Journal of Memory and Language 65 (2011) 247–263

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Cue-dependent interference in comprehension

Julie A. Van Dyke [a,*], Brian McElree [b]

[a] Haskins Laboratories, 300 George Street, New Haven, CT 06511, United States
[b] New York University, 6 Washington Place, New York, NY 10003, United States

ARTICLE INFO

Article history:
Received 19 July 2010
revision received 22 May 2011
Available online 24 June 2011

Keywords:
Retrieval interference
Comprehension
Speed-accuracy tradeoff
Sentence processing

ABSTRACT

The role of interference as a primary determinant of forgetting in memory has long been accepted, however its role as a contributor to poor comprehension is just beginning to be understood. The current paper reports two studies, in which speed-accuracy tradeoff and eye-tracking methodologies were used with the same materials to provide converging evidence for the role of syntactic and semantic cues as mediators of both proactive (PI) and retroactive interference (RI) during comprehension. Consistent with previous work (e.g., Van Dyke & Lewis, 2003), we found that syntactic constraints at the retrieval site are among the cues that drive retrieval in comprehension, and that these constraints effectively limit interference from potential distractors with semantic/pragmatic properties in common with the target constituent. The data are discussed in terms of a cue-overload account, in which interference both arises from and is mediated through a direct-access retrieval mechanism that utilizes a linear, weighted cue-combinatoric scheme.

Limitations on memory storage and retrieval have long been recognized as important determinants of language performance (Miller & Chomsky, 1963). That successful language comprehension requires memory is apparent from the fact that common and seemingly simple expressions often contain non-adjacent, grammatically-dependent constituents. For example, the subject of a sentence can be separated from its matrix verb by relative clauses of different types and lengths, or the direct object for a sentence-final verb can be fronted to the beginning of the sentence. These types of expressions challenge comprehenders to establish syntactic and semantic relations that span multiple words, phrases, or even clauses, which necessitates accessing constituents that a comprehender is no longer actively processing. Hence, a full understanding of the memory processes that support language comprehension requires an investigation into the computational properties of the retrieval mechanism that enables these long-distance dependencies to be created.

Significant inroads in this respect have already been made. For example, while a variety of retrieval mechanisms are theoretically possible, including serial or parallel search mechanisms (Sternberg, 1966, 1975), the evidence overwhelmingly points to the use of an associative direct-access mechanism, in which the content of the cues available at retrieval enables immediate access to the target (i.e., content-addressable retrieval). For example, McElree, Foraker, and Dyer (2003, Exp. 1) contrasted the speed of resolving subject–verb dependencies with no material intervening, such as *The editor laughed*, to sentences in which 1 or 2 subject- or object relative clauses intervened between the subject and verb. They found that interpretation of the subject–verb dependency occurred at an exceptionally fast rate when the dependent elements were adjacent to one another, consistent with the idea that no retrieval was required in order to associate the subject with its verb. However, the speed of accessing a *distant* noun phrase (NP) to bind as subject to its verb was constant for each of the non-adjacent constructions, which contain varying numbers of intervening words, discourse items, and hierarchically embedded constituents. These results point to a direct-access retrieval mechanism, for

* Corresponding author. Address: Haskins Laboratories, 300 George Street, New Haven, CT 06511, United States. Fax: +1 203 865 8963.
E-mail address: jvandyke@haskins.yale.edu (J.A. Van Dyke).

0749-596X/$ - see front matter
doi:10.1016/j.jml.2011.05.002

which the distance between the retrieval site and the target would be irrelevant, unlike for a search mechanism, which would be affected by the amount of information to be examined during the search (McElree, 2006). Notably, these results are also consistent with results found in basic memory studies, where both neuroimaging and behavioral investigations of experimental variables that are diagnostic of the retrieval process, such as manipulations of item recency and size of the memory set all point to direct-access retrieval as the primary means for restoring passively held information into active memory (e.g., (Öztekin, McElree, Staresina, & Davachi, 2008; Öztekin, Davachi, & McElree, in press; McElree, 2006; McElree & Dosher, 1989).

The current work extends these investigations by examining the properties that cause a direct-access retrieval mechanism to fail. Research in the memory domain suggests two well-investigated sources for such failure: First, a cue-driven operation can fail to recover a sought-after memory if the cues used at retrieval do not sufficiently overlap with how the event was encoded into memory. Indeed, Tulving and Pearlstone (1966) pointed out that much of what we commonly view as memory loss—a memory no longer being *available*—is in fact more properly viewed as failures in *accessibility*. Subsequently, Tulving (1979) formulated the cue-dependant nature of accessibility into the *encoding specificity principle*, which states "[t]he probability of successful retrieval of the target item is a monotonically increasing function of information overlap between the information present at retrieval and the information stored in memory" (p. 408).

Even when the cues used in retrieval do sufficiently overlap with the contents of the sought-after memory, a cue-driven operation can fail to recover the 'correct' representation if the retrieval cues also match, even partially, the contents of other items in memory. When cues are also strongly associated with other representations in memory, it creates a condition of *cue-overload* that engenders *retrieval interference* (e.g., Nairne, 2002; Watkins & Watkins, 1975; Öztekin & McElree, 2007). An incorrect representation may be retrieved if available cues do not match the desired target as well as they do other items in memory.

Given the evidence that sentence comprehension relies upon a cue-driven, direct-access operation (e.g., McElree, 2000; McElree et al., 2003), it is natural to expect retrieval interference to be a key determinant of whether comprehension is successful (e.g., Van Dyke & Lewis, 2003; Lewis, Vasishth, & Van Dyke, 2006). Indeed, several recent studies have identified these effects. Gordon and colleagues (Gordon, Hendrick, & Johnson, 2001, 2004; Gordon, Hendrick, & Levine, 2002) and Fedorenko, Gibson, and Rohde (2006a) examined how referential properties of non-target nouns in memory affect the resolution of long distance dependencies in a sentence. In a dual task paradigm, participants were asked to remember three nouns for a subsequent recall test administered after reading a sentence and answering a comprehension test. They observed reduced comprehension accuracy and slower reading times for the sentence when the nouns in the memory set matched those in the sentence in their referential type, which was taken as support for the claim that similarity-based interference disrupts comprehension.

Van Dyke and McElree (2006) investigated whether the locus of these interference effects was in encoding/storage or in retrieval. Keeping the encoding context constant, we manipulated the retrieval cues at the point at which the dependency was resolved so that they either uniquely identify the target NP in the sentence or overlapped with the nouns in the memory set. As discussed above, the latter creates retrieval interference through cue-overload. For example, participants studied a list of three items such as *table–stove–truck* and then read the sentence *It was the boat that the guy who lived by the sea sailed* or the sentence *It was the boat that the guy who lived by the sea fixed.* In this example, all items in the memory list are fixable but none are typically "sail-able". Because the retrieval cues provided by the verb *sailed* are not strongly associated with potential competitors in memory, interference at retrieval should be substantially less than when *fixed* is used as the crucial verb. Consistent with a retrieval-based account, longer reading times were observed with verbs compatible (*fixed*) as compared to incompatible (*sailed*) with the nouns in memory, a difference that disappeared when the sentences were read without first memorizing the distractors in the memory set. Although there is some evidence suggesting that interference may affect encoding operations in comprehension (Desmet, De Baecke, Drieghe, Brysbaert & Vonk, 2006; Gordon et al., 2002), our results clearly implicate interference at the retrieval stage, which specifically arises from retrieval cues provided by the final verb. Moreover, this is consistent with the bulk of experimental evidence in the memory domain, which indicates that the locus of the interference effect is at retrieval, with little or no affect on memory encoding or storage (e.g., Dillon & Bittner, 1975; Gardiner, Craik, & Birstwistle, 1972; Tehan & Humphries, 1996).

These dual-task studies demonstrate that elements in memory engender interference if they share semantic–pragmatic properties with the cues used at retrieval. Other studies, using different methodologies, have demonstrated that interference is modulated by syntactic properties of the elements in memory (Van Dyke, 2007; Van Dyke & Lewis, 2003). For example, Van Dyke and Lewis (2003) examined reading times and comprehension accuracy for sentences such as (1a) and (1b), in which a relative clause intervenes between the subject NP *the lady* and the verb *moaned*.

(1a) *The pilot remembered that the lady who said that the seat was smelly yesterday afternoon moaned about a refund.*
(1b) *The pilot remembered that the lady who was sitting in the smelly seat yesterday afternoon moaned about a refund.*

Comprehension accuracy and self-paced reading times were detrimentally affected when the NP in the intervening relative clause (*the seat*) was syntactically a subject, as in (1a), as compared to when it was an object in a prepositional phrase, as in (1b). Van Dyke (2007) observed similar interference effects in online eye-tracking measures: At the critical verb (*moaned*), longer first pass and regression path times were observed for (1a) than (1b). This finding aligns with those in Van Dyke and McElree (2006) indicating that interference has its primary affect at retrieval. Crucially, the differences observed with contrasts such as (1a&b) indicate that grammatical information is among the set of cues that drive retrieval in comprehension, as a competitor can engender interference even when it does not match required semantic/pragmatic constraints (e.g., *moan* usually requires an animate subject). This suggests that parsing operations use grammatical cues from a verb such as *moaned* to locate appropriate attachment sites for integrating the verb phrase into the existing parse, including cues that identify previously occurring NPs that can serve as its subject.

Van Dyke (2007) also investigated the effect of semantic interference by contrasting sentences of the form in (1a&b) when they contained intervening nouns that either matched or mismatched the semantic/pragmatic constraints of the verb (*man* versus *seat*). Semantic effects were present even when the noun was in a syntactic position incompatible with it serving as the subject of the final verb, as in (1b). In eye-tracking measures for accurate trials, syntactic interference effects emerge as slowed first pass and regression path times at the critical verb (*moaned*), whereas semantic interference effects emerged as slowed regression path times for semantic interference at sentence end.

Goals of the present investigation

We aim to extend the studies of Van Dyke (2007) and Van Dyke and Lewis (2003) in several ways to address three fundamental issues concerning the nature of the cue-driven retrieval operation used in comprehension and how that operation may constrain comprehension.

