

This article was downloaded by: [108.18.231.14]

On: 11 December 2013, At: 17:27

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## The Quarterly Journal of Experimental Psychology

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/pqje20>

### Going the distance: Memory and control processes in active dependency construction

Matthew W. Wagers<sup>a</sup> & Colin Phillips<sup>b</sup>

<sup>a</sup> Department of Linguistics, University of California, Santa Cruz, CA, USA

<sup>b</sup> Department of Linguistics, Neuroscience and Cognitive Science Program, University of Maryland, College Park, MD, USA

Accepted author version posted online: 29 Oct 2013. Published online: 09 Dec 2013.

**To cite this article:** Matthew W. Wagers & Colin Phillips, The Quarterly Journal of Experimental Psychology (2013): Going the distance: Memory and control processes in active dependency construction, The Quarterly Journal of Experimental Psychology, DOI: [10.1080/17470218.2013.858363](https://doi.org/10.1080/17470218.2013.858363)

**To link to this article:** <http://dx.doi.org/10.1080/17470218.2013.858363>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

# Going the distance: Memory and control processes in active dependency construction

Matthew W. Wagers<sup>1</sup> and Colin Phillips<sup>2</sup>

<sup>1</sup>Department of Linguistics, University of California, Santa Cruz, CA, USA

<sup>2</sup>Department of Linguistics, Neuroscience and Cognitive Science Program, University of Maryland, College Park, MD, USA

Filler-gap dependencies make strong demands on working memory in language comprehension because they cannot always be immediately resolved. In a series of three reading-time studies, we test the idea that these demands can be decomposed into active maintenance processes and retrieval events. Results indicate that the fact that a displaced phrase exists and the identity of its basic syntactic category both immediately impact comprehension at potential gap sites. In contrast, specific lexical details of the displaced phrase show an immediate effect only for short dependencies and a much later effect for longer dependencies. We argue that coarse-grained information about the filler is actively maintained and is used to make phrase structure parsing decisions, whereas finer grained information is more quickly released from active maintenance and consequently has to be retrieved at the gap site.

**Keywords:** Parsing; Psycholinguistics; Syntax; Memory; Unbounded dependencies.

An adequate description of the language comprehension architecture must specify how the parser encodes and retrieves information and the decision principles for using or disposing of those encodings—what Lewis (2000) decomposes into *memory* and *control* processes. The comprehension of filler-gap dependencies, as exemplified by the relative clause in Example 1 (where NP = noun phrase, CP = complementizer phrase, VP = verb phrase), has provided a fruitful paradigm for understanding the interaction between the memory and control processes of the parser. There are several reasons for their interest. First, filler-gap dependencies are unbounded: The distance between filler and gap can be arbitrarily large, which Example 2 illustrates. Consequently a representation of the filler must be durably encoded

and reliably retrievable so that it can survive spans of material during which it is not being processed.

- (1) One-clause filler-gap dependency:  
[<sub>NP</sub> The **stones** [<sub>CP</sub> that the pilgrim [<sub>VP</sub> placed \_\_\_\_ on the cairn]]] were smooth.
- (2) (a) Two-clause dependency:  
The **stones** [that the monk advised the pilgrim [to place \_\_\_\_ on the cairn]] ...  
(b) Three-clause dependency:  
The **stones** [that the scriptures advised the monk [to tell the pilgrim [to place \_\_\_\_ on the cairn]]] ...

Second, filler-gap dependencies are “island sensitive”; there are certain domains that filler-gap

Correspondence should be addressed to Matthew Wagers, Department of Linguistics, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA. E-mail: [mwagers@ucsc.edu](mailto:mwagers@ucsc.edu)

The authors gratefully acknowledge Shravan Vasishth for helpful analytical advice and discussion.

Supplemental material is available via <http://people.ucsc.edu/~mwagers>

dependencies cannot cross, which Example 3 illustrates for phrases in subject position (where the asterisk denotes ungrammaticality).

- (3) \*The **stones** [that [cairns of \_\_\_\_] impressed the pilgrim] were smooth.

Ross (1967) identified a number of these domains, which he dubbed “islands”, including adjuncts, coordinate structures, relative clauses, *wh*-clauses, and factive clauses. A diverse array of restrictions thus exist on where a gap may be found. Syntactic theories have attempted to explain the diversity of island domains as the surface manifestation of fewer, more abstract rules (Chomsky, 1977, 1987; Manzini, 1992; Rizzi, 1990). From the perspective of the parser it is sufficient to observe that not every argument position can be a gap site. This suggests that the control processes for dependency construction must incorporate a number of contingencies.

Finally, filler–gap dependencies are widespread in natural languages and cut across many grammatical constructions. Constituent questions, relative clauses, topicalizations, and comparatives, among others, all include unbounded dependencies that involve a relation between a constituent at the edge of a sentence and a site of thematic interpretation contained within that sentence. Whether filler–gap parsing is viewed as consisting in specialized routines or simply as a temporally extended version of local dependency formation, the architectural commitments that filler–gap parsing reveals are likely to reflect core properties of the comprehension system.

Real-time filler–gap dependency construction has been studied intensively in the past 20 years, and a notably consistent account of the properties of dependency construction has emerged. In the section entitled “Constructing unbounded dependencies online”, we briefly review the evidence that dependency construction occurs as soon as possible and that it respects the constraints of the grammar. Then, in “Memory resources for filler–gap dependencies” we turn to our main question, involving

how fillers are stored and retrieved from memory during the comprehension process. One psycholinguistic tradition, beginning with Wanner and Maratsos (1978), holds that fillers are maintained in a distinguished memory state until they can be discharged to complete the dependency. This viewpoint can be challenged empirically, as there is little direct evidence that such a distinguished state exists. It can also be challenged conceptually, as some otherwise well-supported memory models do not easily tolerate maintenance of the kind envisioned (McElree, 2006). The alternative is to assume that the filler is not maintained, but that it is retrieved later at its thematic interpretation site.

