ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Memory and Language 48 (2003) 67–91

Journal of Memory and Language

www.elsevier.com/locate/jml

Memory structures that subserve sentence comprehension

Brian McElree,* Stephani Foraker, and Lisbeth Dyer

Department of Psychology, New York University, 6 Washington Place, Room 859, New York, NY 10003, USA

Received 2 August 2001; revision received 24 May 2002

Abstract

Measures of the speed and accuracy of processing sentences with nonadjacent dependencies derived from the response-signal speed-accuracy tradeoff procedure were used to examine the nature of the memory system that underlies sentence comprehension. Three experiments with different sentence structures demonstrated that the accuracy of processing a dependency decreased as more material was interpolated between nonadjacent constituents. However, processing speed was unaffected by the amount of interpolated material, indicating that memory representations for previously processed constituents can be accessed directly. These results suggest that a content-addressable memory system mediates sentence comprehension, in which syntactic and semantic information provide direct access to memory representations without the need to search through extraneous representations. Notably, content-addressability appears to underlie the interpretation of sentence structures that also require the recovery of order information, a type of operation that has been shown to necessitate a slow search process in list-learning experiments (McElree, 2001; McElree & Dosher, 1993).

Keywords: Sentence comprehension; Working memory; Unbounded dependencies; Memory retrieval; Speed-accuracy tradeoff

Successful language comprehension requires access to memory representations outside the current focus of attention. For example, integrating new material into a coherent discourse representation often requires access to representations of earlier passages to resolve anaphoric relations and to draw crucial bridging inferences. At the sentential level, several common syntactic structures contain dependencies between nonadjacent constituents, which likewise require access to representations outside focal attention when processed. Notable examples are sentences with unbounded (or long-distance) syntactic dependencies like (1), where the initial noun phrase (NP) *the novel* is understood as the underlying syntactic and semantic object of the final verb (*embrace*), despite being displaced from the canonical position for direct objects in English:

(1) *This is the novel that the editor hoped the public would embrace.*

Less obvious but perhaps more common are cases like (2), where an embedded clause (*who testified before the committee charged with investigating ethical violations*) interrupts the dependency between a subject (*the senator*) and a verb (*resigned*):

(2) *The senator who testified before the committee charged with investigating ethical violations resigned.*

In both cases, processing of the material interpolated between the initial NP and the final verb will displace the NP from focal attention. When the final verb is encountered, a representation of the initial NP must be retrieved from memory in order to integrate it with the syntactic and semantic properties of the verb. This notion receives empirical support from studies using a probe recognition task (Bever & McElree, 1988; McElree, 2000; McElree & Bever, 1989) and a cross-modal priming task (Nicol & Swinney, 1989; Osterhout & Swinney, 1993; Swinney, Ford, Bresnan, & Frauenfelder,

* Corresponding author.
E-mail address: brian.mcelree@nyu.edu (B. McElree).

PII: S0749-596X(02)00515-6

1988; but see McKoon & Ratcliff, 1994; Nicol, Fodor, & Swinney, 1994). These studies have demonstrated that constituents like *the novel* in (1) are reactivated in working memory (WM) following the processing of the region that resolves the dependency.

Several approaches to language comprehension assert that the complexity of sentence processing is partly determined by the demands that various sentence structures place on memory resources (e.g., Gibson, 1998; Lewis, 1996; Young & Lewis, 1998), and that working memory resources provide a basis for characterizing individual differences in language abilities (Caplan & Waters, 1999; Just & Carpenter, 1992; Waters & Caplan, 1996). However, few studies have directly examined memory retrieval during on-line sentence comprehension and little is known about the underlying retrieval mechanisms used in comprehension. The studies reported here derive measures for the timecourse of resolving nonadjacent dependencies like (1) and (2) as a means of more directly examining the nature of the memory structures that enable sentence comprehension.

Basic research investigating retrieval processes in WM indicates that memory representations can be accessed in one of two ways (McElree, 1996, 1998, 2001; McElree & Dosher, 1989, 1993). Some types of memory structures can be directly accessed with a content-addressable retrieval operation, while other structures require a slower search process. Recovering an item from memory appears to be mediated by a content-addressable mechanism. In contrast, recovering order or relational information from memory appears to require a relatively slow search process. The timecourse measures reported here suggest that basic binding operations that enable processing of long-distance dependencies like (1) or nonadjacent dependencies like (2) are mediated by memory representations that are directly accessible. Based on these observations, we argue that sentence processing is subserved by content-addressable memory structures in which syntactic and semantic constraints provide the retrieval cues that enable direct access to needed representations. We argue that the memory representations that subserve language comprehension are content-addressable even when order information is required, in contrast to other domains in which order information requires a search process.

Retrieval mechanisms

The defining property of a content-addressable retrieval process is that information (cues) in the retrieval context enables direct access to relevant memory representations, without the need to search through extraneous memory representations. Most episodic memory models (e.g., Gillund & Shiffrin, 1984; Hintzman, 1984; Murdock, 1982; see Clark & Gronlund, 1996, for a review) and many semantic memory models (e.g., Hinton, 1989; Kawamoto, 1988; Plaut, 1997; Seidenberg & McCelland, 1989) posit that memory representations are retrieved with a content-addressable operation. Content-addressable operations can be implemented in memory models with rather diverse storage architectures, including those with highly structured localized representations and those with highly distributed representations (for a review, see Clark & Gronlund, 1996).

The alternative to a content-addressable mechanism is a search process, in which memory representations are searched by location or some other organizing principle (Gillund & Shiffrin, 1984). On grounds of efficiency, a search process, particularly an unconstrained search process, is generally viewed as an implausible means of retrieving an item from the vast amount of information in long-term memory. Efficiency arguments carry less weight for the retrieval of recent events, however, if those events are represented in a more limited-capacity WM system. Indeed, traditional models of the retrieval of short-term representations have argued that the primary means of gaining access to an item in WM is through a search process with either a serial architecture (e.g., Sternberg, 1966, 1975; Theios, 1973; Treisman & Doctor, 1987) or a parallel architecture (e.g., Murdock, 1971; Townsend & Ashby, 1983).

Discriminating between mechanisms

Search and direct access mechanisms can be contrasted empirically by examining the effect that interpolated material has on the speed of retrieval (McElree, 1996, 1998, 2000, 2001; McElree & Dosher, 1989, 1993). With a search mechanism, retrieval speed will slow as potentially interfering material is added to the memory system. In contrast, additional material need not affect retrieval speed in a memory system with a direct access mechanism. Interpolated material may decrease the quality of a memory representation (through loss of storage strength, distinctiveness, or related notions), making the representation less likely to be recovered in any particular context, but a direct-access mechanism nonetheless enables representations of differing quality or strength to be retrieved in equal time (e.g., McElree & Dosher, 1989; Ratcliff, 1978).

McElree and Dosher (1989) (see also McElree, 1996, 1998; McElree & Dosher, 1993; Wickelgren, Corbett, & Dosher, 1980) examined retrieval operations in short-term item recognition, the paradigmatic case for studying how representations in WM are accessed. Using a speed-accuracy tradeoff procedure (SAT; see below), McElree and Dosher found that the number of items in memory (set size) and the number of items interpolated between study and test (recency of the test probe) adversely affected the probability of retrieving an item.

Specifically, the accuracy of recognition judgments decreased as the test probe was drawn from less recent serial positions. Crucially, however, neither set size nor recency affected the speed of retrieval. Search models, whether serial or parallel in form, are incompatible with findings that representations of differing availability are nevertheless equally accessible (for detailed predictions, see McElree & Dosher, 1989). These data indicate that access to an item's representation in memory is direct, and suggest that retrieval from WM is mediated by a content-addressable mechanism.

Direct access appears to be a general property of item retrieval. The same timecourse patterns are found in both supra- and sub-span lists (Wickelgren et al., 1980, see McElree, 2001), indicating that direct access is a property of both short- and long-term representations. Additionally, direct access is evident in the recognition of an item that is part of a hierarchically coded group (McElree, 1998), as well as when recognition is based on component properties (e.g., phonological and semantic properties) of the memory representation (McElree, 1996). The latter provides particularly strong evidence that direct access arises from a content-addressable retrieval operation. Recognition based on properties of the coded memory, rather than a complete representation, requires the redintegration of the studied item from related cues in the retrieval context alone.

However, not all types of information appear to be recoverable with a content-addressable retrieval process. Studies indicate that the retrieval of relational information, including temporal order (McElree, 2001; McElree & Dosher, 1993) and positional (Gronlund, Edwards, & Ohrt, 1997) information, requires a slow serial search. For example, McElree and Dosher (1993) found that both accuracy and retrieval speed were dependent on recency. Temporal order information was examined with a judgment of recency task, in which subjects were presented two test probes from a short list and asked to select the item that occurred more recently. As additional items were interpolated between study and test, both accuracy and speed decreased. McElree (2001) found the same pattern in an *n*-back task, a task that requires restricting (positive) responses to a particular ordinal position in a sequence instead of an overt judgment of order. The systematic slowing of retrieval speed with additional interpolated information found in both studies indicates that relational information is retrieved with a search process that begins with the most recent item. One account of these findings assumes that ordered representations in memory are serially scanned from the most recent, moving backwards in time (for specific models, see Hacker, 1980; McElree, 2001; McElree & Dosher, 1993). Alternatively, the systematic slowing of retrieval could indicate that order information is reconstructed at retrieval by a serial chaining process (cf. Lewandowsky & Murdock, 1989), in which the last item on the list is used as a cue to recover the next item on the list, and so on until the required information is recovered. Both accounts are compatible with the timecourse data, and are markedly distinct from a content-addressable process in which cues at retrieval provide direct access to the required information.

Memory structures in on-line comprehension

Which type of retrieval mechanism subserves the recovery of memory representations in sentence comprehension? Order information has been argued to be crucial in language processing (e.g., Lashley, 1951). It may be particularly important in languages like English in which word order is the primary means of specifying the argument role of a constituent, and even more so for processing sentences with nonadjacent arguments in which a representation of a constituent from a specific position in the sentence must be recovered (e.g., the initial positions in 1 and 2). To the degree to which order information is required, basic studies of short-term retrieval would suggest that representations of sentential constituents would have to be accessed with a search mechanism.

McElree (2000) investigated the question of whether the on-line retrieval of information in sentence processing is mediated by search or content-addressable mechanisms by using SAT procedures to derive measures of the speed and accuracy of the processing of sentences like (3)–(5):

(3) *This was the book that the editor admired.*

(4) *This was the book that the editor who the receptionist married admired.*

(5) *This was the book that the editor who the receptionist who quit married admired.*

In each case, the fronted NP (*the book*), often termed a *filler* item, must be assigned to the missing argument position, often termed a *gap,* in the direct object position of the final matrix verb (*admired*). The sentences varied in the amount of material that needed to be processed before the gap in constituent structure. In sentences like (3), a matrix subject (*the editor*) alone separated the filler and gap. In sentences like (4), an object relative clause (*who the receptionist married*) attached to the matrix subject was interpolated between the filler and gap. In (5), an additional subject relative clause (*who quit*) was attached to the subject of the prior relative clause, further increasing the amount of material between the filler and gap. Following other work (McElree, 1993; McElree & Griffith, 1995, 1998), readers were required to discriminate acceptable from unacceptable filler-gap relationships (unacceptable versions of 3–5 replaced the final matrix verb with a verb like *amused,* which requires an animate filler NP). Discriminating acceptable from unacceptable bindings of the filler items to the last word in the sentence (the matrix verb) ensured that readers fully interpreted the strings, and provided a relatively direct and unintrusive measure of when the filler was accessed and assigned to the gap in constituent structure. McElree (2000) used the SAT procedure (see below), which required readers to respond at one of six times ranging from 50 to 3000 ms after the onset of the critical matrix verb (*admired* in 3–5). This provided measures of how the interpretation of a sentence unfolded over time, and of the ultimate probability that the reader was able to understand the strings well enough to discriminate acceptable from unacceptable bindings.

McElree (2000) found that overall response accuracy was adversely affected by the amount of interpolated material. There are at least two nonmutually exclusive explanations for this finding. Consistent with standard findings in memory research and particular models of memory for sentence processing (Lewis, 1996), the probability of accessing a representation of the filler item may decrease as the amount of interpolated material increases. Or, the likelihood of misanalyzing a string could increase directly with the length and complexity of the string. [McElree (2000) presented independent evidence that the former was at least partly responsible for the observed differences in accuracy; see General discussion.]

Notably, however, there was no evidence to suggest that the speed of binding the displaced filler item to an argument position varied across structures like (3)–(5). In order to discriminate an acceptable from unacceptable sentence, readers had to retrieve a representation of the clefted NP, assign it to the argument position licensed by the verb (*admired* or *amused*, above), and semantically interpret the verb phrase. If retrieving the NP had required a search, overall judgment speed measured from the onset of the final verb would have been systematically slower across increasing amounts of interpolated material. However, processing speed did not vary across (3)–(5), indicating that readers could access the filler item equally quickly when zero, one, or two embedded clauses separated the filler and gap. This pattern suggests, perhaps surprisingly, that a search mechanism was not used to access a representation of the filler item. The timecourse profiles are analogous to what is observed in simple item recognition, and markedly distinct from the patterns observed in the recovery of order information (McElree, 1996, 1998, 2001; McElree & Dosher, 1989, 1993). The data suggest that the memory representation of the filler item is directly accessible when the matrix verb is encountered. Prima facie, that pattern is consistent with the idea that memory representations in sentence processing are content-addressable.