The effect of interference on retrieval processes

The studies discussed above demonstrate that interference modulates the difficulty of sentence comprehension, and one study in particular, Van Dyke and McElree (2006), demonstrated that it does so when retrieval is required. However, beyond being consistent with the received view that interference produces cue-overload, which affects the *probability* (or *availability* in Tulving's terms) of accessing target items, extant research has not addressed how cue diagnosticity (i.e., the ability to uniquely identify a target item) affects the *accessibility* of information. Namely, slowed reading time at the retrieval site could be consistent either with an efficient direct-access retrieval mechanism that has failed to identify the correct target (the target is *unavailable*) or with a slowed retrieval mechanism, perhaps operating in qualitatively different manner because the target is *inaccessible* to cue-based retrieval. To distinguish these possibilities, we utilize the response-signal speed-accuracy tradeoff (SAT) procedure, which provides conjoint measures of speed and accuracy (further discussion of the advantages of this method are available in Dosher (1979), McElree (2006), Wickelgren (1977), Reed (1973, 1976). While this method has been used previously to investigate retrieval mechanisms in a variety of sentence constructions (e.g., Martin & McElree, 2008; McElree et al., 2003; Foraker & McElree, 2007), it has thus far not been used to investigate interference effects.

Different types of interference

Classically, two types of interference have been extensively investigated in basic memory research: *Proactive interference* (PI), where recall or recognition of a target is impaired by items processed *before* the target, and *retroactive interference* (RI), where target recall or recognition is impaired by items processed *after* a target. Although it is possible that the two forms of interference may affect memory operations in qualitatively different ways, several lines of research converge in suggesting that both PI and RI engender cue-overload at retrieval (Crowder, 1976). Both types of interference have been found to be operative in sentence comprehension. However, with the exception of a recent study of verb phrase ellipsis (Martin & McElree, 2009; see Discussion), they have been investigated under rather different experimental conditions, which preclude straightforward assessments of their relative impact on comprehension and whether they affect the same underlying processes. Specifically, studies such as Gordon et al. (2002), Fedorenko, Gibson, and Rohde (2006b), and Van Dyke and McElree (2006) induced PI by requiring participants to study and remember items presented in a list before reading a critical sentence—hence the interfering items were not actually within the sentence itself. In contrast, studies such as Van Dyke (2007) and Van Dyke and Lewis (2003) induced RI with sentence-internal item interpolated between a nonadjacent dependency. We investigated PI and RI effects on comprehension under comparable conditions where potentially interfering items are all sentence internal, occurring either before or between dependant constituents.

Contrasting PI and RI effects can also provide additional evidence concerning the nature of the retrieval operation used in comprehension, specifically whether a search or direct-access operation is used. The key prediction of a search operation is that retrieval time will increase as more competitors are added to the search path (McElree & Dosher, 1989). In contrast, additional competitors need not affect retrieval speed in a content-addressable system with a direct-access operation. With the exception of a recent investigation of ellipsis (Martin & McElree, 2009), extant timecourse studies of nonadjacent dependencies have used RI configurations, where different numbers of competitors are interpolated between the retrieval site and the target item (McElree, 2000; McElree et al., 2003). That additional competitors reduced accuracy but did not increase retrieval time in these studies provides evidence against a large class of search operations. However, RI configurations cannot test for the presence of a forward search, where comprehenders start at the beginning of a sentence or discourse and search forward for a constituent to resolve the dependency. Such a search may be plausible in situations where integrating a verb with its previously occurring subject are difficult, as evidenced by eye-movement studies in which regressions to the beginning of a sentence have been observed (e.g., Frazier & Rayner, 1982). In order to examine both forward and backward search operations, the two experiments presented here employ a design which examines both PI and RI configurations. If a forward search is used, then processing time should be slower under PI than RI configurations. A backward search predicts the opposite pattern, with processing being slower under RI than PI configurations. In contrast, no difference in processing time is predicted for a direct-access operation, as cues at retrieval make direct contact with relevant representations by virtue of their content.

Cue combinatorics

A complete understanding of the constraints that memory operations place on comprehension requires a detailed specification of what types of cues drive retrieval and how those cues combine at the retrieval site. Studies such as Van Dyke (2007) and Van Dyke and Lewis (2003) suggest that available morphosyntactic and semantic–pragmatic information guide retrieval in comprehension. An open question is how different types of information are combined. In many perceptual and cognitive domains, cues are often assumed to combine in a linear, weighted fashion (e.g., Trommershäuser, Landy, & Körding, in preparation). However, nonlinear schemes might be more efficacious in domains where there are asymmetric dependencies in the processing of different types of information. The relationship between syntactic and semantic processing may be an example of when a nonlinear scheme of this sort is warranted. Although the question of how different types of cues are assembled to create a retrieval probe during language comprehension has previously not be investigated, there is a large literature surrounding the issue of whether syntactic information takes precedence over semantic information in compositional operations (cf. Pylkkänen and McElree (2006) for a review). For example, "syntax-first" approaches propose that semantic operations are, to a large degree, conditioned on the syntactic structures generated by syntactic processes (e.g. Ferreira & Clifton, 1986; Frazier & Rayner, 1982; Pickering & Traxler, 1998; Clifton et al., 2003). However, others have shown that different kinds of information, such as lexical information (e.g., Swinney, Zurif, & Nicol, 1989; Gibson, 2006), semantic/plausibility information (e.g., Kim & Osterhout, 2005; Trueswell, Tanenhaus, & Garnsey, 1994), or local discourse context (Altmann & Kamide, 1999; Grodner, Gibson, & Watson, 2005), can guide interpretation processes, independently of syntactic information. Although this work has focused on how information that does not need to be retrieved is weighted when making *local* attachment decisions, it points to the possibility that the language processing mechanism will give preference to certain types of information in certain contexts (i.e., non-obligatory adjunct attachments appear to be more influenced by non-syntactic information than do argument attachments, cf. Frazier and Clifton (1996) for extensive discussion). Although much additional research will be required to fully understand how information is combined during retrieval, the current set of studies is meant to provide an initial foundation, asking whether syntactic cues and semantic/plausibility cues are relied onto the same extent in differing syntactic contexts. Using a similar experimental design and logic to the studies in Van Dyke (2007) and Van Dyke and Lewis (2003), we investigated this issue by examining the effects of a pragmatically appropriate competitor with syntactically appropriate (Experiment 1) and inappropriate properties (Experiment 2). Because interference is engendered by elements in memory that match the cues used at retrieval, we used the amount of interference generated by a particular element as an indicator of what cues were used in retrieval and how that cue was modulated by other cues.

In addition to applying the SAT methodology to investigate these issues, were also conducted eye-tracking versions of both experiments in order to provide converging evidence. This permits us to address a disadvantage of the SAT procedure, namely, that it requires an overt response; in this application, a judgment of acceptability. Eye-tracking measures allow us to assess how the differences found in SAT are expressed in naturalistic reading, as eye-tracking does not impose unnatural memorial or response demands on participants. Reading time measures also afford a perspective from which to compare the current research to previous work, that latter having only used eye-tracking and self-paced reading as online measures.

Experiment 1: influence of appropriate syntactic cues

Experiment 1 examines RI and PI when the interfering item matches the syntactic cues from the verb (i.e., they are grammatical subjects). In this respect, this experiment extends Van Dyke and Lewis (2003), which observed syntactic interference effects, but only in RI contexts, and Van Dyke (2007) which observed semantic and syntactic interference effects, again only in RI contexts. We manipulated the semantic properties of distractor elements in a manner similar to those in previous studies, but using different syntactic constructions. The examples in (2) illustrate the conditions used to investigate semantic interference in an RI context. In both (2a) and (2b), comprehenders must retrieve a subject for the matrix verb *compromised* when it is encountered at the sentence final position. We expected that intervening NPs would engender RI. Because it is a plausible subject for the verb and will have been encoded in the sentence as a grammatical subject, the NP *the judge* was expected to produce RI in attempts to retrieve the appropriate subject (*the attorney*), despite being in an inappropriate (hierarchical) position in the sentence. The NPs *the motion* in (2a) and *the witness* in (2b) should also produce retroactive interference (RI). However, if semantic–pragmatic properties of the matrix verb are used as retrieval cues, the semantic properties of the NPs should modulate the RI effects. Specifically, RI should be reduced with the NP *the motion* in (2a) because, unlike *the witness* in (2b), it lacks the semantic properties (e.g., animacy) to serve as the subject for the verb *comprise*.

(2a) *The attorney who the judge realized had declared that the motion was inappropriate compromised.*
(2b) *The attorney who the judge realized had declared that the witness was inappropriate compromised.*

Van Dyke (2007) found that mismatching semantic properties of the type in (2a) did indeed decrease interference at retrieval. We attempted to replicate this finding and to investigate whether the same effect would be found in a PI context, where the distractors were processed before the target subject (*attorney*), as in (3).

(3a) *The judge who had declared that the witness was inappropriate realized that the attorney for the case compromised.*
(3b) *The judge who had declared that the motion was inappropriate realized that the attorney for the case compromised.*

Based on SAT studies of distance effects in resolving long-distance dependencies (e.g., McElree, 2000; McElree et al., 2003), we expected that RI conditions in (2) would be associated with lower accuracies than the PI conditions in (3), as the subject–NP will have been less recently processed. Of particular interest was whether the addition of a semantically-matching distractor engenders interference in both contexts, and of comparable magnitudes. Interference effects may be less in PI than RI conditions because, at the retrieval site, the subject–NP is more recent and the competitor NPs are less recent than in RI conditions. Differential affects of the recency of both the target and competitors are consistent with direct-access retrieval operation in which the quality (strength, fragility, or related notions) of the respective memory representations determines the probability of successful retrieval.

Experiment 1A: SAT method

Participants

Twenty participants from the New York University community were recruited for the study. All were native American-English speakers with no history of reading difficulty. They were paid a total of $70 at the completion of five experimental sessions ($10/h). One participant was excluded from the analysis because he did not complete all five sessions.

