-Early	2
First clause/NP	−68 (465)	−58 (475)	−81 (441)	−77 (440)
Second clause/NP	−42 (482)	−66 (444)	−61 (465)	−64 (446)
Critical clause	−90 (408)	−74 (428)	−66 (436)	−41 (458)
Next two words	−69 (370)	−101 (394)	−80 (440)	−69 (429)
Completion	−99 (407)	−41 (513)	−43 (501)	−31 (498)

clauses with the nominalized form of the verb had faster residual RTs than clauses with the verb itself. One possible explanation for this effect is that more interpretation takes place at a verb than at a subject NP, such as thematic role assignment (e.g., "the detective suspected" vs. "the suspicion"). But this explanation is difficult to evaluate, because the materials are necessarily different in this comparison. There was no effect of the category of the second SC-taking word (Noun 2/Verb 2; e.g., "knowledge"/"knew") in the first region (Fs < 1), nor was there any interaction between the factors (Fs < 1). The lack of Noun 2/Verb 2 category effects is expected because the materials do not yet differ along this dimension in the first region.

In the second clause/NP region, another 2×2 ANOVA revealed a main effect of Noun 2/Verb 2 category ($F1(1,62) = 4.538$, $MS_{\text{within}} = 2625$, $p < .04$; $F2(1,39) = 3.876$, $MS_{\text{within}} = 3398$, $p = .056$), but not for Noun 1/Verb 1 category ($F1(1,62) = 2.218$, $MS_{\text{within}} = 2040$, $p = .142$; $F2 < 1$). As in the first region, the nominal version (Noun 2) was read faster than the verbal version (Verb 2), plausibly for the same reason as in the first region. There was also a suggestion of an interaction between the factors ($F1(1,62) = 3.106$, $MS_{\text{within}} = 2316$, $p = .083$; $F2(1,39) = 1.996$, $MS_{\text{within}} = 3129$, $p = .166$). Numerically, the Verb 1/Verb 2 (zero predicted verbs) condition is the slowest of the four, with little numerical difference among the other three conditions. One possible interpretation of these data is that there is interference at the site of thematic interpretation, the verb, when there is a similar verb just before it (cf., Gordon et al., 2001). In particular, the Verb 1/Verb 2 condition consists of two SC-taking verbs in a row and their similarity could have led to processing difficulty when the second verb was interpreted. No similar effect may occur at the nouns in the Noun 1/Noun 2 (two predicted verbs) condition because there is less thematic interpretation at nouns than at verbs.

Recall that the storage cost hypothesis predicts that the second region should be read slower when the first region was nominalized. This prediction was not confirmed, but the presence of the verbal interference effect in the second region may have masked any effects of structural storage cost here. In particular, the one-early (Noun 1/Verb 2) condition was predicted to be slower than the zero predicted verbs (Verb 1/Verb 2) condition in the second clause/NP region, but the reverse numerical pattern appeared in the residual RTs, significant by participants only in the planned pair-wise comparison ($F1(1,62) = 6.211$, $MS_{\text{within}} = 1862$, $p < .02$; $F2(1,39) = 2.731$, $MS_{\text{within}} = 3095$, $p = .107$), possibly because of the interference of the two verbs in the zero predicted verbs condition. In addition, the two predicted verbs (Noun 1/Noun 2) condition was not significantly slower than the one-late (Verb 1/Noun 2) condition in the planned pair-wise comparison (Fs < 1), but this lack of an observed difference may be because this region is very short and thematically simple, consisting of a single NP and complementizer (e.g., "the knowledge that"). More thematically complicated material may be needed in order to measure a difference here, as in the critical clause.

Reading times in the critical region

A 2×2 ANOVA revealed a significant main effect of Noun 1/Verb 1 category ($F1(1,62) = 24.63$, $MS_{\text{within}} = 2043$, $p < .0001$; $F2(1,39) = 23.17$, $MS_{\text{within}} = 2006$, $p < .0001$), and a significant main effect of Noun 2/Verb 2 category ($F1(1,62) = 14.53$, $MS_{\text{within}} = 1923$, $p < .0004$; $F2(1,39) = 10.58$, $MS_{\text{within}} = 1201$, $p < .003$), but no interaction (Fs < 1). These main effects are the predicted syntactic storage effects in which the noun conditions are slower than the verb conditions during the critical clause. Note that the interference effect that was present in the second clause/NP region is no longer present in the critical region. In particular, having two similar SC-taking verbs in the first two regions of the zero predicted verbs condition did not cause processing difficulty in the critical region. Instead, the zero predicted verbs condition was the fastest condition to process. The lack of interference effects in the critical region is possibly because the verbs in the embedded region are quite different from those in the first two clauses, which were always verbs that take SCs. Thus, the residual RT effects in the critical region are likely not attributable to interference effects.

We also analyzed the critical region using an omnibus 1-way ANOVA of the four conditions. The ANOVA (0, 1-Late, 1-Early, and 2 predicted verbs) revealed significant differences among the conditions ($F1(3,186) = 14.07$, $MS_{\text{within}} = 1883$, $p < .0001$; $F2(3,117) = 12.51$, $MS_{\text{within}} = 1583$, $p < .0001$). Given the significance of this analysis,

we were able to perform planned pair-wise comparisons in the region. As predicted by the storage cost hypothesis, the zero predicted verbs condition was significantly faster than both single predicted verb conditions and the two predicted verbs condition (vs. one-late: $F1(1,62) = 4.612$, $MS_{\text{within}} = 1888$, $p < .04$; $F2(1,39) = 4.168$, $MS_{\text{within}} = 1162$, $p < .05$; vs. one-early: $F1(1,62) = 13.58$, $MS_{\text{within}} = 1317$, $p < .0006$; $F2(1,39) = 14.37$, $MS_{\text{within}} = 1410$, $p < .0006$; vs. two predicted verbs: $F1(1,62) = 32.85$, $MS_{\text{within}} = 2332$, $p < .0001$; $F2(1,39) = 33.97$, $MS_{\text{within}} = 1586$, $p < .0001$). Furthermore, both single verb predicted conditions were significantly faster than the two predicted verbs condition (vs. one-late: $F1(1,62) = 13.97$, $MS_{\text{within}} = 2410$, $p < .0004$; $F2(1,39) = 12.36$, $MS_{\text{within}} = 2137$, $p < .002$; vs. one-early: $F1(1,62) = 11.90$, $MS_{\text{within}} = 1719$, $p < .002$; $F2(1,39) = 5.103$, $MS_{\text{within}} = 1580$, $p < .03$). Finally, the one-late and one-early conditions were not significantly different from each other by participants, though marginally so by items ($F1 < 1$; $F2(1,39) = 3.267$, $MS_{\text{within}} = 1620$, $p = 0.078$). In all of the comparisons described for the critical region, identical significance patterns were found for raw RTs except for non-significant items analyses for the comparison between the zero predicted verbs and one-late conditions ($F2(1,39) = 2.547$, $MS_{\text{within}} = 1698$, $p = .119$) and the comparison between the one-early and two predicted verbs conditions ($F2(1,39) = 2.772$, $MS_{\text{within}} = 3079$, $p = .104$).

In summary, the pair-wise comparisons show us a 3-tiered pattern in which the condition with zero predicted verbs across the critical region displays the fastest reading times, the conditions with one predicted verb display slower reading times no matter when they are initiated, and the condition with two predicted verbs gives rise to the slowest reading times.

Reading times in the post-critical region

In the analysis of the post-critical region, it was discovered that the second word was actually the sentence-final word in 5 items of the one-late condition. Excluding these items from the analysis produced a mean residual (raw) RT, by participants, of −101 (394) ms/word for the one-late condition and −69 (429) ms/word for the two predicted verbs condition. The one-late condition had significantly faster residual RTs than the two predicted verbs condition in the post-critical region, both by participants and items ($F1(1,62) = 9.518$, $MS_{\text{within}} = 3366$, $p < .004$; $F2(1,34) = 8.287$, $MS_{\text{within}} = 2403$, $p < .007$). This effect was as predicted by the storage cost hypothesis. No other comparisons were made in this or later regions because of substantial item differences.

Discussion

As predicted by an on-line storage cost theory, reading times in the critical region increased proportionally with the number of predicted verbs that the sentence processor needed to maintain: zero predicted verbs fastest, followed by one predicted verb, with two predicted verbs slowest. These results provide evidence for a syntactic prediction cost for predicted verbs, even without overly complex items. In particular, the zero predicted verbs condition was read significantly faster than either of the one predicted verb conditions. None of these sentence types is more than singly nested, and none violate Kimball's (1973) principle of two sentences. Thus, it appears that storage costs apply for predicted verbs even in relatively uncomplex materials. The results in the post-critical region also confirmed the prediction of the storage cost theory. Finally, the amount of time that the dependencies were stored did not significantly affect the measured reading times, corroborating similar evidence from Gibson et al. (in press).

One prediction of the storage cost theory was not confirmed: that reading times should be slower for the one-early condition as compared with the zero predicted verbs condition in the region preceding the critical region. But the lack of an observed effect in this direction can be partially accounted for by verbal interference in this region. In the zero predicted verbs condition there are two similar SC verbs back-to-back, which may have resulted in slower processing during the second verb. The other comparison in this region involves a very short region consisting of only an NP and a complementizer. It is possible that a more thematically complex region (i.e., containing a verb) may be necessary to measure the predicted effect.

Although one possible interpretation of the results of Experiment 1 is in terms of syntactic storage costs, another possibility must be considered: that the higher reading times in the nominal conditions could have been due to a temporary ambiguity in these conditions between a SC of the noun and a RC modifying the noun. This possibility is unlikely for two reasons. First, there is currently no evidence that a temporary ambiguity between two syntactic representations ever leads to slower reading times during an ambiguous region. In fact, the current evidence supports the opposite hypothesis: that temporary ambiguity may in fact speed up reading times, as people are free to follow whichever interpretation they happen upon initially (Traxler, Pickering, & Clifton, 1998). Second, our items were lexically biased towards the SC interpretation, with the consequence that the RC interpretation was probably rarely considered. Nevertheless, the interpretation of the results in Experiments 1 would be strengthened with a comparison involving unambiguous control sentences. One of the goals of Experiment 2 was to evaluate this possibility, by testing unambiguous complement clause structures against temporarily ambiguous ones using nouns similar to those used in Experiment 1.

Another goal of Experiment 2 was to further investigate syntactic storage costs by examining another

kind of predicted head/incomplete dependency, with the goal of clarifying which aspects of syntactic expectations might be stored in on-line sentence comprehension.