We have previously observed that the locus of dependency construction appears to shift from occurring at the verb to a non-verb-adjacent gap site when the length of the filler–gap dependency is increased (Wagers & Phillips, 2009). This shift in timing with respect to heads in the input is difficult to explain on either a pure maintenance or a pure retrieval account. In a series of three experiments, we explore under what conditions the timing of dependency construction appears to shift from early, pregap sites to late, postgap sites. We find that for early dependency formation, coarse-grained, categorical details are reliably present and more robust than finer grained lexically anchored detail. We propose a hybrid maintenance-retrieval model to account for these new data, in a way that combines the insights of the two major perspectives on unbounded distance dependency formation.

## Constructing unbounded dependencies online

A fairly detailed understanding of how filler–gap processing proceeds has been compiled. Existing findings primarily inform us about control processes—that is, those processes that determine when and under what circumstances filler–gap dependency formation is attempted. Intuitively, processing a filler–gap dependency requires

<sup>1</sup> We use the terms *gap* and *gap position* in a theory neutral way: None of our discussion turns upon whether the tail of an unbounded dependency is an empty category or not.

recognizing the filler, identifying a gap position, and linking the filler to the gap position.<sup>1</sup> An important finding regarding filler-gap processing is that comprehenders attempt to construct dependencies in advance of unambiguous information about the gap position. That is, comprehenders anticipate the location of potential gap sites and predictively construct a relation with the filler. This mechanism is referred to as *active* dependency formation, and the supporting evidence comes from a broad variety of measures and languages (for review see Phillips & Wagers, 2007). One type of evidence comes from a reading-time slowdown in sentences like (4a). In this sentence, the comprehender encounters an overt direct object, “us”, where a gap can be expected. Reading times at that word are elevated relative to reading times for the same word in a closely matched sentence that lacks a filler-gap dependency, where no gap is expected (4b). This effect is termed the *filled gap effect* (Crain & Fodor, 1985; Lee, 2004; Stowe, 1986).

- (4) (a) My brother wanted to know **who** Ruth will bring us home to \_\_\_ at Christmas.  
 (b) My brother wanted to know if Ruth will bring us home to Mom at Christmas.

Another type of evidence comes from processing disruptions at the verb when the filler would be a semantically implausible argument of the verb, as in (5). Traxler and Pickering (1996) found elevated reading times for implausible filler-verb combinations at the verb, and in similar configurations Garnsey, Tanenhaus, and Chapman (1989) observed enhancement of the N400, an event-related potential (ERP) component modulated by factors associated with lexicosemantic integration or expectancy.

- (5) That’s the **pistol/garage** with which the heartless killer shot the hapless man \_\_\_ yesterday afternoon.

It is because processing disruptions precede direct evidence for the actual location of the gap that filler-gap processing has been called “active”. In other words, comprehenders posit gaps eagerly, preferring to construct a dependency early that

might subsequently have to be retracted, rather than wait for unambiguous evidence of the gap site. Though both the filled-gap effect and the semantic anomaly effects observed to date point to this conclusion, it is important to recognize that these effects derive from different underlying sources of information. For the filled-gap effect, it is the existence of a dependency (Stowe, 1986) or the syntactic category of the filler (Lee, 2004) that underlies the contrast. For the semantic anomaly effect, the integration of often idiosyncratic lexical features of the filler, the verb, and their relation to real-world knowledge underlies the effect.

An important question is whether active dependency formation is restricted by grammatical constraints on potential gap sites. A number of studies have therefore examined whether island constraints guide the comprehender. Several ERP studies have shown that a processing disruption occurs if the parser is holding an unsatisfied *wh*-phrase when it encounters the edge of an island domain (Kluender & Kutas, 1993; McKinnon & Osterhout, 1996) or is forced to posit a gap inside an island (Neville, Nicol, Barss, Forster, & Garrett, 1991). Behavioural studies have further shown that active dependency formation effects are absent inside island domains. For example, Stowe (1986) found no filled gap effect inside a subject noun phrase. Likewise, Traxler and Pickering (1996) found no plausibility effect inside a relative clause island, a finding recently replicated in second-language learners (Omaki & Schulz, 2011). Evidence from additional behavioural studies (Bourdages, 1992; McElree & Griffith, 1998; Pickering, Barton, & Shillcock, 1994) is consistent with these results. It seems that comprehenders recognize island domains and subsequently refrain from positing gaps inside them. More recently, studies have shown that comprehenders are sensitive in real time to important subtleties of the grammar of filler-gap dependencies, including where parasitic gaps may be licensed (Phillips, 2006) and where across-the-board extraction must occur (Wagers & Phillips, 2009). In sum, these studies provide evidence about the control processes guiding comprehension: The comprehender integrates grammatical knowledge in

guiding its expectations about where it may complete the filler-gap dependency.

A relatively straightforward description thus emerges for the control processes that support filler-gap dependency construction. First, comprehenders attempt to terminate an open filler-gap dependency wherever the grammar allows. Second, the comprehension mechanism does not “wait and see” where the actual gaps in a sentence are, but posits the gaps wherever they might occur, based on the partial input and the upcoming structure it entails.

### Memory resources for filler-gap dependencies

The temporal delay between the encoding of the filler and potential sites in the sentence where it can be integrated may be substantial. More importantly, this delay may be occupied by potentially irrelevant processing operations. For example, in Example 6, after the comprehender processes the filler, she must (at least) assemble the subject phrase, attach its predicate, build that verb's complement clause, and link the main clause subject to its thematic role assigner in the embedded clause—all of this before the gap site is encountered.

- (6) **Which shrines** is the famous monk most likely to have visited \_\_\_\_?

Because it participates in nonlocal dependencies, the encoding of the filler has to be sufficiently durable to survive arbitrary spans of material during which it is not being processed. Note that this problem is true both of cases where there is considerable hierarchical distance between the filler and the gap site, and also of cases where the hierarchical distance is minimal, and the temporal delay is nonetheless substantial, as in Example 7, where the subject is modified by a prepositional phrase.

- (7) **Which shrines** did the monk from the ancient monastery visit \_\_\_\_?