Materials

Thirty-six item sets were constructed following the paradigm demonstrated in Table 1. The three manipulations created a 2 (interference type – PI versus RI) × 2 (semantic properties of the distractor – Matched versus Unmatched) × 2 (acceptability – Yes versus No) design resulting in 288 experimental sentences. In order to minimize the effects of repeating the experimental structures, presentation was spaced over four 1.5 h sessions on subsequent days. In each of these sessions, participants saw a list that contained two experimental conditions from each item set (72 items – 1 acceptable and 1 unacceptable), together with 280 filler items from unrelated experiments. The experimental conditions in each session were chosen randomly from each set, with the restriction that all eight conditions were presented over the four sessions. Each session contained an equal number of acceptable and unacceptable sentences. Acceptability was manipulated be exchanging the last word of the sentence, which was the critical word initiating retrieval of the distal subject (see Table 1). Examples of the filler items, with acceptable and unacceptable alterations in italics are as follows: The architect admired the blueprint, but *the funding did not materialize/the entrepreneur did not melt.* The center spotlight left the prima ballerina. *She breathed./It dimmed.*

Procedure

The materials were randomized within a session and presented on a personal computer running E-prime (Schneider, Eshman, & Zuccolotto, 2002), which recorded button press responses and latencies. For each trial, participants read a sentence presented in a phrase-by-phrase RSVP format. The phrases are illustrated by the slash marks in Table 1. Three hundred milliseconds before the appearance of the last word, a series of 17 tones began (100 ms in duration, occurring every 350 ms). These continued during the presentation of the critical word and afterwards, for a total time span of 5950 ms. Participants began pressing both the ''YES'' and ''NO'' keys as soon as the tones began, and continued pressing in the rhythm of the tones until they ended. While participants are pressing both keys, the last word appeared, and they indicated whether the word constitutes an acceptable completion of the sentence by continuing to press only the key that indicates their answer (i.e., they would stop pressing the ''NO'' key if they thought the sentence was acceptable, or the ''YES'' key if they thought the sentence was unacceptable).

Participants were trained in this procedure for 45 min in a separate session 1 day prior to the experimental sessions in order to familiarize themselves with the task.

Table 1
Example materials from Experiment 1 (high syntactic interference conditions). Acceptable and unacceptable conditions are represented by substituting the appropriate verb as the final word of the sentence. Slashes represent phrasal chunks displayed during SAT presentation.

Proactive interference condition	
1. Semantic mismatch	The judge / who had declared that / the motion / was inappropriate / realized that the attorney / in the case / [compromised/entwined].
2. Semantic match	The judge / who had declared that / the witness / was inappropriate / realized that the attorney / in the case / [compromised/entwined].
Retroactive interference conditions	
3. Semantic mismatch	The attorney / who the judge realized / had declared that / the motion / was inappropriate / [compromised/entwined].
4. Semantic match	The attorney / who the judge realized / had declared that / the witness / was inappropriate / [compromised/entwined].

During this training, participants responded to acceptable and unacceptable sentences that were similar to the filler items that occurred in the actual experiment. They were trained to respond in the rhythm of the tones and received error messages when they responded either too quickly, too slowly, or out of sync. They were also instructed that they could change their response if their assessment of the sentence changed, so that if they initially thought a sentence was acceptable, and then later realized it was unacceptable they could switch from the "YES" to the "NO" key as long as the response tones were still sounding. This provided a record of accuracy at each response time over the 5650 ms period, enabling us to fully measure how the interpretation of the sentence unfolded over time.

Data analysis

All analyses were performed on the individual participants' data to reduce variance and avoid averaging artifacts in the shape of the functions. Consistent patterns across participants were summarized by analyses of the average data. Accuracy at each response delay was computed using a standard d' measure ($d' = z(\text{hits}) - z(\text{false alarms})$) in order to correct for response bias. A "hit" is an acceptable response to an acceptable sentence and a "false alarm" was an acceptable response to an unacceptable sentence. The scaling was done using the yoked acceptable and unacceptable continuations for each experimental condition (indicated by the bracketed words in Table 1).

Potential differences in asymptote, rate, and intercept were evaluated by fitting the d' accuracies at each response delay (t) with an exponential approach to a limit:

$$d'(t) = \lambda(1 - e^{-\beta(t-\delta)}), \quad \text{for } t > \delta \text{ else } 0. \tag{6}$$

In this equation, accuracy is a function of three parameters, corresponding to the three phases of the SAT curve: λ serves to estimate the asymptotic level of performance, δ estimates the intercept or discrete point in time where performance begins to rise from chance, and β indexes the rate at which accuracy grows from chance to asymptote. Hierarchically nested models were fit to the data, ranging from a null model (in which the four experimental conditions in the interference type × distractor type manipulations were fit to a single asymptote, rate, and intercept) to a fully saturated 12-parameter model, in which each of the four conditions was fit with a unique set of parameters. The equation in (6) was fit to the data with an iterative hill-climbing algorithm (Reed, 1976), similar to STEPIT (Chandler, 1969), which minimizes 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 subjects.

In addition, we performed inferential tests of significance (repeated-measures 2 (interference-type) × 2 (intervener type) ANOVA) on the individual participants' observed data and on the fitted parameter estimates for each of the candidate models described in the Results section.

SAT results

Fig. 1 shows the average (over participants) d' data as a function of processing time for PI and RI constructions with the two distractor types, together with a plot of the best fitting 4λ–1β–1δ model. Table 2 presents the parameter values for both the average data and for individual subjects. Asymptotic accuracy was affected by both the interference manipulation and by the semantic properties of the intervening noun phrase. This conclusion was supported by a repeated-measures ANOVA on the mean of the last four d' values, which provided an empirical estimate of asymptotic accuracy. The PI-Unmatched condition yielded the highest asymptote (3.08 d' units), followed by the PI-Matched condition (2.92 d' units), then the RI-Unmatched condition (2.85 d' units) and the RI-Matched condition (2.58 d' units). The main effect of interference type was significant, $F_1(1, 18) = 10.17$, $p < .006$, $\eta_p^2 = .361$; $F_2(1, 33) = 5.95$, $p < .03$, $\eta_p^2 = .153$. The main effect of the semantic manipulation was also significant, $F_1(1, 18) = 7.10$, $p < .02$, $\eta_p^2 = .283$; $F_2(1, 33) = 4.43$, $p < .05$, $\eta_p^2 = .118$. The interaction was not significant. Hierarchical modeling of the data began with a null 1λ–1b–1d model, which assigned a common asymptote, rate, and intercept to all four conditions. This fit produced an adjusted-R^2 of .976 for the average data, ranging from .593 to .947 across the 19 participants. We next fit a 2λ–1β–1δ model to the data, where one common asymptote was assigned to the PI conditions and one to the RI conditions. The model produced an adjusted-R^2 of .986 for the fit of the average data, ranging from .685 to .951 for individual participants' data. All but 5 participants showed an increase in adjusted-R^2 for this model as compared to the 1λ–1β–1δ model, (average adjusted-R^2 increase = .023; minimum = −.004; maximum = .092). A paired t-test indicated that the parameter estimates for the two asymptotes were significantly different, $t_1(18) = 3.40$, $p < .005$. The average of the parameter

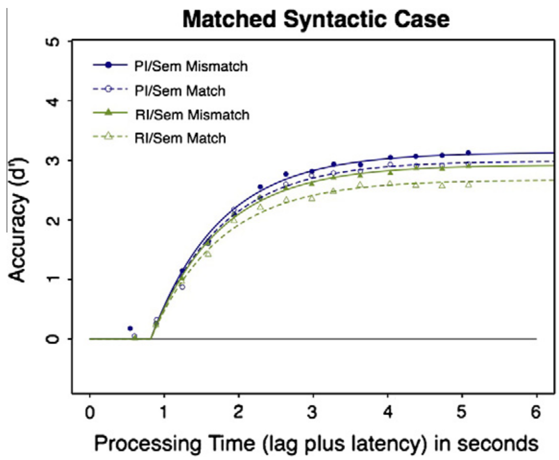

Fig. 1. SAT results from Experiment 1.

Table 2
Experiment 1A: parameter estimates for the average data and individual participants for the 4λ–1β–1δ (PI = proactive interference; RI = retroactive interference; UnM = Unmatched; M = Matched).

	Adjusted R^2	Asymptotes				Rate (ms)	Intercept
		PI–UnM	PI–M	RI–UnM	RI–M		
Average	.995	3.13	2.99	2.92	2.67	1.07	.818
S1	.944	3.13	2.74	1.91	2.27	2.81	1.39
S2	.913	3.05	3.44	3.41	2.65	1.83	.759
S3	.961	2.90	3.27	3.36	2.97	1.82	.704
S4	.946	3.80	3.04	3.13	2.79	1.05	.840
S5	.937	3.56	3.60	3.90	3.29	1.03	.902
S6	.897	2.10	2.12	1.12	1.23	1.25	.791
S7	.950	3.62	3.26	3.56	3.06	.811	.885
S8	.964	3.90	3.73	3.10	3.31	2.29	1.04
S9	.890	1.95	1.81	2.28	1.81	2.27	.937
S10	.958	3.57	2.10	3.70	2.57	.640	1.33
S11	.926	3.78	3.24	3.56	2.87	1.05	.621
S12	.951	2.91	2.67	2.32	2.32	1.26	.740
S13	.948	3.46	3.42	3.32	3.87	1.24	.770
S14	.934	2.88	2.76	2.33	1.43	1.46	1.01
S15	.979	3.09	3.15	2.59	2.95	2.03	.895
S16	.960	3.74	3.75	3.81	3.18	.944	1.10
S17	.932	3.59	3.58	3.30	3.14	.586	1.63
S18	.943	2.35	2.30	2.48	2.32	1.37	.816
S19	.950	2.85	3.03	2.99	3.09	1.39	.761

estimates for the PI conditions was 3.15, and the average for the RI conditions was 2.87.

By comparison, a 4λ–1β–1δ model, which assigned separate asymptotes to the four conditions produced adusted-R^2 values of .995 for the average data, ranging from .890 to .979 for individuals. Parameter estimates for the average data were 3.13 for the PI-Unmatched condition, 2.99 for the PI-Matched condition, 2.92 for the RI-Unmatched condition and 2.67 for the RI-Matched condition. All participants but 1 showed an increase in the adjusted-R^2 for the 4λ–1β–1δ model over the 1λ–1β–1δ model (average increase = .07, ranging from −.001 to .303). This model was also a better fit than the 2λ–1β–1δ model for all but one participant's data (average increase = .045, minimum = −.001, maximum = .212). As suggested by the d' analysis, an ANOVA on the asymptotic estimates for the 4λ–1β–1δ model indicated significant effects of interference type, $F_1(1, 18) = 9.00$, $p < .008$, $\eta_p^2 = .333$ and the semantic manipulation, $F_1(1, 18) = 6.80$, $p < .02$, $\eta_p^2 = .274$. The interaction was not significant, $F < 1$. The average λ estimates across subjects were 3.17 for the PI-Unmatched condition, 3.00 for the PI-Matched condition, 2.96 for the RI-Unmatched condition, and 2.69 for the RI-Matched condition.

