

A content-addressable pointer mechanism underlies comprehension of verb–phrase ellipsis [☆]

Andrea E. Martin, Brian McElree ^{*}

Department of Psychology, New York University, 6 Washington Place, 8th floor, New York, NY 10003, USA

Received 22 November 2006; revision received 18 June 2007

Available online 20 August 2007

Abstract

Interpreting a verb–phrase ellipsis (VP ellipsis) requires accessing an antecedent in memory, and then integrating a representation of this antecedent into the local context. We investigated the online interpretation of VP ellipsis in an eye-tracking experiment and four speed–accuracy tradeoff experiments. To investigate whether the antecedent for a VP ellipsis is accessed with a search or direct-access retrieval process, Experiments 1 and 2 measured the effect of the distance between an ellipsis and its antecedent on the speed and accuracy of comprehension. Accuracy was lower with longer distances, indicating that interpolated material reduced the quality of retrieved information about the antecedent. However, contra a search process, distance did not affect the speed of interpreting ellipsis. This pattern suggests that antecedent representations are content-addressable and retrieved with a direct-access process. To determine whether interpreting ellipsis involves copying antecedent information into the ellipsis site, Experiments 3–5 manipulated the length and complexity of the antecedent. Some types of antecedent complexity lowered accuracy, notably, the number of discourse entities in the antecedent. However, neither antecedent length nor complexity affected the speed of interpreting the ellipsis. This pattern is inconsistent with a copy operation, and it suggests that ellipsis interpretation may involve a pointer to extant structures in memory.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Verb–phrase ellipsis; Sentence processing; Speed–accuracy tradeoff

Introduction

Natural language often contains dependencies that span several words, phrases, or even clauses. To interpret expressions with nonadjacent dependencies, language comprehenders must, at a minimum, identify the site of the dependency, access a representation of an earlier-processed constituent, and then integrate that constituent into the local structure.

We report five experiments that investigate the processing of sentences with nonadjacent dependencies stemming from a verb–phrase ellipsis (VP ellipsis). We investigate how a representation of the antecedent of

[☆] The authors thank Steven Frisson, Gregory Murphy, Ilke Öztekin, Liina Pykkänen, and Julie Van Dyke for their assistance in different aspects of this project, as well as three anonymous reviewers for helpful comments on an earlier version of this work. This research was supported by a National Science Foundation Grant (BCS-0236732) awarded to B.M. and a National Science Foundation Graduate Research Fellowship (2006025605) awarded to A.E.M.

^{*} Corresponding author. Fax: +1 212 995 4349.

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

the VP ellipsis is accessed at the elision site, and how it is then integrated into the local structure. Consistent with investigations of other nonadjacent dependencies (McElree, 2000; McElree, Foraker, & Dyer, 2003), our findings suggest that the antecedent for an ellipsis is content-addressable and can be directly accessed without the need for a search through irrelevant memory representations. We also report findings that are inconsistent with claims that the retrieved antecedent is “copied” into the local structure. Rather, our results suggest that a pointer-like mechanism is used to interpret the VP ellipsis.

Verb-phrase ellipsis

VP ellipsis is the omission of a verb phrase that is necessary for a complete formal representation of the sentence. The sentence in (1) is an example:

- (1) *The pedestrian called a cab, and the bellhop did too.*

Here, comprehenders must interpret the expression *the bellhop did too* in a manner that can be paraphrased as “the bellhop called a cab too”. Hence, the phrase *did too* must receive its interpretation from the interpretation of the earlier verb phrase, *called a cab*.

How do comprehenders accomplish this task? Presumably, comprehenders would have all the information required to interpret an ellipsis if a representation of the antecedent were actively maintained in focal attention. However, in many (perhaps most) instances of ellipsis, the processing of material intervening between the antecedent and the ellipsis will displace the antecedent from the comprehender’s current focus of attention. When this is the case, comprehenders must access a representation of an appropriate constituent in working memory. Once accessed, this information can be used to interpret the VP ellipsis by “calling on” the elided information.

Our studies focus on two component processes involved in interpreting an ellipsis: how comprehenders access an appropriate antecedent in memory and how information is retrieved from the antecedent once it has been accessed. Recent work has addressed issues concerning the latter (e.g., Arregui, Clifton, Frazier, & Moulton, 2006; Frazier & Clifton, 2001, 2005; Murguía, 2004), and our studies build on this research. To our knowledge, no studies have directly investigated how an antecedent representation is accessed. However, this issue has been investigated in studies of the processing of other types of nonadjacent dependencies (McElree, 2000; McElree et al., 2003).

Accessing an antecedent

Ellipsis can be ambiguous. For example, the sentence *John knew Jane read the author’s new novel, but Bill didn’t* could be interpreted as either “Bill didn’t read

the author’s new novel” or “Bill didn’t know that Jane read the author’s new novel.” Although there are important issues concerning how an antecedent is selected when more than one is possible, we focused on the processing of (largely) unambiguous structures in order to investigate basic mechanisms used to access an antecedent representation in memory.

Research on retrieval processes has identified two basic ways in which working memories can be accessed (see McElree, 2006, for a review). Recovering some types of information requires a search process. Several studies have demonstrated that temporal and spatial order information are recovered with a serial search mechanism (Gronlund, Edwards, & Ohrt, 1997; McElree, 2001, 2006; McElree & Doshier, 1993). To date, research has not delineated all the circumstances in which a serial search process might be required. However, it has established that accessing an item representation in memory—viz., retrieving item information—does not require a serial search process. Instead, current evidence indicates that item information is *content-addressable* (McElree, 1996; 1998; 2000; 2006; McElree and Doshier, 1989, 1993), contrary to some early models of short-term memory retrieval (e.g., Sternberg, 1975; Theios, 1973; Treisman & Doctor, 1987). The defining property of a content-addressable representation is that information (cues) in the retrieval context can provide direct-access to the memory representation, without the need to search through extraneous memory representations. Content-addressability can be implemented in models with rather diverse storage architectures, including those with localized representations and those with highly distributed representations (Clark & Gronlund, 1996).

Which type of retrieval process is operative in language comprehension? Inasmuch as the hierarchical structure of a sentence is often encoded in the order of constituents within a string, and predominantly so in languages such as English, one might predict that a serial search is required to access the antecedent for most types of nonadjacent dependencies (McElree et al., 2003). For example, if the dependency requires an antecedent with a particular morphological feature, or if the dependency requires the antecedent to have a specific syntactic and semantic role associated with a particular sentence position, then comprehenders might need to serially search their memory representation of the input to find the required antecedent (McElree et al., 2003).

There is of course a large class of possible search mechanisms. At one extreme, one could envision a relatively low-level serial search, in which an ordered representation of the input is scanned in either a forwards or backwards fashion, with each scanned constituent being sequentially evaluated for its degree of match to the search criteria (e.g., matches the required morpho-syntactic and/or semantic-pragmatic properties needed for

the antecedent). This is the type of search operation that is been found to mediate the recovery of order information in unstructured memory lists (Gronlund et al., 1997; McElree, 2001, 2006; McElree & Doshier, 1993). Alternatively, the search mechanism may be more sophisticated, with the search acting on a more structured representation of the input (McElree et al., 2003). For example, only constituents in certain positions or with certain properties may be iteratively evaluated for their suitability as antecedents. More sophisticated search operations require memory representations with some degree of content-addressability in order to constrain the candidate set of constituents.

If the memory representations formed during comprehension are fully content-addressable, then it is possible that comprehenders consider only the constituent (in the case of an unambiguous expression) or constituents (in the case of an ambiguous expression) that are fully compatible with all properties needed to resolve the dependency. In this case, retrieval is mediated by a direct-access rather than search operation. That is, there is direct contact between the retrieval cues at the site of the dependency and the antecedent such that retrieval “serves up” the correct memory representation by virtue of its content, thereby obviating the need for a search through other constituents in memory. This is the type of mechanism that has been argued to underlie the retrieval of item information from both short- and long-term memory (McElree, 2006; McElree & Doshier, 1989).

The key prediction of a search process is that search time should increase as more information is added to the memory representation. This is true of a serial (iterative) search, as in the examples outlined above, but also true of searches with some degree of parallel processing. Serial models typically predict that serial time increases linearly with the number of items searched, whereas parallel models typically predict nonlinear increases (see Townsend & Ashby, 1983 for specific reaction time predictions and McElree & Doshier, 1989 for speed–accuracy tradeoff predictions). In contrast, if memory representations are fully content-addressable and directly accessible with cues provided at the retrieval site, then retrieval speed will be unaffected by the amount of information in memory.

McElree, Foraker, and Dyer (2003; also McElree, 2000) investigated whether a search process was operative in the processing of two types of common nonadjacent dependencies. The experiments examined the speed and accuracy of processing sentences with filler-gap dependencies such as (2), in which a filler item, *the book*, must be associated with a gap in the direct object position of the final verb, *admired*, and sentences with nonadjacent subject-verb relations such as (3), in which a relative clause intervenes between a matrix subject, *the editor*, and a matrix verb, *laughed*.

- (2) *This was the book that the editor admired.*
 (3) *The editor that the book amused laughed.*

Applying the logic above leads to a prediction that processing speed at the site of the dependency [the final verb in (2) and (3)] should systematically slow as more material intervenes between the two dependent elements. For example, if resolving the dependencies between the verb and its object in (2) or the subject and verb in (3) requires a search—either through a representation of linear surface structure or through a more interpreted representation—then it should take more time to access the relevant constituent when more information is held in memory, which should slow overall interpretation time. McElree et al. found that increasing the amount of interpolated material reduced the probability of computing an acceptable interpretation, but, crucially, it did not affect the speed of comprehension. The same pattern was found in cases where successful interpretation required resolving two dependencies in one of two possible orders, a situation in which the recovery of order information was essential to the interpretation (McElree et al., 2003, Experiment 3).

McElree et al. argued that these results are inconsistent with the type of serial retrieval process that has been found to underlie the recovery of order information. Indeed, the timecourse findings are inconsistent with a large class of search mechanisms. One possible exception might be a forward serial search, in which the comprehender starts at the beginning of the sentence and searches forward for a constituent to resolve the dependency. No effect of interpolated material is predicted by a forward search if the item to-be-retrieved from memory is in a sentence-initial position, as it was in the studies of McElree (2000) and McElree et al. (2003). However, Van Dyke and McElree (2007) have compared in two studies the processing of sentences such as (4) and (5).

- (4) *The assistant who had said that the visitor was important forgot that the client at the office objected.*
 (5) *The client who the assistant forgot had said that the visitor was important objected.*

In sentences such as (4), the subject of the final verb (*objected*) is in an embedded rather than sentence-initial position, as in (5). A forward serial search would predict longer processing time for (4) as compared to (5), as comprehenders would need to search through intervening material that includes (at least) two possible noun phrases (*assistant* and *visitor*) in the former. A backward serial search would predict the opposite pattern of differences. Van Dyke and McElree (2007) found that sentences such as (5), where there was material interpolated between the beginning and end of the

dependency, produced lower levels of accuracy than sentences such as (4), where the additional material occurs before the first element of the dependency. These results suggest that retroactive interference has a more detrimental effect on sentence processing than proactive interference. Crucially, however, there was no difference in processing speed between sentences such as (4) and (5), a result that is inconsistent with *both* a forward and a backward serial search.

Collectively, the evidence is most consistent with the idea that a content-addressable memory system underlies both the binding of a filler to a gap and the binding of a subject to a verb. We assume that in such a system various sources of information available at the point where a dependency must be resolved serve to provide direct access to the relevant representation in memory. These sources may include morpho-syntactic and semantic information, as well as pragmatic and discourse information. In contrast to a mechanism that searches (either in a serial or parallel fashion) a structured memory representation for a constituent that matches the required morpho-syntactic, semantic, referential, and pragmatic properties, a direct-access mechanism uses those properties to reintegrate the constituent needed to resolve the dependency. Direct access can be implemented in different general memory models (see Clark & Gronlund, 1996). In sentence comprehension, the evidence for direct access has motivated parsing models in which a cue-based retrieval mechanism mediates the creation of grammatical dependencies during parsing, and parsing success depends on the extent to which required constituents can be retrieved from working memory (e.g., Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; Van Dyke, 2002; Van Dyke & Lewis, 2003). The findings are also consistent with dynamical models postulating representations where grammatical features are distributed over several nodes (e.g., Tabor, Galantucci, & Richardson, 2004; Tabor & Hutchins, 2004; Vosse & Kempen, 2000), which likewise assume content-addressable representations and direct-access retrieval processes.

If sentence comprehension is generally mediated by directly-accessible content-addressable memory structures then manipulations of distance should likewise not affect the speed of processing elliptical expressions. However, ellipsis differs from filler-gap and subject-verb dependences in at least one important way. In the latter two structures, the constituent to be retrieved from memory is marked in syntax as one that must be integrated with subsequent material: A subject must agree with and be integrated with a verb, and the filler in a filler-gap construction has no role in the sentence until the gap site is identified. Consequently, comprehenders can anticipate that the constituent will be required in later operations, and they may assign it some special status in memory. Indeed, parsing models often assume that

these types of constituents are held in specialized stacks or buffers, and some of these mechanisms can mimic properties of a direct-access operation (see McElree et al., 2003).

In contrast, the antecedent of a VP ellipsis is fully integrated in its local context, and comprehenders cannot routinely anticipate that it will need to be retrieved downstream. Given its lack of special status, the recovery of an antecedent for VP ellipsis provides an important test case for content-addressability in comprehension. Experiments 1 and 2 extend studies investigating the effects of distance on processing ellipsis (e.g., Murphy, 1985, discussed below) in ways that provide a strong test of whether distance engenders differences in the speed of interpreting ellipsis. We test the claim that the representation of the antecedent for an ellipsis is likewise content-addressable, and that comprehenders use available morpho-syntactic, semantic, and pragmatic constraints at the ellipsis site as retrieval cues for accessing the antecedent representation.

Recovering antecedent information

Once an antecedent representation for a VP ellipsis has been accessed, how is it interpreted? A central question in the linguistic analysis of ellipsis has been whether or not interpretation of the ellipsis requires that a fully articulated syntactic structure be present at the elision site (Frazier & Clifton, 2001; 2005; 2006; Murguia, 2004).

One argument that it might be is that ellipses often contain variables that need to be reinterpreted at the elision site, such as the reflexive *himself* in (6).

- (6) John needed to motivate himself, but Bill didn't.

The preferred interpretation of (6) is that “Bill didn't need to motivate himself,” rather than “Bill didn't need to motivate John.” Crucially, the former requires reinterpreting the reflexive *himself* to be coreferent with Bill, which could require copying the syntactic structure of the antecedent into the elision site (Nunes, 1995; cf. Murguia, 2004). Other arguments rest on whether the grammaticality of an ellipsis is determined by whether or not the syntactic structure assumed to be present in the ellipsis site is identical in form to the antecedent (Frazier & Clifton, 2005). Presumably, if interpretation requires syntactic structure at the ellipsis site, then the antecedent should have an identical syntactic form; non-parallel forms should be either ungrammatical or require additional repair operations to be interpreted (Arregui et al., 2006; Frazier & Clifton, 2005).