There are at least two architectural possibilities for how the task of integrating a filler with its gap

host can be accomplished. Under a *maintenance view*, the filler is stored in a short-term memory location until the point at which it can be integrated directly with a gap host. Under a *retrieval view*, the filler is not purposely kept in short-term memory and must be retrieved and reactivated when the parser posits a gap. According to the maintenance account, the filler is continuously accessible to the parser but potentially exacts an overhead cost by consuming the comprehender's short-term memory capacity. According to the retrieval account, there is no taxing maintenance but there is a possibility for error later when the filler has to be reactivated. Both kinds of account can be found in the literature, though there has been relatively little attention paid to distinguishing them (with the notable exception of McElree, Foraker, & Dyer, 2003). In part this is because experimental design, methodologies, and linguistic manipulations have varied between studies, making direct comparisons difficult. Secondly, despite a diversity of measures and the fact that a key property of filler-gap dependencies is their unboundedness, a relatively narrow range of dependency lengths has been examined. As we argue in a moment, at very short dependency lengths the two accounts are difficult to distinguish empirically.

In the remainder of this section, we review some of the major accounts of the memory resources deployed in filler-gap comprehension. Then we lay out the logic of our study, whose goal is to disentangle maintenance versus retrieval accounts of filler-gap processing by examining the effects that dependency length has on differing linguistic manipulations.

One influential maintenance proposal holds that the filler phrase is encoded in a distinguished memory register until it can be assigned a thematic role in the sentence. Wanner and Maratsos (1978) proposed an augmented transition network (ATN) model of parsing, which included a separate buffer, called the HOLD cell, for displaced NPs. When an NP is identified as the head of a relative clause, the contents of the NP are placed in the HOLD cell. The HOLD cell allows the assignment of grammatical

function to be put off until an appropriate context is identified. For example, in an object-extracted relative clause, after a transitive verb has been processed, the ATN attempts to analyse subsequent input as a direct object NP. If it fails to find an overt NP in the input, as would be the case if the clause contains a gap, then the system checks to see whether or not the HOLD cell is empty; if it is not empty, the system retrieves its contents, treating them as input.

This parsing sequence does not qualify as an active strategy, because a gap is not postulated unless the parser fails to encounter a verb's lexical argument. But that is a fact about the parser's control structure, which can be modified independently of the memory architecture. Indeed Frazier (1987) suggested that the nonemptiness of the HOLD cell itself could serve as a signal to postulate a gap before lexical arguments are encountered. This mechanism is formulated as the *active filler strategy*, which states that once a filler has been encountered, the parser is "immediately predispose[d]" to rank gaps more highly than lexical arguments (Frazier & Flores D'Arcais, 1989). Frazier and Flores d'Arcais (1989) suggest that this implies a filler that is not "inert". At predicates following the filler, the parser considers gap analyses before analyses with an overt lexical argument because the unintegrated filler is effectively always an input. Considered in this way, dependency completion may be active because the filler is available in the processing workspace before subsequent inputs, and it therefore effectively out-competes incoming categories for attachment.

There have been many demonstrations that filler-gap dependencies exact a cost on the comprehender (Gibson, 1998, 2000; King & Just, 1991; Sprouse, Wagers, & Phillips, 2012; Wanner & Maratsos, 1978), a fact sometimes taken as evidence for the maintenance view. Several electrophysiological studies have also been brought to bear on the question of memory load during an open dependency. Both King and Kutas (1995) and Fiebach, Schlesewsky, and Friederici (2002) found that in object-extracted filler-gap

dependencies the averaged electroencephalography (EEG) record shows a sustained anterior negativity (SAN). This effect is modulated by performance and participants' memory span and has been implicated in explicit memory load tasks (Ruchkin, Johnson, Canoune, & Ritter, 1990). For these reasons, the presence of a sustained anterior negativity in open filler-gap dependencies has been interpreted as a direct reflection of the memory load consumed by actively maintaining the filler. However, this interpretation may be undercut by the fact that the SAN does not reflect a cumulative effect that accrues at each word, but instead derives mainly from the first words of the dependency (King & Kutas, 1995; Phillips, Kazanina, & Abada, 2005). Importantly, it remains an open question what the actual content of the filler is when the dependency is unsatisfied. In discussing the SAN, Fiebach et al. (2002) point out that the electrophysiological effect does not choose between alternative accounts of what precisely is being maintained in a privileged state. It could be a full semantic or syntactic representation of the filler, or perhaps just a few features. Alternatively, it may not be the content of the filler that is represented at all, but rather the prediction for a category that allows completion of the dependency, as in dependency locality theory (Gibson, 2000).

In contrast to a pure maintenance account, and supporting a retrieval account, there is evidence that semantic features of the filler are not maintained and must be reactivated later into a state suitable for integration. Swinney and colleagues have shown that a filler's semantic associates are primed in a cross-modal lexical decision immediately following the introduction of the filler in the sentence (Nicol, Fodor, & Swinney, 1994; Nicol & Swinney, 1989). However this priming does not persist, declining across the sentence. Priming is once again observed following the verb. McElree (2001) reports a similar pattern using a probe recognition task. The fact that priming is not observed in dependency-medial positions suggests either that the filler contents are not actively maintained or that, in their maintained form, they cannot influence lexical

decision.<sup>2</sup> These findings demonstrate that some properties of the filler are not equally available throughout the span of the dependency, in support of a retrieval view.