Subsequent fits aimed at evaluating the effect of the manipulations on processing speed. The data do not suggest any differences in intercept (cf. Fig 2) and models which expressed dynamics differences in intercept did not fit the average data, nor the majority of participants, better than those expressing differences in the rate parameter. Consequently, only models varying SAT rate are described here. Three models were tested. Model A was a 4λ–2β–1δ model, assigning a common rate to the PI conditions, and a second rate to the RI conditions. Model B was also a 4λ–2β–1δ model, assigning a common rate to the Unmatched conditions and a second rate to the Matched conditions. Model C assigned a separate rate to each experimental condition (4λ–4β–1δ). None of these models produced significant improvement in adjusted-R^2 for the average data: Models A and C had adjusted-R^2 of .996, and Model B had an adjusted-R^2 of .995, matching that of the 4λ–1β–1δ model. ANOVA tests on the parameter estimates in these models were all insignificant. Consequently, we concluded that there were no dynamics differences among the conditions in this experiment.

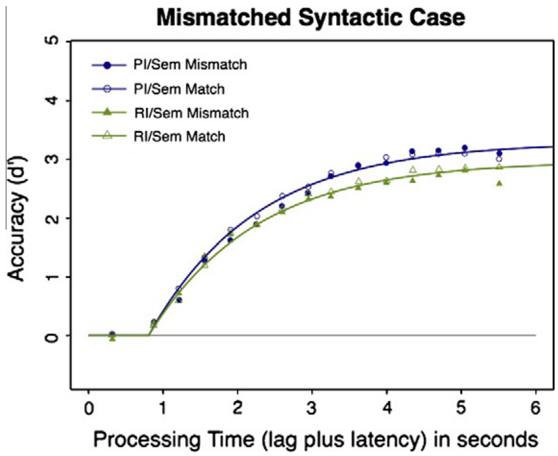

Fig. 2. SAT results from Experiment 2.

Experiment 1B: eye-tracking method

Participants

Forty-eight students from NYU participated in the experiment, receiving partial course credit. All were native speakers of American English with normal or corrected to normal vision. No participants in the eye-tracking study had previously participated in the SAT version of the study.

Materials

Sentences were as shown in Table 1, except that a prepositional phrase spillover region was added after the critical verb in each sentence (i.e., The judge who had declared that the motion was inappropriate realized that the attorney in the case compromised *during the negotiations*). Unacceptable versions were not tested. These items were interleaved with filler items comprised of sentences from several unrelated experiments.

Procedure

We monitored eye-movements using a SensoMotoric Instruments Eyelink I head-mounted eye-tracker sampling at 500 Hz. Screen resolution was set at 1600 × 1200 pixels. Sentences were presented in a fixed-width font, with each letter 18 pixels wide and 33 pixels high. No more than 80 characters were presented on one line of text. Stimulus presentation was controlled using the Eyetrack 0.7.8 software, available at http://www.psych.umass.edu/eyelab/. The stimuli appeared on a CRT monitor approximately 71 cm from the participant's eyes, where 1° of visual angle corresponded to 2.7 characters. A chin rest was used to reduce head movement. Sentences were presented on two lines; line breaks for the PI conditions occurred just prior to the verb *realized* in Table 1 example, and after *realized* in the RI conditions. Cloze comprehension questions querying the long distance dependency were presented, but only on 50% of the items in order to shorten experiment time. As an example, the cloze question for the items in Table 1 was "The _____ compromised." Participants were required to click on one of two nouns occurring below the cloze question to indicate their answer. The incorrect response was randomly chosen from either of the remaining two nouns in the sentence not occurring in the spillover region (e.g., *judge* or *witness* in Table 1). Overall accuracy was 81.5%. Further statistical analyses are not reported due to the low number of trials per subject. Data were analyzed using software available at the website mentioned above.

Analysis

We report results for two regions of interest. We refer to the region containing the main verb as the critical region; this is the region where the distant subject must be retrieved in order to complete the long distance dependency. The spillover region refers to the region directly following the critical region (final region). Four measures are presented: first pass reading time, corresponding to the sum of all fixations in a region starting with the first fixation until the reader's gaze exits the region either to the left or the right; regression path time, corresponding to the sum of all fixations from first entering a region until moving past it to the right (including re-fixations of regions to the left); total reading time, corresponding to the sum of both the first-pass fixations and all subsequent fixations in a region after the eyes have exited that region and returned, including rereading time originating from regions before or after the current region; and first-pass regressions out, corresponding to the proportion of trials on which the reader made at least one regression out of the region after first entering it. Table 3 presents means calculated after applying a winsorizing procedure in which extreme values were replaced by three times the interquartile range across all subjects for each condition. This procedure tends to distort the data less than other procedures based on standard deviation scores (Sturt, Pickering, & Crocker, 1999). Prior to this procedure, the data were trimmed by excluding all fixations less than 100 ms and greater than 1200 ms. Maximum cutoffs were 1600 ms for first pass reading, 10,000 for regression path, and 3000 ms for total reading time. These procedures affected less than 4% of the data. In response to descriptive statistics revealing a high degree of skewness in all dependent measures (first pass: skewness > 1.38; regression path: skewness > 2.55; total time: skewness > 1.68), we conducted statistical analyses on the log transform of reading time variables. Assuming that long reading times reflect real difficulties and not measurement error, utilizing this transformation has the advantage of creating a more normal distribution while not throwing away meaningful data. Analyses on the untransformed data produced a similar pattern of results.

Table 3
Experiment 1B winsorized reading times (in milliseconds) and proportion of regressive eye-movements for each region for each dependent measure. Standard error in parentheses, participants as the random factor.

	Critical region	Spillover region
First pass		
PI/Sem. mismatch	265 (8)	453 (17)
PI/Sem. match	270 (8)	489 (20)
RI/Sem. mismatch	306 (11)	423 (17)
RI/Sem. match	304 (9)	399 (17)
Regression path		
PI/Sem. mismatch	297 (12)	1400 (97)
PI/Sem. match	297 (11)	1471 (93)
RI/Sem. mismatch	412 (21)	2317 (168)
RI/Sem. match	454 (26)	2332 (173)
Total time		
PI/Sem. mismatch	421 (19)	627 (25)
PI/Sem. match	441 (20)	636 (27)
RI/Sem. mismatch	611 (29)	720 (30)
RI/Sem. match	692 (33)	738 (34)
Proportion of regressions		
PI/Sem. mismatch	.05 (.02)	.48 (.04)
PI/Sem. match	.05 (.02)	.56 (.03)
RI/Sem. mismatch	.17 (.03)	.60 (.03)
RI/Sem. match	.20 (.03)	.63 (.03)

Eye-tracking results

First pass

For the critical region, the effect of interference type was significant, $F_1(1, 47) = 7.73$, $p < .008$, $\eta_p^2 = .14$; $F_2(1, 35) = 10.73$, $p < .003$, $\eta_p^2 = .23$. The PI conditions were read more quickly than the RI conditions (268 ms versus 304 ms). In the spillover region, the effect was also significant, $F_1(1, 47) < 13.85$, $p < .001$, $\eta_p^2 = .23$; $F_2(1, 35) = 22.53$, $p < .0001$, $\eta_p^2 = .39$, but the direction was reversed. The PI conditions were read more slowly than the RI conditions (471 ms versus 411 ms). This is due to an interaction, in which the semantic manipulation resulted in faster times for the RI conditions, but slower times for the PI conditions. This interaction was significant in the analysis by items, $F_2(1, 35) = 4.35$, $p < .05$, $\eta_p^2 = .11$ and marginal in the analysis by subjects, $F_1(1, 47) = 3.16$, $p < .09$, $\eta_p^2 = .06$. The effect of the semantic manipulation was not significant for this measure in either region, Fs < 1.05.

First pass regressions

Interference type affected the proportion of first pass regressions in the critical region, $F_1(1, 47) = 35.48$, $p < .0001$, $\eta_p^2 = .43$; $F_2(1, 35) = 37.66$, $p < .0001$, $\eta_p^2 = .52$. PI conditions invoked fewer regressions out of this region than RI conditions (5% versus 18%). A similar effect, but smaller in magnitude, was observed in the spillover region (53% versus 62%), $F_1(1, 47) = 7.70$, $p < .01$, $\eta_p^2 = .14$; $F_2(1, 35) = 4.45$, $p < .05$, $\eta_p^2 = .11$. Neither the effect of the semantic manipulation nor the interaction were significant in either region.

Regression path

The interference manipulation affected regression path reading times in the critical region, producing shorter times for the PI condition than the RI condition (297 versus 433). The effect was significant both by participants, $F_1(1, 47) = 28.03$, $p < .0001$, $\eta_p^2 = .37$, and by items, $F_2(1, 35) = 35.62$, $p < .0001$, $\eta_p^2 = .50$. The effect of the semantic manipulation did not reach significance, $F_{1\&2} < 1.29$, nor did the interaction, $F_{1\&2} < 1$. The interference effect was also present in the spillover region. Again, PI conditions had shorter regression path times than RI conditions (1436 versus 2325) and the effect was significant both by participants, $F_1(1, 47) = 37.57$, $p < .0001$, $\eta_p^2 = .44$ and by items $F_2(1, 35) = 23.24$, $p < .0001$, $\eta_p^2 = .40$. No other effect was suggested in either analysis for this region.

Total time

The interference manipulation was also significant in the critical region, again producing shorter total reading times for the PI conditions than the RI conditions (431 versus 651). The effect was present in both the analysis by participants, $F_1(1, 47) = 55.72$, $p < .0001$, $\eta_p^2 = .54$, and by items, $F_2(1, 35) = 60.15$, $p < .0001$, $\eta_p^2 = .63$. The effect of distractor type was also significant, with semantically unmatched distractors producing shorter total reading times than when the distractor matched the semantic properties of the target noun (516 versus 566). The effect was significant both by participants, $F_1(1, 47) = 4.61$, $p < .04$, $\eta_p^2 = .09$ and by items, $F_2(1, 35) = 4.62$, $p < .04$, $\eta^2 = .12$. The interaction was not significant, Fs < 1.27.