Current studies

The McElree (2000) results are suggestive, but their import is limited by the fact that only one type of structure was examined. If on-line retrieval is in fact mediated by a content-addressable memory system, and this type of memory system is the rule rather than exception in sentence processing, then the pattern reported in McElree (2000) should be observed with different types of interpolated material and with different sentence structures containing other types of binding relations. Moreover, the same pattern should be found in structures in which order information is unquestionably crucial to the interpretation of the sentence. Three experiments exploring these issues are reported. As in the McElree (2000) study, the nature of the retrieval mechanism was addressed by examining what effects interpolated material had on the speed and secondarily the accuracy of processing nonadjacent constituents.

In the first experiment, we modified the nature of the interpolated material to reflect increasing distance in a hierarchical syntactic structure, rather than the linear, surface distance employed in McElree (2000). This modification enabled us to examine whether a more linguistically sophisticated search mechanism was used to access constituents in memory.

In a second experiment, we examined whether the same pattern implicating direct access would be evident in the processing of a nonadjacent subject-verb dependency. Cleft constructions like those in (3)–(5) have been argued to place the clefted constituent in focus (e.g., Gundel, 1999) and to recruit specialized processing strategies (e.g., *Active Filler Strategy*, Clifton & Frazier, 1989; Fodor, 1995; see below). The results reported in McElree (2000) could reflect processes that are particular to the processing of this type of structure rather than the general means through which memory representations are accessed on-line.

Finally, in a third experiment, we examined constructions that required readers to bind two filler items to the direct and indirect object positions of a (ditransitive) matrix verb, explicitly comparing these constructions to others that involved only one filler-gap relation. The order of the filler items in the two gap constructions determined the acceptability of the constructions, so they provided a strong case for testing whether there are circumstances in which order information is recovered by a search process. The contrast between single and double gap constructions also served to demonstrate that the experimental procedure is sensitive to the temporal properties of on-line comprehension, and, in particular, to the timecourse of binding a constituent to an argument position.

Experiment 1

The contrasts reported in McElree (2000), examples of which are given in (3)–(5), varied the distance between the filler (*the book*) and the gap position by adding one or two embedded relative clauses following a matrix subject (*the editor*). Relative clauses of this type increase the distance between the filler and gap in terms of number of words and clauses intervening between filler and gap in surface structure, but not necessarily in terms of distance in a hierarchical syntactic representation. Fig. 1A schematically illustrates the hierarchical phrase structure relations in strings like (3)–(5). Embedded relative clauses (shown as dashed triangles) do not alter the distance between the dislocated filler item and the argument position along the right edge of the (hierarchical) syntactic tree (denoted by the dashed arrow). If the gap in syntactic structure initiated a search along this edge of the parse tree only, then the embedded relative clause structures would not be expected to impact search time. Linguistic principles preclude dependencies between elements within the relative clause and a rightward positioned gap associated with the matrix verb, so it is neither arbitrary nor unreasonable to assume that a search might be constrained to the right edge of a parse tree.

To test whether hierarchical distance affects the speed and accuracy of binding a dislocated argument, the first

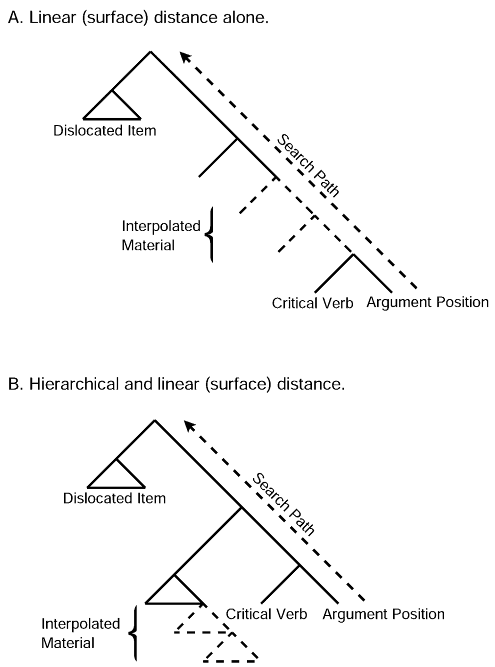

Fig. 1. Schematic tree diagrams illustrating the hierarchical syntactic relations in the materials used in McElree (2000) (A) and in Experiment 1 (B). Hypothetical search path along the right edge of the parse tree is shown as dashed line.

experiment contrasted constructions with embedded complement clauses like those in examples (6)–(8):

(6) *It was the scandal that the celebrity relished.*

(7) *It was the scandal that the model believed the celebrity relished.*

(8) *It was the scandal that the model believed that the journalist reported that the celebrity relished.*

Fig. 1B schematically illustrates that embedded complement clauses (*the model believed...*) increase both the surface distance between the filler (*the scandal*) and the gap position following the matrix verb (*relished*) and the distance along the right edge of the parse tree. Thus, if a gap in constituent structure initiates a search of memory for the filler item but syntactic principles constrain the search to the right edge of a constructed parse tree, we should observe a systematic slowing of processing speed across constructions (6)–(8), as well as any adverse effects that distance might have on the probability (or accuracy) of recovering the correct interpretations of the sentences.

As in previous work (McElree, 2000; but also McElree, 1993; McElree & Griffith, 1995, 1998), we required readers to discriminate acceptable from unacceptable sentences to ensure that the strings were processed to a significant depth. Table 1 gives an example of the conditions used in the experiment. Unacceptable versions of the primary experimental contrasts replaced the final matrix verb with a verb that mismatched the semantic properties of the clefted filler item. In Table 1, for example, the verb *relished* in T1, T3, and T5 (corresponding to examples (6)–(8) above) was replaced with the verb *panicked*, which requires an animate direct object. To judge T1, T3, and T5 as acceptable and T2, T4, and T6 as unacceptable, the reader had to bind the clefted item to the argument position of the verb and process the relation between the verb and its argument to a depth sufficient to detect the semantic anomaly in the latter three cases.

In addition to the three primary experimental conditions and their associated unacceptable forms, 10 other types of sentence structures were included to control for and disallow various decision strategies or heuristics. Using only structures like T1–T6 would enable readers to ignore the interpolated material and focus exclusively on the clefted item and the final verb. To disallow such a simple strategy, we included cases like T7 and T8, in which the anomaly occurs in the first or second interpolated relative clause (see also T13 and T15). Additionally, we included acceptable and unacceptable constructions like T9–T16 without an initial clefted NP, so readers would not pay undue attention to the clefted element.

We expected that the representation for the filler item in memory would be adversely affected by the amount of interpolated material, given prior findings (McElree, 2000), standard findings in memory research, and

Table 1
Constructions used in Experiment 1

Construction type	Acceptability	Example
No interpolation	Acceptable	T1. It was the scandal that the celebrity relished
No interpolation	Unacceptable	T2. It was the scandal that the celebrity panicked
One interpolated clause	Acceptable	T3. It was the scandal that the model believed that the celebrity relished
One interpolated clause	Unacceptable	T4. It was the scandal that the model believed that the celebrity panicked
Two interpolated clauses	Acceptable	T5. It was the scandal that the model believed that the journalist reported that the celebrity relished
Two interpolated clauses	Unacceptable	T6. It was the scandal that the model believed that the journalist reported that the celebrity panicked
Controls for relative clause processing		
One interpolated clause	Unacceptable	T7. It was the scandal that the model amused that the celebrity relished
One interpolated clause	Unacceptable	T8. It was the scandal that the model believed that the journalist amused that the celebrity relished
Additional controls		
No interpolation	Acceptable	T9. The scandal panicked the celebrity
No interpolation	Unacceptable	T10. The scandal relished the celebrity
One interpolated clause	Acceptable	T11. The model believed that the scandal panicked the celebrity
One interpolated clause	Unacceptable	T12. The model believed that the scandal relished the celebrity
One interpolated clause	Unacceptable	T13. The model ambled that the scandal panicked the celebrity
Two interpolated clauses	Acceptable	T14. The model believed that the journalist reported that the scandal panicked the celebrity
Two interpolated clauses	Unacceptable	T15. The model believed that the journalist ambled that the scandal panicked the celebrity
Two interpolated clauses	Unacceptable	T16. The model believed that the journalist reported that the scandal relished the celebrity

particular models of memory for sentence processing (e.g., Lewis, 1996). As more information is processed between the point at which the filler item is read and the point at which it is bound to the argument position of the verb, there is an increase in the likelihood that the representation of the filler will decay, that it will be displaced from memory, or that the cues provided by the verb and related linguistic elements will not be sufficient to retrieve the filler item from memory.

However, discriminating between content-addressable and search processes requires measures of processing speed that are unaffected by differences in the likelihood of recovering the relevant memory representation. Unfortunately, simple timing measures like reaction time, reading time, or eye movement measures are affected by both factors (see McElree, 1993; McElree & Nordlie, 1999; McElree & Griffith, 1995, 1998). One solution to this problem is to derive a function that measures how the accuracy of processing varies with processing time (Wickelgren, 1977), so that both speed and accuracy can be jointly measured and separately assessed. Here, the response-signal SAT procedure was used to construct such functions.

Fig. 2 illustrates the procedure. Sentences were visually presented one word at a time (250 ms/word). Readers were required to make binary (yes/no) acceptability decisions at one of 6 randomly determined times (either 50, 300, 500, 800, 1200, or 3000 ms) after the onset of the final word in the string. In the contrasts of primary interest (T1–T6 in Table 1), the final word was the crucial matrix verb that specified the argument position of the filler item. The cue to respond was signaled by a brief (50 ms, 1000 Hz) tone, and readers were trained to respond within 300 ms of the tone. Readers were required to respond at the tone even if processing of the string was not fully completed and the response had to be based on partial information or, in the limit, a guess. This feature of the SAT procedure minimizes the decision processes that are involved in reaction time or other timing tasks, in which a participant must select a criterion for responding that balances the tradeoff between speed and accuracy (Dosher, 1979; Ratcliff, 1978; Wickelgren, 1977). To further control for response biases (tendency to differentially respond "yes" or "no"), accuracy is measured in d' units by scaling the z-transformation of the hit rate for an acceptable string against the z-transformation of the false alarm rate for the corresponding unacceptable string. The accuracy at various response times provides a relatively direct estimate of the likelihood that processing is completed at that time. The range of response signals was selected to chart the full timecourse of processing, from times when accuracy was near chance to times when accuracy reached an asymptotic level.

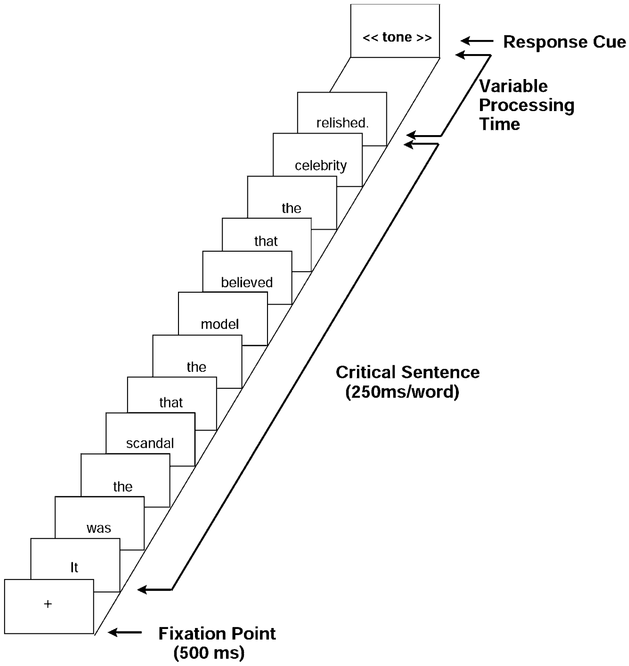

Fig. 2. Sample trial sequence illustrating the speed-accuracy tradeoff (SAT) variant of the acceptability judgment task.

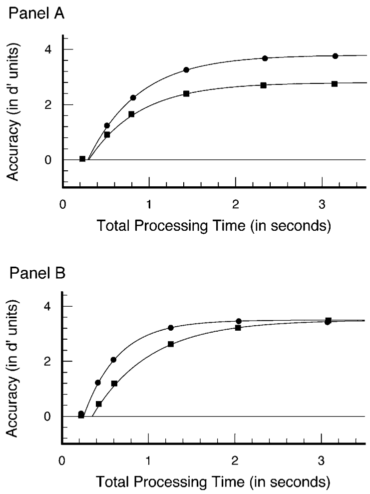

Fig. 3. Hypothetical SAT functions illustrating two conditions that differ by SAT asymptote only (A) or by SAT intercept and rate (B).

Fig. 3 presents hypothetical SAT functions illustrating how different SAT timecourse patterns can discriminate between alternative retrieval processes. Consider first the expected result that interpolating more material between the filler and gap position decreases the accuracy of responding. Recall that this could be because there is a lower probability that a representation of the filler is available when the verb is processed and/or because there is a higher probability of misanalyzing material up to and including the final verb. If additional material decreases only the overall accuracy of responding, the corresponding functions will differ in asymptotic level alone. Panel A depicts two hypothetical conditions that differ in this manner.