Our primary focus is on whether a representation of the antecedent needs to be copied into the elision site, whatever the form of the representation might be. If copying is taken as a real-time operation in comprehension, then a straightforward prediction is that processing

time should increase as the amount of material contained within the antecedent increases. This follows from the intuitive assumption that it should take more time to copy more information (Frazier & Clifton, 2001). Experiments 3–5 test this prediction by examining whether the speed and accuracy of interpreting ellipsis with simple VP antecedents consisting of a verb and simple noun phrase (e.g., ... *understood Roman mythology*) differs from ellipses with lengthier VP antecedents consisting of a verb and complex noun phrase (e.g., ... *understood Rome's swift and brutal destruction of Carthage*). Experiment 5 tests this prediction by contrasting antecedents that contain variables and differing degrees of syntactic complexity.

If an antecedent is not copied into the elision site, how then might the VP ellipsis be interpreted? Several researchers have argued that working memory can include pointers to larger chunks of information in longer-term memory (e.g., Ericsson & Kintsch, 1995; Ruchkin, Grafman, Cameron, & Berndt, 2003). We pursue an alternative hypothesis to a structure-sensitive copy operation that instead views VP ellipsis as a pointer to a preexisting memory structure. Rather than requiring comprehenders to copy structure from memory into the workspace of ongoing processes, we suggest that a clause containing the ellipsis [e.g., *the bellhop did too* in example (1)] might be interpreted by a pointer that links it to the antecedent representation that has been accessed in memory [e.g., ... *called a cab* in example (1)].

Frazier and Clifton (2001, 2005) suggest an alternative hypothesis to a canonical copy mechanism that is similar in some respects to a pointer hypothesis. They argue that basic structure-building operations in comprehension are sensitive to complexity, because building more syntactic structure typically requires more costly syntactic “inferences” (Frazier & Clifton, 2001, pp. 1–2). However, ellipses are thought to exploit a specialized mechanism, dubbed “cost-free” copy- α (where α is the antecedent) in Frazier and Clifton (2001) or *structure sharing* in Frazier and Clifton (2005). They suggest that increasing antecedent complexity may not engender differential processing costs with this type of mechanism because the number of inferences needed to identify the ellipsis site remains constant (assuming the ellipsis is unambiguous), regardless of the amount of structure that needs to be shared.

Frazier and Clifton (2001) speculate that their hypothesized operation copy- α could be implemented as a pointer mechanism in which the ellipsis site points to the left corner of the antecedent's syntactic representation (see also Murguía, 2004). Our proposal differs from Frazier and Clifton's account in that it does not assume that a pointer necessarily directs comprehenders to a syntactic representation. Although some evidence suggests that it might (Frazier & Clifton, 2001, 2005), a pointer mechanism is equally compatible with alterna-

tive views that ellipses are a type of discourse anaphora (see Garnham, 2001) or that they can be interpreted by establishing coherence relations based on semantics and discourse properties alone (Kehler, 2002). In these cases, the pointer would simply point to a more fully interpreted discourse representation, and the interpretation of variables, such as the reflexive in (6), would require reanalysis at a conceptual level.

Frazier and Clifton (2001) report the absence of complexity effects in VP ellipses consisting of one-clause antecedents (e.g., *Sarah left her boyfriend last May. Tina did too*) and two-clause antecedents (e.g., *Sarah got up the courage to leave her boyfriend last May. Tina did too*). Self-paced reading times on the final sentences with the ellipsis did not differ, despite the fact that the two-clause antecedent was lengthier and perhaps more complex than the one-clause antecedent. This finding provides some evidence against the real-time operation of a canonical copy mechanism. Experiments 3–5 follow up on this initial finding. Importantly, we use an experimental procedure that measures how the interpretation of VP ellipsis unfolds over processing time, which provides a more sensitive test of whether complexity affects processing speed.

Speed–accuracy tradeoff

The reported experiments sought to determine whether a search or content-addressable mechanism is used to access an antecedent for a VP ellipsis, and whether a copy or pointer mechanism is then used to interpret the ellipsis. In both cases, key predictions concern the relative speed of interpreting different ellipses. One might imagine that these predictions could be tested with simple timing measures, derived from either response time or reading time tasks. These measures are useful for assessing whether conditions vary in difficulty, but they are of limited value in testing strong predictions concerning differences in the speed of processing. As an illustration, consider a finding that reading time slows as the distance of the antecedent is increased (Murphy, 1985). One might be tempted to take that finding as evidence for a search mechanism, by interpreting the difference as reflecting the time to search through different amounts of material. However, distance can affect the quality of the antecedent's representation in memory, as a distant antecedent will have been processed less recently and could be subjected to more interference (Foraker & McElree, 2007; McElree, 2000; McElree et al., 2003). There are several reasons why a poor memory representation could engender longer reading times. On some trials, the antecedent may not be successfully retrieved at the elision site, which could cause interpretation to fail or could require the comprehender to initiate a costly reanalysis process. Even if retrieval failures are rare, a poorly represented anteced-

ent may not adequately support interpretive operations, and this may result in a less meaningful interpretation. For these reasons and perhaps for others, a reading time difference alone may not reflect underlying differences in the time to access an antecedent. A similar logic applies to investigations of antecedent complexity.

A second reason why reading time measures might not be optimal is that they do not afford much experimental control over the depth to which participants process a sentence. There is a growing body of literature indicating that readers can sometimes underspecify an interpretation (Christianson, Williams, Zacks, & Ferreira, 2006; Pickering, McElree, Frisson, Chen, & Traxler, 2006; Poesio, Sturt, Artstein, & Filik, 2006; Sanford & Sturt, 2002). Here, the concern is that readers may not fully interpret the ellipsis at the regions of interest. To encourage participants to read for understanding, researchers often present comprehension questions after reading, and sometimes conditionalize reading times on comprehension performance. However, comprehension questions are of limited value, as questions are administered after reading times for the region of interest have been collected. In our application, for example, one could fail to detect a distance or complexity effect at the ellipsis region if subjects underspecified the interpretation until the comprehension question forced a more complete interpretation.

As a solution to both concerns, we used the response-signal speed–accuracy tradeoff (SAT) procedure to examine the effects of distance and complexity on VP ellipsis interpretation. The primary benefit of this procedure is that the speed and the accuracy of processing can be measured conjointly within a single task (e.g., Doshier, 1979; Reed, 1973, 1976; Wickelgren, 1977). We had participants read sentences presented phrase by phrase and, at designated points, decide (yes/no) whether the passage was sensible. We used a multiple-response variant of the SAT procedure that has been used in several investigations of language processing (e.g., Bornkessel, McElree, Schleuwsy, & Friederici, 2004; Foraker & McElree, 2007; McElree, 1993; McElree, Pykkänen, Pickering, & Traxler, 2006): Participants were trained to respond to an auditory response signal presented at 18 times after the onset of a crucial expression, here a VP ellipsis. Crucially, the first response signal onset occurred 300 ms before the onset of the VP ellipsis, and thus subjects were required to respond before processing of the crucial expression had begun. The subsequent sampled times (0–6000 ms) enabled us to fully measure how the interpretation of the VP ellipsis unfolded over time. For each sampled point, we constructed a d' measure of accuracy by scaling correct responses to sensible elliptical expressions (hits) against incorrect responses to control expressions with nonsensical VP ellipsis interpretations (false alarms). This scaling provided a measure of the ability of participants to

discriminate acceptable from unacceptable interpretations.

Fig. 1 presents illustrative SAT functions— d' accuracy versus processing time—for two hypothetical conditions. Characteristically, the functions show a period of chance performance ($d' = 0$), a period of increasing accuracy, and an asymptotic period during which further processing does not improve performance. In our studies, the time-course functions for each participant were fit with an exponential approach to a limit, which enabled us to quantify how the interpretation of the different ellipses unfolded over time:

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

The parameter λ , which estimates the asymptote of the function, measures the highest level of discrimination reached with maximal processing time, and hence yields a basic measure of processing accuracy. Differences in asymptote alone are illustrated in Fig. 1A. Conditions that vary in asymptote differ in the likelihood that a meaningful interpretation can be assigned to each type of expression or that the interpretation of the expressions differs in their overall degree of acceptability. Here, the asymptotes index how successful compreh-

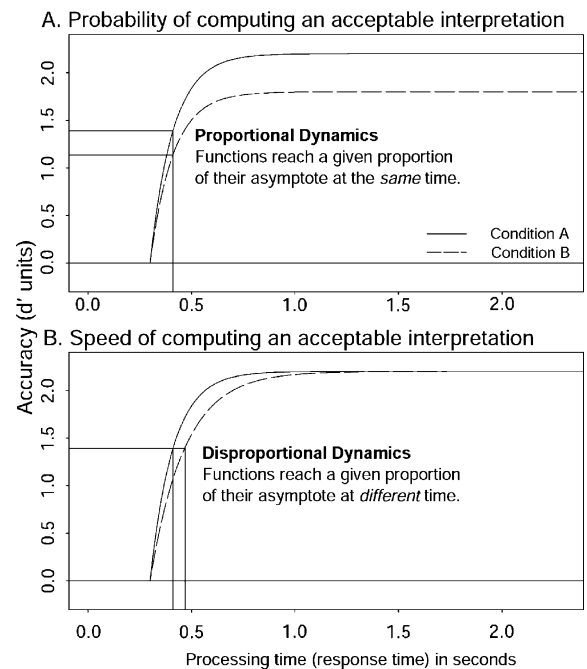


Fig. 1. Hypothetical SAT functions illustrating two conditions that differ by SAT asymptote only (A) or SAT rate (B). The intersection of the horizontal and vertical lines shows the point in time (abscissa) when the functions reach two-thirds of their respective asymptote (ordinate). When dynamics are proportional (A), the functions reach the two-thirds point at the same time.

enders were at retrieving an antecedent for the ellipses. Increasing distance or complexity should lower asymptotic accuracy if they decrease the quality of the antecedent's representation in memory, making the antecedent less likely to be retrieved from memory or reducing the quality of the retrieved information. In a related study of pronoun interpretation, Foraker and McElree (2007) suggested that lower asymptotic performance can be generally construed as differences in the *availability* of information in memory essential to forming coherent interpretations of the anaphoric expression, whether the asymptotic differences reflect failures to recover the antecedent, the inherent quality of the retrieved information, or a mixture of both.

The principle advantage of the speed–accuracy trade-off procedure is that it enables one to measure and compare the speed of interpretation of conditions that may also differ in overall accuracy. Thus, we can determine the relative speed of interpreting an expression on the respective proportion of trials that readers succeed in computing a sensible interpretation. The intercept (δ) and rate (β) of the function provide joint measures of the speed of processing, indexing how quickly accuracy accrues to its asymptotic level. The parameter δ estimates the intercept of the function, or the point at which participants are first sensitive to the information necessary to discriminate acceptable from unacceptable ellipses (i.e. d' departs from 0, chance performance). The parameter β estimates the rate at which accuracy grows from chance to asymptote. Fig. 1B illustrates two conditions that differ in rate. If one ellipsis can be interpreted more quickly than another, the SAT functions will differ in rate, intercept, or some combination of the two parameters (e.g., Bornkessel et al., 2004; McElree, 1993; McElree & Nordlie, 1999; McElree et al., 2006). Whether speed differences are expressed in rate or intercept depends on the mean and variance of the difference in the time it takes to compute the different interpretations. In some contexts, the locus of the effect can be theoretically important (e.g., McElree & Doshier, 1993). However, the predictions we tested are based on general differences in speed of processing, which can be assessed by effects on either parameter. Importantly, whether differences are expressed in rate, in intercept, or in both parameters, the associated functions will display disproportional dynamics, reaching a given proportion of their asymptote at different times. This is illustrated by the intersection of the horizontal and vertical lines in Fig. 1, which show the point in time (abscissa) when the functions reach two-thirds of their respective asymptote (ordinate). When processing speed is identical, as in (A), the functions reach this point at the same time, shown by the vertical line. When processing speed varies, as in (B), the functions reach a given proportion of their respective asymptotes at different times.

Experiment 1

Distance effects in ellipsis have been found in reading time. Murphy (1985) varied the distance between the ellipsis and its antecedent, along with the length of the antecedent and the syntactic parallelism of the antecedent in both surface and deep anaphors (VP ellipsis versus “do it” anaphora; see Sag, 1976; Shapiro, Hestvik, Lesan, & Garcia, 2003). A longer distance between the antecedent and ellipsis slowed reading times, as did increasing the length of the antecedent. Murphy suggested that these distance effects could indicate that a search process is used to access the antecedent in memory, but he emphasized that this type of operation may not be used for all types of anaphora. He suggested that when antecedents are close, surface features of text affect their interpretation; when antecedents are distant, processes based on content and plausibility come into play (see also, Garnham, 2001).

We investigated whether antecedent representations are copied into the elision site in Experiments 3–5. In this experiment, as well as in Experiment 2, we sought to determine whether distance effects reflect the time it takes to access an antecedent in memory. As noted, if comprehenders need to search for an antecedent, then access time should vary with the recency of the antecedent. However, distance may also reflect the quality of the antecedent's representation in memory. As distance increases, the antecedent's representation may decay, or the processing of interpolated material may interfere with its storage or retrieval. Reading time effects can rise from either differences in access time or differences in the quality of the representation that is accessed. The SAT procedure provides data that can discriminate between these accounts.

If distance simply reduces the quality of the antecedent's representation in memory, then it should affect the asymptote of the SAT function (McElree, 2000; McElree et al., 2003). However, if a search process is required to access an antecedent in memory, then increasing the distance between the antecedent and the elision site should also slow the overall interpretation of the ellipses, delaying the intercept (δ) of the SAT function or reducing the rate (β) of approach to asymptote. McElree and Doshier (1989, 1993) presented simulations of the impact of a serial search on SAT intercept and rate for a related manipulation of memory set in a probe recognition task, e.g., Sternberg (1975), and McElree (1993) and McElree and Carrasco (1999) presented related simulations of serial processing in two other domains.

Methods

Participants

Twenty-two native speakers of American English from the New York University community were paid

to participate in the study. They participated in four 1-h sessions, and a 45-min practice session for familiarization with the SAT procedure. All participants were between the ages of 18 and 26.

Materials

Thirty-six sets of eight sentences of the form illustrated in Table 1 were created. The full set of experimental materials is available from the JML online archive. The main contrasts concerned VP ellipsis with a short distance between the antecedent and the ellipsis site, such as (1a), and ellipses with a longer distance between the antecedent and the ellipsis site, such as (2a). Distance was increased by placing the ellipsis site within a complete clause containing passive voice VP (e.g., *everyone at the publishing house was shocked to hear that...*), consisting of 8–11 words, matched across conditions. For each of these conditions, we created a matching unacceptable condition, (1b) and (2b), by replacing the animate subject of the VP ellipsis (e.g., *the critic*) with an inanimate subject (e.g., *the binding*), which would create a pragmatically implausible interpretation when interpreted elliptically (e.g., *the binding admired the author's writing*). These unacceptable conditions were designed to encourage participants to fully process the ellipsis. We reasoned that, to discriminate acceptable from unacceptable sentences, participants would have to process the ellipses at least to the point where they had retrieved the antecedent and interpreted it in the local context.