In the spillover region, the interference manipulation was again significant, $F_1(1, 47) = 11.79$, $p < .002$, $\eta_p^2 = .20$; $F_2(1, 35) = 9.15$, $p < .005$, $\eta_p^2 = .21$. With shorter total reading times for the PI conditions than the RI conditions (636 versus 738). The was no suggestion of an effect of the semantic manipulation nor the interaction in this region, Fs < 1.

Experiment 1A&B: discussion

Experiments 1a and 1b are in complete agreement. Both indicate that retroactive interference is more detrimental than proactive interference during sentence processing, yielding lower asymptotic performance in the SAT task, more regressive eye movements and slower reading times in all reading measures. They are also consistent in showing a main effect of additional semantic/pragmatically matching distractors, which is numerically larger in asymptotic accuracy and total reading times in the RI constructions than in the PI constructions (although there was no interaction). The presence of the effect in the SAT asymptote—and not in rate or intercept—suggests the difficulty is associated with a reduced probability of accessing the target (with no effect on the speed of doing so). This could arise either because the retrieval cues provided by the verb are inadequate to distinguish between semantically similar constituents or because subsequent similar items have a detrimental effect on the quality of the target's representation itself, perhaps via a process of feature overwriting (Nairne, 1990; Oberauer & Kliegl, 2006.) The former assigns the locus of difficulty to retrieval processes in particular, while the latter implicates encoding. We discuss these possibilities in more detail in "General Discussion".

The fact that the semantic effect was observed in total reading times (only) in the *critical* region, is consistent with the suggestion that semantic similarity causes the target to be unavailable for retrieval. Van Dyke (2007) argued that a constituent that matches both the syntactic and the semantic properties of the verb can have a "blocking" effect, so that the correct target cannot be retrieved, even after it becomes apparent that the incorrect interpretation was constructed.[1] This may lead to increased re-reading in the target region, as readers struggle to make sense of their incorrect, yet irreparable interpretation.

One unexpected aspect of the eye-tracking data was the significant cross-over interaction in the first pass reading times in the final (spillover) region. Here, the semantic manipulation produced the expected longer reading times in the PI condition, consistent with the idea that increasing the number of semantically appropriate distractors would make processing difficult. This effect was expected at the critical region, however, and it is possible that the increase seen here reflects a true spillover from that region. Some support for this idea comes from the small (and insignificant) slowdown for the additional distractor condition observed in the critical region for the PI conditions (265 ms versus 270 ms). The apparent *speedup* associated with the semantic manipulation in the RI conditions is more diffi-

[1] Van Dyke (2007) further suggested that this "blocking" phenomenon may underlie what has been termed "Good-Enough" representations (e.g., Ferreira, Ferraro, & Bailey, 2002; Christianson, K., Hollingworth, A., Halliwell, J., & Ferreira, F. (2001)). There, the claim is that initial interpretations constructed in ambiguous contexts may be incompletely dismantled, allowing the original (incorrect) thematic assignments to linger even after disambiguation. Van Dyke suggested that reanalysis may be incomplete when there are insufficient retrieval cues for accessing the alternative representations (see also Van Dyke & Lewis, 2003), leaving the reader "stuck" with only the incorrect interpretation.

cult to understand. Such a speedup may be related to the higher proportion of regressions out of the previous region in the RI conditions as compared to the PI conditions, during which participants may have completed retrievals and/or reanalyses that would enable them to judge whether they understand the sentence or else give up because it is too difficult. The former might occur if participants incorrectly took the distractor as the subject of the critical verb and did not realize their error, while the latter might occur if readers have attempted to repair an incorrect interpretation (perhaps even rereading the critical region), but are unable to revise their interpretation because access to the correct target is blocked. In either case the reduced time in the spillover region seems to reflect a desire to move quickly to the next trial, at least for some proportion of participants. Further replication will be necessary to distinguish these possibilities and clarify whether this reading time decrease is a reliable effect.

The SAT data provide additional information regarding the type of retrieval mechanism employed during sentence processing. As no differences in retrieval speed were observed, this suggests that grammatical dependents are accessed via a direct access mechanism. This is contrary to the predictions of a search mechanism, whether it be a forward search from the beginning of the sentence or a backward search for the appearance of the critical verb. In both cases a search mechanism would be affected by the *number* of semantic distractors and we observed no difference in speed between the unmatched semantic conditions, which have only two nouns that fit the semantic requirements of the critical verb, and the matched semantic conditions, where there are three semantically appropriate nouns. This conclusion is taken up in more detail in the General Discussion.

Experiment 2: semantic interference with inappropriate syntactic cues

In comprehension, the cues that drive retrieval must include morpho-syntactic constraints on the type of constituent that is needed to resolve the dependency (see Martin & McElree, 2008, 2009). Experiment 2 investigates whether the findings from Experiment 1 hold in contexts where the distracting NPs do not fit the syntactic cues from the critical verb. The sentences used in Experiment 1 were changed so that the distracting subject–NP that varied in its semantic–pragmatic fit now occurred in an object position, creating what we refer to as a low syntactic interference context (Van Dyke & Lewis, 2003). We maintain the PI/RI manipulation, with (4a&b) instantiating RI and (5a&b) containing PI:

(4a) *The attorney who the judge realized had rejected the witness in the case compromised.*
(4b) *The attorney who the judge realized had rejected the motion in the case compromised.*
(5a) *The judge who had rejected the witness realized that the attorney in the case compromised.*
(5b) *The judge who had rejected the motion realized that the attorney in the case compromised.*

Again, (a) versions have high semantic–pragmatic interference, as there are three NPs that match the semantic cues for an appropriate subject of *compromised,* whereas (b) versions only have two such NPs. At issue was whether we would observe the same effect of the manipulation of semantic–pragmatic properties of this NP when it has syntactic properties that mismatch the type of constituent needed at the retrieval site. If semantic–pragmatic and syntactic cues were combined in a weighted linear fashion, then we would expect to see an effect of the semantic–pragmatic manipulation, although perhaps not as large as what was observed in Experiment 1. However, if syntactic cues serve a gating function, then the expectation was that manipulating the semantic–pragmatic properties of a syntactically inappropriate NP would not engender different amounts of retrieval interference. This is because unlike in Experiment 1, distractors can be eliminated from consideration based on their mismatching syntactic cues, despite semantic properties that match the semantic cues of the critical verb. Simply put, if cues about the syntactic context where the appropriate noun should be found are more important, then we may see no effect of the semantic manipulation.

As with Experiment 1, we utilized both SAT and eye-tracking methodologies, with identical methods of presentation and analysis. For the SAT study, 20 participants from the New York University community were recruited. All were native American-English speakers with no history of reading difficulty. They were paid a total of $70 at the completion of five experimental sessions ($10/h). For the eye-tracking study, an additional 48 students from NYU participated in the experiment, receiving partial course credit. All were native speakers of American English with normal or corrected to normal vision. Accuracy on comprehension questions was 79% overall.

Table 4 presents the experimental materials used for this experiment. As in Experiment 1, the eye-tracking version of the experiment used the same materials as the SAT experiment, except that a spillover region containing a prepositional phrase was appended to the end of each sentence (i.e., the judge who had rejected the motion realized that the attorney in the case compromised *during the negotiations*). Unacceptable sentences were not presented, and filler sentences were drawn from other experiments.

Experiment 2A: SAT results

Fig. 2 shows the average (over participants) d' data as a function of processing time for PI and RI constructions with the two intervener types. Asymptotic accuracy was affected by the interference manipulation but not by the semantic properties of the intervening noun phrase. This conclusion was supported by a repeated-measures ANOVA on the mean of the last four d' values, which provided an empirical estimate of asymptotic accuracy. The PI-Unmatched condition yielded the highest asymptote (3.15 d' units), followed by the PI-Matched condition (3.07 d' units), then the RI-Unmatched condition (2.69 d' units) and the RI-Matched condition (2.84 d' units) and. The effect of interference type was significant: $F_1(1, 19) = 11.03$,

Table 4
Example materials from Experiment 2 (low syntactic interference conditions). Acceptable and unacceptable conditions are represented by substituting the appropriate verb as the final word of the sentence. Slashes represent phrasal chunks displayed during SAT presentation.

Proactive interference condition	
1. Semantic mismatch	The judge / who had rejected / the motion / realized that the attorney / in the case / [compromised/entwined].
2. Semantic match	The judge / who had rejected / the witness / realized that the attorney / in the case / [compromised/entwined].
Retroactive interference conditions	
3. Semantic mismatch	The attorney / who the judge realized / had rejected / the motion / in the case / [compromised/entwined].
4. Semantic match	The attorney / who the judge realized / had rejected / the witness / in the case / [compromised/entwined].

$p < .004$, $MSE = .214$; $F_2(1, 35) = 9.20$, $p < .005$, $MSE = .35$, with the average asymptote for the PI conditions being 3.11 and the average for the RI conditions being 2.76. The effect of distractor type was not significant, $F < 1$, and the interaction was marginal in the analysis by subjects, $F_1(1, 19) = 3.30$, $p < .09$, $MSE = .079$; $F_2(1, 35) < 1$.

Hierarchical modeling of the data began with a null 1λ–1β–1δ model, which assigned a common asymptote, rate, and intercept to all four conditions. This fit produced an adjusted-R^2 of .979 for the average data, ranging from .799 to .944 across the 20 participants. By comparison, a 2λ–1β–1δ model, which assigned separate asymptotes to the two interference conditions, produced adjusted-R^2 values of .992 for the average data, ranging from .869 to .953. Parameter estimates for the averaged data are 3.29 for the PI conditions, 2.97 for the RI conditions; Table 5 presents estimates for individual subjects. All participants except 2 showed an increase in the adjusted-R^2 for the 2λ–1β–1δ model over the 1λ–1β–1δ model (average increase = .025, ranging from −.002 to .113). As suggested by the d' analysis, there was a significant effect of interference type on the asymptotic parameter estimates, $t_1(19) = 3.67$, $p < .002$. The average λ estimates were 3.36 for the PI conditions and 3.02 for the RI conditions. Models with additional asymptotes provided worse fits to the data; for example, a 4λ–1β–1δ model with separate parameters for each condition, like the model fit to the Experiment 1 data, produced an adjusted-R^2 value of .943 for the mean data (ranging from .888 to .962 for individuals).

Table 5
Experiment 2A: parameter estimates for the average data and individual participants' data for the 2λ–1β–1δ model. (PI = proactive interference; RI = retroactive interference).