Experiment 2

Experiment 1 provided evidence demonstrating a measurable increase in reading times for storing predictions of expected verbs or incomplete syntactic dependencies. These results are consistent with a storage cost theory that is sensitive to all kinds of incomplete dependencies, or with a storage theory that is sensitive only to incomplete clausal dependencies, such as Kimball's (1973) principle of two sentences. Experiment 2 was designed to examine whether incomplete dependencies other than predicted verbs are also associated with storage costs. In particular, we focus on the filler-gap dependency between a wh-pronoun and its expected wh-trace. Wanner and Maratsos (1978) presented evidence from a dual-task experiment that suggested a storage cost associated with keeping track of the dependency between a wh-filler and its gap. In particular, Wanner & Maratsos found that people were worse at recalling lists of names that were presented between the wh-filler and its verb than they were at recalling names presented in a control condition, following a subject-extracted wh-filler and its verb. In addition, there is an increasing ERP literature, starting with Kluender and Kutas (1993), pointing to a brain response associated with holding a wh-filler in memory while processing intervening material (see also Fiebach, Schlesewsky, & Friederici, 2002; Harris, 1998; King & Kutas, 1995). The current experiment tests for the existence of on-line storage costs via reaction times in normal reading, without a secondary task as in Wanner & Maratsos's work. To test the possible processing costs of storing incomplete wh-dependencies/predicting wh-traces, participants were presented with unambiguous forms of noun-modifying SC and RC structures, as in (8):

(8) a. SC Structure, Unambiguous:
The claim alleging [that **the cop who the mobster attacked** ignored the informant] might have affected the jury.
b. RC Structure, Unambiguous:
The claim [which **the cop who the mobster attacked** ignored ___] might have affected the jury.

In sentence (8a), the SC clause "that the cop who the mobster attacked ignored the informant" is a complement of the verb "alleging," which unambiguously attaches to the matrix subject NP "the claim." In sentence (8b), the clause "which the cop who the mobster attacked ignored" is an RC modifying the matrix subject NP. Critically, an RC includes a wh-dependency between its specifier (e.g., "which") and a position inside the RC, in this case, the object of the verb "ignored." An SC includes no such dependency. Thus, if storing the incomplete wh-dependency/predicted wh-trace of the RC has an associated cost, reading times for the embedded subject NP "the cop who the mobster attacked"—the critical region of interest for this experiment—are expected to be slower in the unambiguous RC condition in (8b) than in the unambiguous SC condition in (8a). This is the critical comparison for this experiment with respect to syntactic storage cost. Note that the critical region has the same structure and is in a very similar syntactic position in both conditions, so that integration costs and other factors are controlled.

One other factor that is potentially different between the RC and SC conditions is that the unambiguous RC condition may be interpreted non-restrictively. The wh-pronoun "which" often initiates a non-restrictive RC, although usually with a comma, which is not present in these items. In contrast, the SC condition will very likely be interpreted restrictively. Although restrictiveness is therefore a possible difference between the conditions, it is not a problematic one for this design because such a difference only works against the predictions of the storage cost theories. In particular, non-restrictive RCs are processed more quickly in a null context than restrictive RCs (e.g., Grodner, Gibson, & Watson, in press), possibly because they implicate smaller discourse structures than restrictive RCs in a null context (Altmann & Steedman, 1988; Crain & Steedman, 1985). Thus, if we find evidence in support of the storage cost hypothesis, it will be in spite of differences in restrictiveness across the conditions.

Note that the predicted-head storage cost hypothesis makes different predictions depending on whether or not there are empty categories (wh-traces) mediating wh-filler-gap dependencies (cf., Gibson & Hickok, 1993; Pickering & Barry, 1991). In particular, if there are such empty categories, then a predicted-head storage cost theory predicts increased complexity for the RC structure compared to the SC control structure in the critical region because of the additional predicted empty category which will eventually be placed in the object position of the embedded verb "ignored." But if there are no such empty categories, then the predicted-head storage theory predicts no difference between the two structures in the critical region. In contrast to the predicted-head storage cost theory, the incomplete-dependency storage theory predicts increased complexity over the critical region of the RC structure whether or not there are empty categories mediating wh-dependencies because there is always an extra dependency associated with the wh-filler. Although the theories make potentially different predictions, we will continue to restrict our attention to a storage cost theory based on predicted heads. We will

therefore initially assume that there are empty categories associated with wh-fillers. We return to potential ways of distinguishing alternative storage cost hypotheses in General Discussion.

A second goal of Experiment 2 was to evaluate an issue that is relevant to the interpretation of the results of Experiment 1: whether temporarily ambiguous SC items might be read more slowly than their unambiguous SC counterparts because of the presence of a possible RC interpretation. To investigate this possibility, temporarily ambiguous forms of the SC and RC sentence forms in (8) were also constructed, as in (9):

(9) a. SC Structure, Ambiguous:
The claim [that **the cop who the mobster attacked** ignored the informant] might have affected the jury.
b. RC Structure, Ambiguous:
The claim [that **the cop who the mobster attacked** ignored ___] might have affected the jury.

The items were constructed so that the head nouns were generally biased towards a SC interpretation over an RC interpretation (72% SC-bias across the items using completion norms from Kennison, 2000; 63% SC-bias using completion norms in Pearlmutter & Mendelsohn, 1999). Thus, the lexical biases in the items supported the SC interpretation, approximately as strongly as the items in Experiment 1. Structural factors also favor the SC structure over the RC structure. In particular, DLT storage costs favor the SC structure because it involves the prediction of one less syntactic head, the wh-trace associated with the relative pronoun. The principle of Minimal Attachment (Frazier, 1979)—an alternative syntactic structural heuristic to the DLT—also favors the SC structure over the RC structure because the SC structure involves the application of fewer syntactic rules.

If people generally follow the SC interpretation in the temporarily ambiguous conditions, as predicted by both the lexical biases and the syntactic heuristics, then reading times during the critical region should pattern with the unambiguous SC condition. In contrast, if there is some competition from the RC interpretation, or if the RC reading is more costly and is sometimes followed, then reading times during the critical region should be slower than those for the unambiguous SC condition, and possibly as slow or slower than those for the unambiguous RC condition. Finally, because the SC-biases of these items are weaker than in the materials in Experiment 1, if we find that the ambiguous SC items pattern with the unambiguous SC items in the current experiment, then it is reasonable to assume that the SC items in Experiment 1 would probably be processed like unambiguous SC items also.

Methods

Participants

Forty-eight participants from the MIT community who did not take part in Experiment 1 were paid for their involvement. All were native speakers of English and were naïve as to the purposes of the study.

Materials and design

Twenty sets of sentences were constructed with four conditions each, structure type (SC/RC) crossed with ambiguity (ambiguous/unambiguous), following the form of (8) and (9). These items were constructed based on items from Pearlmutter and Mendelsohn (1999), with an additional RC modifying the embedded subject NP in order to make the critical region longer and more difficult. Each item in a set began with the same matrix subject NP (e.g., "the claim"), which could take an SC. The word "that" followed the NP for the ambiguous forms of the SC and RC structures, "implying that" or some equivalent for the unambiguous SC, and "which" for the unambiguous RC. The embedded clause followed, consisting of an embedded subject NP (e.g., "the cop") and an RC modifying this NP (e.g., "who the mobster attacked"). The SC conditions continued with a transitive verb and its object (e.g., "ignored the informant"), whereas the RC conditions continued with the transitive verb alone completing the top level RC (e.g., "ignored"). Finally, a matrix verb phrase (e.g., "might have affected the jury") completed the sentences.

The ambiguous SC/RC-taking nouns were selected to be biased on average towards the SC in both sentence completion norms conducted by Kennison (2000) as well as Pearlmutter and Mendelsohn (1999) in order to make it more likely that participants would follow the SC interpretation in the temporarily ambiguous conditions. Appendix B provides a complete list of the stimuli along with the individual SC-completion percentage for each head noun. Although there was an overall bias for the SC interpretation over all the items, there was a range of biases in the individual items (72% on average, ranging from 38 to 90% in Kennison, 2000; and 63% on average, ranging from 44 to 93% in Pearlmutter & Mendelsohn, 1999).

The target sentences were split into four lists balancing all factors in a Latin-Square design. Each list was combined with 112 fillers of various types, including 32 sentences from an experiment which was an early version of Experiment 1. The stimuli were pseudo-randomized independently for each participant such that at least one filler item separated any two targets.

Procedure

The procedure was the same as in Experiment 1, except that the comprehension task that followed each item was a simple yes-no question about the contents of the preceding sentence.

Table 3
Mean (standard error) comprehension question performance in percent correct as a function of clause type and ambiguity by participant

	SC	RC
Ambiguous	76 (3.1)	77 (2.8)
Unambiguous	77 (2.9)	79 (2.5)

Results

Four participants' data were omitted from all analyses because of poor comprehension question performance (<67% accuracy overall).

Comprehension question performance

Overall, the comprehension questions for the experimental items in Experiment 2 were answered correctly in 77% of the trials. The percentages of correct answers for each condition are presented in Table 3. A 2 × 2 ANOVA (SC/RC Structure × Ambiguous/Unambiguous) revealed no main effects and no interactions (Fs < 1).

Reading times

Only items with correctly answered comprehension questions were analyzed. Residual RT data points that were greater than 3 standard deviations from the mean were excluded from the analysis, affecting less than 1.8% of the data for Experiment 2 overall. Fig. 2 presents the mean residual RTs (ms/word) across the four conditions in this experiment. For the purpose of region presentations and data analysis, we divided the materials into six regions, as shown in (10):

The first region consisted of the matrix subject NP. The second region occurred only in the unambiguous SC condition, and consisted of the verb which disambiguated the clause following the matrix subject as an SC. The third region consisted of the word "that" or "which." The fourth region consisted of the subject of the embedded clause together with the following RC. This is the critical region of analysis for the storage cost hypothesis. The fifth region consisted of the embedded verb phrase, which consisted of a single verb in the RC conditions, or a verb plus its direct object in the SC conditions. The sixth and final region consisted of the main clause verb and the following two words. Table 4 presents the mean RTs per word for these regions, in both residual and raw forms.