Additionally, we included an equal number of acceptable and unacceptable, near and distant control conditions, without ellipsis in the final phrase, such as the (a) and (b) versions of (3) and (4). These sentences had the same lexical content as (1) and (2), except that a final

word was added to the final clause to block an elliptical interpretation. These sentences were included to reduce any tendency for participants to anticipate the occurrence of an ellipsis from the initial form of the sentence, as well as to block participants from predicting the acceptability of a sentence based on the animacy of the subject of the second clause. In these sentences, inanimate subjects in the second clause should be associated with a positive response, whereas inanimate subjects should be associated with a negative response, exactly opposite the pattern in (1) and (2).

In each of the four sessions, a participant read 72 experimental sentences, two conditions per item, counterbalanced within and across sessions. Therefore, participants saw every item in every condition, but at different points in the experiment. Conditions were counter-balanced across sessions such that participants saw an equal number of each condition in each session, though the item used to represent that condition varied. In order to vary which item was used to represent a given condition in a session systematically, two conditions within an item were yoked together and presented in the same session. These pairs were then shuffled through the 36 items. Conditions 1a and 2b of a given item appeared together in the same session, as did conditions 1b and 3a of the same item, conditions 2a and 4b of the same item, and 3b and 4a of the same item. Critical trials, including unelided controls, constituted 25% of each session, and were presented randomly among the remaining 75%, none of which was elided. The fillers were multi-clause sentences, with equal numbers of acceptable and unacceptable (underlined) versions: *The dancer who had wondered if the performer was entertaining heard that the director of the school smiled/hatched.*

Table 1
Example materials used in Experiment 1

Near antecedent, ellipsis

- 1a. The editor/ admired the author's writing,/ but the critics/ did not.
1b. *The editor/ admired the author's writing,/ but the binding/ did not.

Distant antecedent, ellipsis

- 2a. The editor/ admired the author's writing,/ but everyone/ at the publishing house/ was shocked to hear that/the critics/ did not.
2b. *The editor/ admired the author's writing,/ but everyone/ at the publishing house/ was shocked to hear that/the binding/ did not.

Near control (no ellipsis)

- 3a. The editor/ admired the author's writing,/ but the binding/ did not last.
3b. *The editor/ admired the author's writing,/ but the critics/ did not rip.

Distant control (no ellipsis)

- 4a. The editor/ admired the author's writing,/ but everyone/ at the publishing house/ was shocked to hear that/the binding/ did not last.
4b. *The editor/ admired the author's writing,/ but everyone/ at the publishing house/ was shocked to hear that/ the critics/ did not rip.

* Denotes an unacceptable sentence; /, denote phrase breaks in the phrase-by-phrase presentation method.

Procedure

Stimulus presentation, timing, and response collection were all carried out on a personal computer using software with millisecond timing. A trial began with a 500-ms fixation point presented at the center of the screen. Sentences were presented in a phrase-by-phrase controlled presentation manner, 335 ms per number of words in the phrase. A 50-ms, 1000 Hz tone served as the response signal. The first response signal occurred 300 ms before the onset of the sentence final phrase (which included the elliptical phrase in the experimental conditions). After the onset of the final phrase, 17 more response signals occurred, 350 ms apart, while the final phrase remained on the screen. The response signals continued until 6 s after the onset of the final phrase, for a total of 18 response signals. Participants were trained to respond to the tone. At each tone, participants were instructed to synchronize their responses to the tones, responding within 200 ms of each tone. They were instructed to simultaneously press both the “yes” and “no” keys as an initial (undecided) response, and then to switch to one key of the two keys when information on the acceptability of the sentence became available. They were also encouraged to modulate their responses if their judgment changed during the trial.

Participants first completed a 45-min practice session in order to familiarize themselves with the task. They were trained on pressing and switching responses rhythmically across the sampling period to ensure that they were practiced at modulating their responses, and until they became comfortable with the response requirements and could make a response within 200 ms. The experimental sessions consisted of four 1-h sessions on subsequent days. Between-trial intervals were participant controlled, and there were two mandatory breaks each session.

Data analysis

Comprehension accuracy was calculated using a standard d' measure, $d' = z(\text{hits}) - z(\text{false alarms})$, where a “hit” was an “acceptable” response to an acceptable sentence and a “false alarm” was an “acceptable” response to an unacceptable sentence. The d' scores provide a measure of the participant’s ability to discriminate acceptable from unacceptable structures, uncontaminated by response biases.

A hierarchical model-testing scheme was used to determine whether conditions differed in asymptote (λ), rate (β), or intercept (δ) in Eq. (1). Exponential model fits of the data ranged from a null model in which all functions were fit with a single asymptote, rate, and intercept parameter (a $1\lambda-1\beta-1\delta$ fit) to a fully saturated (a $2\lambda-2\beta-2\delta$ fit) model in which each condition was fit with a unique asymptote, rate, and intercept. For each participant and the averaged data, separate parameters

were allotted to the different conditions if they systematically improved the fit of the SAT function to the observed d' data. The exponential function in Eq. (1) was fit to the data with an iterative hill-climbing algorithm (Reed, 1976), which minimized the squared deviations of predicted values from observed data. Fit quality was assessed by an adjusted- R^2 statistic—the proportion of variance accounted for by the fit, adjusted by the number of free parameters (Judd & McClelland, 1989)—and by an evaluation of the consistency of the parameter patterns across the individual participant fits. Additionally, we performed inferential tests of significance computed over individual participants’ d' data, and the fitted parameter estimates for each of the candidate models detailed in the Results section. We report 95% confidence intervals (CIs) around the mean difference for paired comparisons of interest.

Results and discussion

Fig. 2 presents the average (across participants) d' values as a function of processing time, along with the best-fitting exponential model described below. Inspection of Fig. 2 suggests that distant antecedents were less accurately processed than near antecedents. As an initial

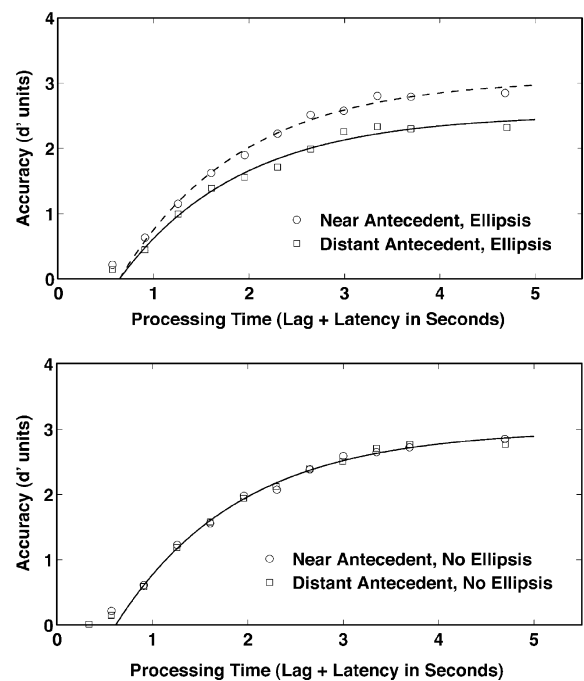


Fig. 2. Average d' accuracy (symbols) as a function of processing time (lag of the interruption cue plus latency to response) for Near and Distant Elided conditions (top) and Near and Distant Unelided conditions (bottom) from Experiment 1. Smooth curves show the best-fitting exponential fit (see text).

means of determining whether there were reliable differences in asymptotic performance as a function of distance, we averaged the d' values for each subject (and, for an item analysis, by each item) in each condition from 3.5 to 6 s post-initial response cue in order to derive an empirical estimate of asymptotic accuracy. Responses to elided sentences with near antecedents were on average .47 d' units higher in accuracy than responses to elided sentences with distant antecedents (95% CI = .27–.67 d' units). A paired t -test¹ on these values showed that this difference in asymptotic accuracy was significant, $F_1(1,21) = 23.67$, $p < .001$ and $F_2(1,35) = .06$, $p = .81$; $\min F'(1,35) = .05$, $p = .81$. However, this was not the case for unelided sentences. Accuracy for near unelided sentences was on average .03 d' units lower than accuracy for distant unelided sentences (95% CI = -.21–.15 d' units), and this difference was not significant.

Competitive fits of the exponential equation also yielded clear evidence that distance modulated asymptotic performance: models that did not allocate separate asymptotes for near versus distant ellipsis sentences produced poor fits to the empirical SAT data, and they left systematic residuals. In fits of the average data, allocating separate asymptotes to each ellipsis condition increased the adjusted- R^2 from .951 observed with a null $1\lambda-1\beta-1\delta$ model to .991. This $2\lambda-1\beta-1\delta$ model improved the quality of the fits of the individual participants' data, systematically increasing the adjusted- R^2 values over what was observed with a $1\lambda-1\beta-1\delta$ model (ranging from .812 to .982 as compared to .627–.971). In the average data, the asymptote for sentences with near antecedents was estimated to be 3.06, while the estimate for the sentences with distant antecedents was 2.52. Across participants, the average difference in asymptotic (λ) estimates was 0.57 d' units (95% CI = .37–.76 d' units), which was significant, $F(1,21) = 36.6$, $p < .001$.

The differences in asymptote indicate that distant antecedents were less accurately retrieved than near antecedents, or that the quality of the retrieved information was poorer for distant antecedents, leading to a less acceptable interpretation. If distance also affected the speed of processing the ellipsis, then it should have engendered differences in either rate (β) or intercept (δ). Crucially, however, allocating separate rate or intercept parameters to conditions with near and distance antecedents did not improve adjusted- R^2 . In fits of average data, a $2\lambda-2\beta-1\delta$ model resulted in an adjusted- R^2 of .989 and a $2\lambda-1\beta-2\delta$ model resulted in an adjusted- R^2 of .990, as compared to the .991 values observed with the simpler $2\lambda-1\beta-1\delta$ model. Importantly, there were no consistent trends across subjects in either the rate or

intercept parameters when they were allowed to vary, and t -tests on the parameter estimates were not significant. Hence, there was no evidence to suggest that distance affected processing speed, and therefore, no evidence that distant antecedents were retrieved more slowly than near ones.

For completeness, we also compared the functions for control conditions without ellipses in the final region. As inspection of the lower panel in Fig. 2 suggests, there were no differences evident in the control conditions. Consequently, the best fit for these functions was a simple $1\lambda-1\beta-1\delta$ model, adjusted- $R^2 = 0.994$. All t -tests on the parameter estimates for models that varied one of the SAT parameters were not significant. This suggests that the distance effect evident in the ellipsis conditions is related to the availability of the antecedent, not due to general differences between the short and long sentence forms.

The time-course profiles are identical to what has been found for the processing of other nonadjacent dependencies (McElree, 2000; McElree et al., 2003): Distance affects the likelihood that an appropriate antecedent can be recovered from memory, thereby lowering asymptotic accuracy, but it does not affect the speed with which an antecedent representation can be accessed. Because no differences were found in processing time for distant and near antecedents, it suggests that a search process was not used to access the antecedent for the VP ellipsis. This pattern is consistent with a content-addressable process, which enables representations of differing quality to be recovered with similar dynamics (McElree, 2006).

Experiment 2

Our SAT findings suggest that distance effects on reading time, such as the whole-sentence reading time differences reported in Murphy (1985), might reflect the quality of the antecedent representation in memory, not the time it takes to search for an antecedent. Specifically, as distance increases, the availability of the antecedent representation in memory may decrease, either because representations have had more opportunity to decay, or because the processing of interpolated material interferes with the storage or retrieval of the antecedent.

Experiment 2 examined eye-movement patterns during the reading of variants of the (acceptable) materials used in Experiment 1. There were two primary purposes. First, we wished to verify that our materials show reading time effects comparable to what was observed in Murphy's study. Second, we wished to explore how the observed speed-accuracy tradeoff differences are expressed in more natural reading situations, and to determine how our time-course findings align with more conventional eye-tracking markers of difficulty in sentence processing.

¹ In order to calculate $\min F'$ values for the contrasts, we computed the F -statistic as the square of the t -statistic.

Different properties of eye movements during reading have been argued to reflect when different operations or particular constraints are operative during comprehension (see Liversedge, Paterson, & Pickering, 1998; Pickering, Frisson, McElree, & Traxler, 2004; and Rayner, 1998 for reviews). Often, a distinction is drawn between effects that occur on early measures, such as how long a region is first fixated, and effects that emerge on later measures, such as how many times or for how long a region is refixated (see Pickering et al., 2004). In a related domain, for example, Garrod and colleagues have used such a distinction as a basis for proposing a two-stage model of anaphoric processing (Garrod & Sanford, 1990; Garrod & Terras, 2000; Sanford, Garrod, Lucas, & Henderson, 1983). They suggest that the first process in anaphoric resolution is a bonding stage, which is argued to be an automatic and low-level process, driven by lexical information, that establishes a superficial link between a pronoun (or other anaphoric device) and its referent. Difficulty at this stage is thought to be reflected in early eye-tracking measures. The second stage, which they label resolution, is largely concerned with integrating the bond with the discourse context and the overall discourse model. Difficulty in integration is thought to be reflected in later eye movement measures.

In a recent study of the effects of prominence on pronoun resolution, Foraker and McElree (2007) found that early and late effects in eye movements during reading correlated with SAT differences in dynamics (rate and intercept) and asymptote (respectively). The prominence of an antecedent for a pronoun (whether the antecedent was syntactically clefted or not) affected the asymptotes of the SAT time-course functions, with less prominent antecedents resulting in lower asymptotes. Foraker and McElree argued that prominence affects the quality of the antecedent representation in memory, not the speed with which it can be accessed. In an eye-tracking study, prominence did not affect early measures, but instead several late eye-movement measures. Following the logic of Garrod and colleagues (Sanford et al., 1983; Garrod & Sanford, 1990; Garrod & Terras, 2000), this may suggest that prominence affects late stages of integration, not the initial stage where the antecedent is bound to the pronoun. Difficulty at the integration stage follows naturally from the assumption that prominence affects the quality of the antecedent's representation, as a poorer quality representation may be more difficult to integrate into the discourse context than a higher quality or richer representation.

In contrast to the effects of prominence, Foraker and McElree found that the inherent ambiguity of a pronoun (whether it was gendered or not) affected the dynamics of the time-course functions, with more ambiguous pronouns slowing the rate of approach to asymptote. In an eye-tracking study, pronoun ambiguity affected early

measures; specifically, more ambiguous pronouns caused more first-pass regressions (the probability of leaving a region on the saccade following a first-pass fixation in order to regress to earlier parts of the sentence; see below). Measures from both procedures converge in suggesting that pronoun ambiguity affects the speed of accessing an antecedent to form an initial bond with the pronoun.

Although ample evidence suggests that asymptotic and dynamics effects in SAT reflect distinct aspects of processing,² we do not suppose that all SAT dynamics effects will correspond to effects in early eye-tracking measures and that all SAT asymptotic effects will correspond to effects in late eye-tracking measures. The mapping between operations and measures is many-to-one and too indirect to expect perfect correspondence in all cases. Nonetheless, interpreting pronouns and ellipses require similar operations of accessing and integrating antecedent representations. Hence, it is reasonable to expect the same type of relationships in ellipses as observed in Foraker and McElree (2007).