	Adjusted R^2	Asymptotes		Rate	Intercept (ms)
		PI	RI		
Average	.992	3.29	2.97	.694	.805
S1	.953	3.86	3.41	.802	1.25
S2	.904	1.86	2.16	.834	1.11
S3	.912	3.16	2.08	.757	1.48
S4	.932	3.55	3.42	1.99	.664
S5	.882	3.18	3.23	1.41	.808
S6	.935	2.28	2.13	.897	1.46
S7	.908	3.80	3.25	2.59	.738
S8	.869	2.89	2.32	.975	.885
S9	.927	4.54	4.16	.286	.817
S10	.900	2.97	2.58	.582	1.94
S11	.931	4.27	3.64	.430	1.16
S12	.907	3.88	2.76	1.10	1.20
S13	.902	2.80	3.00	1.34	.842
S14	.899	2.69	2.65	1.55	.768
S15	.947	4.08	3.81	.461	.845
S16	.922	4.30	3.51	.580	1.24
S17	.946	3.29	3.56	.616	1.31
S18	.937	3.72	3.54	2.00	.704
S19	.923	2.78	2.70	1.16	1.06
S20	.943	3.26	2.49	1.04	1.29

Subsequent fits aimed at evaluating the effect of the manipulations on processing speed. We evaluated models that assigned either a separate rate parameter to the two interfering conditions (Model A: 2λ–2β–1δ), a separate rate parameter to the matched and mismatched distractor conditions (Model B: 2λ–2β–1δ), and a separate rate parameter for each of the four conditions (Model C: 2λ–4β–1δ). We also evaluated the analogous models where these separate parameters were assigned to the model intercept (i.e., Model A′: 2λ–1β–2δ assigns a single rate, but separate intercepts for the two interfering conditions). Likewise, we evaluated Model B′ (2λ–1β–2δ), assigning separate intercepts for the matched and mismatched distractor conditions, and Model C′ (2λ–1β–4δ).

Neither model with separate speed parameters for the two interference conditions (A and A′) produced substantial improvement over the single speed (2λ–1β–1δ) model. Model A had an adjusted-R^2 of .993 for the average data (.001 improvement over the single rate model), but a paired t-test on the rate parameters for Model A was only marginally significant, $t_1(19) = 1.73$, $p < .10$, suggesting that the extra parameter was not capturing significant variance across participants. Indeed, this model was better than the single-rate model for only 9 out of the 20 participants (average improvement = .0, minimum = −.02, maximum = .01). In contrast, a paired t-test on the intercept parameters for Model A′ did produce a significant effect, $t_1(19) = 2.71$, $p < .02$, but the adjusted-R^2 for this model was identical to that of the single speed model (.992). We concluded that this model did not produce a better fit for the data than the single intercept model because it produced improved adjusted-R^2 values for only seven participants (average increase = .001; minimum = −.002; maximum = .01). For all other participants the adjusted-R^2 was either identical to the model with a single rate and intercept (2λ–1β–1δ), or worse. Similarly, the models that assigned separate speed parameters for the Matched and Unmatched distractor conditions (Models B and B′) did not produce a better fit to the average data than the single speed model (Model B adjusted-$R^2 = .990$; Model B′ adjusted-$R^2 = .992$). In addition, a paired t-test comparing the different parameters for the two conditions was not significant for either model, $ps > .22$.

Both the models that assigned separate rate or intercept parameters to the four conditions (Models C and C′)

produced a small improvement (.001) in adjusted-R^2 for the average data over the single-speed model. Average improvement for individual participants was .01 for the model with different rate parameters (minimum = −.02; maximum = .03), and .002 for the model with different intercept parameters (minimum = −.005; maximum = .020. However 2 (interference type) × 2 (distractor type) ANOVA testing for Model C on the rate parameter estimates across individuals produced no significant effects, $ps > .15$, suggesting that the additional parameters were not capturing significant variation. In contrast, ANOVA testing on the intercepts for Model C′ produced a significant effect of interference on the intercept parameters, $F_1(1, 19) = 7.00$, $p < .02$, but no effect of the number of distractors. Although this model did produce an increased adjusted-R^2 value for the average data of .001, and a mean increase of .002 across participants (min = −.005, max = .015), this was true for only 12 of the 20 participants. Since the latter model did not produce a consistent improvement over the model with no dynamics differences, we concluded that there was not sufficient evidence in the data to suggest dynamics differences were present.

Experiment 2B: eye-tracking results

As in Experiment 1, analyses were performed on log transformed data due to skewness (first pass, skewness > 1.69; regression path, skewness > 2.94; total time, skewness > 1.46). Table 6 reports reading times after submission to the winsorizing procedure described above. We report the same regions of interest and the same dependent measures as in the previous experiment.

First pass

In the critical region, the effect of interference was significant, $F_1(1, 47) = 8.50$, $p < .006$, $\eta_p^2 = .15$; $F_2(1, 35) = 5.28$, $p < .03$, $\eta_p^2 = .13$. This effect was only marginal in the analysis by items on the untransformed data, $F_2(1, 35) = 3.45$, $p < .08$, $\eta_p^2 = .09$, but the effect was significant in the analysis by subjects for untransformed data. The PI conditions were read more quickly than the RI conditions (271 versus 288 ms). Neither the semantic manipulation, nor the interaction were significant, $Fs < 1.08$. There were also no significant effects in the spillover region, all $Fs < 1$.

First pass regressions

There were no significant effects in the critical region. The effect of interference was marginal in the spillover region, $F_1(1, 47) = 3.09$, $p < .09$, $\eta_p^2 = .06$; $F_2(1, 35) = 2.89$, $p < .10$, $\eta_p^2 = .08$. The RI conditions produced a higher proportion of regressions in the final region than the PI conditions (.54 versus .48).

Regression path

In the critical region, there were no significant effects, $Fs < 1$. In the spillover region, the effect of interference was significant, $F_1(1, 47) = 13.31$, $p < .001$, $\eta_p^2 = .22$; $F_2(1, 35) = 7.96$, $p < .01$, $\eta_p^2 = .19$. This was due to slower reading times in the RI conditions than in the PI conditions (1539 versus 2088). No other effects were significant, $Fs < 1$.

Total times

In the critical region, the effect of interference was significant, $F_1(1, 47) = 24.48$, $p < .0001$, $\eta_p^2 = .34$; $F_2(1, 35) = 12.91$, $p < .001$, $\eta_p^2 = .27$ This is due to the RI conditions being read more slowly than the PI conditions (504 versus 422 ms). Neither the effect of the semantic manipulation nor the interaction was significant in this region. In the spillover region, the interference effect was marginal in the analysis by subject, $F_1(1, 47) = 3.42$, $p < .08$, $\eta_p^2 = .07$, and did not reach significance in the analysis by items, $F < 1.94$. These effects came out more strongly in the analyses on untransformed data, where the effect was significant in the analysis by subjects, $F_1(1, 47) = 4.94$, $p < .04$, $\eta_p^2 = .10$; and marginal in the analysis by items, $F_2(1, 35) = 2.91$, $p < .10$, $\eta_p^2 = .08$. As in the critical region, RI conditions were read more slowly than the PI conditions (747 versus 681).

Table 6
Experiment 2B, winsorized reading times (in milliseconds) and proportion of regressive eye-movements for each region for each dependent measure. Standard error in parentheses, participants as the random factor.

	Critical region	Spillover region
First pass		
PI/Sem. mismatch	268 (9)	489 (19)
PI/Sem. match	275 (9)	473 (17)
RI/ Sem. mismatch	291 (9)	499 (19)
RI/Sem. match	284 (9)	458 (17)
Regression path		
PI/Sem. mismatch	329 (14)	1483 (98)
PI/Sem. match	330 (14)	1596 (110)
RI/Sem. mismatch	329 (13)	2081 (150)
RI/Sem. match	343 (16)	2095 (152)
Total time		
PI/Sem. mismatch	402 (17)	696 (28)
PI/Sem. match	442 (22)	665 (26)
RI/Sem. mismatch	503 (22)	765 (30)
RI/Sem. match	505 (22)	728 (32)
Proportion of regressions		
PI/Sem. mismatch	.08 (.02)	.49 (.04)
PI/Sem. match	.09 (.02)	.47 (.04)
RI/Sem. mismatch	.04 (.01)	.53 (.04)
RI/Sem. match	.08 (.02)	.54 (.04)

Experiments 2A&B: discussion

Both experiments were consistent with Experiments 1A&B in showing that RI is more detrimental than PI, causing both lower asymptotic values in the SAT task and slower reading times. The SAT results were also consistent with those in Experiment 1A in showing no indication of retrieval speed differences associated with type of interference or number of semantic distractors. As noted above, this gives further support for the use of a direct (content-addressable) access mechanism in sentence processing.

Of particular note is the lack of any effects associated with the semantic manipulation in either Experiment 2A or 2B. This null result, observed in two different experimental paradigms, when compared to the positive results observed in the same paradigms in Experiments 1A&B,

strongly suggests that syntactic context may affect the extent to which semantic interference influences comprehension. We discuss this further below.

General discussion

The reported experiments investigated the effects of semantic and syntactic competitors on memory retrieval in proactive interference (PI) and retroactive interference (RI) configurations. We contrasted PI and RI effects in comprehension to extend prior investigations of the retrieval operations to a broader class of search models. We manipulated the semantic and syntactic overlap between competitors in the sentence and the to-be-retrieved constituent to investigate the unexplored issues of what cues drive retrieval and how those cues combine.