The pre-asymptotic portion of the SAT function measures processing speed or dynamics, jointly specified by the intercept of the function (when accuracy departs from chance, $d' = 0$) and the rate at which accuracy grows from intercept to asymptote. The intercept measures the minimum time needed to form an interpretation that would serve to discriminate acceptable from unacceptable forms. The rate of the SAT function reflects either the rate of continuous information accrual if processing is continuous or the distribution of finishing times if processing is discrete or quantal (Dosher, 1976, 1979, 1981, 1982, 1984; Meyer, Irwin, Osman, & Kounois, 1988). In either case, differences in intercept or rate implicate underlying differences in the speed of processing. This situation is depicted in Panel B of Fig. 3, where the functions are associated with different intercepts and rates of rise to a common asymptote.

If access to the filler's representation requires a search process when the matrix verb is encountered, then the SAT intercept and/or rate of will systematically slow as more material is interpolated between the filler and gap. Rate or intercept differences can arise from factors other than retrieval speed; for example, they might arise from differences in the inherent complexity of computing various interpretations or from an increasing tendency to misanalyze a string followed by reanalysis (McElree, 1993). However, any substantial difference in the time to retrieve the filler will engender differences in some combination of rate and intercept. For example, McElree and Dosher (1993) found that a search process engendered differences in SAT intercepts that were as large as 500 ms in short lists of six words. In contrast, equal intercepts and rates (Panel A) suggest that verb and related information provide direct access to the filler item without potentially interfering effects of interpolated material.

Method

Participants

SAT studies are designed to collect stable functions for individual participants, and each subject's data is analyzed separately. To this end, eight students from New York University served as participants in the experiment. Each one participated in 10 1-h sessions, plus a 1-h practice session for familiarization with the SAT procedure. All participants were native English speakers, and were paid for their participation in the experiment.

Materials

Ten sets of 384 sentences were generated. Each set was composed of 24 instances of the 16 sentence types (six acceptable and 10 unacceptable) listed in Table 1. Across the 10 sets, the 240 instances were of the same word length as illustrated in the table. The 240 instances were recombined into 10 experimental sets, one for each of the 10 sessions, so as to minimize the repetition of content material within a session. For example, each one of the first 10 strings from Table 1 was assigned to one of the 10 sets, and then each of the remaining sentences was assigned to one of 6 of the sets. The assignment was such that an equal number of each of the 16 forms in Table 1 was assigned to each of the 10 experimental sets. No attempt was made to match the filler phrases or the crucial matrix verbs in terms of frequency or letter length, as the contrasts of interest, T1–T6 in Table 1, all contain the same crucial items, with the exception of the interpolated material. The order of presentation within a session was randomized.

Procedure

Stimulus presentation, timing, and response collection were all carried out on a personal computer using software with millisecond timing. Fig. 2 illustrates a trial in the experiment. A trial began with a 500 ms fixation point (a small filled square) presented in the center of the screen. Words were presented one after another for 250 ms each, with a period appended to the final word of a string. A 50 ms, 1000 Hz tone sounded at one of six response lags, either 50, 300, 500, 800, 1200, or 3000 ms after the onset of the final word in the string. Participants were trained to respond "yes" or "no" at the tone by pressing one of two designated keys on the keyboard. After a response, visual feedback on the latency to respond to the tone was given. The participants were informed that responses longer than 300 ms were unacceptably long and that responses shorter than 100 ms should be regarded as anticipations. Both the sentences and the response lags were randomized within a session.

Participants were instructed to respond "yes" if the string formed an acceptable English sentence, and "no" otherwise. During the practice session, they were told that some of the sentences were complex and may be difficult to understand, but nevertheless were meaningful sentences. They were given examples of multiply embedded sentences to illustrate the point. Each participant performed 10 1-h sessions using one of the 10 sets of materials. The order of materials was randomized across participants.

Data analysis

All analyses were performed on the individual participants' data. Consistent patterns across subjects were summarized by analyses of the average data. To correct for response bias, scores were computed by scaling the z-score of the probability of saying "yes" to acceptable strings against the z-score of the probability of saying "yes" to corresponding unacceptable strings at each lag. Accuracy (d') at each lag was plotted and analyzed as a function of the lag of the response tone plus the average latency to respond to the tone. Including latency ensures that any condition-specific differences in latency are factored into the estimates of processing accuracy at each lag (see McElree & Dosher, 1989, 1993). Potential differences in asymptote, rate, and intercept were assessed by fitting the accuracies at various processing times (t) with an exponential approach to a limit:

$$d'(t) = \lambda(1 - e^{-\beta(t-\delta)}) \quad \text{for } t > \delta, \quad \text{else } 0,$$

where λ reflects the asymptote of the function, δ denotes the intercept or discrete point in time when accuracy departs from chance, and β indexes the rate at which accuracy grows to asymptote. Hierarchically nested models were fit to the data, ranging from a null model, in which all three crucial conditions were fit with a single asymptote (λ), rate (β), and intercept (δ), to a fully saturated (nine parameters) model, in which each experimental condition was fit with a unique set of parameters. Eq. (1) was fit to the data with an iterative hill-climbing algorithm (Reed, 1973, 1976), similar to STEPIT (Chandler, 1969), which minimized the squared deviations of predicted values from observed data. Fit quality was assessed by an adjusted-R^2 statistic—the proportion of variance accounted for by the fit adjusted by the number of free parameters (Judd & McClelland, 1989)—and by an evaluation of the consistency of the parameter estimates across the participants.

Results and discussion

Fig. 4 shows the average (over participants) d' data as a function of processing time for constructions in which no complement clause intervened between the filler and gap (triangles), one complement clause intervened between the filler and gap (circles), and two complement clauses intervened between the filler and gap (squares). Asymptotic accuracy decreased as more material was interpolated between the filler and gap. That conclusion was supported by an ANOVA on the d' values at the longest response time (3000 ms), $F(2,16) = 6.04$, $MSe = .219$, and by competitive fits of Eq. (1). A $1\lambda - 1\beta - 1\delta$ (null) model, in which all three conditions were fit with a common set of parameters, produced an adjusted-R^2 value of .959 in the average data, ranging from .804 to .936 across the nine participants. In contrast, a $3\lambda - 1\beta - 1\delta$ model, with separate asymptotes (λs) for each construction type, produced an adjusted-R^2 value of .976 in the average data, ranging from .878 to .942 across participants. The common rate parameter (β) was estimated at 2.43 (ranging from 6.41 to 1.24 ms across participants), and the common intercept (δ) at 437 ms (ranging from 315 to 569 ms across participants). All but two participants showed an increase in adjusted-R^2 for the $3\lambda - 1\beta - 1\delta$ over the $1\lambda - 1\beta - 1\delta$ model. Equally important, this model yielded a consistent set of asymptotic estimates for average data and individual participants' data. The estimates for the average data were 2.86, 2.61, and 2.34 for conditions with 0, 1, 2 embedded clauses (respectively), and, consistent with the empirical differences in d' at the longest lag, the difference in parameter estimates across participants was significant, $F(2,16) = 8.16$, $MSe = .086$. These asymptotic differences may partly reflect loss of the filler item resulting from the interfering effect of processing the complement clauses and partly an increased tendency to misanalyze strings of longer and more complex structure.

Despite systematic asymptotic differences, there was no evidence that processing speed at the critical verb varied across the three constructions. Two facets of the analysis compel that conclusion. First, when the average data and individual participants' data were fit with models that further varied the rate parameter (a $3\lambda - 3\beta - 1\delta$ model), the intercept parameter (a $3\lambda - 1\beta - 3\delta$ model), or both parameters (a fully saturated $3\lambda - 3\beta - 3\delta$ model), the resulting adjusted-R^2 values were lower than the $3\lambda - 1\beta - 1\delta$ model for all but two participants. The decrease in adjusted-R^2 indicates that the additional parameters were not accounting for any systematic variance in the data. Second, when these more embellished models were applied to the data, they did not yield a consistent ordering of β or δ estimates in the average data or across the data of individual participants. Consequently, statistical tests on the resulting parameter estimates were all nonsignificant. Both facts indicate that the additional parameters were not accounting for systematic variance in the data. The timecourse profiles suggest that interpolated material does not affect the speed with which the antecedent filler item can be accessed and processed. The smooth functions in Fig. 4 show the best fitting $3\lambda - 1\beta - 1\delta$ model for the average data.

Additional properties of the data were examined to check whether participants were performing the task in the expected manner, and to bolster the conclusions drawn above. Included in the set of materials were strings like T7 and T8 in Table 1, which were ill-formed at regions other than the final verb. These were included

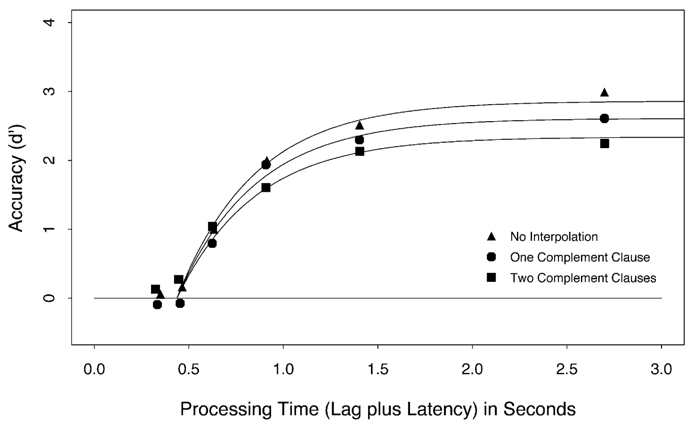

Fig. 4. Average d' accuracy (symbols) as a function of processing time (lag of the response cue plus latency to respond to the cue) for judgments of constructions with no embedded clauses (triangles), one embedded complement clause (circles), and two embedded complement clauses (squares). Smooth curves show the best fitting $3\lambda - 1\beta - 1\delta$ exponential model (see text of Experiment 1).

to encourage participants to fully process the strings and to discourage a strategy that only evaluated the fit between the clefted filler item and the final verb position. Correct rejection rates for these strings were quite high; specifically, 83.9, 81.7, 79.1, 81.2, 82.9, and 82.1% on average for lags 1–6, respectively. The comparable rejection rates for the three strings (T2, T4, and T6) that were ungrammatical at the final verb were 40.8, 41.7, 65.5, 82.3, 88.0, and 90.8%. Asymptotic accuracy was higher in the latter set, which might be expected, as the anomaly immediately preceded the judgment. (When the anomaly occured at earlier regions, participants might have recognized it but have less confidence in rejecting the string after processing acceptable regions.) It is clear that the substantially above-chance rejection rates for the T7 and T8 strings indicate that participants were not ignoring the interpolated material in the clefted constructions. Notable also is the fact that correct rejection rates were clearly asymptotic at the first (50 ms) lag, which demonstrates that participants were well aware of the anomaly before the end of the string, and hence were processing the strings incrementally.

One concern about the procedure employed here is that, because participants performed the task for several hours, they might have developed specific strategies to deal with the materials that might have masked underlying timecourse differences. To address this issue, we analyzed the average data from the first and last sessions. (There were too few trials in each session to examine the patterns for individual participants.) Overall, accuracy improved and speed increased across the 10 sessions. For example, the average asymptotic d' levels in the first session were 2.43, 3.15, and 2.09 for 0, 1, and 2 complement clause constructions (respectively), as compared to 3.62, 3.43, and 2.93 for the final session. Overall, accuracy improved by .75d' units. Similarly, the rate (β) was estimated at 1.74 and the intercept (δ) at 439 ms in the first session, as compared to 2.77 and 355 ms (respectively) in the final session. Clearly, processing speed increased across the sessions. Crucially, however, there was no evidence for *differences* in the speed of processing among the crucial conditions in the first session. For example, a $1\lambda - 1\beta - 1\delta$ model yielded an adjusted-R^2 of .794, which was improved to .822 by a $3\lambda - 1\beta - 1\delta$ model. However, adjusted-R^2 decreased for the $3\lambda - 3\beta - 1\delta$ model (.809), the $3\lambda - 1\beta - 3\delta$ model (.803), and the fully saturated $3\lambda - 3\beta - 3\delta$ model (.805). As with the full set of data, there was no evidence to suggest that multiple sessions masked any possible dynamics differences between conditions, due to the amount of interpolated material. The timecourse patterns suggest that the likelihood of binding a filler to the gap decreases with more interpolated material; however, the time to resolve the dependency, when it can be resolved, remains constant. If a search process were used to access the filler item, then processing time should have increased with more structure between the filler and gap (cf. McElree & Dosher, 1993). These data show the same pattern as the data reported in McElree (2000), which used contrasts with relative instead of complement clauses. The difference is, as we have noted, that the latter also vary in hierarchical (syntactic) distance. There is no evidence from the current study to motivate the notion of a search process constrained by linguistic principles. As with McElree (2000), the data suggest that syntactic and semantic constraints provide direct access to a representation of the relevant dislocated item.