Specifically, if distance affects the quality of the antecedent representation, then it may engender differences in late eye-tracking measures, which are thought to reflect difficulty in integration. If longer distances decrease the quality of the antecedent representation, then we might expect longer second-pass and total times on the antecedent in ellipses with distant as compared to near antecedents. This prediction follows from the intuitive idea that subjects may often need to reprocess a distant antecedent because the antecedent's representation in memory is too poor to support interpretation when initially encountering the ellipsis site.

Conversely, if a search process were required to find an antecedent in memory, then increased distance should engender elevated reading times at the point when the reader first encounters the cues that signal that retrieval of an antecedent is required, viz., at the ellipsis site or in the adjacent spillover region. Of course, it is possible that these first-pass effects at the ellipsis site could co-occur with second-pass and total times on the antecedent, if distant antecedents also affect the quality of the antecedent's representation in memory. However, effects on second-pass and total times on the antecedent without concomitant first-pass effects at the ellipsis site are inconsistent with a search model.

² Analogous to the dissociation between eye-tracking and SAT measures in Foraker and McElree (2007), Bornkessel et al. (2004) found in complementary SAT and event-related potential (ERP) studies that asymptotic effects in SAT were associated with modulations of the amplitude of a single ERP component (P600), whereas SAT dynamics (rate) differences were associated with the presence or absence of an ERP component (N400).

Methods

Participants

Forty native speakers of American English from the New York University community were paid to participate in the study. They were between the ages of 18–26 and all had normal or corrected-to-normal vision. They participated in one 60-min session.

Materials

The materials were variants of the 28 items used in Experiment 1, illustrated in Table 2. A spillover region was added to each sentence, such as *even though the first book won an award* or *due to the controversial nature of the research*. The unacceptable conditions necessary for Experiment 1 were not included in Experiment 2 due to the change in method. The control conditions were modified to control the animacy of the embedded subject, the source of the acceptability manipulation in Experiment 1. Fillers constituted 87.5% of the stimuli presented in the session and were acceptable multi-clause sentences like the fillers used in Experiment 1.

Procedure

We monitored the movements of the participants' right eyes during reading 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 font, with each letter 18 pixels wide and 33 pixels high. No more than 80 characters were presented on one line of text. 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. Yes/no-comprehension questions were asked for 50% of the items. After determining fixation, an automatic procedure pooled short contiguous fixations. This procedure combined all fixations shorter than 80 ms and within one character of each other. Fixations that were shorter than 80 ms but not contiguous were excluded from the analyses, because

presumably little information is extracted (Rayner & Pollatsek, 1989). Fixations longer than 1000 ms were excluded as well.

Reading times were analyzed by region. We defined our regions of interest in the following way, using as an example the sentence, *The editor admired the author's writing, but the critics did not, even though his first book won an award*: (1) antecedent verb region (*admired*), (2) complement (or object) of the antecedent region (*the author's writing*), (3) pre-ellipsis region (*but the critics...*), (4) ellipsis region (*did not*), (5) disambiguation region (only for nonelided conditions), and (6) spillover region (*even though*). The following eye-movement based dependent measures were used: (1) *first-pass reading time*, or the time spent fixating a region initially, excluding re-fixations, (2) *first-pass regressions out*, or the proportion of trials in which at least one regression out of a region occurred during first reading of the region, (3) *second-pass reading time*, or the time spent in that region only after the eyes have moved out of that region to the right, (4) *total time*, or the sum of all fixations within a region, and (5) *regression path*, or the sum of all fixations in a region from first entering that region until moving to the right of that region. Measures (3) and (4) are sensitive to later aspects of processing.

Results and discussion

All participants scored well on comprehension questions, with the average percent correct being 92%. We excluded trials on which track loss occurred (~1%). Reading times over three standard deviations from the mean (subject × condition mean) were excluded from the analysis. The reading time measures at the antecedent regions were analyzed with a series of repeated measures 2 (Distance: Near or Distant) × 2 (Ellipsis: Elided or Nonelided) ANOVA, treating both subjects (F_1) and sentences (F_2) as a random-factor. These ANOVAs were followed up by pairwise comparisons to locate the source of effects. Table 3a reports these F -values, along with $\min F'$ values of the antecedent regions. In the text,

Table 2

Example materials used in Experiment 2

Near antecedent, ellipsis

1. The editor admired the author's writing, but the critics did not, even though his first book won an award.

Distant antecedent, ellipsis

2. The editor admired the author's writing, but everyone at the publishing house was shocked to hear that the critics did not, even though his first book won an award.

Near control (no ellipsis)

3. The editor admired the author's writing, but the book did not do well, even though his first book won an award.

Distant control (no ellipsis)

4. The editor admired the author's writing, but everyone at the publishing house was shocked to hear that the book did not do well, even though his first book won an award.

Table 3a
 Statistical analysis of the eye movement patterns in Experiment 2 (Distance × Ellipsis analysis of variance on antecedent regions)

Measure	Pair	Mean diff. (Std. error)	Source	df	F ₁	df	F ₂	df	minF'		
<i>Verb</i>											
First-pass	Near vs. Distant, Ellipsis	7 (11.18)	Distance	1	.163 (2225.53)	1	.001 (1381.11)	1, 27	0		
			Ellipsis	1	.109 (2893.96)	1	.398 (703.16)	1, 57	.08		
			Distance × Ellipsis	1	.291 (2712.54)	1	.083 (3014.14)	1, 42	.06		
Second-pass	Near vs. Distant, Unelided	−1 (10.36)	Within-subject error	39		27					
			Near vs. Distant, Ellipsis	60 (22.58)	Distance	1	15.642 (7553)**	1	10.64 (6789.615)*	1, 58	6.33*
					Ellipsis	1	.501 (6028)	1	.48 (5157.853)	1, 63	.24
					Distance × Ellipsis	1	.274 (5245.76)	1	.202 (3661.182)	1, 59	.11
			Within-subject error	39		27					
Total time	Near vs. Distant, Unelided	48 (11.4)	Distance	1	12.692 (10367.99)**	1	8.203 (8850.66)*	1, 57	4.98*		
			Ellipsis	1	.163 (8506.87)	1	.19 (5759.86)	1, 65	.08		
			Distance × Ellipsis	1	4.04 (10791.12)	1	.238 (7806.88)	1, 30	.22		
			Within-subject error	39		27					
			Near vs. Distant, Ellipsis	68 (29.34)	Distance	1	.116 (2565.24)	1	.015 (7884.91)	1, 34	.01
Ellipsis	1	1.165 (4205.51)			1	.008 (8359.92)	1, 27	0			
Distance × Ellipsis	1	.564 (2087.76)			1	.128 (3675.61)	1, 39	.1			
<i>Complement</i>											
First-pass	Near vs. Distant, Ellipsis	8 (9.85)	Distance	1	.116 (2565.24)	1	.015 (7884.91)	1, 34	.01		
			Ellipsis	1	1.165 (4205.51)	1	.008 (8359.92)	1, 27	0		
			Distance × Ellipsis	1	.564 (2087.76)	1	.128 (3675.61)	1, 39	.1		
Second-pass	Near vs. Distant, Unelided	−2 (11.65)	Within-subject error	39		27					
			Near vs. Distant, Ellipsis	106 (35.66)	Distance	1	8.197 (19545.13)**	1	12.412 (7977.22)**	1, 66	4.93*
					Ellipsis	1	3.176 (8913.46)	1	2.512 (6425.87)	1, 60	2.51
					Distance × Ellipsis	1	5.553 (13545.86)*	1	4.53 (9234.23)*	1, 61	2.49
			Within-subject error	39		27					
Total time	Near vs. Distant, Unelided	20 (19.57)	Distance	1	8.01 (21764.44)*	1	6.356 (16668.35)*	1, 60	3.54†		
			Ellipsis	1	3.75 (15137.06)†	1	1.193 (11795.84)	1, 44	.9		
			Distance × Ellipsis	1	5.437 (17514.84)*	1	3.83 (13362.2)†	1, 58	2.24		
			Within-subject error	39		27					
			Near vs. Distant, Ellipsis	115 (36.95)	Distance	1	.116 (2565.24)	1	.015 (7884.91)	1, 34	.01
Ellipsis	1	1.165 (4205.51)			1	.008 (8359.92)	1, 27	0			
Distance × Ellipsis	1	.564 (2087.76)			1	.128 (3675.61)	1, 39	.1			
<i>Complement</i>											
First-pass	Near vs. Distant, Ellipsis	8 (9.85)	Distance	1	.116 (2565.24)	1	.015 (7884.91)	1, 34	.01		
			Ellipsis	1	1.165 (4205.51)	1	.008 (8359.92)	1, 27	0		
			Distance × Ellipsis	1	.564 (2087.76)	1	.128 (3675.61)	1, 39	.1		
Second-pass	Near vs. Distant, Unelided	−2 (11.65)	Within-subject error	39		27					
			Near vs. Distant, Ellipsis	106 (35.66)	Distance	1	8.197 (19545.13)**	1	12.412 (7977.22)**	1, 66	4.93*
					Ellipsis	1	3.176 (8913.46)	1	2.512 (6425.87)	1, 60	2.51
					Distance × Ellipsis	1	5.553 (13545.86)*	1	4.53 (9234.23)*	1, 61	2.49
			Within-subject error	39		27					
Total time	Near vs. Distant, Unelided	20 (19.57)	Distance	1	8.01 (21764.44)*	1	6.356 (16668.35)*	1, 60	3.54†		
			Ellipsis	1	3.75 (15137.06)†	1	1.193 (11795.84)	1, 44	.9		
			Distance × Ellipsis	1	5.437 (17514.84)*	1	3.83 (13362.2)†	1, 58	2.24		
			Within-subject error	39		27					
			Near vs. Distant, Ellipsis	115 (36.95)	Distance	1	.116 (2565.24)	1	.015 (7884.91)	1, 34	.01
Ellipsis	1	1.165 (4205.51)			1	.008 (8359.92)	1, 27	0			
Distance × Ellipsis	1	.564 (2087.76)			1	.128 (3675.61)	1, 39	.1			

* $p < .05$.

** $p < .001$.

† Marginal, $p < .07$.

we report 95% confidence intervals calculated using the interaction error term from the ANOVA (Masson & Loftus, 2003). The reading time measures at the succeeding regions (pre-ellipsis, ellipsis, and spillover) where lexical content is not directly comparable between elided and unelided conditions were analyzed with a series of paired *t*-tests (Near versus Distant Antecedent conditions, using the square of the *t*-statistic to calculate $\min F'$). The means are reported in Table 3b. In order to analyze regression patterns, we performed a repeated measures ANOVA on regression path duration in all regions. Table 3c reports the *F*-values, along with $\min F'$ values, of interest for this measure.

Pre-ellipsis region, ellipsis site, and spillover

For the Near and Distant conditions, the pre-ellipsis regions differed in length (e.g., *...but the critic* versus *...but everyone at the publishing house was shocked to hear that the critics*). Not surprisingly, reading times in the Near condition were significantly shorter than the Distant condition in all measures. Of critical concern are differences at the ellipsis site and spillover region. No significant differences were found in these regions in first-pass, first-pass regressions out, second-pass, or total time measures.

Antecedent: Verb and complement regions

There were no significant first-pass effects found in either the verb region of the antecedent or the following complement region. However, differences emerged in

both regions as a function of the distance manipulation in later reading measures.

In second-pass reading times, there was a main effect of distance at the verb region (see Table 3a). Pairwise comparisons revealed that both Elided and Unelided conditions were responsible for this effect: Second-pass times in the Near Elided condition were on average 60 ms shorter than the Distant Elided condition (136 vs. 196, 95% CI = 37–83 ms), and times for the Near Unelided condition were 48 ms shorter than the Distant Unelided condition (134 vs. 182, 95% CI = 25–71 ms). In the complement region, there was a significant interaction in second-pass time between distance and ellipsis, coupled with a significant main effect of distance (see Table 3a). Pairwise comparisons revealed that these effects were driven only by the Elided conditions. Times for the Near Elided condition were 106 ms shorter than times for the Distant Elided condition (119 vs. 225, 95% CI = 69–143 ms). The Near Unelided condition had a nonsignificant 20 ms advantage over the Distant Unelided condition (136 vs. 156, 95% CI = –17–57 ms). Main effects of distance simply show that readers spent more time re-reading the verb and complement regions of the antecedent in long as compared to short sentences. However, the significant interaction in the complement regions shows that this effect is partly driven by processes involved with interpreting the ellipsis.

The same pattern was evident in total time measures. There was a main effect of distance at the verb region (see Table 3a). Pairwise comparisons revealed that the

Table 3b
Mean and standard error of the mean for all regions in Experiment 2 (Distance \times Ellipsis)

Measure	Condition	Mean (Std. error)				
		Verb	Complement	Pre-ellipsis	Ellipsis	Spillover
First-pass (ms)	Near, elided	275 (8.86)	274 (12.56)	423 (20.59)	274 (12.55)	323 (16.05)
	Distant, elided	282 (12.6)	282 (14.26)	835 (50.58)	282 (14.25)	330 (19.7)
	Near, unelided	282 (14.56)	268 (12.97)	436 (22.82)	268 (12.97)	318 (16.56)
	Distant, unelided	281 (11.56)	266 (12.44)	798 (44.95)	266 (12.44)	340 (22.71)
Second-pass (ms)	Near, elided	136 (12.7)	119 (13.68)	151 (16.21)	64 (9.69)	75 (12.49)
	Distant, elided	196 (13.92)	225 (38.37)	290 (39.44)	85 (12.64)	94 (13.83)
	Near, unelided	134 (25.08)	136 (23.13)	132 (18.48)	81 (12.49)	66 (10.46)
	Distant, unelided	182 (19.83)	156 (22.57)	206 (31.33)	73 (13.49)	63 (8.64)
Total time (ms)	Near, elided	411 (15)	393 (17.46)	573 (24.12)	338 (15.12)	398 (22.35)
	Distant, elided	479 (31.25)	508 (43.65)	1125 (71.91)	367 (20.35)	424 (22.57)
	Near, unelided	416 (23.8)	404 (17.46)	567 (30.2)	349 (19.81)	384 (22.75)
	Distant, unelided	464 (24.09)	421 (27.54)	1003 (56.1)	338 (17.77)	403 (25.64)
First-pass regressions-out (proportion)	Near, elided	12 (2)	20 (3)	4 (1)	12 (3)	5 (1)
	Distant, elided	10 (2)	21 (2)	5 (1)	11 (2)	9 (2)
	Near, unelided	12 (2)	19 (3)	3 (1)	6 (2)	10 (1)
	Distant, unelided	8 (1)	21 (3)	5 (1)	9 (2)	9 (2)
Regression-path duration (ms)	Near, elided	345 (14.77)	638 (32.39)	465 (30.29)	359 (25)	361 (20.68)
	Distant, elided	348 (19.77)	661 (42.68)	919 (58.99)	383 (38.2)	507 (57.08)
	Near, unelided	342 (17.9)	641 (37.65)	455 (23.31)	306 (16.43)	445 (37.71)
	Distant, unelided	350 (24.92)	641 (38.88)	860 (51.85)	395 (57.89)	418 (25.42)