The possibility that no filler contents are actively maintained is supported by the notion that there is a fairly stringent limit on the amount of information that can concurrently occupy short-term memory. There is good empirical support for such a constraint (Broadbent, 1958; Cowan, 2001; McElree & Doshier, 1989), which is argued by many authors to correspond to an architectural component, the “focus of attention” (McElree, 2006; Oberauer & Kliegl, 2006). The focus of attention is sufficiently limited so as to effectively exclude information unassociated with local processing events, like maintaining a filler phrase while other constituents of the sentence are being interpreted. If the filler’s contents are not available in the focus of attention, then they must be retrieved to be integrated with the verb (Lewis & Vasishth, 2005; McElree et al., 2003). Retrieval accounts must thus answer the question of how the correct constituent is retrieved from memory once a suitable verb is reached. Several sources for these cues are supported in the literature. Cues for retrieval could in principle derive from specific lexical or contextual information supplied by the verb and other heads near the gap site. For example, a verb like *amuse* selects for an animate internal argument, and therefore animacy is a plausible cue for retrieval of the filler. The evidence that we review below suggests that some but not all kinds of verb-particular information constrain the retrieval of the filler. The cues used to retrieve the filler could also derive from grammatical rules or parsing routines, leading to cues that are independent of particular sentences. For example, filler phrases are typically (but not always) of category NP, and they (always) occupy a specifier position high in

the clause. Thus those cues could be used alongside the lexically specific cues.

An attractive possibility is that a verb’s selectional restrictions provide the necessary cues to retrieve the filler. The semantic anomaly studies discussed above (Garnsey et al., 1989; Traxler & Pickering, 1996) provide a useful constraint on this idea. In those studies, an anomalous combination of verb and filler was detected either at the verb or shortly thereafter. In Traxler and Pickering’s (1996) study, the plausibility contrast was evident in first fixations. This suggests that even if a filler is not a good object for the verb, it may nonetheless be retrieved at the verb position. Caution is required here, as previous studies have not entirely distinguished between purely selectional restrictions and violations of real-world expectations. For example, the verb “punish” requires an animate theme, and thus “punish the rock” is ill-formed. In contrast, a phrase like “shoot the garage” (Traxler & Pickering, 1996) is anomalous because it is unusual, but not impossible. Based on the absence of plausibility mismatch effects, Boland, Tanenhaus, and Garnsey (1995) found that active dependency completion was blocked for implausible verb-object combinations if the verb was a verb like “remind”, and the displaced object was an inanimate noun like “movie”. Because verbs like “remind” can combine with both a NP object and a clause, the parser may thus be influenced by the possibility that the filler could be combined with a plausible host in the upcoming clause. In this way, the semantic anomaly effect could be blocked. Pickering and Traxler (2001) provided data convergent with Boland et al. (1995) and argued that these results favour a mechanism by which selectional features could be used to filter dependency completion. In the example given, the filler “movie” lacks the animacy feature that “remind” selects for in its

<sup>2</sup>McKoon and Ratcliff (1994) object to the reactivation interpretation of cross-modal lexical priming in these sentential contexts on other grounds. They argue, instead, that the goodness-of-fit of the lexical decision target is an important determinant of the reaction time (RT) patterns in these experiments. Thus lower RTs are observed in postverbal positions not because of priming from a reactivated filler, but because the lexical decision target is itself a noun that could occur in the postverbal position. The probe recognition study of McElree (2001) is less vulnerable to this objection, since the probe words were adjectives, and the task was to determine whether a synonymous word was encountered in the prior context.

NP argument. Consequently the parser blocks combination with the filler early.

Subcategorization information may also be used as a cue, and indeed Van Dyke and McElree (2006) have suggested more generally that syntactic cues may gate semantic cues in the retrieval of constituents from memory. A crucial case to test this claim is intransitive verbs, like “arrive”, which do not take a direct object. Presently evidence remains somewhat mixed on whether fillers give rise to active effects at verbs that are purely intransitive (Omaki, Lau, Davidson White, & Phillips, 2010; Staub, 2007). However, Frazier and Clifton (1989) and Pickering and Traxler (2003) have both found evidence that fillers give rise to active effects for verbs that are preferred intransitives but nonetheless optionally transitive (cf. Stowe, Tanenhaus, & Carlson, 1991). This suggests that active dependency completion can lead the parser to postulate gaps that the verb’s argument structure statistics might otherwise disprefer.

The idea that lexically specified information alone cues retrieval meets a further challenge in some recent data from Wagers and Phillips (2009). They compared the semantic anomaly effect for short versus long dependencies. Dependencies were serially lengthened by adding a five-word prepositional phrase inside the subject. In conditions with short dependencies there was an immediate anomaly effect at the verb, whereas in conditions with long dependencies the anomaly effect was absent at the verb and was not reliable until after the gap site (an indirect object gap that occurred a few words later in the sentence). This finding is surprising because one might initially assume that effects that are triggered by verb retrieval cues ought to consistently occur at the same position—that is, near the verb. If the retrieval occurs much later than the verb, the question arises how the verb-specific information itself survives in the focus of attention. The shift in the locus of the effect raises the possibility that active

dependency processing occurs only at very short distances and, possibly, that previous studies may have systematically undersampled longer distance dependencies. Past studies of filler–gap dependency construction have indeed been heavily skewed to very short distances. A sample of 21 experiments<sup>3</sup> performed between 1985 and 2004 revealed that, on average, 2.9 words or 1.8 constituents linearly intervened between filler and gap. These experiments spanned different paradigms: 10 self-paced reading studies, 5 eye-tracking studies, 3 sensicality monitoring studies, 3 cross-modal priming studies. The strong bias to examine short dependencies calls into question the generality of the idea that filler–gap dependency formation reliably occurs at verbs that intervene between fillers and gaps. If there is variation in the timing of filler–gap dependency completion, then the richness of the set of retrieval cues available to the comprehender might vary depending on when the retrieval actually occurred with respect to the verb.