PI and RI effects on comprehension

Retrieval speed

The key prediction of a search operation is that retrieval time depends on the number of competitors in the hypothesized search path. PI conditions placed competitors before the to-be-retrieved constituent, while RI conditions place competitors between the to-be-retrieved constituent and retrieval site. Consequently, retrieval time is predicted to be longer in PI than RI configurations if a forward search is operative, while the opposite ordering is predicted if a backward search is operative. Crucially, there was no evidence from either SAT study that processing speed (SAT intercept or rate) varied across PI and RI conditions. This is consistent with past time-course studies of retrieval in comprehension, showing that retrieval speed is unaffected by number of competitors interpolated between a dependent constituent and the retrieval site (Foraker & McElree, 2007; Martin & McElree, 2008, 2009; McElree, 2000; McElree et al., 2003). The time-course patterns observed here are inconsistent with a backward search operation, the type of retrieval operation that has been found to mediate the recovery of relational information in other domains (McElree, 2001, 2006; McElree & Dosher, 1993). Notably, the current experiments extend past research by demonstrating that the time-course profiles are also inconsistent with a forward search, which predicts retrieval speed in PI configurations should be slower than RI configurations. Caution must be exercised when attempting to reject the full class of possible search models on the basis of the current set of manipulations, however. Aside from the directionality of the search, time-course predictions for any search operation depend on how the search set is constructed (Martin & McElree, 2008). Determining the nature of the search set is relatively straightforward in basic studies of memory retrieval using unstructured lists: In those cases where a search operation appears to be operative (e.g., the recovery of order information), the evidence indicates that the search set consists of a linear sequence of available representations of the list, where availability is a direct function of the recency of encoding (e.g., McElree, 2001, 2006; McElree & Dosher, 1993). However, sentence processing requires encoding of the syntactic and semantic relations between sentential elements, which can be expected to result in highly structured representations that might be searched in a manner substantially different from a linear sequence of items. Thus, it is important to acknowledge that our PI/RI manipulations are confounded with different hierarchical relationships between the relevant nouns (i.e., presence in a subordinate versus main clause and simple versus complex subject structure) and further experimentation will be required to disentangle this issue. Previous research exploring how the geometry of hierarchical structure within a sentence affects the search path does give reason to be confident that the current conclusions will ultimately prove correct, however. McElree et al. (2003) contrasted sentences that varied the distance of the search path through either a hierarchically encoded or a linear representation and as in the current studies, found no evidence for either type of search operation.

Retrieval accuracy

Although PI and RI configurations did not differentially affect retrieval speed, asymptotic accuracy was higher in PI conditions than RI conditions in both SAT experiments. These differences in accuracy demonstrate that competitors interpolated between the initial encoding and the retrieval of a target constituent have a more disruptive effect on comprehension than those encoded before the target constituent. This result may arise either because the competitors make it more difficult to retrieve the target or because they reduce the quality of the target representation during encoding. In the current task, either factor would serve to lower the accuracy of an associated acceptability judgment. A number of related pieces of evidence help to distinguish these two explanations.

First, in the analogue reading studies (Experiments 1B and 2B), the differences in SAT asymptotic accuracy were reflected in the time readers spent on the region containing the retrieval site, with longer first-pass, regression path, and total reading times in RI than PI conditions. Additionally, when competitors shared semantic and syntactic properties with the target constituent (Experiment 1B), readers more frequently regressed out of the critical region in RI than PI configurations. The locus of these findings suggest that readers found this region more difficult to interpret in RI than PI configurations, as would be expected if they were less likely to adequately retrieve a representation of the constituent needed to resolve the dependency at the critical region.

Additional support for a retrieval locus to the observed effects comes from basic research on forgetting (Bjork, 2003). From this perspective, lower levels of accuracy for RI as compared to PI configurations are not surprising because the to-be-retrieved noun phrase is more recent in PI conditions, and hence likely to have a higher fidelity representation that may be easier to retrieve. A recency explanation which implicates *decay* as the causal factor distinguishing recent from distant competitors would not, however, provide a sufficient account of all asymptotic differences. Notably, it would not explain why different NPs (e.g., *the witness* and *the motion* in Experiment 1) in otherwise identical sentences differ in asymptotic accuracy in both PI and RI conditions. This type of difference implicates interference as a determinant of retrieval failure, as it assigns a causal role to item-specific properties of the NPs themselves. Consistent with basic memory research, which has increasingly looked away from decay-based explanations and towards accounts that emphasize interference (e.g., Anderson & Neely, 1996; Crowder, 1976; see Nairne (2002) for a review), the asymptotic differences between PI and RI conditions can be readily explained by the fact that competitors provide a stronger source of interference in RI configurations because they are more recent than competitors in the PI configurations.

Several models have been proposed to explain the mechanism through which interference has its effect. For example, Nairne's (1990) Feature model exemplifies one possible approach in which items are represented as vectors of features with values of +1 and −1. Individual features are probabilistically overwritten (assigned a null value) if that feature occurs in a subsequent item. This captures the intuition that shared features lose their ability to contribute to the distinctiveness of a representation. Simulations with this model have shown it to account for many signature phenomena in the serial recall literature, including modality effects, effects of articulatory suppression, and phonological similarity.

In its original form, this model would not be able to account for the observed accuracy differences in PI and RI conditions, however, because feature overwriting was assumed to be limited to adjacent items only. Oberauer and Kliegl (2006) have proposed a model with extended overwriting among all items held concurrently in working memory. Retroactive interference arises because the features of item n have a degrading effect on item $n - 1$, reducing the probability of accessing that item because each overwritten feature represents a potential retrieval cue that has been rendered ineffective. This means that the quality of the representation of *attorney* in an RI sentence such as (2b, repeated here) would be quite poor, due to the presence of two other nouns that share its syntactic and semantic features. In contrast, the representation of *attorney* in a PI sentence (e.g., 3a) would be much better, since the only subsequent noun (*case*) to overwrite its features has limited semantic overlap. This will yield a higher probability for retrieving attorney in the PI sentences as compared to the RI conditions.

(2b) The attorney who the judge realized had declared that the witness was inappropriate compromised.
(3a) The judge who had declared that the witness was inappropriate realized that the attorney for the case compromised.

It is important to note that these feature overwriting accounts both assume that interference has its primary effect on memory storage (or encoding): A representation loses distinctiveness at retrieval because some of its stored features have been overwritten by the subsequent encoding of similar items. This approach differs from the *cue-overload* account (e.g., Nairne, 2002; Watkins & Watkins, 1975; Öztekin & McElree, 2007) discussed in the Introduction, which attributes the detrimental effects of interference to the loss of the distinctiveness of cues at retrieval rather than to memory overwriting. We believe that a *cue-overload* approach provides a better account of our data, as well as the data from other interference studies.

This view comports with the results of studies that have specifically aimed to determine the locus of interference effects, all of which have favored the retrieval view. The most compelling evidence is that adverse effects of interference can be eliminated by changing cues at retrieval, both in comprehension (Van Dyke & McElree, 2006) and basic memory studies (Dillon & Bittner, 1975; Gardiner et al., 1972; Tehan & Humphries, 1996; Öztekin & McElree, 2007). A related effect is evident here. The semantic interference effect found in Experiment 1 was notably absent in Experiment 2, where the crucial competitor, either *the witness* or *the motion*, occurred in an object position rather than subject position, as in 4a–b repeated here.

(4a) *The attorney who the judge realized had rejected the witness in the case compromised.*
(4b) *The attorney who the judge realized had rejected the motion in the case compromised.*

Although proponents of *feature overwriting* could argue that *the witness* will overwrite fewer features in *the attorney* when it is encoded as a grammatical object rather than as a grammatical subject (e.g., 2b), this account would still predict that *the witness* would overwrite more features than would *the motion* in (4b) due to shared animacy features with the target, which would lead to higher retrieval probability in (4a). In contrast, a *cue-overload* approach, in which interference is conditioned on the cues at retrieval, provides an organic account of the context-dependent effects of interference seen in our experiments. Because the cues driving retrieval at the verb *compromised* presumably reflect the fact that it requires an NP grammatically encoded as a subject, neither *the witness* nor *the motion*, when encoded as objects, would be expected to reduce the distinctiveness of *the attorney* if the grammatical constraints of the verb are strongly weighted at retrieval. We continue this discussion below.

Syntactic and semantic interference: cue combinatorics

Cue-based retrieval models typically operate via global-matching, and in some applications all retrieval cues are thought to combine equally and simultaneously at retrieval time (Clark & Gronlund, 1996). If cues combine equally in sentence comprehension, then regardless of the syntactic context, distractors that match the semantic/pragmatic cues at retrieval should engender greater interference than distractors that mismatch those cues. In our materials, a distractor like *witness* should engender more interference than a distractor like *motion* when the semantic/pragmatic cues from the context and a verb like *compromised* drive retrieval. Experiments 1A and 1B demonstrated that this was case. However, no such effect was found in Experiments 2A and 2B, when the distractors appeared in an object rather than subject position. The differences between the first and second experiments provide prima facie evidence that syntactic cues are used in retrieval; otherwise, we should have observed comparable differences between semantic/pragmatic matching and mismatching distractors in the two experiments. More importantly, however, the absence of any effect of distractor type in the second experiment indicates that semantic/pragmatic and syntactic cues are not combined in an additive fashion. Rather, the latter appears to nullify the detrimental effect of the former.

The pattern reported here indicates, minimally, that syntactic cues are given greater weighting than semantic cues, so that distractors with inappropriate syntactic features have very low retrieval probability even when they have appropriate semantic/pragmatic properties. While it has not been the practice to apply differential weighting for different types of cues in the memory domain, the notion that certain sources of information are valued differently during language tasks has been explored in language research. For example, the Competition Model proposed by MacWhinney, Bates, and colleagues (e.g., MacWhinney & Bates, 1989; MacWhinney, Bates, & Kliegl, 1984) proposes that the frequent and reliable sources of information in a language (e.g., word order, agreement, case marking, etc.) play a dominant role in determining interpretation. Cues regarding the particular syntactic role that a noun phrase holds within a sentence could naturally join such a list, as such information is diagnostic of the particular relationships between agents in each individual linguistic expression. Moreover, there is a large body of evidence (discussed in the Introduction), pointing to the primacy of syntactic structure in determining interpretation. While this latter research has not investigated the role of syntactic cues in retrieval contexts, it sets a strong precedent for the notion that syntactic information may be given preferential value.

Combinatorially, either a linear or a non-linear cue weighting scheme could achieve the results observed here. In a linear scheme, cues are combined in a weighted fashion, as they are assumed to be in many cognitive and perceptual domains (e.g., Trommershäuser et al., in preparation), with extreme differential weights on syntactic and non-syntactic cues, so that the latter would have little or no measurable effect when a distracting element was syntactically inappropriate. A nonlinear scheme, in contrast, places a much stronger constraint on the retrieval architecture and on the contents of the search space, implying that syntactic cues serve a gating function, so that *only* candidates in memory with matching syntactic properties are considered. Distinguishing between these two possibilities is difficult because a linear scheme can mimic a nonlinear scheme if the weighting differential were sufficiently large. It is possible to rule out a nonlinear scheme, however, if distractors with inappropriate syntactic properties are found to engender retrieval interference. There are two suggestive findings in the literature, which might be interpreted in such a fashion, although neither provides definitive evidence.