Reading times in the critical region

Across the critical region "the cop who the mobster attacked," a 2 × 2 ANOVA (SC/RC Structure × Ambiguous/Unambiguous) revealed a significant main effect of ambiguity ($F1(1,43) = 5.524$, $MS_{within} = 2415$, $p < .03$; $F2(1,19) = 12.65$, $MS_{within} = 650.2$, $p < .003$), a significant effect of structure by items only ($F1(1,43) = 1.750$, $MS_{within} = 2078$, $p = .193$; $F2(1,19) = 4.778$, $MS_{within} = 1389$, $p < .05$), and an interaction that was marginally significant by participants and significant by items ($F1(1,43) = 3.874$, $MS_{within} = 2114$, $p = .056$; $F2(1,19) = 7.534$, $MS_{within} = 882.3$, $p < .02$). Subsequent planned pair-wise comparisons showed that residual RTs were slower for the unambiguous RC condition than any of the other three conditions. Most importantly, unambigu-

(10)

The claim	(alleging)	that/which	**the cop who the mobster attacked**	ignored (the informant)	might have ...
Matrix Subject	(verb-ing)	that/which	Critical Clause	Embedded VP	Matrix VP

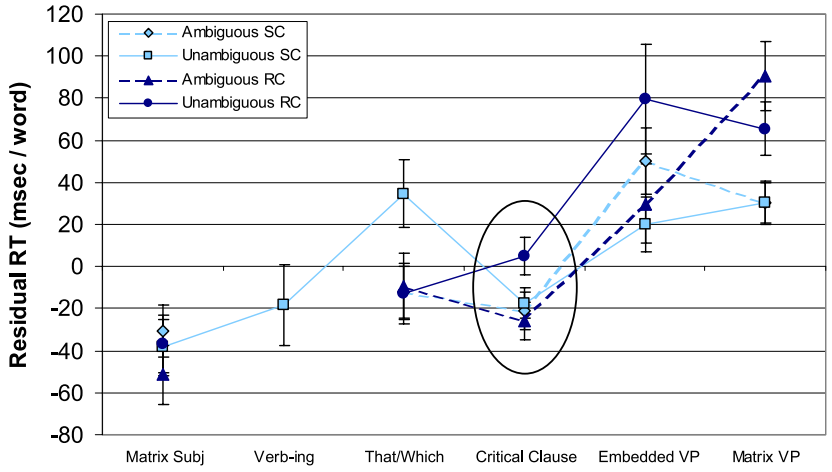

Fig. 2. Plot of mean (standard error) residual RTs per word by region by participants in Experiment 2.

Table 4
Mean residual RTs (ms/word) by participants as a function of SC/RC structure type and ambiguity, for the six main regions (raw RTs in parentheses)

	SC		RC	
	Ambiguous	Unambiguous	Ambiguous	Unambiguous
Matrix Subject	−31 (348)	−39 (348)	−51 (338)	−37 (349)
Verb-ing		−18 (402)		
That/Which	−12 (355)	35 (400)	−9 (359)	−13 (371)
Critical Clause	−21 (358)	−18 (360)	−26 (353)	5 (382)
Embedded VP	50 (437)	20 (411)	30 (430)	80 (491)
Matrix VP	30 (409)	30 (397)	91 (470)	66 (445)

ous RC sentences had significantly longer residual RTs than the unambiguous SC sentences as predicted by the storage cost theories ($F1(1,43) = 5.136$, $MS_{within} = 2214$, $p < .03$; $F2(1,19) = 9.593$, $MS_{within} = 1385$, $p < .006$). In addition, unambiguous RC sentences were significantly slower than ambiguous SC sentences ($F1(1,43) = 6.537$, $MS_{within} = 2364$, $p < .02$; $F2(1,19) = 13.97$, $MS_{within} = 1061$, $p < .002$) as well as ambiguous RC sentences ($F1(1,43) = 7.332$, $MS_{within} = 2893$ $p < .01$; $F2(1,19) = 18.77$, $MS_{within} = 789.9$, $p < .0005$). For all other comparisons, Fs < 1. In all of the comparisons described above for the critical region, similar significance patterns were found for raw RTs, except for the comparison between unambiguous RC and unambiguous SC sentences, which was marginal in the items analysis ($F1(1,43) = 4.578$, $MS_{within} = 2321$, $p < .04$; $F2(1,19) = 4.135$, $MS_{within} = 2370$, $p = .056$).

One potential alternative explanation of the difference between the unambiguous RC and unambiguous SC conditions during the critical region is that people may process the RC continuation more slowly because of the lexical bias in the SC nouns that were used, favoring an SC interpretation. That is, the difficulty that people experienced in processing the RC structure could have been due to a preference to have an SC following the head noun, rather than an RC, an effect that could potentially exist through the RC itself. To test this potential explanation, we divided the items into two groups: those with a strong bias toward SC completions (9 items, averaging 75% SC completions using the Pearlmutter & Mendelsohn norms[4]) and those with no bias toward either SC or RC completions (11 items, averaging 53% SC completions using the Pearlmutter and Mendelsohn norms). If the lexical-preference explanation of the RC difficulty is correct, then we should expect a stronger effect in the items that are more strongly SC-biased and we should see little to no effect in the equi-biased set of items. Contrary to this prediction, there was a numerical trend in the reverse direction. Despite having half as much data overall, and half as many items, the effects that were observed with all items still remained in the equi-biased subset of 11 items. In particular, the unambiguous RC condition was read slower than the unambiguous SC condition in this set of 11 items, marginally by participants and significant by items ($F1(1,40) = 3.110$, $MS_{within} = 4844$, $p = .085$; $F2(1,10) = 6.613$, $MS_{within} = 1935$, $p < .03$), and were slower than either of the ambiguous conditions (vs. ambiguous SC: $F1(1,40) = 3.756$, $MS_{within} = 4469$, $p = .060$; $F2(1,10) = 4.814$, $MS_{within} = 1954$, $p = .053$; vs. ambiguous RC: $F1(1,40) = 5.032$, $MS_{within} = 6241$ $p < .05$; $F2(1,10) = 16.63$, $MS_{within} = 1011$, $p < .003$). (Note that the data of three additional subjects had to be omitted in these analyses because of insufficient correct responses in at least one condition). In contrast, many of the comparisons in the more strongly SC-biased item set were not quite significant, although the means were numerically in the predicted direction. For example, the comparison between the unambiguous SC and unambiguous RC conditions was numerically present, but was not significant by participants or items (Fs < 1.7). These results provide strong evidence against the lexical-bias interpretation of the effects in the critical region.

We did one further analysis in order to test the lexical-bias explanation of the elevated RTs in the unambiguous RC condition. We tested for a correlation between the SC-bias of an item, as provided by the Pearlmutter and Mendelsohn (1999) completion norms in Appendix B, and the RT difference score for each item, as determined by subtracting the mean residual RT during the critical region of the unambiguous SC condition from the unambiguous RC condition. Contrary to the prediction of the lexical-bias explanation, there was a negative correlation of 0.27 ($r^2 = .07$; $p > .24$) between SC-bias and the RT difference score in the region. This correlation was non-significant and was in the reverse direction to that predicted by the lexical-bias hypothesis.

[4] The Kennison (2000) norms could not be used to evaluate our items because many of the SC-biased nouns were not included in Kennison's set.

Table 5
Residual (raw) reading times during the embedded RC region as a function of condition

	SC	RC of the Critical
Ambiguous	−10 (372)	−23 (356)
Unambiguous	−9 (369)	16 (397)

A second potential alternative account of the slower processing of the unambiguous RC condition compared to the unambiguous SC condition is that the RCs are necessarily object-extracted ("which the cop…ignored") and hence have a rare English word order (OVS), whereas the SCs have regular SVO word order. Dynamic models, such as simple recurrent networks (e.g., Elman, 1991), predict an advantage of regular (the predominant main clause word order) over irregular word order, irrespective of storage cost (e.g., MacDonald & Christiansen, 2002). Word order regularity is therefore a confounding factor when analyzing a region containing the embedded subject.[5]

To test the storage cost hypothesis independent of the word-order regularity hypothesis, we conducted an analysis of the RC following the embedded subject NP. The word order regularity hypothesis makes no predictions in this portion of the critical region, because the word order is equally unusual in both conditions. In contrast, the storage cost hypothesis predicts that RTs should be slower for the unambiguous RC condition compared to the unambiguous SC condition. RTs for the most embedded RC region are presented in Table 5:

A 2×2 ANOVA conducted on this region revealed a main effect of ambiguity ($F1(1,43) = 5.909$, $MS_{\text{within}} = 3084$, $p < .02$; $F2(1,19) = 9.096$, $MS_{\text{within}} = 927.3$, $p < .008$), no significant effect of structure ($F1 < 1$; $F1(1,19) = 1.718$, $MS_{\text{within}} = 1611$, $p = .206$), and a significant interaction ($F1(1,43) = 7.012$, $MS_{\text{within}} = 2320$, $p < .02$; $F1(1,19) = 11.96$, $MS_{\text{within}} = 997.7$, $p < .003$). Planned pair-wise comparisons showed that the unambiguous RC condition was read more slowly than any of the other conditions (vs. the unambiguous SC condition: $F1(1,43) = 5.695$, $MS_{\text{within}} = 2547$, $p < .03$; $F2(1,19) = 9.665$, $MS_{\text{within}} = 1355$, $p < .006$; vs. the ambiguous RC condition: $F1(1,43) = 10.77$, $MS_{\text{within}} = 3200$, $p < .003$; $F2(1,19) = 22.15$, $MS_{\text{within}} = 912.4$, $p < .0003$; vs. the ambiguous SC condition: $F1(1,43) = 4.684$, $MS_{\text{within}} = 3374$, $p < .04$; $F2(1,19) = 7.048$, $MS_{\text{within}} = 1480$, $p < .02$). For all other comparisons, Fs < 1.8, $ps > .19$. These results were predicted by the storage cost hypothesis, but not by the word-order regularity hypothesis.

[5] We would like to thank Lars Konieczny for bringing this issue to our attention, and for suggesting ways to disentangle the predictions of the different models.

Reading times in the post-critical region

The comparisons involving the temporarily ambiguous sentences suggest that the SC interpretation was being followed during the critical region because RTs for the temporarily ambiguous sentences were very similar to those for the unambiguous SC sentences. One potential implication of this hypothesis is that RTs should be highest for the ambiguous RC sentences in the matrix VP region when it is discovered that the target structure is an RC, not an SC. In a 2×2 ANOVA (SC/RC Structure × Ambiguous/Unambiguous) at the matrix VP region, there was a significant main effect of structure in which the combined RC conditions are slower than the combined SC conditions ($F1(1,43) = 20.82$, $MS_{\text{within}} = 4810$, $p < .0001$; $F2(1,19) = 15.54$, $MS_{\text{within}} = 3374$, $p < .001$). This demonstrates the difficulty of the RC relative to the SC, plausibly because of the additional integration that is required here for the RC (the object position of the verb back to the relative pronoun) as compared to the SC (Gibson, 1998, 2000). All other effects were not significant (Fs < 1.5, ps $> .23$). Furthermore, pair-wise comparisons revealed that the ambiguous RC condition was significantly slower than ambiguous SC condition ($F1(1,43) = 11.64$, $MS_{\text{within}} = 6858$, $p < .002$; $F2(1,19) = 11.66$, $MS_{\text{within}} = 2769$, $p < .003$). Although the ambiguous RC condition was numerically slower than the unambiguous RC condition, a pair-wise comparison between the two revealed no significant difference (Fs < 1.7, ps $> .20$), possibly because of a ceiling effect on RTs: RTs are very long here, reflecting the integrations that are needed at the verbs in a doubly nested structure. It may be difficult to measure ambiguity effects on top of these complexity effects as well as any potential end-of-sentence effects for this final clause.