### The current study

The research reviewed above presents two different views of how unbounded dependencies are completed. On one view, the filler phrase is actively maintained throughout the course of the sentence. By virtue of this (potentially costly) maintenance, a pressure arises to rapidly complete the dependency, and attachments are attempted at grammatically licit gap sites. On the other view, maintaining the filler phrase exceeds the system’s capacity, and it must be removed from the focus of attention until it is reactivated at a potential gap host. Although discussions of unbounded dependency processing have tended to adopt the language of *either* the maintenance *or* the retrieval/reactivation view, there is a natural compromise (articulated by Fiebach et al., 2002, for example). Namely, maintenance could apply to some but not all features of the filler encoding, with reactivation applying

<sup>3</sup>Tanenhaus, Stowe, and Carlson (1985); Stowe (1986); Swinney, Ford, Frauenfelder, & Bresnan (1988); Frazier and Clifton (1989); Boland et al. (1995); Pickering, Barton, and Shillcock (1994); Traxler and Pickering (1996); Clahsen and Featherston (1999); McElree (2000); Sussmann and Sedivy (2003); Aoshima, Phillips, and Weinberg (2004); Conklin, Koenig, and Mauner (2004); Lee (2004).

to its unmaintained components. The present study seeks to explore a hybrid maintenance/retrieval view by testing what aspects of a filler phrase contribute to active dependency formation at longer dependency lengths and what aspects become attenuated.

A decomposition of filler-gap dependency formation into more fine-grained operations would seem to be necessary to answer the questions about memory contents raised in Section 1.2. In the present set of studies, we test active dependency formation at different serial lengths, as in Wagers and Phillips (2009), but we also use three different indices of dependency formation, each of which presumably draws upon different sources of information: the filled-gap effect, the semantic anomaly effect, and an index based on subcategorization.

## EXPERIMENT 1: THE FILLED-GAP EFFECT

### Rationale

In the first experiment, we used the filled-gap effect design to test whether active dependency formation persists over long dependency lengths.

We used a modified version of the filled-gap design devised by Lee (2004), which manipulates the need for active gap search while holding constant the presence of a filler in memory. Lee contrasted displaced noun phrases (NPs) with displaced prepositional phrases (PPs). In unmarked word order, only displaced NPs can occupy subject or object position. Relative to reading a displaced PP, reading a displaced NP should cause difficulty if subject or object position is filled by another phrase. Example 8 illustrates a simplified version of our experimental contrast.

- (8) (a) The **stones** which the pilgrim toppled  
the cairn for \_\_\_\_ ...  
(b) The **stones** for which the pilgrim  
toppled the cairn \_\_\_\_ ...

In both Example 8(a) and 8(b), it is possible to form a dependency at the critical verb (“topple”).

But only when the displaced category is a PP (8b) can readers unambiguously predict the gap site. If the comprehender hypothesizes that the gap occupies the direct object position at the verb in 8(a), then encountering the overt NP “the cairn” should disconfirm that hypothesis. Meanwhile, the PP category of the filler (8b) should prevent the comprehender from hypothesizing that there is a gap in direct object position. Consequently processing the direct object is predicted to be easier in 8(b) than in 8(a). Lee’s (2004) design represents an improvement over previous filled-gap effect designs, because it contrasts sentences that both have filler-gap dependencies and identical interpretations.

We created contrasts like those in Example 8 for three different dependency lengths: a short single-clause dependency, a longer single-clause dependency, and a longer biclausal dependency. In this design, the filled-gap effect becomes a test of whether category information is maintained over the course of a long dependency. A sample materials set is given in Table 1.

### Method

#### *Participants, materials, and procedure*

Participants were 36 native speakers of English from the University of Maryland community. Each received \$10 for taking part in the experiment, which lasted 45 minutes.

This experiment crossed the factors *filler category* and *length* in a  $2 \times 3$  factorial design. Filler category was either “NP” or “PP”. Length was “short”, “+PP”, or “+CP” (where CP stands for complementizer phrase, the syntactic category of an embedded sentence). In the short condition, a two-word subject and an adverb separated the filler from the verb. In +PP conditions, the subject was modified by a five-word prepositional phrase (PP). Finally, in +CP conditions, the sentence inside the short condition relative clause was embedded under a verb like “say” or “think”, verbs that do not readily host a NP object. We chose such verbs in order to minimize the possibility of premature dependency completion at the main clause verbs.

Table 1. Sample materials set for Experiment 1

Length	Filler category	Sentence
Short	Item frame: "... [T]he chemicals ... _____ ... might still become contaminated ... "	
	NP	... <b>which</b> the technician carefully prepared <u>the clean tubes</u> for ____ while wearing a mask ...
	PP	... <b>for which</b> the technician carefully prepared <u>the clean tubes</u> ____ while wearing a mask ...
	+PP	... <b>which</b> the technician at the medical research facility carefully prepared <u>the clean tubes</u> for ____ while wearing a mask ...
+CP	Item frame: "... [T]he chemicals ... _____ ... might still become contaminated ... "	
	NP	... <b>which</b> the technician at the medical research facility carefully prepared <u>the clean tubes</u> ____ while wearing a mask ...
	PP	... <b>for which</b> the technician at the medical research facility carefully prepared <u>the clean tubes</u> ____ while wearing a mask ...
	+PP	... <b>which</b> the technician at the medical research facility carefully prepared <u>the clean tubes</u> for ____ while wearing a mask ...
+CP	Item frame: "... [T]he chemicals ... _____ ... might still become contaminated ... "	
	NP	... <b>which</b> the young biologist said that the technician carefully prepared <u>the clean tubes</u> for ____ while wearing a mask ...
	PP	... <b>for which</b> the young biologist said that the technician carefully prepared <u>the clean tubes</u> ____ while wearing a mask ...
	+PP	... <b>which</b> the young biologist said that the technician carefully prepared <u>the clean tubes</u> for ____ while wearing a mask ...

Note: Each of the six complete sentences in this set is derived by filling in the underscored blank in the item frame with the phrases in the Length  $\times$  Filler cells. PP = prepositional phrase; CP = complementizer phrase; NP = noun phrase. Filler phrases are indicated in bold font, the critical filled-gap region is indicated by double underlining, and the gap is indicated by an underscore. In the short length conditions, a five-word preamble also preceded the item frame to control for the ordinal position of the critical region. For this set, the preamble was "The biologist was distressed that ...".

Like the +PP condition, the embedding clause was composed of five words. Table 1 illustrates one full materials set. The full materials list, for all experiments, is given in the supplementary materials.