Experiment 2

How general are the results reported in Experiment 1 and in McElree (2000)? One concern is that the key contrasts in both experiments involved cleft constructions (e.g., *It was the book that*...; *These are the books that*...). These constructions are argued to focus the reader/listener on the clefted element (*the books*), often with the purpose of contrasting it with another element in the discourse. How linguistically focused items differ from unfocused items in memory is not well understood, but studies have demonstrated that focused items are more memorable than unfocused items (e.g., Birch & Garnsey, 1995). It has been suggested that some forms of linguistic focus might reflect what is in the current focus of attention (Gundel, 1999), and there is a body of empirical evidence from studies of working memory with word lists that indicates that the focus of attention is distinct from other temporary memory representations (Cowan, 1995; Dosher, 1981; McElree & Dosher, 1989, 1993; McElree, 1996, 1998; Wickelgren et al., 1980; for a review, see McElree, 2001). If linguistic focus is equivalent to the focus of attention, and if information in focal attention is more accessible than other memory representations, then the pattern of results found in Experiment 1 and in McElree (2000) may not generalize to constructions involving other nonadjacent dependencies. For example, if a clefted item remains within the focus of attention across the processing of intervening material, then it might be accessible without searching through potentially conflicting memory representations. However, unfocused items, in as much as they are not actively maintained in focal attention, may require a search process.

A related perspective holds that clefted structures recruit processing strategies that are not representative of all the types of memory operations needed in comprehension. Cleft constructions are a type of long-distance or filler-gap dependency that have been argued to be processed by a specialized operation dubbed the *Active Filler Strategy* (Clifton & Frazier, 1989; Fodor, 1995): Readers are thought to actively postulate a gap for a marked filler item—in this case the clefted NP—in every legal syntactic position. A possible construal of this strategy reduces to the focus-based notion outlined above: Readers may actively maintain a representation of the dislocated constituent in focal attention until a syntactic-semantic role has been found, thereby circumventing the retrieval processes that would otherwise be needed to restore the constituent to active processing were it not held in focal attention.

We tested the generality of prior results by examining the speed and accuracy of binding a subject NP to a matrix verb in more standard main clause constructions, which do not contain syntactic devices for focusing a constituent and are not thought to invoke specialized processes like the Active Filler Strategy. Table 2 illustrates the full set of contrasts that were used. In the simplest case, we examined the speed and accuracy of binding of a subject phrase (*the book*) to a matrix intransitive verb (*ripped*) in constructions like *The book ripped,* in which the two elements are adjacent in surface structure. T1 and T3 in Table 2 are examples of acceptable structures of this type. As before, readers were required to discriminate acceptable from unacceptable bindings. T2 and T4 are examples of the corresponding unacceptable constructions, in which the verbs in T1 and T3 were exchanged to produce an unacceptable binding.

Table 2
Constructions used in Experiment 2

Construction type	Acceptability	Example
No interpolation	Acceptable	T1. The book ripped
No interpolation	Unacceptable	T2. The book laughed
No interpolation	Acceptable	T3. The editor laughed
No interpolation	Unacceptable	T4. The editor ripped
One object relative	Acceptable	T5. The book that the editor admired ripped
One object relative	Unacceptable	T6. The book that the editor admired laughed
One object relative	Unacceptable	T7. The book that the editor amused ripped
One object relative	Acceptable	T8. The editor that the book amused laughed
One object relative	Unacceptable	T9 The editor that the book amused ripped
One object relative	Unacceptable	T10. The editor that the book admired laughed
Prepositional phrase plus object relative	Acceptable	T11. The book from the prestigious press that the editor admired ripped
Prepositional phrase plus object relative	Unacceptable	T12. The book from the prestigious press that the editor admired laughed
Prepositional phrase plus object relative	Unacceptable	T13. The book from the prestigious press that the editor amused ripped
Prepositional phrase plus object relative	Acceptable	T14. The editor of the prestigious journal that the book amused laughed
Prepositional phrase plus object relative	Unacceptable	T15. The editor of the prestigious journal that the book amused ripped
Prepositional phrase plus object relative	Unacceptable	T16. The editor of the prestigious journal that the book admired laughed
Object relative plus subject relative	Acceptable	T17. The book that the editor who quit the journal admired ripped
Object relative plus subject relative	Unacceptable	T18. The book that the editor who quit the journal admired laughed
Object relative plus subject relative	Unacceptable	T19. The book that the editor who quit the journal amused ripped
Object relative plus subject relative	Acceptable	T20. The editor that the book that won the award amused laughed
Object relative plus subject relative	Unacceptable	T21. The editor that the book that won the award amused ripped
Object relative plus subject relative	Unacceptable	T22. The editor that the book that won the award admired laughed
Two object relatives	Acceptable	T23. The book that the editor who the receptionist married admired ripped
Two object relatives	Unacceptable	T24. The book that the editor who the receptionist married admired laughed
Two object relatives	Acceptable	T25. The editor that the book that the journalist wrote amused laughed
Two object relatives	Unacceptable	T26. The editor that the book that the journalist wrote amused ripped

Following prior studies, additional material was interpolated between the subject and matrix verb to increase the number of potentially competing memory representations. The examples in T5, T6, T8, and T9 are analogous to examples T1–T4 except that an object relative clause (*. . .that the editor admired . . .*;*. . .that the book amused. . .*) was placed between the subject and matrix verb. Two extra unacceptable contrasts were added, examples T7 and T10, in which there was an anomalous relationship between the transitive verb within the relative clause (e.g., *amused*) and the main subject NP (*The book*). These constructions were intended to encourage readers to fully process the interpolated material.

The remaining conditions added further material between the matrix subject and verb to test whether a search process was needed to retrieve the subject NP. In all of these conditions, eight words (two potential NP competitors) intervened between the subject and verb, as compared to zero and four words (0–1 relative clauses; 0 or 1 NP competitor) for the conditions exemplified by T1–T4 and T5–T10, respectively. However, in addition to examining the impact of interfering material, the remaining conditions were designed to examine how different degrees of syntactic complexity might affect performance on the task. In previous studies, interpolating more material between the crucial constituents decreased asymptotic accuracy. One potential explanation for these decrements in accuracy is that a relevant constituent may be less likely to be retrieved with more interfering material (Lewis, 1996; McElree, 2000; also see General discussion). However, a lower level of accuracy may also result from an increased probability of misanalyzing the sentence, particularly in complex structures with embedded clauses. If the asymptotic levels are partly determined by the likelihood of successfully processing the sentence, then complexity may exert an influence in constructions with otherwise identical numbers of potentially interfering words.

T11–T16 are examples of conditions in which a prepositional phrase attached to the matrix subject (*. . .from the prestigious press*;*. . .of the prestigious journal*) was added before an object relative clause. Because the prepositional phrase combines with the subject NP to form a complex nominal, we anticipated that it would have a small (if any) impact on processing accuracy or speed. T17–T22 are examples of conditions in which an object relative clause with an embedded subject relative clause (*. . .that the editor who quit the journal admired*;*. . .that the book that won the award amused*) was placed between the matrix subject and verb. Relative to a prepositional phrase, the additional relative clause was predicted to substantially increase processing complexity, as the reader must temporarily suspend processing of the object relative clause in order to process the embedded subject relative clause. Finally, conditions exemplified by T23–T26 added two object relative clauses (*. . .that the editor who the receptionist married admired*;*. . .that the book that the journalist wrote amused*) between the matrix subject and verb. Because object relative clauses are more difficult to process than subject relative clauses (e.g., Holmes & O'Regan, 1981), we anticipated that the two object relative clause sentences would engender the lowest levels of accuracy and have the greatest potential to impact the speed of processing. Indeed, two embedded relative clauses approaches what is typically assumed to be the upper bound on the ability to process embedded structures (e.g., Lewis, 1996).

As with examples T7 and T10, constructions like T13, T16, T19, and T22, which have an anomalous relationship between an embedded verb and the subject NP, were included to encourage readers to fully process the interpolated material.

The advantage of the constructions used in Table 2 is that they measure the speed and accuracy of an essential form of binding operation in sentence processing, namely the binding of a subject to a matrix verb in the absence of any special syntactic device like clefting. However, the materials in Table 2 are in other respects less optimal contrasts for testing search models than the cleft constructions used in prior experiments. Unlike cleft constructions, the initial NP in all sentences except T1–T4 in Table 2 is involved in two nonadjacent dependencies; the initial NP is both the subject of the final (matrix) verb and the direct object of the penultimate (object relative clause) verb. One potentially undesirable effect of this is that the subject NP might be primed from binding it to the object position of the penultimate verb. This could attenuate potential effects of linear distance across the contrasts. Nevertheless, it is still the case that the number of elements in memory systematically increased from 0 to 8 words, and search processes are generally predicted to slow with the number of elements in memory.

Method

Participants

Five students from New York University served as participants in the experiment. Each participated in 10 1-h sessions, plus a 1-h practice session for familiarization with the SAT procedure. All participants were native English speakers and were paid for their participation in the experiment. None had participated in Experiment 1.

Materials

Ten sets of 624 sentences were generated. Each set was composed of 24 instances of the 26 sentence types (10 acceptable and 16 unacceptable) listed in Table 2. All 240 instances were of the same word length as illustrated in the Table. Following the same randomization procedures as used in Experiment 1, the 240 instances were recombined into 10 experimental sets, one for each of the 10 sessions, minimizing the repetition of content material within a session. The order of presentation within a session was randomized.

Procedure

The experimental parameters (stimulus presentation and timing) and procedures were the same as described in Experiment 1. Also, as before, each participant performed 10 1-h sessions using one of the 10 sets of materials, and the order of materials was randomized across participants.

Results and discussion

Fig. 5 shows the average (over participants) d' data as a function of processing time for the five conditions of interest: (1) the condition with no interpolation (NO) (open squares), formed by scaling the hit rate for acceptable constructions like T1 and T3 in Table 2 against the false alarm rate for unacceptable constructions like T2 and T4; (2) the condition with one interpolated object relative clause (ORC) (filled squares), formed by scaling the hit rate for acceptable constructions like T5 and T8 against the false alarm rate for unacceptable ones like T6 and T9; (3) the condition with a prepositional phrase and an interpolated object relative clause (PP + ORC) (filled circles), formed by scaling the hit rate for acceptable constructions like T11 and T14 against the false alarm rate for unacceptable ones like T12 and T15; (4) the condition with an interpolated object relative clause with an embedded subject relative clause (ORC + SRC) (filled triangles), formed by scaling the hit rate for acceptable constructions like T17 and T20 against the false alarm rate for unacceptable ones like T18 and T21; and (5) the condition with two embedded object relative clauses (ORC + ORC) (filled diamonds), formed by scaling the hit rate for acceptable constructions like T23 and T25 against the false alarm rate for unacceptable ones like T24 and T26.

As is evident in Fig. 5, asymptotic accuracy decreased as more material was interpolated between the matrix subject and verb, as it did in prior studies. An ANOVA on the d' values at the longest response time (3000 ms) was significant, $F(4, 16) = 11.42$, $MSe = .146$. Pairwise comparisons using Tukey's LSD procedure ($\alpha = .05$) showed that accuracy was higher with no material intervening between the subject and verb (NO, $d' = 2.84$) than with one or more intervening clauses, although the difference was marginal ($p < .066$) with only one intervening object relative clause (ORC). The asymptotic accuracy for the ORC condition (average $d' = 2.49$) was not significantly higher than the PP + ORC condition (average $d' = 2.24$, $p = .257$), but it was significantly higher than constructions with two intervening clauses. The average asymptotic accuracy was higher for the PP + ORC condition than for the ORC + SRC condition (d's of 2.24 vs 1.70), but the difference was not significant ($p = .133$). The PP + ORC condition was significantly higher than the 2ORC condition ($d' = 1.42$). Finally, the 2ORC condition engendered significantly lower asymptotic levels than the ORC + SRC condition.

These asymptotic differences indicate that it is not just the amount of information between the subject and verb but also the complexity of that information that affects the likelihood of successfully interpreting the string. The clearest evidence for this claim is the differ-

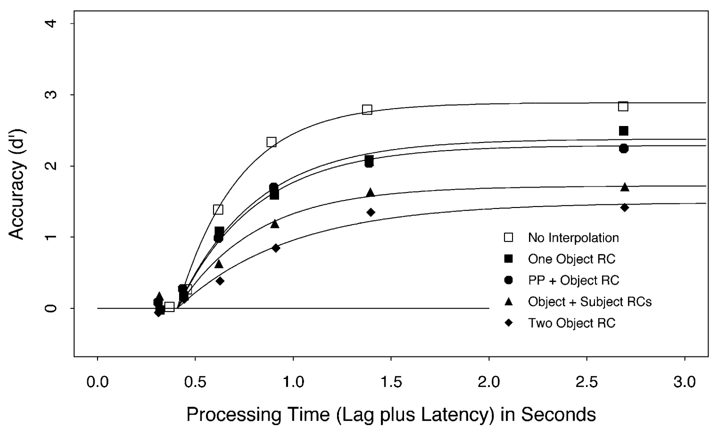

Fig. 5. Average d' accuracy (symbols) as a function of processing time (lag of the response cue plus latency to respond to the cue) for judgments of constructions with no embedded clauses (open squares), one object relative clause (filled squares), a prepositional phrase plus object relative clause (filled circles), an object relative clause and subject relative clause (filled trianges), and two object relative clauses (filled diamonds). Smooth curves show the best fitting $5\lambda - 3\beta - 1\delta$ exponential model (see text of Experiment 2). (RC, relative clause; PP, prepositional phrase).

ence between the ORC + SRC and 2ORC conditions, which contain exactly the same number of nouns, verbs, and function words, but differ only in terms of underlying grammatical structure denoted by word order changes.