Table 3c
Statistical Analysis of Regression-path duration in Experiment 2

Experiment 2, Distance × Ellipsis		Analysis of variance at the Ellipsis site and Spillover region		Regression-path duration (cumulative)		Pair		Mean Diff. (Std. Error)	Source	df	F ₁ (MSe)	df	F ₂ (MSe)	df	minF ²
<i>Ellipsis site</i>															
Near vs. Distant, Ellipsis	Distance	1	2.23(566.19.51)	1	.974(40943.84)	1,49	.68	24(44.64)	Distance	1	2.23(566.19.51)	1	.974(40943.84)	1,49	.68
									Ellipsis	1	.425(38591.7)	1	.512(27808.08)	1,65	.23
									Distance × Ellipsis	1	.839(50026.5)	1	.854(32265.99)	1,64	.42
									Within-subject error	39		27			
Near vs. Distant, Unelided	Distance	1	2.78(50674.53)	1	3.40(27383.64)	1,66	1.52	146(58.76)	Distance	1	2.78(50674.53)	1	3.40(27383.64)	1,66	1.52
									Ellipsis	1	.002(58479.9)	1	.001(23753.79)	1,52	0
									Distance × Ellipsis	1	7.14(41984.49)*	1	5.06(27065.65)*	1,58	2.96
									Within-subject error	39		27			
<i>Spillover region</i>															
Near vs. Distant, Unelided	Distance	1	2.78(50674.53)	1	3.40(27383.64)	1,66	1.52	-27(34.34)	Distance	1	2.78(50674.53)	1	3.40(27383.64)	1,66	1.52
									Ellipsis	1	.002(58479.9)	1	.001(23753.79)	1,52	0
									Distance × Ellipsis	1	7.14(41984.49)*	1	5.06(27065.65)*	1,58	2.96
									Within-subject error	39		27			

* Denotes $p < .05$

Elided and Unelided conditions differed significantly from each other. Overall, the verb region was fixated 68 ms longer in the Distant Elided condition than the Near Elided condition (411 vs. 479, 95% CI = 35–101 ms). In the Unelided conditions, the corresponding difference was 48 ms (416 vs. 464, 95% CI = 15–81 ms). At the complement region, there was a significant interaction between distance and ellipsis, as well as a main effect of distance. Pairwise comparisons revealed that only Elided conditions differed significantly from one another. The complement region was fixated 115 ms longer in the Distant Elided condition than the Near Elided condition (393 vs. 508, 95% CI = 72–156 ms). In the Unelided conditions, the corresponding difference was 17 ms (404 vs. 421, 95% CI = -25–59 ms), a nonsignificant difference.

Regression path duration analysis

In regression path times, there was a main effect of distance in the pre-ellipsis region due to the distance manipulation. Otherwise, there were no significant effects in any region, including the ellipsis site, until the spillover region. In the spillover region, there was a reliable interaction between Distance and Ellipsis (see Table 3c). Pairwise comparisons showed that in the Near Elided condition, the spillover region was read 146 ms faster than in the Distant Elided condition (361 vs. 507, 95% CI = 80–212 ms). This interaction suggests that the effects found in later measures (i.e., second-pass and total time) are due to regressions launched after first processing the ellipsis site.

Summary

Distance had no reliable effects on reading times at the ellipsis site or the spillover region. This finding is inconsistent with any search process in which access time depends on the number of competing representations in memory. If readers had to search through representations of past material to find an antecedent, then, when more material intervened between the antecedent and the ellipsis, one would have expected elevated reading times on the ellipsis site or the spillover region, particularly in early measures, such as first-pass time. These reading time measures converge with the findings in Experiment 1 of equal SAT dynamics at the ellipsis site for Near and Distant conditions in suggesting that readers can access the antecedent for an ellipsis directly.

We found that readers spent more time rereading a distant antecedent than a near one. This finding suggests that the quality of the retrieved representation of the antecedent is reduced when comprehenders process more material before the ellipsis site. When the quality of the retrieved information is not sufficient to support interpretation, readers may have to regress back to the antecedent region to construct an acceptable interpretation. This finding aligns with the demonstrably lower

asymptotic performance for Distant as compared to Near conditions in Experiment 1. Both experiments indicate that distance only affects the quality of the antecedent, and it does not affect the time needed to access the representation of the antecedent. Experiments 1 and 2 are fully consistent with the hypothesis that an ellipsis is resolved with the same type of content-addressable process that has been argued to mediate the resolution of other types of nonadjacent dependencies (McElree, 2000; McElree et al., 2003).

Experiment 3

Experiments 1 and 2 demonstrate that comprehenders can access antecedents directly. Experiments 3–5 investigate how an accessed antecedent is interpreted at the ellipsis site. We test whether interpreting an ellipsis requires a representation of the antecedent to be copied into the ellipsis site, whether or not the copied representation contains detailed syntactic information.

The canonical interpretation of a literal copy mechanism is that copying more information should take more time. One could simply assert that “copying” does not require time, but we suggest that in that case, the notion “copy” is no longer explanatory. A real-time copy operation predicts that processing time should increase as the amount of material contained *within* the antecedent increases. Extant data on this issue are somewhat mixed. Murphy (1985) found that short antecedents, such as *Jimmy swept the floor*, had shorter reading times at the ellipsis site (*Later, his uncle did too*) than longer antecedents, such as *Jimmy swept the floor behind the chairs free of hair and cigarettes*. Whole sentence reading times for sentences with the VP ellipsis were 244 ms longer for long antecedents than for short ones. However, this effect disappeared when a sentence intervened between the sentence containing the antecedent and the sentence containing the VP ellipsis. Murphy suggested that these findings are consistent with the deployment of a copy operation when the surface form of the antecedent is still available in short-term memory, as when no additional sentence intervened between the antecedent and the ellipsis.

In contrast, Frazier and Clifton (2001) did not find a complexity effect in VP ellipses consisting of one-clause antecedents (e.g., *Sarah left her boyfriend last May. Tina did too*) and two-clause antecedents (e.g., *Sarah got up the courage to leave her boyfriend last May. Tina did too*). Whole-sentence reading times on the final elliptical sentences did not differ, despite the fact that the two-clause antecedents were lengthier than the one-clause antecedents. It is unclear why Frazier and Clifton did not find an effect comparable to what was reported by Murphy (1985), given that their VP ellipsis occurred immediately after the antecedent VP. It is possible that

Murphy’s materials simply had a larger difference in complexity.

However, even if some types of complexity do engender reliable reading time differences, these differences alone do not provide strong support for a copy mechanism. As we argued, reading time effects do not uniquely reflect differences in the underlying speed of processing, which is the key prediction of a copy mechanism. Rather, those effects can simply reflect the quality of the retrieved antecedent representation. More complex structures may be retrieved less accurately, or at a lower probability, which may make it more difficult for comprehenders to fully interpret VP ellipsis with complex antecedents.

To test the copy hypothesis and to contrast it with a pointer hypothesis, Experiments 3–5 examine whether two types of complexity manipulations affect the speed of interpreting VP ellipsis. Experiments 3 and 4 contrast antecedents containing a verb and a simple noun phrase (e.g., *The history professor [understood Roman mythology]. . .*) with lengthier antecedents containing a verb and complex noun phrase (e.g., *The history professor [understood Rome’s swift and brutal destruction of Carthage]. . .*). Like Murphy (1985), we increased the complexity of the verb’s complement rather than following Frazier and Clifton’s strategy of embedding a verb phrase within another verb phrase, which can introduce temporary ambiguities in interpretation. Because Murphy did not find complexity when an additional sentence intervened between the antecedent and the ellipsis, we placed the VP ellipsis in a second clause immediately following the antecedent [*. . . but the over-worked students attending summer session did not*].

Manipulating complexity by varying the length of an antecedent necessarily confounds complexity with other factors. If one simply increases antecedent length in one condition, as in the studies of Murphy (1985) and Frazier and Clifton (2001), then the overall length of the sentences prior to the ellipsis differs. This is problematic if one finds a nominal complexity effect, as Murphy did, because it may be due to an effect of general sentence length on interpretation, particularly when the ellipsis appears at the end of the sentence when substantial “wrap-up” effects occur (Just & Carpenter, 1980; Mitchell & Green, 1978; Rayner, Kambe, & Duffy, 2000). Alternatively, one can balance overall length by interpolating additional information between the antecedent and the ellipsis in the simple conditions. However, the distance from the end of the antecedents to the ellipsis then differs between conditions, with the end of the complex antecedent being closer to the ellipsis than the end of the simple antecedent. As distance favors complex conditions, this strategy runs a risk of masking any complexity effects.

In this experiment, we adopted the first strategy of controlling overall sentence length by interpolating addi-

tional material between the antecedents and ellipsis in the simple antecedent condition. The SAT procedure was used to contrast the full time-course of processing simple and complex antecedents. At issue was whether complexity affects the intercept or rate of the SAT time-course functions, which is the key prediction of a copy mechanism in which a longer antecedent slows interpretation time. Experiment 4 also used the SAT procedure, but adopted the strategy used in other studies of forgoing matching on overall sentence length in favor of controlling distance between the antecedent and the ellipsis.

Methods

Participants

Twenty-two native speakers of American English from the New York University community were paid to participate in the study. They participated in four 1-h sessions, and one 45-min practice session for familiarization with the SAT procedure. All participants were between the ages of 18 and 26.

Materials

Thirty-six sets of eight sentences of the form illustrated in Table 4 were created. The eight-sentence forms were modeled on the materials for Experiment 1, except that the length of the antecedent material was manipulated instead of the distance between the antecedent and the ellipsis. We manipulated complexity by replacing a simple noun phrase complement, such as *Roman mythology* in (1a) and (1b), with a complex one, such

as *Rome's swift and brutal destruction of Carthage* in (2a) and (2b). To offset the increase in length, we added material before the ellipsis in the simple condition [e.g., *but the principal was displeased to learn that the over-worked students...* in (1a) and (1b) versus *but the principal knew the over-worked students...* in (2a) and (2b)], the conditions were closely matched in overall length. Unacceptable versions were again created by replacing the animate subject of the VP ellipsis (e.g., *students*) with an inanimate subject (e.g., *books*). We again included an equal number of acceptable and unacceptable simple and complex control conditions, without an ellipsis in the final phrase.

In each of the four sessions, a participant read 72 experimental sentences, two conditions per item, counter-balanced within and across sessions. Therefore, participants saw every item in every condition, but at different points in the experiment. As with Experiment 1, conditions were counter-balanced across sessions such that participants saw an equal number of each condition in each session, though the item used to represent that condition varied. Again, two conditions within an item were yoked together and presented in the same session. These pairs were then shuffled through the 36 items. Conditions 1a and 2b of a given item appeared together in the same session, as did conditions 1b and 3a of the same item, conditions 2a and 4b of the same item, and 3b and 4a of the same item. Overall, the conditions in Table 4 made up only 25% of the items within a session, with the remaining 75% of the items consisting of fillers of the type used in Experiment 1.

Table 4
Example materials used in Experiment 3

Simple antecedent, ellipsis

- 1a. The history professor/ understood Roman mythology,/ but the principal/ was displeased to learn that/ the over-worked students/ attending summer session/ did not.
1b. *The history professor/ understood Roman mythology,/ but the principal/ was displeased to learn that/ the overly worn books/ used in summer session did not.

Complex antecedent, ellipsis

- 2a. The history professor/ understood Rome's swift and brutal/ destruction of Carthage/, but the principal knew the over-worked students/ attending summer session/ did not.
2b. *The history professor/ understood Rome's swift and brutal/ destruction of Carthage/, but the principal knew the overly worn books/ used in summer session/ did not.

Simple control (no ellipsis)

- 3a. The history professor/ understood Roman mythology,/ but the principal/ was displeased to learn that/ the over-worked students/ attending summer session/ looked mystified.
3b. *The history professor/ understood Roman mythology,/ but the principal/ was displeased to learn that/ the overly worn books/ used in summer session/ looked mystified.

Complex control (no ellipsis)

- 4a. The history professor/ understood Rome's swift and brutal/ destruction of Carthage/, but the principal knew the over-worked students/ attending summer session/ looked mystified.
4b. *The history professor/ understood Rome's swift and brutal/ destruction of Carthage/, but the principal knew the overly worn books/ used in summer session/ looked mystified.

* Denotes an unacceptable sentence; / denote phrase breaks in the phrase-by-phrase presentation method.

Procedure and design

The same procedure and analysis were used in Experiment 3 as in Experiment 1.

Results and discussion

Fig. 3 presents the average (across participants) d' values as a function of processing time, along with the fitted exponential model described below. Inspection of Fig. 3 suggests that there were very small differences between simple and complex ellipsis conditions. In terms of asymptotic accuracy, there appears to be a modest advantage for complex conditions over simple conditions. As an initial means of investigating whether these differences were reliable, we again averaged the empirical d' values from 3.5 to 6 s post-initial response cue for each subject in each condition to derive an empirical estimate of asymptotic accuracy. On average, simple elided sentences were .15 d' units lower in accuracy than complex elided sentences (95% CI = $-.39-.09$ d' units), while simple unelided sentences were .15 d' units greater in accuracy than complex unelided sentences (95% CI = $-.03-.35$ d' units). Paired t -tests on the empirical d' values by subjects showed no significant difference between conditions. Likewise, we averaged by items

the empirical d' values for 3.5–6 s post-initial response cue. On average, simple elided sentences were .01 d' units lower in accuracy than complex elided sentences (95% CI = $-.15-.13$ d' units), while simple unelided sentences were .03 d' units greater in accuracy than complex unelided sentences (95% CI = $-.09-.15$ d' units). Paired t -tests on these empirical d' values were also not significant.

Application of the exponential model to the full time-course data likewise indicated that there were small but unreliable differences between the Simple and Complex conditions. A model that allocated separate asymptotes for Simple and Complex conditions, a $2\lambda-1\beta-1\delta$ model, did not substantially improve the quality from a simple (null) $1\lambda-1\beta-1\delta$ model (adjusted- R^2 of .986 versus .987 in fits of the average data). Although there were small asymptotic differences between the Simple and Complex conditions, 2.33 versus 2.28, respectively, the differences in the λ estimates across participants were not significant. Models that varied either the rate (β) or intercept (δ), either with or without concomitant variations in asymptote, likewise did not produce better fits to the data and did not yield any systematic difference in the estimates of either dynamics parameter. Hence, a $1\lambda-1\beta-1\delta$ model provided the best description of the time-course profiles. This model was also found to give the best fit to the control conditions (adjusted- $R^2 = .989$). Fig. 3 presents the $2\lambda-1\beta-1\delta$ fit of the Elided condition to graphically illustrate the small trend toward better performance for the Complex condition.

In summary, there was no evidence that the complexity of the antecedent had a reliable effect on processing speed or asymptotic accuracy. This finding is inconsistent with a real-time copy mechanism, which predicts that the speed of interpreting ellipses should slow as the antecedent becomes more complex. The time-course profiles are consistent with the hypothesis that a pointer mechanism subserves the interpretation of VP ellipsis, as this type of mechanism predicts that both simple and complex antecedents can be interpreted with comparable speed.

We also conducted a study examining the eye-movements of a separate group of 39 subjects during the reading of the acceptable conditions in Table 4, using procedures analogous to Experiment 2. Contrary to the predictions of a copy mechanism, complex antecedents were not more costly to interpret than simple antecedents. In fact, several eye movement measures indicated that sentences with complex antecedents were read faster than sentences with simple antecedents. The advantage for complex antecedents was evident in first-pass, second-pass, and total times at the ellipsis site, and in total times on a 3-word spillover region.