The object NP comprised the critical region in this design because it was where a filled-gap effect could first occur. Object NPs were always three-word sequences of the form "determiner–adjective–noun" or "determiner–noun–noun" (i.e., a noun–noun compound). The relative clause always contained an adverbial phrase in final position (e.g., in the example in Table 1, "while wearing a mask"). This phrase provided a strong signal that a constituent was missing, by means of the sequence *preposition–preposition* (e.g., "for while").

Finally, the short conditions were always preceded by a five-word embedding preamble so that the position of the critical region was ordinarily matched across conditions. For the sample set in Table 1, the short preamble was "The biologist was distressed that ...".

Sentences were presented on a desktop PC using the Linger software package (Rohde, 2003) in a self-paced, word-by-word, moving window paradigm (Just, Carpenter, & Wooley, 1982). Each trial began with a screen presenting a sentence in

which the words were masked by dashes while spaces and punctuation remained intact. Each time the participant pressed the space bar a word was revealed, and the previous word was remasked. A *yes/no* comprehension question appeared all at once on the screen after each sentence. The "f" key was used for "yes", and the "j" key was used for "no". On-screen feedback was provided for incorrect answers. Participants were instructed to read at a natural pace and answer the questions as accurately as possible. In all of the self-paced reading experiments presented here, participants were never informed that sentences would contain grammatical errors. Order of presentation was randomized for each participant. Seven practice items were presented before the beginning of the experiment.

### Analysis

The following analysis procedure was applied to all data sets reported in this paper, so we describe it in detail here.

Both reaction time (RT) and accuracy data were analysed with mixed-effects modelling. We estimated models using the lme4 package (Bates, Maechler, & Bolker, 2011) in the R software environment (R Development Core Team, 2011). Accuracy data were treated with a logistic

model; RT data were logarithmically transformed before being entered into a linear model. Both experimentally manipulated factors were coded with orthogonal contrasts. The two levels of the *filler category* condition were coded with sum contrasts ( $\pm 1$ ; where +1 represents NP fillers). The three levels of *length* were coded using Helmert contrasts: *Short* conditions were compared to the mean of +PP/+CP conditions (reported as the *length.long* coefficient); and +PP conditions were directly compared to +CP conditions (reported as the *length.clause* coefficient). This coding makes sense in view of the questions we posed: Are short dependencies different from long dependencies? Are different kinds of length relevant?

Reading time data were treated for outliers in the following manner, adopting suggestions of Baayen and Milin (2010). Initially no observations were excluded from the RTs. Linear models were fitted to the log reading times, and the residuals were inspected for fit to the normal distribution. Observations with absolute standardized residuals greater than 2 were then excluded. This method thus identifies outliers on the criterion of how much they stress the assumptions of the statistical model. On average, our analyses excluded 5% of observations. Results of untrimmed and residual-trimmed analyses are presented as coefficient tables.<sup>4</sup>

We first fitted models with crossed random factors; these sometimes failed to converge and never presented a different profile of coefficients or *t/z*-values. We report therefore the simpler models with subject/item intercepts only. *P*-values were determined by Markov Chain Monte Carlo sampling of the posterior distribution from the model with no nested experimental factors, using the *pvals.fnc* function in R (Baayen, 2008).

Table 2. *Experiment 1: Comprehension accuracy*

<i>Filler category</i>	<i>Length</i>			<i>Mean</i>
	<i>Short</i>	<i>+PP</i>	<i>+CP</i>	
NP	88 (2)	87 (3)	78 (3)	84
PP	88 (2)	84 (3)	77 (3)	83
Mean	88	86	77	84

*Note:* Accuracy expressed as average percentage correct over participants, with row, column, and grand means. Standard error of the cell means across subjects is reported in parentheses. PP = prepositional phrase; CP = complementizer phrase; NP = noun phrase. *N* = 36.