Model fits of the full timecourse functions provided further evidence of systematic decreases in accuracy with more intervening material and greater complexity. A $1\lambda - 1\beta - 1\delta$ (null) model, in which all five conditions were fit with a common set of parameters, produced adjusted-R^2 values of .798 in the average data ranging from .41 to .908 across the five participants. A $5\lambda - 1\beta - 1\delta$ model, with separate asymptotes (λs) for each construction type, produced adjusted-R^2 values of .984 in the average data ranging from .919 to .964 across participants. All participants showed an increase in adjusted-R^2 for the $5\lambda - 1\beta - 1\delta$ over the $1\lambda - 1\beta - 1\delta$ model. The $5\lambda - 1\beta - 1\delta$ model yielded a consistent set of asymptotic estimates for the average data; the λ estimates (in d' units) for the average data were 3.00 for the NO condition, 2.35 for the ORC condition, 2.25 for the PP + ORC condition, 1.69 for the ORC + SRC condition, and 1.33 for the 2ORC condition. An ANOVA on the λ estimates across participants was significant, $F(4, 16) = 13.7$, $MSe = .156$. Pairwise comparisons of the λ estimates showed a pattern comparable to the pairwise comparisons of the empirical d' values at the longest lag. (The only exceptions were that the λ estimates for the NO condition were significantly higher, as opposed to marginally higher, than the ORC condition, and that the ORC condition was not significantly different from the ORC + SRC condition, $p = .092$.)

There was no evidence in prior experiments that the amount of intervening material slowed processing speed at the critical verb, despite the fact that it had systematic effects on asymptotic accuracy. We found comparable results here. Models that allocated separate rate parameters (a $5\lambda - 5\beta - 1\delta$ model), separate intercept parameters (a $5\lambda - 1\beta - 5\delta$ model), or separate rate and intercept parameters (a fully saturated $5\lambda - 5\beta - 5\delta$ model) to each of the five conditions did not yield a consistent ordering of β or δ estimates in terms of either the number of intervening words (0, 4, or 8) or the number of intervening clauses (0, 1, or 2). However, there were two notable differences in processing speed, which were best expressed in rate (β).

First, the β estimates were consistently faster for the NO condition than for any other condition. For example, the β estimates from fits of the $5\lambda - 5\beta - 1\delta$ model to the average data were 3.18 for the NO condition, 2.27 for the ORC condition, 2.57 for the PP + ORC condition, 2.34 for the ORC + SRC condition, and 1.74 for the 2ORC condition. The NO condition had the fastest β estimate in every participant's $5\lambda - 5\beta - 1\delta$ fit. Second, the β estimate for the 2ORC condition was notably lower than the ORC, PP + ORC, and ORC + SRC conditions. Again, this pattern was evident in the fit of every participant's data. There was no trend evident among the ORC, PP + ORC, and ORC + SRC conditions, and, as in fits of the average data, the differences among the estimates were small and none were found to be reliable.

The dynamics differences were best fit with a $5\lambda - 3\beta - 1\delta$ model,[1] in which one rate parameter was allocated to the NO condition, another parameter to the ORC, PP + ORC, and ORC + SRC conditions, and the third parameter to the 2ORC condition. This model yielded an adjusted-R^2 value of .988 in the average data, which was higher than any other model with the same, smaller, or greater number of parameters. The adjusted-R^2s ranged from .920 to .966 across participants. T tests on the parameter estimates showed that all rate differences were significant: (a) for the NO condition and the (composite) ORC, PP + ORC, and ORC + SRC conditions, $t(4) = 4.44$, $p = .011$; (b) for the (composite) ORC, PP + ORC, and ORC + SRC conditions and the 2ORC condition, $t(4) = 6.51$, $p = .002$; and (c) for the NO condition and the 2ORC condition, $t(4) = 8.79$, $p = .001$. The smooth functions in Fig. 5 show the best fitting $5\lambda - 3\beta - 1\delta$ model for the average data.

The fastest dynamics were found for the NO condition, where the matrix subject and verb were always adjacent in surface structure. Research on WM with tasks in which information is sequentially presented has consistently found faster dynamics for responses to a small set of items that are thought to be the focus of current processing (for a review, see McElree, 2001). For example, in probe recognition or Sternberg (1966) tasks, in which a list of items is sequentially presented for study, followed immediately by a recognition probe, the SAT dynamics for judgments of a test probe that matches the last item studied is 40–50% faster than judgments of test probes that match any other, less recently studied item (McElree, 1996, 1998; McElree & Dosher, 1989; Wickelgren et al., 1980). The standard account of these findings is that the most recently presented item remains within awareness or focal attention when no mental activity intervenes between study and test. As a consequence, the recognition probe can be compared directly

[1] The dynamics difference was also evident in a $5\lambda - 1\beta - 3\delta$ model in which the difference was forced into intercept (δ) instead of rate (β). However, the adjusted-R^2 values for this model were lower than the $5\lambda - 3\beta - 1\delta$ model (.986 versus .988, respectively, in the average data), suggesting that the underlying difference is more likely a consequence of the rate at which information is processed rather than the minimum time needed to bind and interpret the structures. Fits that varied both intercept and rate with a $5\lambda - 3\beta - 3\delta$ model introduced parameter tradeoffs (e.g., a faster rate but a later intercept), suggesting that the $5\lambda - 3\beta - 1\delta$ model provides a sufficient description of the underlying differences.

to the contents of focal attention. Dynamics are faster for this item because, unlike less recent items, no retrieval process is needed to restore it to a state that is amenable to ongoing cognitive operations. Similar principles appear to apply to the NO condition: The matrix subject can be quickly bound to the verb because it has not been displaced from active processing by other items. If correct, this implies that the 85 ms ($1/\beta$ units) difference in rate between the NO condition and the composite rate for the ORC, PP + ORC, and ORC + SRC conditions provides an estimate of the time needed to retrieve a subject NP that has been displaced from focal attention by the processing of intervening material.

No comparable effect was found in Experiment 1 or in the study reported in McElree (2000). However, the relevant constituents in both studies were always separated by at least one NP. The pattern of results is fully consistent with WM studies that have found that the contents of focal attention are displaced by the processing of even one item (e.g., McElree, 1996, 1998, 2001; McElree & Dosher, 1989).

The second dynamics difference, the notably slower rate for the 2ORC condition, might seem to suggest that interpolated material can slow processing down by making the subject NP more difficult to retrieve, as predicted by the search hypothesis. However, two properties of the data are inconsistent with such an account. First, processing did not systematically slow across conditions that increased either the number of words or the number of clauses. There was, for example, no detectable change in dynamics, either rate or intercept, across conditions that had 4 versus 8 interpolated words (viz., a difference between the ORC and PP + ORC conditions) or across conditions that had 1 or 2 interpolated clauses (viz., a difference between the ORC and ORC + SRC conditions). A search model predicts a systematic change across either the number of words or clauses. Second, the ORC + SRC and 2ORC conditions had exactly the same number of words and clauses, yet there were substantial differences in processing speed. Both facts suggest that a search process was not responsible for the slower rate in the 2ORC condition.

Object relative clauses are known to be more difficult to process than subject relatives (e.g., Holmes & O'Regan, 1981), and embedding object relative clauses appears to disproportionately increase processing difficulty (Gibson & Thomas, 1996, 1998). The slower rate evident in the 2ORC condition is more likely a consequence of an increased tendency to temporarily misanalyze the multiple dependencies in these constructions. Correct parsing of the 2ORC constructions requires the reader to associate three subject-NPs and two object-NPs with the argument positions in a string of three verbs at the end of the sentence. For example, in *The book that the editor who the receptionist married admired ripped*, the reader must bind the third and second NP to the first verb to form *the receptionist married the editor,* the second and first NP to the second verb to form *the editor admired the book,* and the first NP to the final verb to form the main clause *the book ripped*. Misanalysis of any one of these dependencies will leave a verb stranded without an argument or produce a semantic anomaly. Timecourse differences of the form seen here could arise from successful reanalysis following misanalysis on a proportion of times (see McElree, 1993). A difference in rate would result from a greater proportion of reanalysis in the 2ORC as compared to the ORC + SRC and other conditions. The lower asymptotes for the 2ORC condition provide independent evidence that these strings are indeed more difficult to process. (In the General discussion, we suggest that misanalysis in cases like the 2ORC constructions may arise from impoverished retrieval cues.)

An alternative account of the slower dynamics for the 2ORC condition is that this type of structure exceeds the capacity of the mechanisms that underlie normal sentence processing. When readers succeed in discriminating acceptable from unacceptable 2ORC constructions, they do so by an altogether different type of operation, one that might be viewed as a form of problem solving rather than a standard sentence processing operation. To account for the slower rate for the 2ORC condition, one need only assume that this "problem solving" operation has a slower timecourse than normal sentence comprehension operations. That assumption may be justified in that normal sentence comprehension is typically assumed to consist of a set of highly practiced and automated routines. The evidence at hand cannot discriminate between these alternative explanations.[2] However, neither explanation, nor the basic finding itself, suggests that the readers needed to search memory for a representation of the subject NP when processing the final matrix verb. As with Experiment 1, there was independent evidence that readers were processing the interpolated material adequately. The correct rejection rates were high for constructions with anomalous relations in the interpolated regions (examples T7, T10, T13, T16, T19, and T22 in Table 2), specifically, 72.2, 76.4, 78.1, 75.3, 75.6, and 80.1% on average for lags 1–6, re-

[2] If standard sentence processing operations could not be used to discriminate acceptable from unacceptable 2ORC constructions, it might be assumed that it would take several trials to develop an effective strategy. Analyses of the data from the first sessions and final sessions showed that participants were much more accurate by the final session, 87% vs 66%. There are clear practice effects here, as there were in the first experiment and in other conditions of this experiment. However, performance is clearly above chance in the first session, and there is no clear evidence that the practice effect was more pronounced in the 2ORC condition than in other conditions.

spectively. The comparable average rejection rates for the constructions that were ungrammatical at the final verb (T6, T9, T12, T15, T19, and T21) were 42.3, 44.5, 63.4, 79.4, 85.1, and 87.7%. We conclude that processing time does not directly vary with the amount of interpolated material, and this suggests that a search process does not underlie basic binding operations in sentence processing, even in cases where the relevant constituent is not in a focused state.

Experiment 3

Research on retrieval from WM with standard (list-learning) paradigms indicates that a search operation is required only when order (or relational) information needs to be recovered from memory (Gronlund et al., 1997; McElree, 2001; McElree & Dosher, 1993). Processing the nonadjacent dependencies used in prior studies implicitly requires such information, as a reader must recover the relevant item from a particular position in the sentence. The absence of an effect of interpolated material on retrieval speed suggests that some other process or some other information enables readers to resolve nonadjacent dependencies without recourse to a search. As with any null result, however, one has to be concerned that the task was not sensitive enough to detect differences or that the experimental manipulations were not strong enough to engender measurable differences. In this experiment, we sought to determine if findings would differ with structures that stress order information to a greater extent, and to verify that the SAT procedure was providing veridical measures of the speed of processing nonadjacent dependencies.

We increased the importance of order information by employing strings whose acceptability depended on the relative ordering of constituents. We required readers to discriminate acceptable expressions like (9) from unacceptable ones like (10):

(9) *This is the album that the stamps were difficult to mount in.*

(10) *These are the stamps that the album was difficult to mount in.*

Both expressions ended with a verb complex containing a ditransitive verb (*mount*), which accepts as arguments a direct object that appears canonically as a postverbal NP and an indirect object that appears canonically as a postverbal prepositional phrase (PP). In both cases, the NPs have been moved leftward from their canonical positions. Consequently, at the verb and associated preposition, the reader must bind both NPs to the appropriate argument positions of the verb. The ordering of the dislocated NPs determined the acceptability of the string; expressions like (9) are semantically well formed, whereas expressions like (10), which reverse the order of the NPs, are semantically and/or pragmatically ill-formed.

We reasoned that contrasts of this form would provide a strong test of a search mechanism, if such a mechanism were needed to recover the relative ordering of the dislocated constituents. The experimental logic followed prior work in that additional information was interpolated between the filler items (*the album* and *the stamps*) and the gap positions (*mount___in___*), as in examples (11) and (12):

(11) *This is the album that the stamps that obviously angered the fussy collector were difficult to mount in.*

(12) *These are the stamps that the album that obviously angered the fussy collector was difficult to mount in.*

Additionally, we contrasted expressions like (9)–(12) with those like (13)–(16), which contained one filler and one gap, or stranded a filler without a gap:

(13) *This is the album that the collector found difficult to spread open.*

(14) *This is the album that the stamps were difficult to spread open.*

(15) *This is the album that the customer who obviously angered the fussy collector found difficult to spread open.*

(16) *This is the album that the stamps that obviously angered the fussy collector were difficult to spread open.*

Like examples (9) and (11), acceptable strings like (13) and (15) began with two NPs (*the album* and *the collector/customer*). However, they ended with a phrasal verb consisting of a verb particle construction (*spread open*) that took the clefted object NP (*the album*) as its sole argument. Unacceptable variants like (14) and (16) used the same phrasal verb construction, but replaced the complement clause (e.g., *found difficult*) with a "tough" (or subject-to-object raising) construction. This resulted in the second NP (*the stamps*) being left stranded without a valid argument position.