Discussion

There were two main results from this experiment. First, reading times were longer for the unambiguous RC condition than for the unambiguous SC condition across the critical embedded clause material. This result supports the hypothesis that there is a cost associated with storing a predicted wh-trace/incomplete dependency across this region, consistent with earlier results from Wanner and Maratsos (1978). The result is predicted by a storage cost theory which associates cost with predicted syntactic heads, as long as there are empty-categories mediating wh-filler-gap dependencies. The result is also predicted by a storage cost theory that associates cost with each incomplete dependency, but is not predicted by a storage cost theory that associates cost only with predicted verbs or incomplete clauses. Additional analyses demonstrated that the results cannot be explained in terms of different degrees of SC vs. RC lexical expectations following the different nouns that were used in the experiment. Furthermore, analyses

of the region consisting of the most embedded RC region suggest that the results are best explained in terms of a storage cost hypothesis, and not a word-order regularity hypothesis (e.g., MacDonald & Christiansen, 2002).

Second, reading times for the temporarily ambiguous SC/RC conditions were significantly faster than for the unambiguous RC condition during the critical region, and did not differ from the unambiguous SC condition in this region. This result supports the hypothesis that people were generally following the SC interpretation of the ambiguity during the ambiguous region, as predicted by both the lexical biases of the particular nouns used, and syntactic heuristics such as minimizing storage costs at an ambiguous choice point. The result provides evidence that suggests that the results of Experiment 1—the increased reading times in SCs of nouns as compared with verbs—were probably not due to the presence of a temporary ambiguity in the nominal SC conditions. Thus, the interpretation of Experiment 1 in terms of syntactic storage costs remains most plausible.

In addition to these two main results, the fact that the ambiguous SC/RC versions were read as fast as the unambiguous SC versions—in fact, the ambiguous versions were read numerically, but non-significantly *faster* than the unambiguous SC versions—provides evidence for the serial parsing hypothesis over the parallel parsing hypothesis within these structures (Gibson & Pearlmutter, 2000; Lewis, 2000). As Lewis (2000) notes, if there is a cost associated with maintaining a structure, then retaining multiple structures should be more costly than retaining a single structure. The experiments presented here provide reading time evidence that there are syntactic storage costs associated with each predicted syntactic head within a structure. In particular, the reading time results suggest that there is a cost for keeping track of the wh-filler dependency/expectation. If people were able to pursue both the SC and the RC structures for the input string simultaneously in a temporarily ambiguous input, then the temporarily ambiguous versions should be processed more slowly than the unambiguous SC versions—perhaps as slowly as the unambiguous RC versions—because of the cost of the retaining the wh-filler dependency in the RC version (cf., Pearlmutter & Mendelsohn, 1999, for data from a different experimental design which is suggestive that both the SC and the RC versions may be pursued simultaneously). Note that even if the human parser represents multiple structures for the input string by sharing their internal structures as much as possible without compromising differences in meaning (e.g., the embedded RC and SC structures in the current experimental items might share the structure for the top-level clause "the claim . . . might have affected . . .") as in the data structures used in chart parsing (Earley, 1970), there still needs to be a component of the RC structure which keeps track of the open wh-filler dependency. Thus, a parallel parsing hypothesis predicts slower reading times in these structures for the temporarily ambiguous versions compared to the unambiguous versions, especially the unambiguous SC. This was definitively not the case in Experiment 2, suggesting that only the SC version was pursued. These results are consistent with gathering evidence that suggests that the human sentence processor is serial in nature, stochastically following the best interpretation at each choice point (Gibson, submitted; Traxler et al., 1998; van Gompel, Pickering, & Traxler, 2000, 2001).

Although the first result—that the unambiguous RC was read more slowly than the unambiguous SC—is as predicted by various storage cost hypotheses, there is an alternative explanation that is worth considering. The design of the items included another RC modifying the embedded subject of the SC/RC (e.g., "who the mobster attacked" in (8) and (9)). The inclusion of the second RC could have caused interference between the two wh-filler pronouns for the RC condition, such that both are seeking an embedded position to be interpreted with (cf., Lewis, 1996; van Dyke & Lewis, 2003). The longer reading times in this region may therefore have been due to this interference, and not necessarily due to one wh-filler dependency/predicted empty category on its own.

To test this possibility, we can analyze the residual RTs of just the embedded subject NP (e.g., "the cop" in (8)) as seen in Table 6.

A comparison of the residual RTs between the unambiguous RC and SC conditions reveals that the RC condition was read more slowly in this region by the items analysis but not significantly by the participants analysis ($F1(1,43) = 1.402$, $MS_{\text{within}} = 4981$, $p = .243$; $F2(1,19) = 4.718$, $MS_{\text{within}} = 2757$, $p < .05$). Because the ambiguous versions patterned with the unambiguous SC condition, we can also compare the unambiguous RC condition to the ambiguous versions in this region. Such a comparison reveals that the unambiguous RC condition was read significantly more slowly than either ambiguous conditions (vs. ambiguous SC: $F1(1,43) = 5.066$, $MS_{\text{within}} = 2868$, $p < .03$; $F2(1,19) = 6.644$, $MS_{\text{within}} = 3726$, $p < .02$; vs. ambiguous RC: $F1(1,43) = 6.426$, $MS_{\text{within}} = 4053$, $p < .02$; $F2(1,19) = 10.06$, $MS_{\text{within}} = 2122$, $p < .006$). These analyses suggest that the difference between SC and RC conditions is not due to the interference of the extra wh-filler in the target region, but is due to storing the wh-filler dependency,

Table 6
Residual (raw) reading times during the embedded subject NP as a function of condition

	SC	RC
Ambiguous	−42 (335)	−50 (328)
Unambiguous	−34 (342)	−16 (361)

whether or not there is an additional filler to follow. But these analyses are only suggestive, because of the non-significance of the unambiguous RC vs. SC analysis by participants. Experiment 3 was conducted in order to address this issue more directly.

Experiment 3

The results of Experiment 2 suggest that there is a storage cost associated with retaining a wh-filler dependency across linguistic material. However, the materials in the unambiguous RC condition in Experiment 2 were complex, consisting of doubly nested RC materials. In this experiment, we seek evidence of storage costs for a wh-filler dependency using simpler materials, without the extra nested RC.

Methods

Participants

Twenty-nine participants from the MIT community who did not take part in Experiments 1 or 2 were paid for their involvement. All were native speakers of English and were naïve as to the purposes of the study.

Materials and design

Sixty sentences were constructed as in (11) with two conditions each: an unambiguous RC condition in which the wh-pronoun "which" followed the initial noun and a temporarily ambiguous SC condition in which the complementizer "that" followed the initial noun:[6]

(11) a. SC Structure:
The announcement [that **the baker from a small bakery in New York City** received the award] helped the business of the owner.
b. RC Structure:
The announcement [which **the baker from a small bakery in New York City** received ___] helped the business of the owner.

The head nouns in the items (e.g., "announcement" in (11)) were heavily biased towards the SC completion. Each of the 19 nouns that were used had at least an 80% SC-bias, averaging 86%, from the norms of Kennison (2000). Appendix C provides a complete list of the stimuli along with the SC-bias percentage for each matrix subject noun.

We elected to use temporarily ambiguous SC structures instead of fully unambiguous SC structures as in Experiment 2 (i.e., not including extra verbs like "stating" or "claiming") because temporarily ambiguous SC structures differ from their unambiguous RC counterparts more minimally than the fully unambiguous SC structures do (i.e., only in the use of the complementizer "that" in place of the relative pronoun "which"). The results from Experiment 2 demonstrated that there is no difference between how people process ambiguous SC structures with SC-biased head nouns and their unambiguous SC counterparts during the target embedded subject region, so the temporarily ambiguous SC structure is likely to be a good control for the RC. In any case, if we do find the storage cost difference that is predicted, then this will provide evidence that the people are following the SC reading in the temporarily ambiguous versions.

The target sentences were split into two lists balancing all factors in a Latin-Square design. Each list was combined with 80 fillers of various types. The stimuli were pseudo-randomized independently for each participant such that at least one filler item separated any two targets. Nineteen different head nouns were used, with at most four items out of the 60 total items containing the same head noun. Appendix C provides a complete list of the stimuli along with the individual SC-completion percentage for each noun. These items were balanced across lists so that people saw at most two of the same head noun in an SC or an RC version.

Procedure

The procedure was identical to Experiment 2.

Results

One participant's data were omitted from all analyses because of poor comprehension question performance (<67% accuracy overall).

Comprehension question performance

Overall, the comprehension questions for the experimental items in Experiment 2 were answered correctly in 86% of the trials. The percentages of correct answers for each condition are presented in Table 7. A pair-wise *t* test revealed no significant effects (Fs < 1)

Table 7
Mean (standard error) comprehension question performance in percent correct as a function of condition

Ambiguous SC	Unambiguous RC
85 (1.6)	86 (1.7)

[6] A large set of items was used in this experiment because of plans to use the same materials later in a brain imaging experiment investigating similar issues. A greater number of trials per condition are needed in brain imaging experiments than in reading time experiments.