## Results

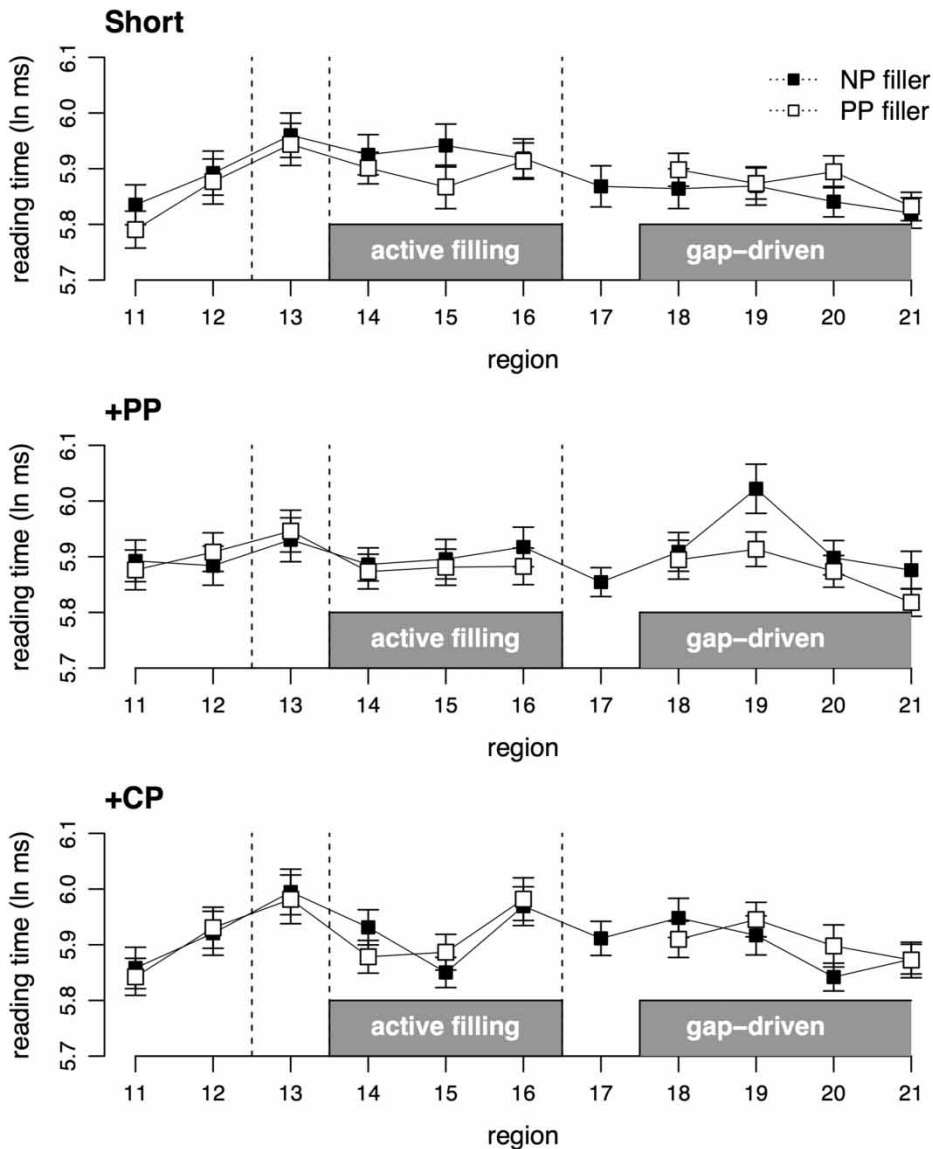
### *Comprehension accuracy*

Mean accuracy was 84%. Comprehension accuracy by condition is reported in Table 2. There was no effect of filler type. The only reliable contrasts were in length: both between short conditions and the average of the long conditions (*length.long*:  $0.58 \pm 0.23$ ,  $z = 2.5$ ,  $p < .05$ ) and between the two long conditions (*length.clause*:  $0.89 \pm 0.24$ ,  $z = 3.8$ ,  $p < .001$ ). A paired comparison between just the short and +PP conditions was not significant. In other words: (a) +CP-lengthened sentences were harder to understand; (b) the filler type did not affect performance on the comprehension questions.

### *Reading times*

Reading time results are reported in Figure 1 as log-transformed RTs from the untrimmed dataset. Words 14–16 are annotated as “active filling regions”, since these occur before direct evidence of the gap position. Words 18–20, occurring after unambiguous evidence in the input for the gap position, are annotated as “gap-driven” regions. We report four regions of analysis: the

<sup>4</sup>In any experiment of modest complexity, a number of implicit choices are made by the analyst, more than the explicit steps reported in a research report. Simmons, Nelson, and Simonsohn (2011)'s recent study illustrates how such flexibility in the experimental pipeline can lead to increased effective false-positive rates. We have attempted to disclose all major steps in the analysis of the data we present. In the supplementary materials, we supply coefficient tables for the models over untrimmed data. On the first author's web site, we supply the raw data accompanied by R scripts used to process it, so that the reader can replicate or modify the analyses reported here.



**Figure 1.** Experiment 1 reading times. Each panel corresponds to a different length condition. PP = prepositional phrase; CP = complementizer phrase. Closed symbols represent noun phrase (NP) fillers, open symbols PP fillers. Reading times are reported in natural log ms (ln ms). Error bars indicate standard error of the cell means.

Sample sentence:

[...] the chemicals (for) which [...] <sub>CP</sub> the technician [...] <sub>PP</sub> carefully<sub>12</sub> prepared<sub>13</sub> the<sub>14</sub> clean<sub>15</sub> tubes<sub>16</sub> (for<sub>17</sub>) while<sub>18</sub> wearing<sub>19</sub> a<sub>20</sub> mask<sub>21</sub> might still become contaminated.

verb (Word 13) and preverb region (Word 12), the active filling region, and the gap-driven region. Coefficient tables for the linear mixed-effects models on this regions are reported in Table 3.

In brief, evidence of active dependency construction was found for all dependency lengths, with some variation in the timing and the size of the effect.

**Table 3.** *Experiment 1: Mixed-effect model coefficient tables*

Region and coefficient names	Residual trimmed dataset				Untrimmed dataset			
	Estimate	SE	t	p <sub>MCMC</sub>	Estimate	SE	t	p <sub>MCMC</sub>
Adverb								
Intercept	5.87	0.051	110	<.001	5.90	0.056	100	<.001
Filler	−0.004	0.008	−0.45	.65	−0.003	0.011	−0.29	.77
Len.long:								
Sht − mean(CP,PP)	−0.049	0.017	−2.9	<.005	−0.026	0.023	−1.1	.26
Len.cls: PP − CP	−0.028	0.020	−1.4	.16	−0.029	0.027	−1.1	.27
Filler: len.long	0.039	0.017	2.3	.02	−0.016	−0.023	−0.27	.50
Filler: len.cls	−0.001	0.020	−0.07	.95	−0.007	0.027	0.70	.81
Verb								
Intercept	5.91	0.051	120	<.001	5.96	0.055	110	<.001
Filler	−0.003	0.009	−0.40	.69	0.002	0.012	0.19	.85
Len.long:								
Sht − mean(CP,PP)	−0.001	0.019	−0.07	.95	−0.011	0.026	−0.43	.66
Len.cls: PP − CP	−0.031	0.022	−1.4	.15	−0.050	0.030	−1.7	.10
Filler: len.long	0.020	0.019	1.1	.29	0.009	0.026	0.34	.74
Filler: len.cls	−0.013	0.022	−0.58	.56	−0.014	0.030	−0.48	.63
Active filling object NP								
Intercept	5.87	0.038	140	<.001	5.91	0.046	130	<.001
Filler	0.01	0.004	2.4	.02	0.009	0.006	1.6	.11
Len.long:								
Sht − mean(CP,PP)	0.003	0.008	0.35	.73	0.008	0.013	0.67	.50
Len.cls: PP − CP	−0.029	0.010	−2.7	.008	−0.03	0.015	−1.9	.06
Filler: len.long	−0.001	0.009	−0.09	.93	0.01	0.013	0.93	.35
Filler: len.cls	0.001	0.010	0.10	.92	0.01	0.015	0.67	.50
Gap-driven								
Intercept	5.86	0.04	156	<.001	5.90	0.041	140	<.001
Filler	−0.006	0.004	−1.5	.15	.001	0.006	0.08	.93
Len.long:								
Sht − mean(CP,PP)	−0.02	0.009	−2.2	.03	−.041	0.013	−3.2	<.005
Len.cls: PP − CP	−0.009	0.011	−0.83	.41	.008	0.015	0.57	.57
Filler: len.long	−0.020	0.009	−2.1	.04	−.024	0.013	−1.8	.07
Filler: len.cls	0.020	0.011	1.8	.07	0.032	0.015	2.2	.03