The first is the data from dual-task paradigm studies (e.g., Fedorenko et al., 2006a; Gordon et al., 2002; Van Dyke & McElree, 2006), described above, where nouns retained in memory while processing a sentence produced proactive interference effects on reading times when they matched the semantic cues from the critical verb at the retrieval site. One might argue that a nonlinear scheme would predict that these nouns should not produce any semantic interference because they were not encoded into a syntactic context and therefore should not resonate with the syntactic cue at retrieval. This challenge to the non-linear account is weakened, however, by the possibility that the *absence* of a syntactic feature may not be considered to be a mismatching feature, making these results orthogonal to the issue of whether syntactic cues categorically restrict the set of potential distractors.

The second relevant study is Van Dyke (2007), which reported semantic interference effects on reading times and comprehension accuracy from sentence-internal distractors with mismatching syntactic cases, as in (10).

(10) The pilot remembered that the lady who was sitting near the smelly [seat/man] moaned about a refund.

These results are at odds with those of Experiment 2, where semantic interference was not observed. A notable difference between the two studies is that the semantic distractor occurred as a direct object in the current study (cf. Table 4), while it occurred as the object of a preposition in the Van Dyke (2007) study. Most grammatical theories make hierarchical distinctions between *core* arguments (i.e., subject, object, sometimes indirect objects) and other modifying adjuncts (or *oblique* arguments), including prepositional phrases (Bresnan, 2001; Chomsky, 1981; Culicover & Jackendoff, 2005; Frazier & Clifton, 1996; Keenan & Comrie, 1977; Perlmutter, 1983; Van Valin & LaPolla, 1997). In these formulations, a core argument, such as a direct object, plays a more prominent role at the interface between syntactic structure and semantic interpretation than adjuncts because the former specifies the thematic relationships that noun phrases play with respect to the meaning of the predicate (i.e. Agent, Patient, Theme, etc.). Consequently, it is possible that the syntactic encoding of adjuncts is less distinctive than arguments, and that these syntactic features are not salient enough to produce a mismatch for the purpose of syntactic gating, making distractors in these positions available to produce interference when those in core argument positions are more easy to rule out. If this is correct, then the distractors in a sentence like (10) are functionally equivalent to the non-sentential nouns in the Van Dyke and McElree (2006) study, having little or no relevant syntactic features with which to be eliminated from the distractor set. This is akin to the challenge presented by the dual-task paradigm data; but it suffers from the same weakness: an alternative valuation scheme may also be at work, where less distinctive syntactic features may produce a greater mismatch by virtue of their absent or weakly represented features.

Thus, there is so far no definitive evidence to rule out a nonlinear scheme for combining retrieval cues, where syntactic cues may serve to restrict the potential distractor set. However, the current evidence does clearly support at least a linear scheme with differential weighting for syntactic and semantic cues, such that syntactic cues are weighted more strongly, giving them the potential for canceling out semantic distractors in highly salient syntactic positions. Further research across a variety of syntactic contexts is needed to determine whether a weighted linear scheme for combining cues at retrieval may hold more generally, or whether a non-linear, gating scheme is at work.

Conclusion

The findings from the reported SAT and eye-tracking experiments add to a growing body of research indicating that the representations formed during sentence comprehension are content addressable and retrieved with a cue-driven, direct-access operation. The susceptibility of this type of retrieval mechanism to interference provides an alternative account of constraints on sentence processing that have traditionally been attributed to WM capacity and resource limits. The reported studies demonstrate that interference lowers retrieval accuracy, but does not affect retrieval speed. Notably, we found that syntactic constraints at the retrieval site are among the cues that drive retrieval in comprehension, and that these constraints may limit potential sources of interference from memory constituents that have semantic properties in common with the target constituent. Minimally, our data indicate a combinatorics scheme in which syntactic constraints are weighted more heavily than semantic constraints in retrieval.

A major implication of the current study is that the retrieval mechanism utilized during language comprehension is (at least) functionally identical to that utilized in non-language domains (i.e., content-addressable retrieval). Since differential weighting schemes are not typically assumed in models of list memory retrieval, assignment of different weights for particular types of features (syntactic, semantic, referential) may be one way that the basic memory mechanism is further specialized for language—even for particular languages, while retaining the same retrieval mechanism for accessing previously encoded information.

Acknowledgments

Portions of this research were supported by NIH/NICHD F32-HD-049215 to New York University (Van Dyke, PI), by NIH/NICHD Grant R01-HD-040353 to Haskins Laboratories (Shankweiler, PI), by NIH/NICHD Grant R21-HD-058944 to Haskins Laboratories (Van Dyke, PI) and by NIH/NICHD Grant R01-HD-056200 to New York University (McElree, PI). We are grateful for the suggestions of Ev Fedorenko and two anonymous reviewers on earlier versions of this manuscript.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247–264.

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Handbook of perception and memory, Memory (Vol. 10* (pp. 237–313). San Diego: Academic Press.

Bjork, R. A. (2003). Interference and forgetting. In J. H. Byrne (Ed.), *Encyclopedia of learning and memory* (2nd ed., pp. 268–273). New York: Macmillan Reference USA.

Bresnan, J. (2001). *Lexical functional syntax*. Oxford: Blackwell.

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

Chomsky, N. (1981). *Lectures in government and binding*. Foris: Dordrecht.

Christianson, K., Hollingworth, A., Halliwell, J., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42*, 68–407.

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

Clifton, C., Jr., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language, 49*, 317–334.

Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.

Culicover, P. W., & Jackendoff, R. (2005). *Simpler syntax*. Oxford: Oxford University Press.

Desmet, T., De Baecke, C., Drieghe, D., Brysbaert, M., & Vonk, W. (2006). Relative clause attachment in Dutch: On-line comprehension corresponds to corpus frequencies when lexical variables are taken into account. *Language and Cognitive Processes, 21*, 453–485.

Dillon, R. F., & Bittner, L. A. (1975). Analysis of retrieval cues and release from proactive inhibition. *Journal of Verbal Learning and Verbal Behavior, 14*, 616–622.

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

Fedorenko, E., Gibson, E., & Rohde, D. (2006a). The nature of working memory capacity in sentence comprehension: Evidence against domain specific resources. *Journal of Memory and Language, 54*.

Fedorenko, E., Gibson, E., & Rohde, D. (2006b). The nature of working memory in linguistic, arithmetic and spatial integration processes. *Journal of Memory and Language, 56*.

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

Ferreira, F., Ferraro, V., & Bailey, K. G. D. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*, 11–15.

Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Availability versus accessibility. *Journal of Memory and Language, 56*, 357–383.

Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.

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

Gardiner, J. M., Craik, F. I. M., & Birstwistle, J. (1972). Retrieval cues and release from proactive inhibition. *Journal of Verbal Learning and Verbal Behavior, 11*, 778–783.

Gibson, E. (2006). The interaction of top-down and bottom-up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language, 54*, 363–388.

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

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language, 51*, 97–114.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science, 13*, 425–430.

Grodner, D., Gibson, E., & Watson, D. (2005). The influence of contextual contrast on syntactic processing: Evidence for strong-interaction in sentence comprehension. *Cognition, 95*, 275–296.

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

Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry, 8*, 63–99.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*, 205–225.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science*, Oct.

MacWhinney, B., & Bates, E. (1989). *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior, 23*, 127–150.

Martin, A. E., & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language, 58*, 879–906.

Martin, A. E., & McElree, B. (2009). Memory operations that support language comprehension: Evidence from verb-phrase ellipsis. *Journal of Experimental Psychology: Learning Memory and Cognition, 35*, 1231–1239.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory. *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. (2006). Accessing recent events. In B. H. Ross (Ed.). *The psychology of learning and motivation* (Vol. 46, pp. 1–2). San Diego: Academic Press.

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., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language, 48*, 67–91.

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In R. Duncan Luce, Robert R. Bush & Eugene Galanter (Eds.)., *Handbook of mathematical psychology* (Vol. 2, pp. 419-491). New York: Wiley.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*, 251–269.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53*, 53–81.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language, 55*, 601–626.

Öztekin, I., Davachi, L., & McElree, B. (In press). Are representations in working memory distinct from those in long-term memory? Neural evidence in support of a single store. *Psychological Science*.

Öztekin, I., McElree, B., Staresina, B. P., & Davachi, L. (2008). Working memory retrieval: Contributions of left prefrontal cortex, left posterior parietal cortex and hippocampus. *Journal of Cognitive Neuroscience, 21*, 581–593.

Öztekin, I., & McElree, B. (2007). Retrieval dynamics of proactive interference: PI slows retrieval by eliminating fast assessments of familiarity. *Journal of Memory and Language, 57*, 126–149.

Perlmutter, D. (Ed.). (1983). *Studies in relational grammar*. University of Chicago Press.

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

Pylkkänen, L., & McElree, B. (2006). The syntax–semantics interface. On-line composition of sentence meaning. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed.. NY: Elsevier.

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.

Schneider, W., Eshman, A., & Zuccolotto, A. (2002). *E-prime: A user's guide*. Pittsburgh, PA: Psychology Software Tools.

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.

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

Swinney, D., Zurif, E., & Nicol, J. (1989). The effects of focal brain damage on sentence processing: An examination of the neurological organization of a mental module. *Journal of Cognitive Neuroscience, 1*, 25–37.

Tehan, G., & Humphries, M. S. (1996). Curing effects in short-term recall. *Memory & Cognition, 24*, 719–732.

Trommershäuser, J., Landy, M. S., & Körding, K. (Eds.). (in preparation). *Sensory cue integration*. New York: Oxford University Press.

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

Tulving, E. (1979). Relation between encoding specificity and levels of processing. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory*. (pp. 405-428). Hillsdale, N.J: Erlbaum.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning & Verbal Behavior, 5*(4), 381–391.

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 407–430.

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A retrieval interference theory of recovery from misanalyzed ambiguities. *Journal of Memory and Language, 49*, 285–413.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language, 55*, 157–166.

Van Valin, Robert D., Jr., & LaPolla, R. (1997). *Syntax: Structure, meaning and function*. Cambridge: Cambridge University Press.

Watkins, O. C., & Watkins, M. J. (1975). Build-up of proactive inhibition as a cue overload effect. *Journal of Experimental Psychology: Human Learning and Memory, 104*, 442–452.

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