As noted, one concern with the design of this experiment is that complex antecedents were closer to the ellipsis site than were simple antecedents. Because distance favors complex antecedents, any positive effect of

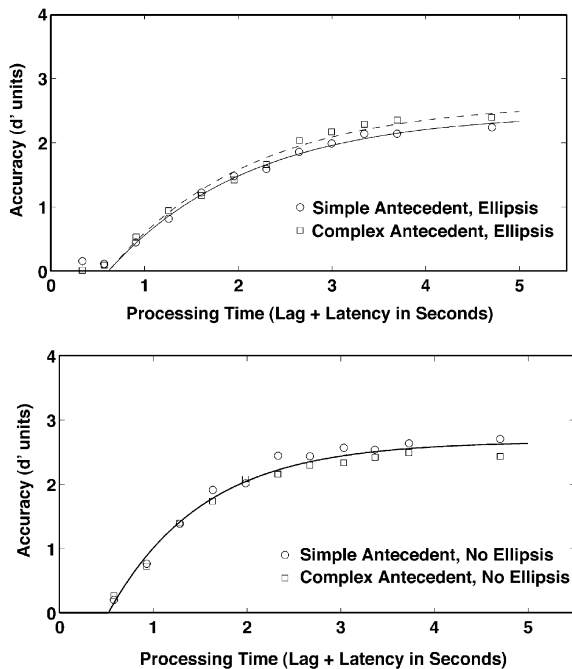


Fig. 3. Average d' accuracy (symbols) as a function of processing time (lag of the interruption cue plus latency to response) for Simple and Complex Elided conditions (top) and Simple and Complex Unelided conditions (bottom) from Experiment 3. Smooth curves show the best-fitting exponential fit (see text).

recency may have masked a complexity effect. The slightly higher asymptotes and the reliably faster reading time for the condition with complex as compared to simple antecedents is consistent with a distance effect, in line with the results of Experiments 1 and 2.

Experiment 4

Here, we once again used the SAT procedure to investigate antecedent complexity, but we adopted the alternative strategy of controlling the distance between the antecedent and the ellipsis. We used the materials of Experiment 3 but, to equate overall sentence length, we eliminated the additional material in the simple conditions that was placed between the antecedent and the ellipsis. Again, if complexity impacts on the processing of ellipsis, we should observe effects on either accuracy, reflected in the SAT asymptotes, or speed, reflected in the SAT intercepts or rates.

This manipulation also has the potential to address whether the advantage for complex antecedents observed in Experiment 3 and the comparable eye-tracking experiment was due to the shorter distance between the antecedent and ellipsis or because complex antecedents were more richly encoded or distinctive in memory. If the latter were the case, we would expect to see the advantage persist. If the former is the case, then the advantage should disappear once distance is equated.

Methods

Participants

Fifteen³ native speakers of American English from the New York University community were paid to participate in the study. They participated in four 75-min sessions, and one 45-min practice session for familiarization with the SAT procedure. All participants were between the ages of 18 and 26.

Materials

The materials were the same as what were used in Experiment 3, except that we eliminated the additional material in the simple conditions that served to equate overall sentence length. Table 5 presents examples of the conditions of interest. As in Experiments 1 and 3, during each of the four sessions, a participant read 72 experimental sentences, two conditions per item, counterbalanced within and across sessions. Therefore, par-

ticipants saw every item in every condition, but at different points in the experiment. Again, conditions were counter-balanced across sessions such that participants saw an equal number of each condition in each session, though the item used to represent that condition varied. Two conditions within an item were yoked together and presented in the same session. These pairs were then shuffled through the 36 items. Conditions 1a and 2b of a given item appeared together in the same session, as did conditions 1b and 3a of the same item, conditions 2a and 4b of the same item, and 3b and 4a of the same item. Experiments 4 and 5 were run together. Overall, the conditions in Table 5 made up only 16% of the items within a session, with the remaining items consisting of Experiment 5 (33%) and fillers of the type used in Experiments 1 and 3 (51%). Thirty percent of the total sentences presented in each session were elided, including material from Experiment 5.

Procedure and design

The same procedure and analysis were used in this experiment as in Experiments 1 and 3.

Results and discussion

Fig. 4 presents the average (across participants) d' values as a function of processing time for distance-controlled, Simple and Complex Ellipsis conditions (top panel) and corresponding nonelided controls (bottom panel), along with the best-fitting exponential model described below. Inspection of Fig. 4 suggests that there were no salient differences between the Simple and Complex conditions in either the elided or nonelided sentences. Statistical tests on the average of the empirical d' values from 3.5 to 6 s post-initial response cue for each subject in each condition confirmed that there were no differences in asymptotic accuracy. Hierarchical exponential model fits and the associated statistical tests of parameter estimates did not show any indication that the full time-course functions differed in either accuracy or dynamics. Consequently, the best fit to the functions for both the Elided and Unelided conditions was a $1\lambda-1\beta-1\delta$ fit model. For the average data, the adjusted- R^2 value was .983 for fits of the ellipsis sentences, and .987 for the unelided sentences.

The data indicate that the complexity of the noun phrase in the VP antecedent had no measurable effects on either the speed or the accuracy of processing the VP ellipsis. This is again consistent with a pointer mechanism but is inconsistent with a copy process. It also suggests that the advantage for complex antecedents in Experiment 3 was likely due to the confounding effects of distance. We note that an alternative account of this advantage is that complex antecedents were more richly represented in memory, and hence facilitated the processing of subsequent ellipsis. If that were the case, the

³ One subject was run twice in the same session due to an error. We also performed the reported analysis without this subject, and found no change in the statistical tests or the pattern of the means. Given this, we have included this subject in the reported data.

Table 5
Example materials used in Experiment 4

Simple antecedent, ellipsis

- 1a. The history professor/ understood Roman mythology,/ but the over-worked students/ did not.
1b. *The history professor/ understood Roman mythology,/ but the overly worn books/ did not.

Complex antecedent, ellipsis

- 2a. The history professor/ understood Rome's swift and brutal/ destruction of Carthage,/ but the over-worked students/ did not.
2b. *The history professor/ understood Rome's swift and brutal/ destruction of Carthage,/ but the overly worn books/ did not.

Simple control (no ellipsis)

- 3c. The history professor/ understood Roman mythology,/ but the overly worn books/ did not last.
3d. *The history professor/ understood Roman mythology,/ the over-worked students/ did not wrinkle.

Complex control (no ellipsis)

- 4a. The history professor/ understood Rome's swift and brutal/ destruction of Carthage,/ but the overly worn books/ did not last.
4b. *The history professor/ understood in Rome's swift and brutal/ destruction of Carthage,/ but the over-worked students/ did not wrinkle.

* Denotes an unacceptable sentence; / denote phrase breaks in the phrase-by-phrase presentation method.

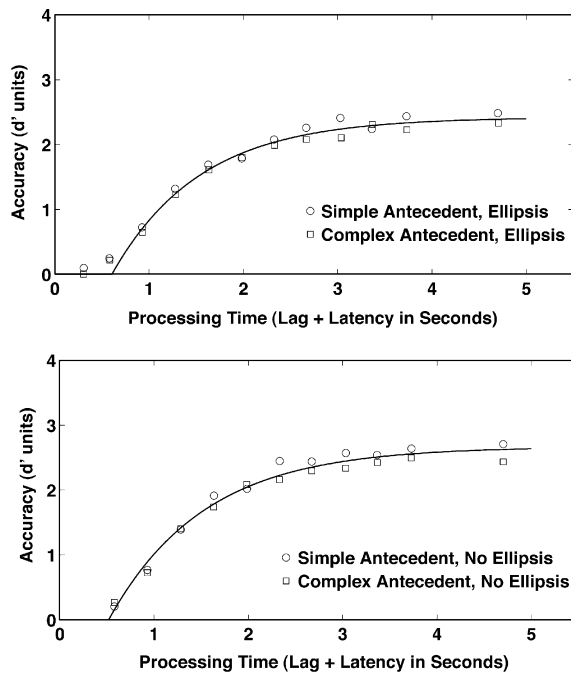


Fig. 4. Average d' accuracy (symbols) as a function of processing time (lag of the interruption cue plus latency to response) for Simple and Complex Elided conditions (top) and Simple and Complex Unelided conditions (bottom) from Experiment 4. Smooth curves show the best-fitting exponential fit (see text).

advantage should have been evident in this experiment as well.

Experiment 5

Experiment 4 suggests that VP ellipses do not require copying the antecedent into the elision site. The absence

of a complexity effect in both the SAT time-course and eye-tracking measures is consistent with our alternative hypothesis that a pointer mechanism might be used to interpret VP ellipses. Both measures are consistent with the reading time results of Frazier and Clifton (2001), who likewise found no measurable effects of complexity. Collectively, these results are at odds with Murphy's (1985) finding of a reading time advantage for simple, recent antecedents. Murphy's results suggest that some types of antecedent complexity may in fact engender a processing cost. However, that result alone does not provide strong support for a copy mechanism because reading time effects do not necessarily reflect underlying differences in the speed of processing, as a copy mechanism predicts.

Given these somewhat mixed results, we used the SAT procedure to investigate another type of complexity manipulation, one that might also be particularly problematic for a pointer mechanism. Such a mechanism may not suffice when interpretation of the elided antecedent contains a variable that must be bound to a constituent in the locality of the ellipsis site.⁴ The reflexives in (6) in the Introduction provide one such case. Other clear cases involve antecedents with covert subject arguments (PRO) and overt pronouns, such as (7) and (8).

(7) *The photographer knew who to hire, but the book designer did not.*

(8) *The photographer knew who he should hire, but the book designer did not.*

In (7), the antecedent is argued to have the linguistic form *...knew who [PRO] to hire*, in which PRO denotes

⁴ We thank Mark Baltin for suggesting this possibility.

the underlying subject of the verb hire. PRO is bound to the matrix subject the *photographer*, such that the meaning of the phrase is *the photographer knew who [the photographer will] hire*. However, when this phrase is taken as the antecedent for the VP ellipsis *did not*, the resulting interpretation is not *the book designer knew who [the photographer will] hire*, but rather *the book designer knew whom [the book designer will] hire*. Hence, the PRO in the antecedent must be reindexed to a new overt subject, *the book designer*.

Pronouns and other overt anaphors have different binding constraints than PRO (Chomsky, 1981; cf. Landau, 2003). In (8), the pronoun *he* in *The photographer knew who he should hire* could be interpreted as coreferent with the *photographer* or as coreferent with some other unmentioned discourse entity. When this VP is taken as the antecedent for the VP ellipsis *did not*, the resulting interpretation could be *the book designer knew who [the photographer should] hire*, *the book designer knew who [the book designer should] hire*, or *the book designer knew who [someone else should] hire*. Hence,

the antecedent could be reindexed to a new overt subject.

As a further means of contrasting whether antecedent complexity affects the speed or accuracy of interpreting ellipsis, and as a means of testing the potential limitations of a pointer mechanism, we contrasted ellipsis with a variable in the antecedent, such as (7) and (8), to ellipsis without variables in the antecedent. Table 6 illustrates the contrasts. Ellipsis with a PRO-variable (1a) and with an overt pronoun (2b) were compared to two (arguably) simpler types of ellipses, (3a) and (4a). In the first case, the antecedent contained the wh-item, *who*, like both (1a) and (2a), but we replaced the variable subject of the wh-clause with an overt noun phrase (e.g., *the painter*). Here, reanalysis of the subject of the relative clause in the antecedent at the elision site is not possible. In the second case, we replaced the wh-clause with a simple direct object complement, analogous to the contrasts used in our other experiments. We added adjectives and adverbs to match the antecedents in overall

Table 6
Example materials used in Experiment 5

Ellipsis with covert subject (PRO)

- 1a. The photographer/ questioned who to hire,/ but the gallery owner/ realized that/ the famous art critic/ did not.
1b. *The photographer/ questioned who to hire,/ but the gallery owner/ realized that/ the famous elephant/ did not.

Ellipsis with overt subject pronoun

- 2a. The photographer/ questioned who she should hire,/ but the gallery owner/ realized that/ the famous art critic/ did not.
2b. *The photographer/ questioned who she should hire,/ but the gallery owner realized that/ the famous elephant/ did not.

Ellipsis with overt noun phrase

- 3a. The photographer/ questioned who the painter should hire,/ but the gallery owner/ realized that/ the famous art critic/ did not.
3b. *The photographer/ questioned who the painter should hire,/ but the gallery owner/ realized that/ the famous elephant/ did not.

Ellipsis with simple verb phrase

- 4a. The photographer/ hastily hired an inexperienced agent,/ but the gallery owner/ realized that/ the famous art critic/ did not.
4b. *The photographer/ hastily hired an inexperienced agent,/ but the gallery owner/ realized that/ the famous elephant/ did not.

No ellipsis with covert subject (PRO)

- 5a. The photographer/ questioned who to hire,/ but the gallery owner/ realized that/ the famous art critic/ did not respond.
5b. *The photographer/ questioned who to hire,/ but the gallery owner/ realized that/ the famous art critic/ did not spray.

No ellipsis with overt subject pronoun

- 6a. The photographer/ questioned who she should hire,/ but the gallery owner/ realized that/ the famous art critic/ did not respond.
6b. *The photographer/ questioned who she should hire,/ but the gallery owner/ realized that/ the famous art critic/ did not spray.

No ellipsis with overt noun phrase

- 7a. The photographer/ questioned who the painter should hire,/ but the gallery owner/ realized that/ the famous art critic/ did not respond.
7b. *The photographer/ questioned who the painter should hire,/ but the gallery owner/ realized that/ the famous art critic/ did not spray.

No ellipsis with simple verb phrase

- 8a. The photographer/ hastily hired an inexperienced agent,/ but the gallery owner/ realized that/ the famous art critic/ did not respond.
8b. *The photographer/ hastily hired an inexperienced agent,/ but the gallery owner/ realized that/ the famous art critic/ did not spray.
-

* Denotes an unacceptable sentence; / denote phrase breaks in the phrase-by-phrase presentation method.

length. We again included corresponding conditions without VP ellipsis (examples 5–8), so that comprehenders could not predict an ellipsis on the basis of the sentence form.

The verb phrases with a wh-clause complement in (1a), (2a), and (3a) in Table 6 are syntactically more complex antecedents than the verb phrase with a simple noun phrase complement in (4a). Hence, a copy hypothesis predicts that the VP ellipsis should be more costly to process in the former cases than the latter. Further, if comprehenders must copy the syntactic form of the antecedent into the ellipsis and then reindex any syntactic variables, then there might be an additional cost associated with interpreting (1a) with a PRO variable and (2a) with an overt pronoun. Given that (2a) is potentially ambiguous between whether it was the photographer, the designer, or someone else who knew who to hire, it is also possible that (2a) might show an additional cost beyond what might be associated with reindexing a subject variable.