Our primary reasons for including single argument structures like (13)–(16) was to demonstrate further that the SAT procedure was sensitive to on-line differences in the timecourse of processing nonadjacent dependencies. One might be concerned that judgments at the end of a sentence are not sensitive to the timecourse of on-line processes associated with binding a dislocated argument, possibly because general sentence wrap-up effects might mask them. The rate differences documented in Experiment 2 provided clear evidence against such a notion. There we found that adjacent arguments were processed faster (higher SAT rate) than nonadjacent ones, demonstrating that the SAT procedure provides a sensitive estimate of the time that is required to retrieve a dislocated constituent from memory. Here we sought an additional form of evidence: Constructions like (9)–(12), which involved the retrieval and assignment of two arguments, should be processed slower than constructions like (13)–(16), which required the retrieval and assignment of one argument only. Additionally, constructions like (13)–(16) were included to discourage readers from paying particular attention to the two argument constructions.

Table 3 shows the main contrasts employed in the experiment. (The conditions in Table 3 represent 60% of the material used in the experiment. The remaining material served as filler constructions, consisting of two conjoined main clauses of variable lengths.) Each acceptable version of the short and long, single and double gap constructions (T1, T3, T5, and T7) consisted of two versions, denoted as (a) and (b), that exchanged or reversed the order of NP arguments (*album* and *stamp*), with appropriate modifications of the final verb complex. This was done to ensure that acceptability could not be predicted from presentation order. As in prior work, we included constructions like (T9) and (T10) with anomalies in the interpolated region to ensure that readers fully processed the middle regions of the longer strings.

In this experiment, the strings were presented in a phrase-by-phrase rather than a word-by-word method. The phrase breaks are indicated by the slashes in Table 3. We initially ran a version of this experiment in a word-by-word fashion, and found that performance was well above chance and nearly asymptotic at the earliest response times (50–300 ms) after the onset of the final preposition (e.g., *in* in the verb complex *mount in*). Apparently, the verb itself provided much of the information needed to assign argument roles to the dislocated NPs. With the fragment *This is the album that the stamps were difficult to mount,* for example, readers apparently assumed that *the stamps* was the direct object argument and anticipated that *the album* was the indirect object. To ensure that our measure included the time to access the relevant NPs, we measured time relative to the onset of the final infinitival verb phrase (e.g., *to mount in*).

Method

Participants

Eight students from New York University served as participants in the experiment. Each participated in 10 1-h sessions, plus a 1-h practice session for familiarization with the SAT procedure. All participants were native English speakers, and were paid for their participation in the experiment. None had participated in Experiments 1 or 2.

Materials

Ten sets of 336 sentences were generated. As in prior studies, each set was composed of 24 instances of the 14 sentence types (eight acceptable and six unacceptable) listed in Table 3. Following the same randomization

Table 3
Constructions used in Experiment 3

Construction type	Acceptability	Example
Short double gap	Acceptable	T1a. This is the album/that the stamps/were difficult/to mount in T1b. These are the stamps/that the album/was difficult/to complete without
Short double gap	Unacceptable	T2. These are the stamps/that the album/was difficult/to mount in
Long double gap	Acceptable	T3a. This is the album/that the stamps/which obviously angered/the fussy collector/were difficult/to mount in T3b. These are the stamps/that the album/which obviously angered/the fussy collector/was difficult/to complete without
Long double gap	Unacceptable	T4. These are the stamps/that the album/which obviously angered/the fussy collector/was difficult/to mount in
Short single gap	Acceptable	T5a. This is the album/that the collector/found difficult/to spread open T5b. These are the stamps/that the collector/found difficult/to spread out
Short single gap	Unacceptable	T6. This is the album/that the stamps/were difficult/to spread out
Long single gap	Acceptable	T7a. This is the album/that the customer/who obviously angered/the fussy collector/found difficult/to spread open T7b. These are the stamps/that the customer/who obviously angered/the fussy collector/found difficult/to spread out
Long single gap	Unacceptable	T8. This is the album/that the stamps/which obviously angered/the fussy collector/were difficult/to spread out.
Control	Unacceptable	T9 This is the album/that the stamps/which obviously despised/the fussy collector/were difficult/to mount in
Control	Unacceptable	T10. These are the stamps/that the album/which obviously despised/the fussy collector/was difficult/to complete without

procedures used in the other experiments, the 240 instances were recombined into 10 experimental sets, one for each of the 10 sessions, minimizing the repetition of content material within a session. These ten sets were combined with 2400 filler strings (half acceptable, half unacceptable), 240 per set. None of these strings used a clefted structure, but all included 3 or 4 NPs and ranged from 16 to 22 words in length. The order of presentation within a session was again randomized.

Procedure

As before, each participant performed 10 1-h sessions using one of the 10 sets of materials, and the order of materials was randomized across participants. The experimental parameters (stimulus presentation and timing) and procedures were the same as described in Experiments 1 and 2, with the following exceptions. The strings were presented in a phrase-by-phrase manner. Each phrase was displayed for 300 ms times the number of words in the phrase. The critical region of the final verb phrase in each condition consisted of three words, displayed for a total duration of 900 ms. The response tone was presented at either 300, 500, 700, 900, 1500, or 3000 ms after the onset of this final phrase. Longer lag times were selected because more reading time was required for a phrase as opposed to a single word.

Results and discussion

Fig. 6 shows the average (over participants) d' data as a function of processing time for the four conditions of interest; the double gap short (solid squares) and long (solid triangles) constructions, formed by scaling the hit rate for acceptable constructions like T1 (short) and T3 (long) in Table 3 against the false alarm rate for the appropriate unacceptable constructions like T2 (short) and T4 (long), and the single gap short (open squares) and long (open triangles) constructions, formed by scaling the hit rate for acceptable constructions like T5 (short) and T7 (long) against the false alarm rate for the appropriate unacceptable constructions like T6 (short) and T8 (long).

An ANOVA on the d' values at the longest response time (3000 ms) with length (short and long) and number of gaps (single and double) as repeated measures factors yielded a significant main effect of the interpolated material, $F(1,7) = 25.7$, $MSe = .346$, and a significant main effect of the number of gaps, $F(1,7) = 5.6$, $MSe = .264$, with no interaction ($F < 1$). On average, asymptotic accuracy was lower by approximately 1 d' unit for a long as compared to a short distance between the filler and gap (2.05 vs 3.10, respectively), and lower by almost one-half a d' unit for double as compared to single arguments (2.36 vs 2.80, respectively). Consequently, model fits of the full timecourse functions with a $1\lambda - 1\beta - 1\delta$ (null) model, in which all four conditions were fit with a common set of parameters, produced a very low adjusted-R^2 value of .470 in the average data, ranging from .297 to .695 across the eight participants. By comparison, a $4\lambda - 1\beta - 1\delta$ model, with separate asymptotes (λs) for each of the four construction types, produced adjusted-R^2 values of .920 in the average data ranging from .712 to .878 across participants. All participants showed a dramatic increase in adjusted-R^2 for the $4\lambda - 1\beta - 1\delta$ model over the $1\lambda - 1\beta - 1\delta$ model. An ANOVA on asymptotic estimates likewise yielded a significant main effect of the interpolated material, $F(1,7) = 38.1$, $MSe = .240$, and a significant main effect of the number of gaps, $F(1,7) = 23.0$, $MSe = .346$, with no interaction ($F < 1$). For the average data, the λ estimates were 3.51 for the short single gap condition, 2.43 for long single gap condition, 2.47 for the short double gap condition, and 1.45 for the long double gap condition.

Subsequent model fits found clear evidence for an effect of the number of gaps on processing speed, with

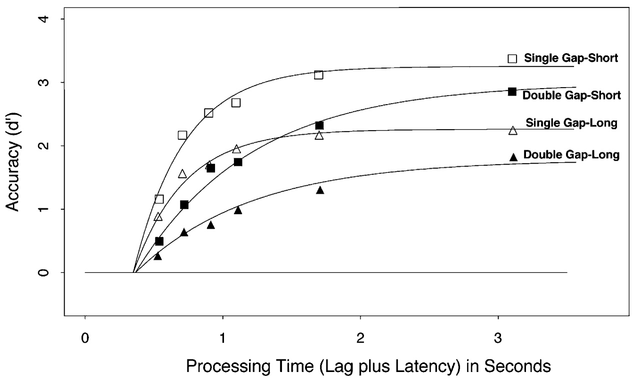

Fig. 6. Average d' accuracy (symbols) as a function of processing time (lag of the response cue plus latency to respond to the cue) for judgments of short (squares) and long (triangles) single gap constructions (open symbols) and double gap constructions (filled symbols). Smooth curves show the best fitting $4\lambda - 2\beta - 1\delta$ exponential model (see text of Experiment 3).

rate being slower for double gaps as compared to single gaps, but no indication that the amount of interpolated material affected processing speed in either condition. (All the dynamics differences were better expressed in rate rather than intercept, so we discuss rate differences only.) A $4\lambda - 2\beta - 1\delta$ model, with separate rates (βs) for the double gap as compared to single gap conditions, yielded an adjusted-R^2 value of .981 in the average data, ranging from .719 to .912 across participants. All participants showed an increase in adjusted-R^2 for this model over the $4\lambda - 1\beta - 1\delta$. The β estimate was 1.15 for the double gap conditions as compared to 2.78 for the single gap conditions in fits of the average data, and every participant showed this ordering of parameter estimates. Consequently, a t-test on the β estimates was significant, $t(7) = 3.36$, $p < .01$. In millisecond units (β^{-1}), the average difference in processing speed was substantial, over 500 ms (869 vs 359 ms for double and single gap conditions).

The analogous $4\lambda - 2\beta - 1\delta$ model, in which separate rates (βs) were allocated to the short and long constructions rather than the single and double gap constructions, yielded an adjusted-R^2 value of .916 in the average data, lower than the simpler $4\lambda - 1\beta - 1\delta$ model. A drop in adjusted-R^2 was evident in six of the eight participants. More importantly, no consistent difference in rate estimates was observed, and a *t* test on the β estimates was nonsignificant, $t(7) = .69$, $p < .51$.

An absence of an overall effect of the amount of interpolated material would not be surprising for single gap constructions, as these are similar in structure to constructions used in McElree (2000) and in certain respects to the constructions in Experiment 1. However, one might have expected the interpolated material to have an impact on double gap constructions, as order information is more critical in these constructions than in the single gap constructions. A $4\lambda - 4\beta - 1\delta$ model was fit to the data files to test this hypothesis. Overall, this model did not improve the adjusted-R^2 over the $4\lambda - 2\beta - 1\delta$ model that allowed rate to vary with the number of gaps, but our primary purpose was to derive rate estimates for each of the four conditions to examine if there were systematic differences among the double gap constructions that might not be evident in the single gap constructions. In the average data, the β estimates were 1.25 for the short double gap condition, .84 for long double gap condition, 2.62 for the short single gap condition, and 3.03 for the long single gap condition. For the double-gap conditions, there is a trend for slower processing in the long as compared to the short conditions, but this trend is only evident in four out of the eight participants. Consequently, neither the main effect of short versus long nor the interaction between the amount of interpolated material and number of gaps was significant in an ANOVA on the rate parameter estimates, $F(1,7) = 1.49$, $MSe = .347$ and $F(1,7) = 2.55$, $MSe = .375$, respectively. However, the main effect of single versus double gaps was significant, $F(1,7) = 5.78$, $MSe = 17.9$.

The single and double gap constructions differed only in terms of the structural analysis mandated by the penultimate phrase (e.g., *were difficult* in T1 versus *found difficult* in T5a) and by whether the final verb was transitive or ditransitive. The rate differences between these constructions indicate that the SAT procedure is sensitive to the temporal aspects of resolving long distance dependencies. Given these differences, it is likely that we would have detected a search process had it been used to access the dislocated constituents. As in other experiments, there was independent evidence that readers paid attention to the interpolated material: Correct rejection rates for the control structures with ill formed relations in the interpolated region (T9 and T10) were high; specifically 85.5, 83.7, 84.1, 81.3, 84.5, and 82.8% across the six lags, as compared to 46.7, 63.7, 70.7, 74.3, 84.6, and 90.1% for constructions that were ill-formed at the final phrase (T2, T4, T6, and T8). The interpolated material would have had measurable effects on SAT dynamics if a search had been used.