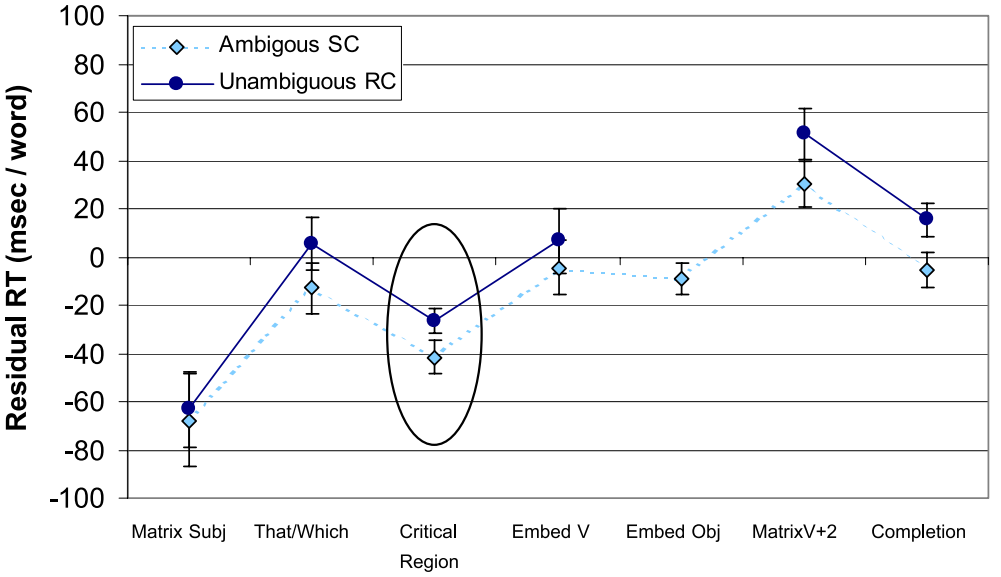

Fig. 3. Plot of mean (standard error) residual RTs per word by region by participants in Experiment 3.

Reading times

Only items with correctly answered comprehension questions were analyzed. Residual RT data points that were greater than 5 standard deviations from the mean were excluded from the analysis, affecting less than 0.6% of the data. Fig. 3 presents the mean residual RTs (ms/word) across the two conditions. The sentences were divided into seven regions for presentation: (1) the matrix subject NP; (2) the word "that" or "which"; (3) the critical region consisting of the embedded subject NP with the attached PP; (4) the embedded verb; (5) the embedded object NP (for the SC condition); (6) the matrix verb and the first two words following; and (7) the final words completing each sentence. Table 8 presents the mean RTs per word for these regions, in both residual and raw forms.

Reading times in the first two regions

Residual RTs at the matrix subject were not significantly different between the two conditions (Fs < 1). At the region consisting of "that" or "which," residual RTs tended to be slower for the RC condition than the SC condition, marginally significant by participants but non-significant by items ($F1(1,27) = 4.068$, $MS_{\text{within}} = 1164$, $p = .054$; $F2(1,59) = 1.603$, $MS_{\text{within}} = 3462$, $p = .211$).

Reading times in the critical region

In the critical region, the RC condition was read significantly more slowly that the SC condition ($F1(1,27) = 8.114$, $MS_{\text{within}} = 400.2$, $p < .009$; $F2(1,59) = 4.793$, $MS_{\text{within}} = 928.5$, $p < .04$) as predicted by the storage cost hypothesis. The raw RTs for the critical region were also significantly different.

Reading times in the post-critical region

Analyses at the embedded verb revealed that residual RTs for the RC condition numerically larger than but not significantly different from the SC condition (Fs < 1). At the region that included the matrix verb plus the next two words, residual RTs for the RC condition were slower than the SC condition, marginally by participants and significant by items ($F1(1,27) = 3.207$, $MS_{\text{within}} = 1950$, $p = .085$; $F2(1, 59) = 4.047$, $MS_{\text{within}} = 3427$, $p < .05$). This result is consistent with the results of Experiment 2. One plausible cause for the difference in RTs at the main verb is due to spill-over from the embedded verb. According to the integration component of the DLT, there are two integrations to be performed at the embedded verb in the RC condition (the subject–verb and object–trace—relative–pronoun integrations) as compared with only a single integration in the SC condition (the subject–verb integration). The two integrations lead to greater complexity on the embedded verb, complexity which may be measured on the following words.

Discussion

The RC condition was read more slowly than the SC condition during the critical region, as predicted by the syntactic storage hypothesis. These results replicate Experiment 2 using less complex materials, without a second embedded RC. Thus, there appears to be an on-line storage cost associated with an incomplete wh-filler dependency or the expectation for an upcoming wh-trace.

Table 8
Mean residual RTs (ms/word) by participants as a function of condition, for the seven regions in Experiment 3 (raw RTs in parentheses)

	Ambiguous SC	Unambiguous RC
Matrix Subject	−68 (357)	−63 (362)
That/Which	−13 (387)	5 (419)
Critical Clause	−42 (370)	−26 (384)
Embedded V	−4 (441)	7 (443)
Embedded Object	−9 (401)	
Matrix V + 2	30 (458)	51 (480)
Completion	−5 (402)	15 (423)

General discussion

There is gathering evidence that the word-by-word interpretation of sentences is sensitive to a number of informational constraints—such as lexical frequency, syntactic knowledge, real-world knowledge, discourse context and intonation—and further constrained by the available working memory resources (Gibson & Pearlmutter, 1998; Tanenhaus & Trueswell, 1995). This paper provides reading evidence from English that the working memory resources, which constrain the representations that are built and/or preferred, are sensitive to the number of predicted syntactic heads or partially processed syntactic dependencies within a linguistic structure. Importantly, the results from the experiments in this paper demonstrate that the human parser is sensitive to syntactic expectations that are derived from the current input (Elman, 1991; Gibson, 1991, 1998), indicating that the parser is top-down in some respects.

The results from Experiment 1 demonstrated that the expectation of each predicted verb is associated with a processing cost as measured by on-line reading times. The results from Experiments 2 and 3 demonstrated that a similar storage cost is associated with keeping track of a wh-filler-gap dependency. These results, especially those of Experiments 1 and 3, demonstrate that storage costs are not restricted to complex doubly nested cases: expectation costs are also measurable in singly embedded structures compared to non-embedded structures. Thus, the expectation costs are not only the result of interference between the same types of expectation, although such interference may exacerbate the effect. It seems then that the English parser keeps track of predicted syntactic heads or incomplete syntactic dependencies, and that there is some cost associated with maintaining these expectations.

Several questions are raised by the results of these experiments. First, what other kinds of syntactic expectations are associated with processing cost? So far, the results of these experiments have demonstrated that there is a cost associated with predicting verbs and wh-traces following verbs. It is an open question how general these expectations are (cf., Shapiro, Zurif, & Grimshaw, 1987, who found effects of expected arguments following a verb, and Rayner & Duffy, 1986, and Schmauder, Kennison, & Clifton, 1991, who found no such effects using better methods and materials). Some preliminary work from our lab suggests that there may also be storage cost associated with predicting arguments of verbs in the region following the verb. In particular, using self-paced reading, we compared the region following obligatory ditransitive verbs like "give" to a similar region following obligatory transitive verbs like "publish," as in (12):

(12) a. Obligatory Ditransitive Verb:
Mary gave **a book which had impressed some critics** who worked for a magazine to a young child.
b. Obligatory Transitive Verb:
Mary published **a book which had impressed some critics** who worked for a magazine.

In the obligatory ditransitive condition in (12a), the complex NP "a book which had impressed some critics who worked for a magazine" is the object of the verb "gave." When an inanimate NP like "a book" follows the verb "gave," a PP goal argument is required further downstream. In contrast, no PP argument of a transitive verb like "published" is possible following the object NP "a book" in (12b). Reading times for the object NP "the book which had impressed some critics" were correspondingly faster during this region for the obligatory transitive condition as compared with the obligatory ditransitive condition, as predicted by storage cost theories. This preliminary result provides further support for storage costs in English.[7]

A second question that arises from our results is whether or not storage costs are the same cross-linguistically. This is also an open question, but there is preliminary evidence from Japanese, a head-final language, that suggests that there is a storage cost for open verbal predictions (Miyamoto, 2002; Nakatani & Gibson, 2003). Results from Chinese also support the existence of storage costs in that language (Hsiao & Gibson, 2003).

A third question that is raised by these experiments is what the nature is of the expectations that the parser is

[7] The result is only preliminary because the items were not properly normed for plausibility differences prior to running the experiment. A subset of 15 of the 20 items in which plausibility was matched between the conditions showed the same significant differences, but this result needs to be replicated on a greater set of items before it is fully convincing.

maintaining. Together, the results of our experiments are potentially consistent with a variety of on-line syntactic storage cost theories. We evaluate our experimental results together with other results in the literature with respect to four potential theories. These theories differ in terms of what types of representations incur storage cost at a parser state as follows: (1) incomplete clauses/predicted verbs (Kimball, 1973); (2) incomplete dependencies, thematic role assignments or case assignments (Gibson, 1991; Hakuta, 1981; Lewis, 1996; Stabler, 1994); (3) incomplete phrase structure (PS) rules (Chomsky & Miller, 1963); or (4) predicted syntactic heads/categories (Gibson, 1998).

The first theory, in which on-line syntactic storage cost is indexed by the number of incomplete clauses or predicted verbs at a parser state (Kimball, 1973), is insufficient to account for the results of both Experiments 2 and 3, because wh-filler dependencies and predicted PP dependencies are associated with storage cost independent of the number or incomplete clauses. Although it is possible that there are multiple different types of syntactic storage that are associated with on-line cost, such a theory is dispreferred to a simpler theory in which the same underlying storage process is responsible for both results.

Under the second class of theories, there is a storage cost for each incomplete syntactic dependency, thematic role assignment or case assignment (Gibson, 1991; Hakuta, 1981; Lewis, 1996; Stabler, 1994). Each of these theories can correctly account for the results of all three experiments presented here. Thus, this class of theories can account for the observed data under a uniform theoretical account. However, such a theory is difficult to reconcile with existing results from the literature with respect to the processing of head-final constructions. Such a theory predicts an increasing processing load in head-final constructions, across the dependents of a head, which appears in final position. For example, in an SOV language like Japanese, processing is predicted to be slower over the processing of embedded elements depending on the number of dependents of the verb. Thus, processing embedded material (e.g., an adverbial or an optional prepositional phrase) should be fastest when preceding verbs with one argument (intransitives), slower for verbs with two arguments (transitives) and slowest for verbs with three arguments (ditransitives). Whereas processing does slow down when an additional verb/clause is expected (e.g., in Japanese, Miyamoto, 2002), there is no evidence in head-final languages for slowed processing across verbal dependents that do not require the prediction of an additional clause. In fact, processing appears to speed up over the additional arguments and modifiers of an upcoming head in SOV languages (for results in Japanese, see Miyamoto, 2002; and Nakatani & Gibson, 2003; for results in Hindi, see Vasishth, 2002; for results in German, see Konieczny, 2000; and Konieczny & Doring, 2003), perhaps because the verb/verb class becomes more predictable as more of its dependents are processed (Konieczny, 1996). As a result, such a storage cost theory based on incomplete dependencies seems unlikely.