There are clearly other dimensions on which the contrasts in (1)–(4) in Table 6 differ. At a referential level, the antecedents in (3a) and (4a) contain an additional discourse entity, viz., the person doing the hiring (3a) and the person that was hired (4a). In contrast, the antecedent in (1a) contains only one entity, viz., the person to be hired. The antecedent in (2a) arguably contains up to two additional entities, depending on the reference of the pronoun. If complexity at the referential level matters, then we should see that conditions such as (3a) and (4a) with overt NPs in the antecedent are more costly to process than conditions such as (1a) (and perhaps (2a)) without these additional arguments in the antecedent. Whether the overt subject pronoun in (2a) is costly to process depends on which co-reference interpretation the subject converges on, which is likely not to be uniform across subjects.

The SAT procedure was again used to assess how these manipulations of the antecedents affect the speed and accuracy of processing the VP ellipsis.

Methods

Participants

Fifteen⁵ native speakers of American English from the New York University community were paid to participate in the study. They participated in four 75-min sessions, and a 45-min practice session for familiariza-

tion with the SAT procedure. All participants were between the ages of 18 and 26.

Materials

Table 6 contains illustrations of the conditions in Experiment 5. The full set of experimental materials is available from the JML online archive. In each of the four sessions, a participant read 144 experimental sentences, two conditions per item, counterbalanced within and across sessions. Therefore, participants saw every item in every condition, but at different points in the experiment. As in the above experiments, conditions were counter-balanced across sessions such that participants saw an equal number of each condition in each session, though the item used to represent that condition varied. Again, two conditions within an item were yoked together and presented in the same session, and we shuffled the pairs through the 36 items. The conditions in this experiment were yoked in the following way: 1a and 4b of a given item appeared together in the same session, as did conditions 1b and 3a, 2a and 3b, 2b and 4a, 5a and 8b, 5b and 7a, 6a and 7b, and 6b and 8a. In order to create unacceptable versions in which the unacceptability results from processing of the ellipsis, we varied the plausibility of the subject for the ellipsis. For example, we used plausible and implausible subjects (underlined) such as the following: (1) *The politician aggressively lobbied the foundation for more funds, but the chief of staff realized that the awkward intern did not.* /The politician aggressively lobbied the foundation for more funds, but the chief of staff realized that the awkward chimpanzee did not. (2) *The puppy regularly played fetch with guests, but the pet owner realized that the senile mutt did not.* /The puppy regularly played fetch with guests, but the pet owner realized that the senile astronaut did not. Because both human and nonhuman animals subjects occurred with equal frequency as plausible and implausible subjects, the humanity of the subject could not be used to predict the acceptability of the sentence. Nine of the items, such as: (3) *The chimpanzee easily deceived the ape in the experiment, but the scientist realized that the striped macaque did not.* /The chimpanzee easily deceived the ape in the experiment, but the scientist realized that the Bishop did not. become unacceptable at the integration of the embedded subject with the complement of the elided verb phrase. The other 27 items become unacceptable in these conditions upon the integration of the embedded subject and the elided verb. Experiments 4 and 5 were run together. Overall, the conditions in Table 6 made up 33% of the items within a session, with the remaining 67% of the items consisting of Experiment 4 (16%) and fillers of the type used in Experiment 1, 3, and 4 (51%). Thirty percent of the total sentences presented in each session were elided, including material from Experiment 4.

⁵ One subject was run twice in the same session due to an error. We also performed the reported analysis without this subject, and found no change in the statistical tests or the pattern of the means. Given this, we have included this subject in the reported data.

Procedure and analysis

The same procedure and analysis were used in this experiment as in the prior SAT experiments.

Results and discussion

Fig. 5 presents the average (across participants) d' values as a function of processing time for the four types of VP ellipsis conditions (top panel) and corresponding NonElided conditions (bottom panel), along with the best-fitting exponential model described below. Inspection of the top panel of Fig. 5 suggests that ellipsis with antecedents that contained an additional referential NP, either an overt subject NP (*questioned who the painter should hire*), Condition (3) or an overt object NP (*hastily hired an inexperienced agent*), Condition (4), were associated with lower accuracy levels than ellipses that only contained a covert or potentially co-referring subject in the antecedent (*questioned who to hire or questioned who she should hire*), Conditions (1) and (2). No corresponding differences are evident in the control conditions without VP Ellipsis in the bottom panel of Fig. 5.

As an initial means of investigating whether the differences among the ellipsis conditions were reliable, we again averaged, by subjects and by items, the empirical

d' values from 3.5 to 6 s post-initial response cue for each condition to derive an empirical estimate of asymptotic accuracy. A repeated measures ANOVA on the empirical d' values for the Elided conditions found a main effect of condition, $F_1(3,42) = 6.32$, $p < .005$ and $F_2(3,105) = 5.95$, $p < .001$; $\min F'(3,123) = 3.06$, $p < .05$. Pairwise comparisons indicated that performance was more accurate in Condition (1) with a covert (PRO) subject than in Condition (3) with an overt NP (.46 d' units, 95% CI = .25–.67) and Condition (4) with a simple VP (.44 d' units, 95% CI = .21–.66). Likewise, performance was more accurate in Condition (2) with an overt pronoun than in Condition (3) (.36 d' units, 95% CI = .04–.68) and Condition (4) (.33 d' units, 95% CI = .05–.62). Conditions (1) and (2) did not differ from each other (.1 d' units, 95% CI = -.19–.4), nor did Conditions (3) and (4) (.02 d' units, 95% CI = -.35–.31).

A repeated measures ANOVA on the empirical d' values for the Unelided conditions found a main effect of condition by subjects but not by items, $F_1(3,42) = 8.56$, $p < .001$ and $F_2(3,105) = 1.92$, $p = .13$; $\min F'(3,140) = 1.57$, $p = .2$. Pairwise comparisons indicated that the pattern differences were substantially different from those observed in the Elided conditions. Performance was significantly better in Condition (3) than in Condition (1) (–.2 d' units, 95% CI = –.36–.05), in Condition (2) (–.25 d' units, 95% CI = –.42–.07), and in Condition (4) (–.37 d' units, 95% CI = –.56–.18). Conditions (1)–(3) did not differ from one another. Hence, Conditions (3) and (4), which contained the additional referential noun phrase, do not pattern together in showing lower levels of performance as they do in the presence of ellipsis. Consistent with this, a repeated measures ANOVA on the empirical d' values of both Elided and Nonelided conditions together revealed an interaction between condition and ellipsis, $F_1(3,42) = 9.16$, $p < .001$ and $F_2(3,105) = 5.78$, $p < .001$; $\min F'(3,140) = 3.54$, $p < .05$, and a main effect of condition, $F_1(3,42) = 5.14$, $p < .01$, and $F_2(3,105) = 1.85$, $p = .14$; $\min F'(3,147) = 1.36$, $p = .26$. The interaction confirms the differential patterns seen in the pairwise tests.

Exponential model fits likewise demonstrated that asymptotic performance was lower for ellipses when the antecedent contained an additional referential NP. Models that did not allow asymptotes to vary across the ellipsis conditions produced poor fits to the empirical SAT data, and left systematic residuals. In fits of the average (over participants) data, a $4\lambda-1\beta-1\delta$ model increased the adjusted- R^2 to .993 from the .952 value observed with the $1\lambda-1\beta-1\delta$ model, and a similar improvement was evident in the fits of the individual participants' SAT functions. In the model fit of the average data, the four asymptotic (λ) parameters were respectively: 3.14, 3.07, 2.66, and 2.65, respectively. Paired t -tests on the asymptotic parameter estimates revealed that Condition (1) was significantly higher than

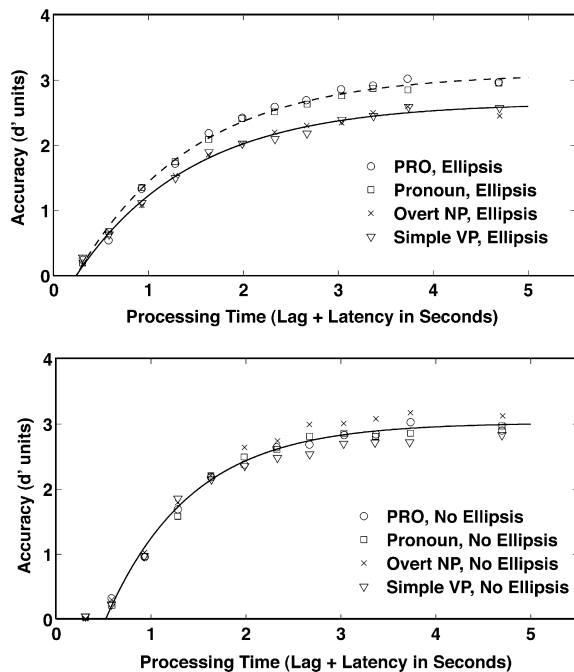


Fig. 5. Average d' accuracy (symbols) as a function of processing time (lag of the interruption cue plus latency to response) for the Elided conditions (top) and Unelided conditions (bottom) from Experiment 5. Smooth curves show the best-fitting exponential fit (see text).

Condition (3) (.48, 95% CI = .15–.81) and Condition (4) (.49, 95% CI = .19–.79), and that Condition (2) was higher than Condition (3) (.41, 95% CI = .08–.74) and Condition (4) (.42, 95% CI = .15–.69). As the estimates for Conditions (1) and (2) and Conditions (3) and (4) did not significantly differ from each other, the pattern could be parsimoniously fit with a $2\lambda-1\beta-1\delta$ model fit, in which one asymptote was assigned to Conditions (1) and (2), where the antecedent contained either an covert subject (PRO) or pronoun, and another asymptote was assigned to Conditions (3) and (4), in which there was an additional referential NP in the antecedent. This $2\lambda-1\beta-1\delta$ model produced an adjusted- R^2 of .993, with the Condition 1–2 asymptote estimated at 3.1 and the Condition 3–4 asymptote at 2.65 (.45, 95% CI = .21–.69).

Models that varied the intercepts (δ) or rates (β) across the different ellipsis conditions did not improve the adjusted- R^2 values. For example, a $2\lambda-2\beta-1\delta$ model and a $2\lambda-1\beta-2\delta$ model yielded comparable adjusted- R^2 values to the simpler $2\lambda-1\beta-1\delta$ model. Importantly, the resultant dynamics parameter estimates were nearly identical in fits of the average data, .81 versus .83 in the $2\lambda-2\beta-1\delta$ model, and .25 and .22 in the $2\lambda-1\beta-2\delta$ model, and there was no systematic trends in the parameter estimates across participants, $t(14) = 1.1$, $p = .29$ and $t(14) = .5$, $p = .62$. More embellished models, including $4\lambda-4\beta-1\delta$, $4\lambda-1\beta-4\delta$, and $4\lambda-4\beta-4\delta$ models, likewise did not improve the quality of the fit, nor did they yield any systematic trend in dynamics parameter estimates. Hence, there was no evidence to suggest that any of the manipulations in (1)–(4) affected the speed of processing an ellipsis. Fits of the $2\lambda-1\beta-1\delta$ model to average data are shown in Fig. 5.

For completeness, we also fit exponential models to the four unelided conditions. Despite some evidence from the analyses of the empirical asymptotic data that Condition 3 produced higher levels of performance than the other three conditions (see above), no statistically reliable trend emerged across the fits of the individual participant data. Fig. 5 presents the fits of a $1\lambda-1\beta-1\delta$ model to the average data.

We found clear evidence that asymptotic accuracy was lower for conditions in which the antecedent contained an additional NP argument, specifically for antecedents in Conditions (3) and (4) that contained a referential NP such as *painter* or *agent* (e.g., *questioned who the painter should hire* or *hastily hired an inexperienced agent*; see Table 6). This effect was observed only in the Ellipsis conditions, indicating that it is linked to processes involved with interpreting an ellipsis. Interestingly, other potential differences in antecedent complexity did not affect the accuracy of ellipsis processing. These included whether or not there was a variable in the antecedent, such as a covert subject (PRO) or an overt pronoun, as well as differences in syntactic complexity between, for example, Conditions (3) and (4),

where the former contains an additional embedded verb in the antecedent.

An additional referential argument in the antecedent can be viewed as a type of complexity effect, although the complexity appears to be at the discourse or referential level, rather than strictly at a syntactic level. Condition (3) has an embedded VP within the antecedent (e.g., *questioned who the painter should hire*), and so one might be tempted to argue that this type of syntactic complexity increased the likelihood of misinterpreting the ellipsis. For example, comprehenders might have sometimes inappropriately taken the embedded VP as the antecedent. However, if syntactic rather than referential complexity was responsible for the lower asymptotic levels, then two patterns should have resulted. First, as noted, Condition (3) should have been more difficult to process than Condition (4), as the former contains an embedded verb in the antecedent whereas the latter contains a simple VP (e.g., *hastily hired an inexperienced agent*). We found no evidence that these conditions differed in any way. Second, Conditions (1) and (2), which also have an embedded VP, either an infinitival VP (e.g., *questioned who to hire*) or an embedded VP with a pronoun as a subject (e.g., *questioned who she should hire*), should have yielded levels of performance comparable to Condition (3). However, both Conditions (1) and (2) engendered higher levels of performance than Conditions (3) and (4).

Hence, it appears that the presence of an additional referential NP in the embedded VP was what lowered accuracy, not syntactic complexity per se. Although it is not entirely clear why an additional discourse entity lowered performance, we suspect that the additional discourse entity increased the difficulty of fully integrating the antecedent into the VP ellipsis. Crucially, whatever the source of the difficulty, we note that this type of complexity did not slow the speed of interpreting VP ellipsis. As such, we found no evidence to support a psychological copy mechanism. The results of this experiment, like those of Experiments 3 and 4, are more consistent with a pointer-like mechanism in which the ellipsis is interpreted without the need to copy or reprocess structure from the antecedent.

General discussion

We investigated two issues in the reported experiments: how the antecedent is accessed at the ellipsis site, and how this information is then used in interpreting a VP ellipsis.

Recovering the antecedent

We manipulated the distance between an antecedent and the ellipsis site to investigate whether a search pro-

cess is required to gain access to antecedent representations in memory. The time-course functions observed in Experiment 1 demonstrated that near antecedents had higher asymptotic levels of performance than distant antecedents. Crucially, however, distance did not affect the estimates for the speed of processing VP ellipsis.

The demonstrably lower levels of accuracy for distant antecedents indicate that the processing of interpolated information disrupted comprehenders' ability to gain access to the previously processed antecedent. This may be the case if interpolated material lowered the quality of the retrieved information associated with the antecedent, if it caused comprehenders to sometimes fail to access any representation of the antecedent, and finally, if it increased the probability that comprehenders directly accessed an inappropriate constituent in memory. If distance had any of these effects or some combination of these effects, comprehenders would be more likely to judge ellipsis with a distant antecedent as less acceptable than one with a near antecedent.