General discussion

When participants are required to determine the temporal or spatial ordering of items in a memory list (Gronlund et al., 1997; McElree & Dosher, 1993) or to locate an items in a particular position in a memory list (McElree, 2001), increasing the amount of material interpolated between study and test systematically reduces accuracy and slows judgment speed. The reduction in accuracy is consistent with well established principles that interpolated material decreases the probability of successfully retrieving the required information, either by decreasing the quality of the memory representation itself or by reducing the distinctiveness of the retrieval cues needed to recover the relevant information. The systematic slowing of retrieval speed suggests that order information is retrieved by a search mechanism, either by scanning a structured set of memory representations or by using local cues to reconstruct the order of events (McElree, 2001; McElree & Dosher, 1993). In both cases, additional information in memory decreases the speed with which the required information is recovered. McElree and Dosher (1993) and McElree (2001) found that even one interpolated item measurably slowed retrieval speed, engendering differences in SAT intercept and/or rate.

Here, and in the experiment reported in McElree (2000), the amount of material interpolated between two constituents (e.g., NP and verb) decreased the accuracy of discriminating acceptable from unacceptable relations between the constituents. Processing additional material increased the likelihood of readers either: (a) failing to retrieve the earlier processed constituent or (b) misanalyzing relations in the sentence. These accounts are not mutually exclusive, and we believe that both factors may have contributed to the observed decrements in accuracy.

Experiment 2 found that accuracy was lower for complex strings with the same number of constituents as simpler strings, suggesting that accuracy was partly determined by the probability of successfully analyzing the relations in the strings. On the other hand, McElree (2000) reported data from a probe recognition study in which interpolated material also decreased the probability that an earlier processed constituent was retrieved from memory. Participants were interrupted with a test probe and required to judge whether it was synonymous with an element in the sentence at one of five points (denoted by numbers 1–5) during the reading of strings like those in (17)–(19):

(17) *It was the fearless passengers* [1] *who the able sailor* [2] *advised* [3] *about the lifeboats* [4] *although the heavy storm was quickly abating* [5].

(18) *The able sailor* [1] *had advised* [2] *the fearless passengers* [3] *about the lifeboats* [4] *although the heavy storm was quickly abating* [5].

(19) *The able sailor* [1] *believed that* [2] *the fearless passengers* [3] *entered the lifeboats* [4] *although the heavy storm was quickly abating* [5].

The crucial trials involved test probes that were synonymous with an element in the initial NP. Accuracy monotonically declined across probe positions 1–5, indicating that the initial NP's representation was less accessible as more information was processed. The only exception to this pattern was between points 3 and 4 for strings like (17), where the clefted NP was bound to the direct object position of the matrix verb (e.g., *advised*). Across this region, accuracy increased in strings like (17) but not in ones like (18) or (19). The increased accessibility after a gap is consistent with other reports that a gap in constituent structure reactivates the dislocated element (e.g., Bever & McElree, 1988; McElree & Bever, 1989; Nicol & Swinney, 1989; Osterhout & Swinney, 1993; Swinney et al., 1988).

Decrements in accuracy as a function of additional processing are not surprising; they are consistent with the claim that standard memory variables affect the difficulty of sentence processing (e.g., Gibson, 1998; Lewis, 1996). For example, Lewis (1996) has argued that processing difficulty is related to the amount of retroactive and proactive inhibition that various sentence structures engender in WM. The important and less intuitive results concern processing speed. Experiment 2 showed that adjacent arguments were processed faster than arguments separated by one or more constituents, and that very complex syntactic structures were processed slower than simpler structures, either because they required more reanalysis or because they recruited alternative procedures. Furthermore, Experiment 3 demonstrated that processing was slower when the sentence required resolving two nonadjacent dependencies rather than one. Crucially, however, across three experiments with different types of sentence structures and syntactic dependencies, there was no evidence to indicate that the amount of interpolated information affected the speed of binding dislocated constituents.

These results are inconsistent with the hypothesis that a search process is needed to access a previously processed constituent prior to binding it to an argument position. Such a hypothesis predicts that the amount of interpolated material should have slowed processing down, engendering systematic timecourse differences similar to what has been found in studies of the retrieval of order or relational information (Gronlund et al., 1997; McElree, 2001; McElree & Dosher, 1993). Instead, readers have direct access to the memory structures that are constructed in the course of processing a sentence. Apparently, some process or information enabled readers to circumvent the slow search process that was needed to recover order information in other domains. Below we outline and evaluate possible explanations for this finding. We first evaluate and reject two structural accounts that have some precedent in the sentence processing literature, and then propose an alternative account that attributes the observed pattern to content-addressable memory structures.

Structural accounts

Direct access may result from special cognitive structures that enable constituents to be maintained in a readily accessible state. There are at least two variants of this type of account that have some support in the literature.

Focus of attention

One would not expect interpolated material to affect processing speed if the relevant constituent were maintained in focal attention until it was bound to an argument position in the sentence. The processing of dislocated constituents (e.g., the clefted NP in Experiments 1 and 3) has been argued to follow an *Active Filler Strategy* (Clifton & Frazier, 1989; Fodor, 1995), which might be construed in this fashion. This strategy has been framed in general terms, with minimal commitments to architectural details. It requires only that parsing operations mark constituents that have yet to be assigned a role in the sentence and posit positions for the marked fillers in syntactically licensed positions. However, one could envision a variant of this strategy in which unassigned constituents were maintained in attention until they were assigned a role, which would circumvent the retrieval processes that would otherwise be needed to restore these constituents to active processing. The idea that focal attention is distinct from more passive working memory states, and that information in focal attention does not require the same retrieval operations that are needed to recover information outside of focal attention, is supported by studies which have shown that access to an item is exceptionally fast when no processing intervenes between study and test (McElree, 1996, 1998, 2001; McElree & Dosher, 1989, 1993; Wickelgren et al., 1980).

Unfortunately, this account is not fully compatible with the results reported here, particularly Experiment 2. Note first that this type of explanation is applicable only to structures like the cleft construction, which are thought to focus the initial NP. Because the dislocated constituent occurs in a marked (nonargument) position, readers would be able to recognize the constituent as one that requires subsequent processing and therefore amenable to the *Active Filler Strategy*. However, the same pattern implicating direct access was found with subject-verb dependencies in Experiment 2, which are not thought to be amenable to such a strategy. Minimally, such an account would have to assume that items other than those explicitly marked for focus and for special processing are routinely maintained in focal attention.

There are also problems extending this account to the processing of sentences with more than one long-distance dependency. To fully account for the contrasts employed here, more than one dislocated item would have to be simultaneously maintained in focal attention; at least two items in the double gap constructions of Experiment 3, and up to three items in the multiply embedded structures of Experiment 2. Some have argued that focal attention has a capacity of 3–4 items (Cowan, 2001), but that claim is controversial and inconsistent with several findings (see McElree, 2001; McElree & Dosher, 2001). More problematic for this account is that some operation would be needed to select the correct constituent if more than one item were maintained in focal attention. Scanning or search operations would be inconsistent with the timecourse data, so this approach would have to presuppose the direct access operation that it was intended to explain.

Additionally, there are data that are incompatible with this account. First, if clefted items were typically maintained in focal attention, then different recognition profiles should have been observed in the probe recognition task reported by McElree (2000). Rather than systematically decreasing with interpolated material, accuracy should have remained flat (and relatively high) until the argument position for the clefted NP was processed (i.e., flat across regions 1, 2, and 3 in example 17). To the contrary, the observed profiles indicate that interpolated material interfered with the representation of the clefted constituent, which suggests that the processing of subsequent material displaces the constituent from focal attention.[3] Second, cross-modal priming studies have found evidence for activation of the displaced item *after* but not *before* the gap position (Swinney et al., 1988), likewise suggesting that the displaced item was not maintained in focal attention but rather restored to active processing once a gap was found.

Finally, we note that the timecourse measures in Experiment 2 showed that processing is exceptionally fast with adjacent arguments. This result is consistent with several memory studies that have found that memory judgments are notably fast when new material has not displaced a target item from focal attention (McElree, 1996, 1998, 2001; McElree & Dosher, 1989, 1993; Wickelgren et al., 1980). Time-course measures in the two domains appear to give convergent estimates that focal attention is limited to the most recently processed unit only (see Garavan, 1998; McElree, 1998, 2001; McElree & Dosher, 2001). This suggests that direct access to previously processed constituents should not be attributed to their maintenance in focal attention. Additionally, these findings cast doubt on approaches which argue that linguistic focus is tantamount to the focus of attention (Gundel, 1999); that, for example, linguistic focus resulting from syntactic devices like clefting (Experiments 1 and 3) is mediated by placing a constituent within focal attention. Collectively, the evidence at hand suggests that direct access is not simply a consequence of a linguistic device or parsing strategy that serves to maintain a constituent in a special focused state. The evidence points to a more general mechanism.

Specialized memory structures

In computational models of natural language processing, nonadjacent dependencies are often parsed with specialized memory structures (stacks, buffers, etc.) designed to hold a constituent until it can be assigned to a structural position in the sentence (e.g., Gazdar & Mellish, 1989). Such a strategy was made popular in the Augmented Transitional Network (ATN) models in the early 1970s (Woods, 1973), and at least one behavioral study purported to find evidence supporting the psychological reality of this type of specialized memory structure (Wanner & Maratsos, 1978).

[3] It might be assumed that interpolated material displaces the dislocated constituent from focal attention on a proportion of trials only, so that binding of the dislocated constituent to a subsequent argument position consists of a mixture of two types of operations, either the binding of constituents to a representation in focal attention or to a representation in WM when the constituent has been displaced from focal attention. However, mixture models of this form also predict substantial differences in SAT rate as the amount of interpolated material increases (see McElree, 2001) and, therefore, are also inconsistent with the observed SAT dynamics.

To process sentences with multiple dependencies (Experiments 2 and 3), a buffer with a capacity larger than one item is needed. Furthermore, to be consistent with the timecourse data, items in the buffer would need to be accessed without a search process. Both constraints are satisfied with a pushdown stack architecture, in which constituents are stored and retrieved in a "last-in first-out" fashion (Fodor, 1978; Kaplan, 1973). A constituent is stored by placing it on the top of the stack, pushing previously stored constituents further down in the stack. Constituents are retrieved by "popping" or removing the item from the top of the stack. An attractive property of this architecture is that it codes order information implicitly: Constituents are retrieved in exactly the reverse order in which they are stored. As such, it provides an explanation for why the recovery of order information in sentence processing appears to be less difficult than in other domains.

In a pushdown architecture, only the constituent on the top of the stack is directly accessible.[4] In the case of sentences with multiple long-distance dependencies, direct access is only possible when the fillers (F) and gaps (G) are in a nested relationship, an example of which is schematically depicted in (20):

(20) $F_1 \quad F_2 \quad F_3 \quad G_3 \quad G_2 \quad G_1$.

The required filler for each gap (denoted by indices) will be at the top of the stack, hence directly accessible, if the previous filler is removed from the stack when it is assigned to the structural role of its gap. Limiting direct access to only nested structures is less severe than might be assumed, however, as many languages, including English, largely conform to what has been described as a nested-dependency constraint (Fodor, 1978, 1995).

All constructions with long-distance dependencies used in the current experiments and in McElree (2000) had nested relationships, so prima facie this type of architecture is consistent with the findings that interpolated material did not affect processing speed, although to account for decreases in accuracy with additional interpolated materials, the architecture would need to be embellished by assuming that there is loss of item or order information as more constituents are added to the stack. There are, however, several reasons to question whether this type of architecture provides an adequate account of the memory structures in comprehension. Fodor (1978) notes that this architecture predicts that nested relationships should be among the easiest of all hierarchical relations to processes. In fact, nested structures are among the most difficult relationships to process. Even granting that other factors might explain the general difficulty associated with processing nested structures, this approach minimally predicts that nested dependencies should be easier to process than crossed dependencies (e.g., *I told them what to buy*), which are at odds with the order in which constituents are removed from the stack. In fact, they are not. Bach, Brown, and Marslen-Wilson (1986) found that nested dependencies in German are harder to process than crossed dependencies in Dutch.

Additionally, there are also reports that sentences with multiple long-distance dependencies are easier to process if the dislocated constituents are more semantically and syntactically distinct from one another (King & Just, 1991; Lewis, 1996; Stolz, 1967). These results are at odds with a pushdown storage structure that uniquely determines which constituent is to be retrieved at any point in the sentence. Semantic similarity effects suggest that binding operations are sensitive to the content of the dislocated constituents.

A stack model could attribute similarity effects to the probability of storing or maintaining a representation in the stack. However, it seems unlikely that the difficulty associated with processing sentences that have similar constituents can be accounted for by differences in storage alone. Consider, for example, a notoriously difficult string like (21):

(21) *The dog the rat the cat chased bit fell.*

When processing the final string of verbs, retention of the NPs or even the order in which they occurred does not seem to be the limiting factor. Indeed, having the full sentence visible at best attenuates but does not eliminate the difficulty, as the reader can verify.

Content-addressability

We suggest that a more plausible explanation for the difficulty engendered by structures like (21) is that retrieval is cue-dependent, and that high similarity renders less effective the usual semantic and syntactic cues that enable access to appropriate memory representations. In (21), for example, the NPs are all plausible subjects and/or direct objects of the verbs. As the verbs do not uniquely cue a NP, it is difficult for a reader to determine which NP should be associated with the argument roles of the verbs.