Under the third theory that we will consider, there is a storage cost for each partially processed PS rule at a parse state (Chomsky & Miller, 1963), assuming a fully connected structure for the input as each word is processed. Such a theory can account for the results of Experiment 1, because there is a partially processed PS rule of the form S ⇒ NP VP (i.e., the subject–verb rule) for each incomplete verbal dependency. Such a theory can also extend to account for the results of Experiments 2 and 3 by adopting the slash notation from now standard phrase-structure grammars. Under such context-free phrase structure rule theories—e.g., head-driven phrase structure grammar (Pollard & Sag, 1994) and its predecessor, generalized phrase structure grammar (Gazdar, Klein, Pullum, & Sag, 1985)—wh-fillers are associated with their underlying role-assigning positions by percolating a "slash" feature through intermediate rules until an empty category can be posited in the target position. Because the slash feature indicates a missing constituent, a PS-rule storage cost theory can account for the results of Experiment 2 if rules with active slash features also contribute to storage cost in addition to partially processed rules. Thus, a storage cost theory based on incompletely processed PS rules can account for the results of all three experiments naturally. However, this theory has difficulty in accounting for the data from head-final languages. In particular, it is standardly assumed that subject and object NPs are connected to the verb through different phrase structure rules, even in SOV languages. That is, the subject rule is something like S ⇒ NP VP, whereas the object rule is something like VP ⇒ NP Verb. If this is so, then storage cost should increase through the VP in transitive sentences, after processing the subject and object, but this does not appear to be the case. Thus, the PS-rule based storage cost theory cannot account for the head-final data in much the same way that the incomplete dependency theory cannot.

The final account that we will evaluate is one in which there is storage cost associated with predicted syntactic heads. This theory can account for the results from all three experiments. In Experiment 1, there is a cost associated with predicting a verb to come. In Experiments 2 and 3, there is a cost associated with predicting an empty category to be associated with the wh-filler. In addition, this theory does not make the undesirable prediction of increasing complexity across the dependents of a single head in a head-final language: all the dependents support the prediction of the same head, and therefore there is no increasing cost. As a result, this theory can account for all of the existing data.

An interesting consequence of this discussion is that the combination of results supports a theory that includes empty categories mediating wh-dependencies over one that does not. That is, if we accept that storage costs are indexing predicted syntactic heads rather than incomplete dependencies (because of the evidence from head-final languages), then the only way to account for the results of Experiments 2 and 3 is to assume the existence of a wh-trace, a phonologically empty syntactic head mediating wh-dependencies (Chomsky, 1965, 1981; Fodor, 1978), as indicated by "[NP t_i]" in (13):

(13) a. SC Structure:
The claim alleging [that **the cop** ... ignored the informant] might have affected the jury.
b. RC Structure:
The claim [which$_i$ **the cop** ... ignored [NP t_i]] might have affected the jury.

If there were no wh-trace (or "gap" category), such that the wh-filler dependency were represented via a direct link between the wh-filler and the verb to come (e.g., Pickering & Barry, 1991), then there would be no additional storage cost for the RC condition as compared to the SC condition through the embedded clause, because the verbal head would already be predicted by the initiation of the embedded clause ("that" in the SC; "which" in the RC). Thus, the three on-line English experiments, in conjunction with existing results from the processing of head-final languages, provide indirect evidence for the existence of empty categories in wh-filler dependencies (cf., Gibson & Hickok, 1993; Gibson & Warren, 2004; Pickering & Barry, 1991). It would of course be more convincing if all of the evidence for this important syntactic representational issue came from one language, but the cross-linguistic evidence is parsimoniously suggestive at the moment. Further clever experimentation within a single language is needed to adequately resolve this issue.

Finally, it is worth reiterating a point that was made in the discussion following Experiment 2. In that experiment, it was shown that the temporarily ambiguous SC/RC versions were read at least as fast as the unambiguous SC versions, and faster than the RC versions. This observation, together with the result that there is a cost for retaining the wh-filler expectation, suggests that people are following the SC structure in the temporary ambiguity in the materials in Experiment 2. This result is consistent with gathering evidence that suggests that the human sentence processor is serial in nature, stochastically following the best interpretation at each choice point (Gibson, submitted; Traxler et al., 1998; van Gompel et al., 2000, van Gompel, Pickering, & Traxler, 2001).

Appendix A. Experiment 1 Materials

In the experimental items below, the four conditions of Experiment 1 can be derived as follows. For the zero incomplete verb dependency condition, choose the verb form for the first two clauses (e.g., "the detective suspected" and "the thief knew" in (1)) and the first completion (e.g., "and so he reported immediately to the museum curator"). For the one-late condition, pick the verb form for the first clause and the nominalized form for the second (e.g., "the knowledge") and the second completion (e.g., "came from an insider"). For the one-early condition, pick the nominalized for in the first clause (e.g., "the suspicion"), the verb form of the second clause (e.g., "the thief knew"), and the third completion (e.g., "worried the museum curator"). For the two incomplete verb dependencies condition, pick the nominalized form of the two initial clauses and the last completion (e.g., "came from an insider worried the museum curator"). The number following each nominalization is the percent SC continuations from the completion norming data of Kennison (2000), where available.

1. (The detective suspected / The suspicion (82%)) that (the thief knew / the knowledge (–%)) that the guard protected the jewels (and so he reported immediately to the museum curator. / came from an insider. / worried the museum curator. / came from an insider worried the museum curator.)
2. (The dictator insisted / The insistence (88%)) that (the country acknowledge / the acknowledgement (60%)) that the army violated the treaty (because he felt enraged. / be made public. / threatened the unstable peace. / be made public threatened the unstable peace.)
3. (The employee realized / The realization (90%)) that (the boss implied / the implication (76%)) that the company planned a layoff (and so he sought alternative employment. / had been unintentional. / caused a panic. / had been unintentional caused a panic.)
4. (The clerk recommended / The recommendation (51%)) that (the customer should complain / the complaint (70%)) that the product was faulty (and so he pointed to Customer Service. / should be filed immediately. / was taken seriously. / should be filed immediately was taken seriously.)
5. (The mother sensed / The sense (60%)) that (the child feared / the fear (66%)) that a monster might eat little boys (and so she kept a light on. / kept the boy up at night. / worried the parents greatly. / kept the boy up at night worried the parents greatly.)
6. (The psychiatrist worried / The worry (73%)) that (the patient felt / the feeling (65%)) that everyone deserved to die (and so he quickly called hospital security. / could not be helped. / concerned the family. / could not be helped concerned the family.)
7. (The lawyer acknowledged / The acknowledgement (60%)) that (the defendant had hinted / the hint (57%)) that the mob bribed the official (and so he went with a new line of questioning. / had been ignored completely. / was recorded by the court. / had been ignored completely was recorded by the court.)
8. (The producer doubted / The doubt (76%)) that (the director would realize/ the realization (90%)) that the actress hated the lead actor (and so he immediately sent a messenger. / might eventually cause trouble. / puzzled the motion picture executives. / might eventually cause trouble puzzled the motion picture executives.)

9. (The counselor implied / The implication (76%)) that (the teacher should know / the knowledge (–%)) that the student had a disability (and so he took matters into his own hand. / should help the school in dealing with the student. / was not given enough emphasis. / should help the school in dealing with the student was not given enough emphasis.)
10. (The author contended / The contention (60%)) that (the publisher predicted / the prediction (90%)) that the novel would be a success (and so he planned to buy a new car. / was nothing more than adulation. / amused the public. / was nothing more than adulation amused the public.)
11. (The banker worried / The worry (73%)) that (the investor sensed / the sense (60%)) that the company was performing poorly (and so he called to offer reassurance. / might influence stock prices. / was well founded. / might influence stock prices was well founded.)
12. (The king doubted / The doubt (76%)) that (the squire felt / the feeling (65%)) that the knight would win the joust (and so he reprimanded the young boy. / was purely emotional. / proved to be true. / was purely emotional proved to be true.)
13. (The historian hypothesized / The hypothesis (57%)) that (the emperor believed / the belief (82%)) that the elixir gave long life (and so he wrote an essay on the topic. / was promoted by alchemists. / was not a big surprise. / was promoted by alchemists was not a big surprise.)
14. (The lifeguard thought / The thought (80%)) that (the swimmer implied / the implication (76%)) that a shark was in the waters (and so he yelled for everyone to get out. / might have been a joke. / made everyone angry. / might have been a joke made everyone angry.)
15. (The family knew / The knowledge (–%)) that (the weatherman predicted / the prediction (90%)) that the storm would cause severe damage (and so they boarded up their windows. / would be accurate. / caused some panic. / would be accurate caused some panic.)
16. (The homeowner suggested / The suggestion (73%)) that (the contractor should request / the request (69%)) that the carpenter use maple instead (and so the project was delayed. / was an excellent decision. / made everyone happy. / was an excellent decision made everyone happy.)
17. (The magician realized / The realization (90%)) that (the assistant feared / the fear (66%)) that the guillotine was actually real (and so he removed the trick from the act. / was not an act. / came almost too late. / was not an act came almost too late.)
18. (The conductor claimed / The claim (81%)) that (the engineer indicated / the indication (80%)) that the train was in good condition (and so the accident was not his fault. / was a lie. / became a front page story. / was a lie became a front page story.)
19. (The brother thought / The thought (80%)) that (the sister hoped / the hope (86%)) that the couple would break up (and so everyone was soon gossiping. / was fruitless. / made everyone laugh. / would be fruitless made everyone laugh.)
20. (The chauffeur suspected / The suspicion (82%)) that (the executive wished / the wish (53%)) that the documents would be shredded (and so he discreetly took them. / was part of the scandal. / would become a huge story. / was part of the scandal would become a huge story.)
21. (The paramedic feared / The fear (66%)) that (the fireman thought / the thought (80%)) that the wound was not serious (and so he injected a strong sedative. / would result in further injury. / was precautionary. / would result in further injury was precautionary.)
22. (The tourist implied / The implication (76%)) that (the waiter suggested / the suggestion (73%)) that the tip was too small (and so the manager apologized profusely. / was inappropriate. / concerned the manager. / was inappropriate concerned the manager.)
23. (The athlete contended / The contention (60%)) that (the trainer claimed / the claim (81%)) that the injury was not too severe (and so the coach put him back in the game. / was gross malpractice. / concerned the general manager. / was gross malpractice concerned the general manager.)
24. (The photographer sensed / The sense (60%)) that (the model worried / the worry (73%)) that the photos were not tasteful (and so he tried to convince her. / could not be helped. / would delay the shoot. / could not be changed would delay the shoot.)
25. (The movers wished / The wish (53%)) that (the family knew / the knowledge (–%)) that the boxes were too heavy (and so they complained loudly while working. / would mean a bigger tip. / went unfulfilled. / would mean a bigger tip went unfulfilled.)
26. (The caddie hoped / The hope (86%)) that (the golfer sensed / the sense (60%)) that the wind was blowing eastwards (and so he pulled out a pitching wedge. / was correct. / turned out to be in vain. / was correct turned out to be in vain.)
27. (The butler recommended / The recommendation (51%)) that (the gardener should insist / the insistence (88%)) that the master relandscape the property (and so he set up a meeting. / should be considered. / fell on deaf ears. / should be considered fell on deaf ears.)
28. (The sailor indicated / The indication (80%)) that (the diver complained / the complaint (70%)) that the tanks were half full (and so the captain ordered new ones brought up. / had been noted already. / worried the admiral. / had been noted already worried the admiral.)
29. (The brewer hoped / The hope (86%)) that (the merchant would sense / the sense (60%)) sense that the beer was made from fine ingredients (and so he poured a sample glass. / meant more business. / was perfectly within reason. / meant more business was perfectly within reason.)
30. (The grocer hinted / The hint (57%)) that (the chef thought / the thought (80%)) that the produce was the best in town (and so he beamed a big smile. / was completely accurate. / seemed arrogant. / was completely accurate seemed arrogant.)
31. (The farmer thought / The thought (80%)) that (the winemaker recommended / the recommendation (51%)) that the grapes should be harvested (and so he warmed up the tractor. / was premature. / had been correct. / was premature had been correct.)
32. (The parishioner insinuated / The insinuation (71%)) that (the bishop knew / the knowledge (–%)) that the priest had a criminal history (and so there was a public outcry. / did not result in action. / made the public angry. / did not result in action made the public angry.)
33. (The jockey feared / The fear (66%)) that (the veterinarian implied / the implication (76%)) that the horse was not fit to race (and so he consulted with the owner. / meant disaster. / was an over-reaction. / meant disaster was an over-reaction.)