One reason why distance may adversely affect the quality of the recovered antecedent information is that the antecedent's representation in memory may have decayed during the time required to process the additional interpolated material. For other types of materials, Gibson and colleagues (e.g., Gibson, 1998, 2000; Grodner & Gibson, 2005) have appealed to decay as an explanation of increased cost of processing sentences that require the integration information across long distances. We cannot fully rule out a simple decay explanation on the basis of our data alone, but we note that research on working memory has increasingly looked from explanations based on decay (or displacement) towards accounts that emphasize interference at retrieval as the major constraint on accessing information in memory (e.g., Anderson & Neely, 1996; Crowder, 1976; see Nairne, 2002 for a recent review). Consonant with these approaches, recent investigations of the role of memory in comprehension have indicated that the quality of the cues at retrieval are a determining factor in the processing of nonadjacent dependencies (Gordon, Hendrick, & Johnson, 2001, 2004; Gordon, Hendrick, & Levine, 2002; Van Dyke, 2002; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006). Hence, we suspect that it is more likely that distance reduced the distinctiveness of the retrieval cues at the ellipsis site. For example, processing interpolated material may introduce into the discourse potential competitors to the antecedent. More generally, it may shift the comprehender's focus away from the antecedent. In turn, the reduction in distinctiveness may hinder the retrieval of antecedent information necessary to fully interpret the ellipsis and may even have inappropriately directed comprehenders to the wrong constituent in memory (McElree et al., 2003).

Reading time studies of ellipsis have found that distance slows whole-sentence reading times (Murphy, 1985). The time-course analyses in Experiment 1 suggest that these reading time effects do not arise from differences in search time, but rather from differences in the quality of retrieved information about the antecedent. If these effects were indeed due to a search process, then we should have observed a slowing of the dynamics of the SAT functions, either a delayed intercept or slower rate (for simulations of a search process, see McElree & Doshier, 1989, 1993).

The eye-tracking measures from Experiment 2 reinforce this interpretation of the reading time effects. We found that longer distances between the antecedent and the ellipsis site did not affect reading time at either the ellipsis site or the spillover region. This pattern is inconsistent with a search mechanism, in that one would expect slower times at these regions if more time was required to search for a remote antecedent. We found that distance engendered longer second-pass time and total time on the antecedent region, significantly so on the antecedent complement region. That comprehenders spent demonstrably more time re-reading distant antecedent regions strongly suggests that the quality of the retrieved memory representation at the ellipsis site was not good enough to enable an adequate interpretation of these ellipses on a proportion of trials, and it may have required comprehenders to regress back to the antecedent region. This pattern is fully consistent with the lower asymptotic level observed in the SAT study of Experiment 1. As the SAT procedure does not allow rereading, comprehenders would likely misjudge a sentence on those trials that would engender, in a reading time procedure, regressions back to the antecedent.

SAT studies of the effects of distance on resolving filler-gap dependencies and subject-verb dependencies (McElree, 2000; McElree et al., 2003) have found the same essential pattern as observed in Experiment 1. As with the results reported here, these studies indicate that comprehenders have direct access to the products of past analyses. As noted, the constituent to be retrieved from memory in these other types of nonadjacent dependencies—the filler item in a filler-gap relation or the subject in a subject-verb relation—is marked by the grammar as having some future role in the sentence. Hence, comprehenders can predict that the constituent will be needed at a later point in the sentence, and they may assign it some special status in memory. However, this is not true of the antecedent for VP ellipsis, which has an independent role in its matrix clause. That the antecedent of an ellipsis displays the same property of direct access as the constituents involved in predictable grammatical relations provides support for a general claim that all representations formed in sentence comprehension are fully content-addressable and retrieved with a cue-driven, direct-access operation.

We note, however, that without contrasting conditions analogous to Van Dyke and McElree (2007) (see Examples (4) and (5)), where the amount of information is varied before the antecedent, we cannot completely rule out an alternative search model in which comprehenders start the search at the beginning of the sentence. Nonetheless, the fact that this type of manipulation has no effect on the processing of other nonadjacent dependencies suggests to us that it is unlikely to have an effect on the speed of processing of ellipsis. Moreover, one can question whether such a search strategy would be effective in language understanding. Ellipses are unbounded and often span across sentences. Hence, unlike a backward or recency-based search, there is no natural starting point for a forward search in real discourses.

Interpretation of ellipsis

Experiments 3–5 manipulated the length or complexity of the antecedent to investigate how, once accessed in memory, an antecedent is interpreted at the ellipsis site. Following logic similar to Frazier and Clifton (2001), we reasoned that comprehenders should be slower to interpret long (or complex) antecedents than short (or simple) antecedents if interpretation requires copying of the antecedent structure into the ellipsis site.

In the SAT studies of Experiments 3 and 4, we contrasted simple- and complex-NP antecedents (e.g., *Roman mythology* versus *Rome's swift and brutal destruction of Carthage*). There was no evidence in either experiment that antecedent length affected the speed of interpreting an ellipsis. In Experiment 3, which controlled overall sentence length but confounded distance from the antecedent, there was a nonsignificant trend for longer antecedents to be interpreted *more* accurately than short antecedents. It is not entirely clear whether the accuracy advantage for long antecedents reflects their closer distance to the ellipsis site, their salience in memory, or perhaps differences in the degree to which comprehenders fully specify the respective interpretations. Whatever the source, however, the observed differences are in a direction opposite to what is predicted by a copy mechanism. Indeed, the advantage for complex antecedents disappeared completely in Experiment 4 when we controlled distance, rather than overall sentence length.

Experiments 3 and 4 indicate that, at best, antecedent length may modulate the accuracy of interpreting an ellipsis, but it has no measurable effect on the speed of processing. This finding is inconsistent with a structure-sensitive copy operation underlying the interpretation of an ellipsis. The timecourse measures are more consistent with approaches that view the VP ellipsis as functioning as a pointer to a preexisting memory structure, where the clause containing the ellipsis is interpreted by a pointer that links it to the antecedent

representation that has been accessed in memory. The *structure-sharing* hypothesis of Frazier and Clifton (2005), where the ellipsis site shares the antecedent's structure, could be implemented with a pointer mechanism. For example, Frazier and Clifton (2001) suggested that the ellipsis points to the left corner of the antecedent's syntactic representation. Structure-sharing may be essential for (re-)binding variables in the antecedent. We note, however, a pointer mechanism is equally compatible with alternative views that treat ellipses as a type of discourse anaphora (see Garnham, 2001) or that argue that ellipsis can be interpreted by establishing coherence relations based on semantics and discourse properties alone (Kehler, 2002). In these cases, the ellipsis simply points to a fully interpreted discourse representation.

In Experiment 5, we tested the viability of a pointer mechanism further by contrasting antecedents containing a simple VP to those with a VP with a wh-clause complement, including ones with variables (either a PRO or an overt pronoun) that may require reindexing within the antecedent at the ellipsis site. We did not find any evidence that these more complex antecedents were measurably slower to process than simpler VP antecedents, which again is inconsistent with the idea that interpretation requires copying the structure of the antecedent into the ellipsis site. Further, structural complexity per se did not affect accuracy. Rather, the number of discourse entities in the antecedent affected asymptotic accuracy when the sentence was elided, with accuracy being demonstrably lower when the antecedent contained an additional discourse referent. This finding suggests that although interpretation is less accurate when multiple discourse referents exist in memory, the number of discourse entities does not affect the speed of interpretation. We take this as evidence that one form of complexity, namely discourse complexity, affects interpretation of ellipsis, but it does not affect the pointer mechanism's processing speed.

Summary

Collectively, our experiments indicate that the antecedent for a VP ellipsis can be accessed in memory without a search through other representations formed before encountering the ellipsis, and that interpretation of an ellipsis can be accomplished without copying the contents of the antecedent into the ellipsis site. The former is consistent with a growing body of empirical evidence suggesting that the memory representations formed during comprehension are content-addressable and retrieved with a direct-access process (McElree et al., 2003; McElree, 2000), as well as with modeling approaches that assume content-addressability (e.g., Lewis & Vasishth, 2005; Lewis et al., 2006; Tabor et al., 2004; Tabor & Hutchins, 2004; Vosse & Kempen,

2000). The latter is consistent with accounts of ellipsis that assume some type of “structure-sharing” between antecedent and ellipsis (Frazier & Clifton, 2005). We suggest that sharing can be implemented by assuming that syntactic and semantic constraints at the ellipsis site serve as a pointer to extant memory representations.

Appendix A. Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.jml.2007.06.010](https://doi.org/10.1016/j.jml.2007.06.010).

References

- Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In *Handbook of perception and memory*. In E. L. Bjork & R. A. Bjork (Eds.). *Memory* (Vol. 10, pp. 237–313). San Diego: Academic Press.
- Arregui, A., Clifton, C. E., Jr., Frazier, L., & Moulton, K. (2006). Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language*, *55*, 232–246.
- Bornkessel, I., McElree, B., Schlesewsky, M., & Friederici, A. D. (2004). Multi-dimensional contributions to garden path strength: Dissociating phrase structure from relational structure. *Journal of Memory and Language*, *51*, 495–522.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Christianson, K., Williams, C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults’ “Good-Enough” interpretations of garden-path sentences. *Discourse Processes*, *42*, 205–238.
- Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, *3*, 37–60.
- Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.
- Doshier, B. A. (1979). Empirical approaches to information processing: Speed–accuracy tradeoff functions or reaction time. *Acta Psychologica*, *43*, 347–359.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.
- Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representation. *Journal of Memory and Language*, *56*, 357–383.
- Frazier, L., & Clifton, C. E. Jr., (2001). Parsing coordinates and ellipsis: COPY α . *Syntax*, *4*, 1–22.
- Frazier, L., & Clifton, C. E. Jr., (2005). The syntax–discourse divide: Processing ellipsis. *Syntax*, *8*, 121–174.
- Frazier, L., & Clifton, C. E. Jr., (2006). Ellipsis and discourse coherence. *Linguistics and Philosophy*, *29*, 315–346.
- Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Hove, East Sussex, UK: Psychology Press.
- Garrod, S., & Sanford, A. J. (1990). Referential processing in reading: Focusing on roles and individuals. In D. A. Balota, G. B. Flores d’Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 465–486). Hillsdale, NJ: Erlbaum.
- Garrod, S., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. *Journal of Memory and Language*, *42*, 526–544.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1–76.
- Gibson, E. (2000). The dependency-locality theory: A distance-based theory of linguistic complexity. In: I. Miyashita, A. Marantz, & W. O’Neil (Eds.), *Image, language, brain: Papers from the first mind articulation symposium* (pp. 94–126). Cambridge, MA, USA: The MIT Press.
- 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. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, *29*, 261–290.
- Gronlund, S. D., Edwards, M. B., & Ohrt, D. D. (1997). Comparison of the retrieval of item versus spatial position information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *23*, 1261–1274.
- Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model comparison approach*. San Diego: Harcourt Brace Jovanovich.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329–354.
- Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA, USA: CSLI Publications.
- Landau, I. (2003). Movement out of control. *Linguistic Inquiry*, *34*, 471–498.
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *2*, 375–419.
- Lewis, R. L., Vasishth, S., & Van Dyke, J. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*, 447–454.
- Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 55–75). Oxford: Elsevier.
- Masson, M., & Loftus, G. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, *57*, 203–220.
- McElree, B. (1993). The locus of lexical preference effects in sentence comprehension: A time-course analysis. *Journal of Memory and Language*, *32*, 536–571.
- McElree, B. (1996). Accessing short-term memory with semantic and phonological information: A time-course analysis. *Memory & Cognition*, *24*, 173–187.
- McElree, B. (1998). Attended and non-attended states in working memory: Accessing categorized structures. *Journal of Memory and Language*, *38*, 225–252.
- McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*, *29*, 111–123.

- McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 27, 817–883.
- McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 46). San Diego: Academic Press.
- McElree, B., & Carrasco, M. (1999). Temporal dynamics of visual search: A speed–accuracy analysis of feature and conjunction searches. *Journal of Experimental Psychology: Human Perception & Performance*, 25, 1517–1539.
- McElree, B., & Doshier, B. A. (1989). Serial position and set size in short-term memory: Time course of recognition. *Journal of Experimental Psychology: General*, 18, 346–373.
- McElree, B., & Doshier, 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.
- McElree, B., & Nordlie, J. (1999). Literal and figurative interpretations are computed in equal time. *Psychonomic Bulletin & Review*, 6, 486–494.
- McElree, B., Pyllkkänen, L., Pickering, M. J., & Traxler, M. (2006). The time course of enriched composition. *Psychonomic Bulletin & Review*, 13, 53–59.
- Mitchell, D. C., & Green, B. W. (1978). The effects of context and content on immediate processing in reading. *Quarterly Journal of Experimental Psychology*, 30, 609–636.
- Murphy, G. L. (1985). Processes of understanding anaphora. *Journal of Memory and Language*, 24, 290–303.
- Murguía, E. (2004). *Syntactic identity and locality restrictions on verbal ellipsis*. Unpublished doctoral dissertation, University of Maryland College Park, College Park, MD, USA.
- Nairne, J. S. (2002). The myth of the encoding–retrieval match. *Memory*, 10, 389–395.
- Nunes, J. (1995). *The Copy theory of movement and linearization of chains in the minimalist program*. Unpublished doctoral dissertation, University of Maryland College Park, College Park, MD, USA.
- Pickering, M. J., Frisson, S., McElree, B., & Traxler, M. J. (2004). Eye movements and semantic composition. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP and beyond*. New York: Psychology Press.
- Pickering, M. J., McElree, B., Frisson, S., Chen, L., & Traxler, M. J. (2006). Underspecification and aspectual coercion. *Discourse Processes*, 42, 131–156.
- Poesio, M., Sturt, P., Artstein, R., & Filik, R. (2006). Underspecification and anaphora: Theoretical issues and preliminary evidence. *Discourse Processes*, 42, 157–176.
- Rayner, K. (1998). Eye movements in reading and information processing: Twenty years of research. *Psychological Bulletin*, 124, 372–422.
- Rayner, K., Kambe, G., & Duffy, S. A. (2000). The effect of clause wrap-up on eye movements during reading. *Quarterly Journal of Experimental Psychology*, 53A, 1061–1080.
- Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. NJ: Prentice-Hall.
- 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.
- Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain Science*, 26, 709–777.
- Sag, I. (1976). *Deletion and logical form*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA.
- Sanford, A. J., Garrod, S. C., Lucas, A., & Henderson, R. (1983). Pronouns without explicit antecedents. *Journal of Semantics*, 303–318.
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, 6, 382–386.
- Shapiro, L., Hestvik, A., Lesan, L., & Garcia, R. (2003). Charting the timecourse of VP-ellipsis sentence comprehension: Evidence for initial and independent structural analysis. *Journal of Memory and Language*, 49, 1–19.
- Sternberg, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology*, 27, 1–32.
- Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50, 355–370.
- Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 431–450.
- Theios, J. (1973). Reaction time measurement in the study of memory processes: Theory and data. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 7, pp. 44–85). New York: Academic Press.
- Treisman, M., & Doctor, E. (1987). Memory scanning: A comparison of the dynamic stack and exhaustive serial scan models with an extension of the latter. *Acta Psychologica*, 64, 39–92.
- Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. New York: Cambridge University Press.
- Van Dyke, J. A. (2002). *Retrieval effects in sentence parsing and interpretation*. Unpublished doctoral dissertation, University of Pittsburgh, Pittsburgh.
- 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, 285–316.
- Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, 55, 157–166.
- Van Dyke, J. A. & McElree, B. (2007). *Similarity-based proactive and retroactive interference reduces quality of linguistic representations*. Poster presented at the CUNY Sentence Processing Conference, San Diego.
- Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, 75, 105–143.
- Wickelgren, W. A. (1977). Speed–accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 41, 67–85.