Cue-dependent retrieval, when coupled with findings indicating direct access, strongly implicates a content-addressable memory system, in which cues at retrieval make contact with related memory structures without a search through extraneous representations. In such a system, the quality of the retrieval cues can affect the probability of retrieving a memory representation but need not affect the speed at which the representation can be retrieved (for a formal model, see Ratcliff, 1978, 1981

[4] Retrieving items from other positions in the stack requires additional operations. Related WM architectures have proposed that less recent items are retrieved with a serial scanning operation that moves down the stack (e.g., Theios, 1973; see McElree & Dosher, 1989 for timecourse predictions for this type of model).

and Ratcliff, Van Zandt, & McKoon, 1999; for applications to WM, see McElree & Dosher, 1989).

Our timecourse findings are fully consistent with this type of memory system. Interpolated material decreases the distinctiveness of the retrieval cues used to recover the relevant representation by adding additional information that alters the retrieval context. The reduction in distinctiveness will engender lower asymptotic levels. We noted that there are at least two reasons why interpolated material decreases asymptotic accuracy, and both are explicable with a content-addressable approach. Shifting the retrieval context and thereby reducing the effectiveness of the retrieval cues may result in a failure to access the required representation. (Note that failure to retrieve a representation is not the same as total loss of the representation from memory.) Alternatively, the shift in retrieval context may result in the retrieval of an inappropriate constituent. This in turn may lead to a misanalysis of the structure of the sentence, which could result in judging an acceptable sentence as unacceptable.

Fully specifying a content-addressable model requires enumerating and explicating the various forms of information that are computed during the course of processing a sentence and that can potentially drive memory access. At present, this is beyond the current state of the field. However, evidence at hand suggests that both semantic and syntactic information are among the forms of information that drive memory access. The former is suggested by the semantic similarity effects in the processing of sentences with multiple dependencies (King & Just, 1991; Lewis, 1996; Stolz, 1967). Two forms of evidence implicate syntactic information.

Lewis (1996) notes that comprehensible Japanese constructions like (22) involve five nonadjacent dependencies, well beyond the number that is typically comprehensible in English:

(22) *John-wa Bill-ni Mary-ga Sue-ni Bob-o syookai sita to it-ta.*

John-TOPIC Bill-DATIVE Mary-NOMINATIVE Sue-DATIVE Bob-ACCUSATIVE introduced say

"John said to Bill that Mary introduced Bob to Sue."

Lewis (1996) argues that (22) is comprehensible because the NPs are marked for case, which increases their distinctiveness in memory relative to a string of five unmarked NPs in English. For example, the verb *syookai* (introduce) requires an accusative (direct object) and a dative (indirect) argument. The accusative constraint uniquely identifies *Bob* as the relevant NP, and dative constraint reduces the set to either *Bill* or *Sue*.

Additionally, there is evidence from on-line sentence processing tasks that binding of dislocated constituents respects what linguists have termed *island constraints* (e.g., see Bourdages, 1992; McElree & Griffith, 1998; Pickering, Barton, & Shillcock, 1994; Stowe, 1986; Traxler & Pickering, 1996). An indefinite amount of lexical material can intervene between a filler and a gap, but there are, nevertheless, restrictions on the type of syntactic structures across which a filler can be associated with a gap. Since Ross (1967), linguists have characterized syntactic constituents that block filler-gap associations as syntactic *islands*. Crucially, island constraints reflect a type of configurational constraint defined over the geometry of the syntactic structure. That on-line binding operations respect this type of constraint suggests that any viable content-addressable model will have to acknowledge the role played by abstract syntactic relations.

There are at least two open issues that require further research. First, it is unclear what role serial order information proper may play in sentence comprehension. If memory structures are content-addressable, then it is conceivable that syntactic and semantic information alone are all that is typically needed to associate dislocated constituents. Serial order information may be used only in impoverished situations like (21) or the 2ORC constructions in Experiment 2 (Examples T23 and T25 in Table 2), where semantic and syntactic cues might not be sufficient to uniquely identify the correct arguments. A serial mechanism of the type observed in the WM studies (McElree, 2001; McElree & Dosher, 1993) may be the only viable means of recovering the correct constituent in such circumstances.

Second, it remains to be determined whether the same types of memory operations mediate the binding of intrasentential or various intersentential constituents. For example, the mechanism identified in this research could be the same one that mediates the resolution of pronouns, or provides access to representations necessary for discourse inferences. What these three experiments make clear is that the memory structures enabling sentence comprehension utilize a direct-access mechanism that is plausibly guided by content-addressable retrieval cues. There is no compelling evidence to suggest that this mechanism is distinct from those used to access intersentential constituents.

Acknowledgment

The research was supported by NIMH Grant MH57458 (awarded to B.M.).

References

Bach, E., Brown, C., & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch. *Language and Cognitive Processes, 1*, 249–262.

Bever, T. G., & McElree, B. (1988). Empty categories access their antecedents during comprehension. *Linguistic Inquiry, 19*, 35–43.

Bourdages, J. S. (1992). Parsing complex NPs in French. In H. Goodluck & M. Rochemont (Eds.), *Island constraints: Theory, acquisition and processing* (pp. 61–88). Dordrecht, Holland: Kluwer Academic Publishers.

Birch, S. L., & Garnsey, S. M. (1995). The effect of focus on memory for words in sentences. *Journal of Memory and Language, 34*, 232–267.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences, 22*, 77–126.

Chandler, J. P. (1969). Subroutine STEPIT—finds local minimum of a smooth function of several parameters. *Behavioral Science, 14*, 81–82.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3*, 37–60.

Clifton, C., Jr., & Frazier, L. (1989). Comprehending sentences with long distance dependencies. In G. N. Carlson & M. K. Tanenhaus (Eds.), *Linguistic structures in language processing* (pp. 273–317). Dordrecht, Holland: Kluwer Academic Publishers.

Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford: Oxford University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185.

Dosher, B. A. (1976). The retrieval of sentences from memory: A speed-accuracy study. *Cognitive Psychology, 8*, 291–310.

Dosher, B. A. (1979). Empirical approaches to information processing: Speed-accuracy tradeoff or reaction time. *Acta Psychologica, 43*, 347–359.

Dosher, B. A. (1981). The effect of delay and interference: A speed-accuracy study. *Cognitive Psychology, 13*, 551–582.

Dosher, B. A. (1982). Sentence size, network distance and sentence retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 173–207.

Dosher, B. A. (1984). Degree of learning and retrieval speed: Study time and multiple exposures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 541–574.

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry, 98*, 427–473.

Fodor, J. D. (1995). Comprehending sentence structure. In L. R. Gleitman & M. Liberman (Eds.), *An invitation to cognitive science: Language* (Vol. 1, pp. 209–246). Cambridge, MA: MIT Press.

Garavan, H. (1998). Serial attention within working memory. *Memory & Cognition, 26*, 263–276.

Gazdar, G., & Mellish, C. (1989). *Natural language processing in prolog*. New York: Addison-Wesley.

Gibson, E. & Thomas, J. (1996). The processing complexity of English center-embedded and self-embedded structures. In: Schütze, C. (Ed.) Proceedings of the NELS 26 sentence processing workshop. MIT Occasional Papers in Linguistics 9. Cambridge, MA.

Gibson, E. & Thomas, J. (1998). The complexity of nested structures in English: Evidence for the syntactic prediction locality theory of linguistic complexity. MIT manuscript.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1–78.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1–67.

Gronlund, S. D., Edwards, M. B., & Ohrt, D. D. (1997). Comparison of the retrieval of item versus spatial position information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1261–1274.

Gundel, J. K. (1999). On different kinds of focus. In P. Bosch & R. van der Sandt (Eds.), *Focus: Linguistic, cognitive, and computational perspectives* (pp. 293–305). New York: Cambridge University Press.

Hacker, M. J. (1980). Speed and accuracy of recency judgments for events in short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 6*, 651–675.

Hinton, G. E. (1989). Implementing semantic networks in parallel hardware. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 191–217). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hintzman, D. L. (1984). MINERVA2: A simulation model of human memory. *Behavior Research Methods, Instruments, and Computers, 16*, 96–101.

Holmes, V. M., & O'Regan, J. K. (1981). Eye fixation patterns during the reading of relative clause sentences. *Journal of Verbal Learning and Verbal Behavior, 20*, 417–430.

Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model-comparison approach*. San Diego: Harcourt Brace Jovanovich.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122–149.

Kaplan, R. (1973) A multi-processing approach to natural language. In: Proceedings of the First National Computer Conference.

Kawamoto, A. (1988). Distributed representations of ambiguous words and their resolution in a connectionist network. In S. L. Small, G. W. Cottrell, & M. K. Tanenhaus (Eds.), *Lexical ambiguity resolution: Perspectives from psycholinguistics, neuropsychology, and artificial intelligence* (pp. 195–228). San Mateo, CA: Morgan Kaufmann.

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

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York: Wiley.

Lewandowsky, S., & Murdock, B. B., Jr. (1989). Memory for serial order. *Psychological Review, 96*, 25–53.

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

McElree, B. (1993). The locus of lexical preference effects in sentence comprehension: A time-course analysis. *Journal of Memory and Language, 32*, 536–571.

McElree, B. (1996). Accessing short-term memory with semantic and phonological information: A time-course analysis. *Memory & Cognition, 24*, 173–187.

McElree, B. (1998). Attended and non-attended states in working memory: Accessing categorized structures. *Journal of Memory and Language, 38*, 225–252.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research, 29*, 111–123.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 817–835.

McElree, B., & Bever, T. G. (1989). The psychological reality of linguistically defined gaps. *Journal of Psycholinguistic Research, 18*, 21–35.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: Time course of recognition. *Journal of Experimental Psychology: General, 118*, 346–373.

McElree, B., & Dosher, B. A. (1993). Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General, 122*, 291–315.

McElree, B., & Dosher, B. A. (2001). The focus of attention across space and across time. *Behavioral and Brain Sciences, 24*, 129–130.

McElree, B., & Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 134–157.

McElree, B., & Griffith, T. (1998). Structural and lexical constraints on filling gaps during sentence processing: A time-course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 432–460.

McElree, B., & Nordlie, J. (1999). Literal and figurative interpretations are computed is equal time. *Psychonomic Bulletin & Review, 6*, 486–494.

McKoon, G., & Ratcliff, R. (1994). Sentential context and on-line lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1239–1243.

Meyer, D. E., Irwin, D. E., Osman, A. M., & Kounois, J. (1988). The dynamics of cognition and action: Mental processes inferred from speed-accuracy decomposition. *Psychological Review, 95*, 183–237.

Murdock, B. B., Jr. (1971). A parallel-processing model for scanning. *Perception and Psychophysics, 10*, 289–291.

Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89*, 609–626.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research, 18*, 5–19.

Nicol, J. L., Fodor, J. D., & Swinney, D. (1994). Using cross-modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1229–1238.

Osterhout, L., & Swinney, D. (1993). On the temporal course of gap-filling during comprehension of verbal passives. *Journal of Psycholinguistic Research, 22*, 273–286.

Pickering, M., Barton, S., & Shillcock, R. C. (1994). Unbounded dependencies, island constraints, and processing complexity. In C. Clifton & L. Frazier (Eds.), *Perspectives on sentence processing* (pp. 199–224). Hillsdale, NJ: Lawrence Erlbaum Associates.

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes, 12*, 765–805.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59–108.

Ratcliff, R., Van Zandt, & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106*, 261–300.

Reed, A. V. (1973). Speed–accuracy trade-off in recognition memory. *Science, 181*, 574–576.

Reed, A. V. (1976). The time course of recognition in human memory. *Memory & Cognition, 4*, 16–30.

Ross, J.R. (1967) Constraints on Variables in Syntax. Doctoral dissertation, Massachusetts Institute of Technology.

Stowe, L. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes, 1*, 227–245.

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

Seidenberg, M. S., & McCelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review, 96*, 523–568.

Sternberg, S. (1966). High speed scanning in human memory. *Science, 153*, 652–654.

Sternberg, S. (1975). Memory-scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology, 27*, 1–32.

Swinney, D., Ford, M., Bresnan, J., & Frauenfelder, U. (1988). Coreference assignment during sentence processing. In M. Macken (Ed.), *Language structure and processing*. Stanford, CA: CSLI.

Theios, J. (1973). Reaction time measurement in the study of memory processes: Theory and data. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 7, pp. 44–85). New York: Academic Press.

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. New York: Cambridge University Press.

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

Treisman, M., & Doctor, E. (1987). Memory scanning: A comparison of the dynamic stack and exhaustive serial scan models with an extension of the latter. *Acta Psychologica, 64*, 39–92.

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

Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review, 103*, 761–772.

Wickelgren, W. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41*, 67–85.

Wickelgren, W. A., Corbett, A. T., & Dosher, B. A. (1980). Priming and retrieval from short- term memory: A speed-accuracy tradeoff analysis. *Journal of Verbal Learning and Verbal Behavior, 19*, 387–404.

Woods, W. (1973). An experimental parsing system for transition network grammars. In R. Rustin (Ed.), *Natural language processing*. Englewood Cliffs, NJ: Prentice-Hall.

Young, R. M., & Lewis, R. L. (1998). The soar cognitive architecture and human working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control.* New York: Cambridge University Press.