34. (The lawyer indicated / The indication (80%)) that (the brother claimed / the claim (81%)) that the will had been tampered with (and so the papers were examined closely. / was completely untrue. / calmed everyone down. / was completely untrue calmed everyone down.)
35. (The aide speculated / The speculation (88%)) that (the president would acknowledge / the acknowledgment (60%)) acknowledge that the situation had become worse (and so a speech was prepared. / would come that night. / was more than rumor. / would come that night was more than rumor.)
36. (The principal acknowledged / The acknowledgment (60%)) that (the librarian complained / the complaint (70%)) that the books were falling apart (and so he poured over the budget for more money. / could not be addressed. / spurred the parents to donate. / could not be addressed spurred the parents to donate.)
37. (The chiropractor realized / The realization (90%)) that (the patient feared / the fear (66%)) that the injury might be incurable (and so he tried to be reassuring. / was causing the unpleasantness. / came as an epiphany. / was causing the unpleasantness came as an epiphany.)
38. (The demonstrator doubted / The doubt (76%)) that (the executive felt / the feeling (65%)) that the environment truly was important (and so he hurled rotten eggs at the car. / was sincere at all. / had only been expected. / was sincere at all had only been expected.)
39. (The authority sensed / The sense (60%)) that (the smuggler hoped / the hope (86%)) that the stash would not be found (and so he searched harder. / was becoming desperate. / had only been natural. / was becoming desperate had only been natural.)
40. (The astronomer requested / The request (69%)) that (the technician should promise / the promise (42%)) that the telescope would get daily maintenance (and so he arranged a schedule. / be in writing. / was unreasonable. / be in writing was unreasonable.)

Appendix B. Experiment 2 Materials

The four conditions of Experiment 2 can be derived as follows. Regarding the parenthesized material following the first NP (e.g., "the reason" in (1)), do as follows: for the ambiguous conditions, include "that"; for the unambiguous SC condition, include the present participle (e.g., "implying") plus "that"; and for the unambiguous RC condition, include "which". Regarding the region following the embedded verb (e.g., "ridiculed"), do as follows: for the SC conditions, include the embedded object (e.g., "the newscaster"); for the RC conditions, omit the object. The number following each item number is the percent SC continuations for the head noun, first from Kennison's (2000) completion norming data where available, and then from Pearlmutter and Mendelsohn (1999) completion norming data.

1. (–, 93%) The reason (implying) (that/which) the comedian who the network fired ridiculed (the newscaster) was kept a secret.
2. (84%, 48%) The revelation (showing) (that/which) the executive who the company employed belittled (the secretary) had to be completely ignored.
3. (90%, 59%) The perception (suggesting) (that/which) the photographer who the magazine hired abused (the model) was an issue at the agency.
4. (76%, 60%) The verification (confirming) (that/which) the ranger who the tourists trusted checked (the canoe) came over the walkie-talkie.
5. (38%, 44%) The conclusion (reporting) (that/which) the spy who the CIA pursued reached (the checkpoint) has not yet been confirmed.
6. (82%, 91%) The belief (asserting) (that/which) the terrorists who the UN denounced held (the hostage) was depressing to the negotiators.
7. (–, 60%) The allegation (stating) (that/which) the senator who the army supported neglected (the danger) was leaked to the press.
8. (65%, 60%) The theory (claiming) (that/which) the chief who the lawyers advised defended (the tribe) was proven to be untrue.
9. (81%, 68%) The claim (alleging) (that/which) the cop who the mobster attacked ignored (the informant) might have affected the jury.
10. (80%, 55%) The discovery (revealing) (that/which) the physicist who the FBI funded publicized (the project) should be the top story.
11. (86%, 84%) The hope (suggesting) (that/which) the teenager who the teachers applauded encouraged (the boy) was shared by both parents.
12. (–, 63%) The accusation (implying) (that/which) the guard who the warden questioned refused (the request) was absurd and completely unfounded.
13. (76%, 71%) The implication (stating) (that/which) the philosopher who the speaker cited refuted (his opponents) had to be considered illogical.
14. (58%, 46%) The threat (suggesting) (that/which) the men who the priest protected disregarded (the laws) might have scared the sheriff.
15. (–, 63%) The rule (stating) (that/which) the kids who the teenager watched should follow (the clown) was part of a game.
16. (78%, 67%) The confirmation (indicating) (that/which) the parent who the teacher recruited made (the costumes) was sent to the PTA.
17. (60%, 54%) The acknowledgment (mentioning) (that/which) the fireman who the station contacted got (the truck) did not reach his partner.
18. (–, 47%) The concern (stating) (that/which) the instructor who the students disliked raised (the grades) was discussed by the board.
19. (43%, 45%) The opinion (indicating) (that/which) the runner who the trainer rejected challenged (his opponent) was good for his reputation.
20. (86%, 73%) The assumption (implying) (that/which) the girl who the woman brought along made (the dress) had misled the sewing teacher.

Appendix C. Experiment 3 Materials

The two conditions of Experiment 3 can be derived as follows. For the unambiguous RC condition, include "which" while for the ambiguous SC condition, include "that." Regarding the region following the embedded verb (e.g., "received" in (1)), do as follows: for the unambiguous SC condition, include the embedded object (e.g., "the award"); for the RC conditions, omit the object. The number following each

item number is the percent SC continuations for the head noun from the completion norming data of Kennison (2000), where available.

1. (82%) The announcement (that/which) the baker from a small bakery in New York City received (the award) helped the business of the owner.
2. (82%) The announcement (that/which) the official from the labor union of the controversial company delivered (the contract) concerned the lawyers of the executives.
3. (82%) The announcement (that/which) the actress from the new movie about the dance industry signed (the autographs) excited the fans of ballroom dancing.
4. (82%) The announcement (that/which) the ranger with the inexperienced tourists in the blue tent received (the injury) worried the supervisor of the park.
5. (86%) The assumption (that/which) the reporter from the controversial magazine with the liberal bias made (the remark) disgruntled some members of the board.
6. (86%) The assumption (that/which) the guard at the new jail with the hardened criminals made (the mistake) caused many problems for the warden.
7. (86%) The assumption (that/which) the parents at the important meeting about the school performance made (the costumes) confused the director of the play.
8. (86%) The assumption (that/which) the colonel from the army base with the secret facilities made (the promise) misled the members of the community.
9. (82%) The belief (that/which) the philosopher from the historical period of the bronze age destroyed (the emperor) interested the class of eager students.
10. (82%) The belief (that/which) the terrorists in the nondescript building with many broken windows held (the hostage) concerned the authorities of Homeland Security.
11. (82%) The belief (that/which) the bishop at the international conference in a quiet town encouraged (the contributions) brought much hope to the community.
12. (81%) The claim (that/which) the advocate for the needy people in the impoverished neighborhood made (a difference) convinced the mayor of the city.
13. (81%) The claim (that/which) the reaction of the dangerous chemicals from the unmarked bottles disproved (the theory) confused the assistant of the researcher.
14. (81%) The claim (that/which) the newscaster from the major network with an unusual logo ridiculed (the politician) destroyed the reputation of the show.
15. (81%) The claim (that/which) the cop with the bushy mustache in the old photo ignored (the informant) influenced the decision of the jury.
16. (80%) The discovery (that/which) the friends of the teenage girl at the department store made (the commotion) surprised the shoppers and the salesmen.
17. (80%) The discovery (that/which) the spy from the secret agency of the national government fabricated (the data) infuriated the officers of the military.
18. (80%) The discovery (that/which) the archeologist from the famous expedition to the Egyptian pyramids ignored (the instructions) angered the sponsor of the project.
19. (80%) The discovery (that/which) the official from the biology lab at the secret facility publicized (the project) made the headlines of the newspaper.
20. (100%) The fact (that/which) the countries in the western hemisphere with the popular presidents ignored (the treaty) angered the environmentalists of the world.
21. (100%) The fact (that/which) the veteran from the European country in the World War recorded (some stories) aided the research of the historian.
22. (100%) The fact (that/which) the attorney on the strong defense of the innocent victim presented (the evidence) aided the decision of the jury.
23. (86%) The hope (that/which) the charity for the tragedy survivors from the impoverished town gave (the money) warmed the hearts of the families.
24. (86%) The hope (that/which) the fireman at the remote station in a dangerous district inspired (the teenagers) encouraged the parents of the delinquents.
25. (86%) The hope (that/which) the book about the complicated politics of the developing world gave (constructive criticisms) started serious discussions among the politicians.
26. (80%) The indication (that/which) the student with a Master's degree from a prestigious university received (the fellowship) preceded the awarding of the certificates.
27. (80%) The indication (that/which) the waiter at the Italian restaurant in the Main Square ignored (the request) upset the patron and the manager.
28. (80%) The indication (that/which) the hacker from a secret network in a European country sent (the virus) concerned the head of internet security.
29. (88%) The insistence (that/which) the chief from an Indian reserve with a prosperous casino maintained (the traditions) became a point of serious contention.
30. (82%) The misconception (that/which) the designer of the unique building in the arts district created (the memorial) confused the critics from the newspapers.
31. (82%) The misconception (that/which) the scientists from the large organization with lots of money developed (the toxins) alarmed the president of the country.
32. (82%) The misconception (that/which) the runner at the track meet for the state championship challenged (the decision) tarnished the reputation of the team.
33. (90%) The perception (that/which) the magician in the Vegas showcase at the impressive hotel created (the display) fascinated the crowd of eager tourists.
34. (90%) The perception (that/which) the photographer from the fashion magazine with a global audience ignored (the model) fascinated the readers of the article.
35. (100%) The possibility (that/which) the doctor from the cardiology department at the teaching hospital ignored (the warnings) shocked the head of the university.
36. (100%) The possibility (that/which) the boy with the basketball logo on the oversized shirt discussed (the secret) started an argument in the family.
37. (100%) The possibility (that/which) the instructor from the state university with the rich history raised (the grades) troubled the dean of the department.
38. (90%) The realization (that/which) the monk in the old monastery on the mountain top achieved (true enlightenment) delighted the members of the order.
39. (90%) The realization (that/which) the spy from the security agency of the secretive nation achieved (the objective) validated the years of hard work.

40. (90%) The realization (that/which) the kids at the birthday party in the beautiful park brought (the gifts) aroused the memories of happier times.
41. (86%) The requirement (that/which) the children from the expensive preschool in the rural suburbs should meet (the standards) satisfied the members of the PTA.
42. (86%) The requirement (that/which) the secretary at the software company with the famous CEO should pass (the test) satisfied the manager of the firm.
43. (86%) The requirement (that/which) the principal from the magnet school in the inner city must consider (the applicants) decreased the chances of early admissions.
44. (86%) The requirement (that/which) the prospector with the dangerous equipment for the gold mine should meet (the sheriff) ensured the safety of the townsfolk.
45. (84%) The revelation (that/which) the patient in the rehabilitation program for continued substance abuse shared (their happiness) lifted the spirit of the group.
46. (84%) The revelation (that/which) the clown on the circus tour from the Laughing Academy experienced (multiple seizures) upset the performers in the show.
47. (84%) The revelation (that/which) the victim in the emergency room at a London hospital experienced (serious shock) surprised the nurse at the desk.
48. (84%) The revelation (that/which) the executive from the large company with the clean record belittled (the secretary) surprised the employees in the office.
49. (90%) The rumor (that/which) the foreman at the successful bakery beside the grocery store overheard (the negotiations) shocked the stockholders of the company.
50. (90%) The rumor (that/which) the philanthropist from the wealthy family in the tiny town created (the problem) hurt the charity for needy families.
51. (90%) The rumor (that/which) the candidate in the long debate before the close election disputed (the bill) influenced the opinions of the voters.
52. (90%) The rumor (that/which) the senator in the special committee on the terrorist activities ignored (the warnings) ruined any possibility for a reelection.
53. (88%) The speculation (that/which) the politician from the important district in New York state questioned (the reports) infuriated the voters of the electorate.
54. (88%) The speculation (that/which) the daughter of the school headmaster in the conservative neighborhood offered (the bribe) shocked the elders of the community.
55. (82%) The suspicion (that/which) the girl on the speeding train to the college town developed (some homesickness) worried the mother and the father.
56. (82%) The suspicion (that/which) the man in the old building with the drug dealers shared (the money) confused the judge and the jury.
57. (82%) The suspicion (that/which) the woman at the resort town on the Mediterranean coast ignored (the warning) disconcerted the mother of the boyfriend.
58. (80%) The thought (that/which) the mother of the newborn child at the suburban clinic ignored (the advice) worried the nurse and the doctor.
59. (80%) The thought (that/which) the patient with symptoms of depression at the new clinic suppressed (the rage) frightened the nurse of the psychiatrist.
60. (80%) The thought (that/which) the man in the support group from the small town shared (his feelings) encouraged the members with similar problems.

References

Abney, S. P., & Johnson, M. (1991). Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research, 20*, 233–250.

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

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunnen, & A. Zwicky (Eds.), *Natural language parsing*. Cambridge, UK: Cambridge University Press.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht, The Netherlands: Foris.

Chomsky, N. (1986). *Barriers. Linguistic inquiry monograph 13*. Cambridge, MA: MIT Press.

Chomsky, N., & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.). *Handbook of mathematical psychology* (Vol. 2, pp. 269–321). New York: Wiley.

Earley, J. (1970). An efficient context-free parsing algorithm. *Communications of the Association of Computing Machinery, 6*, 451–455.

Elman, J. L. (1991). Distributed representations, simple recurrent networks and grammatical structure. *Machine Learning, 7*, 195–225.

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

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language, 47*, 250–272.

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

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

Gazdar, G., Klein, E., Pullum, G. K., & Sag, I. A. (1985). *Generalized phrase structure grammar*. Oxford, UK: Basil Blackwell.

Gibson, E. (1991). *A Computational Theory of Human Linguistic Processing: Memory Limitations and Processing Breakdown.* Unpublished doctoral dissertation. Pittsburgh, PA: Carnegie Mellon University.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1–76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In Y. Miyashita, A. Marantz, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95–126). Cambridge, MA: MIT Press.

Gibson, E. (submitted). The interaction of top-down and bottom-up statistics in syntactic ambiguity resolution. Department of Brain and Cognitive Sciences, MIT.

Gibson, E., Desmet, T., Grodner, D., Watson, D., & Ko, K. (in press). Reading relative clauses in English. *Cognitive Linguistics*.

Gibson, E., & Hickok, G. (1993). Sentence processing with empty categories. *Language and Cognitive Processes, 8*, 147–161.

Gibson, E., & Pearlmutter, N. (1998). Constraints on sentence comprehension. *Trends in Cognitive Science, 2*, 262–268.

Gibson, E., & Pearlmutter, N. (2000). Distinguishing serial and parallel parsing. *Journal of Psycholinguistic Research, 29*, 231–240.

Gibson, E., & Thomas, J. (1997). The complexity of nested structures in English: Evidence for the syntactic prediction locality theory of linguistic complexity. Unpublished paper. Cambridge, MA: MIT Press.

Gibson, E., & Warren, T. (2004). Reading time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax, 7*, 55–78.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 1411–1423.

Grodner, D., Gibson, E., & Tunstall, S. (2002). Syntactic complexity in ambiguity resolution. *Journal of Memory and Language, 46*, 267–295.

Grodner, D., & Gibson, E. (in press). Consequences of the serial nature of linguistic input. *Cognitive Science*.

Grodner, D., Gibson, E., & Watson, D. (in press). The influence of contextual contrast on syntactic processing: Evidence for strong-interaction in sentence comprehension. *Cognition*.

Hakuta, K. (1981). Grammatical description versus configurational arrangement in language acquisition: The case of relative clauses in Japanese. *Cognition, 9*, 197–236.

Harris, A. (1998). Electrophysiological indices of syntactic processing difficulty. PhD thesis, MIT.

Holmes, V. (1973). Order of main and subordinate clauses in sentence perception. *Journal of Verbal Learning and Verbal Behavior, 12*, 285–293.

Hsiao, F., & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition, 90*, 3–27.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processing in reading comprehension. *Journal of Experimental Psychology: General, 111*, 228–238.

Kennison, S. M. (2000). Two assessments of American usage frequencies for ninety-seven sentence complement-taking nouns. *Journal of Psycholinguistic Research, 29*, 249–264.

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

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: the role of working memory. *Journal of Memory and Language, 30*, 580–602.

King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7*, 376–395.

Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience, 5*, 196–214.

Konieczny, L. (1996). Human sentence processing: A semantics-oriented parsing approach. Doctoral Thesis. IIG-Berichte, 6-96. Albert-Ludwigs-University, Freiburg, Germany.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research, 29*(6), 627–645.

Konieczny, L., & Doring, P. (2003). Anticipation of clause-final heads: Evidence from eye-tracking and SRNs. In Proceedings of the ICCS/ASCS-2003 joint international conference on cognitive science. Sydney, Australia 13–17 July 2003.

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research, 25*, 93–115.

Lewis, R. (2000). Falsifying serial and parallel parsing models: Empirical conundrums and an overlooked paradigm. *Journal of Psycholinguistic Research, 29*, 241–248.

MacDonald, M., & Christiansen, M. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1999). *Psychological Review, 109*, 35–54.

Miller, G. A., & Isard, S. (1964). Free recall of self-embedded English sentences. *Information and Control, 7*, 292–303.

Miyamoto, E. (2002). Case markers as clause boundary inducers in Japanese. *Journal of Psycholinguistic Research, 31*, 307–347.

Nakatani, K., & Gibson, E. (2003). An on-line study of Japanese nesting complexity. In Poster presented at the 16th annual CUNY conference on human sentence processing. Cambridge, MA: MIT Press, March 2003.

Pearlmutter, N.J., & Mendelsohn, A. (1999). Serial versus parallel sentence comprehension. Northeastern University, Department of Psychology, manuscript submitted for publication.

Pickering, M., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes, 6*, 229–259.

Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. Chicago, IL: University of Chicago Press.

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.

Schmauder, A. R., Kennison, S. M., & Clifton, C. (1991). On the conditions necessary for obtaining argument structure complexity effects. *Journal of Experimental Psychology—Learning, Memory and Cognition, 17*, 1188–1192.

Shapiro, L. P., Zurif, E., & Grimshaw, J. (1987). Sentence processing and the mental representation of verbs. *Cognition, 27*, 219–246.

Stabler, E. P. (1994). The finite connectivity of linguistic structures. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 303–336). Hillsdale, NJ: Erlbaum.

Stolz, W. S. (1967). A study of the ability to decode grammatically novel sentences. *Journal of Verbal Learning and Verbal Behavior, 6*, 867–873.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language and communication* (pp. 217–262). San Diego, CA: Academic Press.

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

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

van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language, 49*, 285–316.

van Gompel, R., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 621–648). Oxford: Elsevier.

van Gompel, R., Pickering, M. J., & Traxler, M. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language, 45*, 225–258.

Vasishth, S. (2002). *Working memory in sentence comprehension: Processing Hindi center embeddings.* Unpublished doctoral dissertation, Ohio State University, Columbus, OH.

Wanner, E., & Maratsos, M. (1978). An ATN approach in comprehension. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 119–161). Cambridge, MA: MIT Press.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition, 85*, 79–112.

Yngve, V. H. (1960). A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society, 104*, 444–466.