*Note:* Left: models on residual-trimmed data. Right: models on untrimmed data. PP = prepositional phrase; CP = complementizer phrase; NP = noun phrase; len = length; cls = clause; sht = short; MCMC = Markov Chain Monte Carlo. In filler contrasts, NP conditions are positive (+1); in length.long contrasts, short conditions are +2/3; in length.clause contrasts, +PP conditions are (+1/2). *P*-values are reported rounded to 2 decimal places, except where they are less than .005. Dark shading highlights coefficients, excluding the intercept, with a *p*-value less than .05; light shading for .05 ≤ *p* < .10. *SE* = standard error.

No effects were observed on the verb, though there was a tendency for PP conditions to be read more quickly. This suggests that we had a good baseline for evaluating the filled-gap effect in the critical active filling region. An interesting pattern of reading times in the adverb deserves comment, though it does not seem to impinge upon our

interpretation of the filled-gap effect: Short conditions were read more quickly than either long condition, and there was an interaction with filler category. Inspection of the means suggests that this reflects a slowdown for PP fillers in the PP length condition ( $t = -1.9$ ,  $p < .05$ ). Our ability to interpret this interaction confidently is limited,

since items were not matched prior to the adverb (a consequence of our design). The difference may, however, reflect the fact that two PP constituents were recently processed (the filler, and the subject postmodifier). If we incorporate reading times from the previous word as a covariate, the pairwise difference between the two filler conditions in +PP-length conditions is no longer significant ( $t = -1.3$ ,  $p = .19$ ), suggesting that spillover is a relevant concern.

The critical region of interest is the object NP, which we dubbed the “active filling” region. Here we observed a significant, positive difference between NP conditions and PP conditions. We interpret this as a filled-gap effect, following Lee (2004). Overall, CP conditions were also read more slowly than PP conditions. Crucially, there was no interaction of the filler effect with either of the length contrasts. Inspection of the word-by-word means in Figure 1 makes it clear that, for each of the length conditions, there was a more punctate effect that varied in onset: Region 15 for short conditions, Region 16 for +PP-length conditions, and Region 14 for +CP-length conditions. The effect sizes for each of the maximal pairwise comparisons, expressed as difference in mean log RT normalized by standard deviation log RT (Cohen's  $d$ ), was 0.16 for short conditions, 0.08 for PP conditions, and 0.14 for CP conditions.

Finally, the gap-driven region was defined as the three words following unambiguous evidence for the gap. Overall, short conditions were read more quickly. There was no overall effect of filler category, though there was a nonsignificant tendency for NP filler conditions to be read more quickly. However, an interaction between filler category and length amplified the NP filler speed-up in the short condition and reversed it in the +PP-length conditions: This is apparent in the RT “spike” in Figure 1 for the NP conditions.

## Discussion

In this experiment, we tested whether participants actively constructed the filler-gap dependency at all dependency lengths by contrasting NP fillers

with PP fillers. This contrast tests the availability of syntactic category information across the three dependency lengths. NP-filler sentences were expected to be read more slowly than PP-filler sentences in the direct object region if the dependency was constructed actively—that is, if the dependency was constructed prior to direct evidence for the position of the gap. This filled-gap effect was found for all dependency lengths. Participants thus actively constructed the filler-gap dependency, even for long biclausal or PP-extended monoclausal dependencies. The filler-gap distances tested in this experiment greatly exceeded those used in most studies completed to date using this paradigm: A total of 9 words and 5 or 6 major constituents linearly intervened between the filler and the critical verb for both +PP- and +CP-length conditions, compared with an average of 2.9 words in previous studies. We can thus conclude more confidently that filler-gap dependency construction itself remains active across clause boundaries. It appears that sensitivity to the filler's grammatical category is retained when comprehenders actively posit gaps, since participants reliably distinguished the NP-filler and PP-filler conditions.

The filled-gap effect was observed as a punctate effect, whose onset varied across conditions. That is, the entire direct object region was not rendered more difficult, but one particular word was. Interestingly, this effect appeared earliest for the +CP-length dependency on the determiner of the direct object. It appeared in the middle word of the direct object for short dependencies (an adjective or a noun) and on the last word of the direct object for +PP-length dependencies (a noun). Due to the limited temporal resolution of the self-paced reading technique, we cannot draw clear conclusions about the finer time-scale dynamics of the process. However, in every condition pair, the effect occurred during the processing of the direct object, so we can still conclude that the gap was posited before direct evidence of the missing constituent's location. One possibility for explaining the variation in timing is that different sets of processing events immediately precede the verb in the different conditions. If the status or difficulty of parsing decisions made prior to the

VP onset affects the point when the processor begins to evaluate a hypothesized filler–gap dependency, then variation in the observed timing of the filled-gap effect would be expected. However, it is worth emphasizing that there was only an effect of filler category and no interaction with length.

These results provide clear evidence of active dependency formation for all dependency lengths. However, there was also a reading time slowdown in NP-filler sentences in the postgap region of +PP-length conditions. This observed contrast was unanticipated. It could be taken as evidence that linking the gap with the NP filler is more difficult when the subject of the clause also contains a PP. A possible explanation for this difficulty stems from the fact that, once the direct object analysis has been proven incorrect, a gap-containing PP may be predicted. It may be that establishing an expectation for this PP could be more difficult in sentences that already contain a similar phrase (i.e., in the +PP-length sentences). This suggestion is plausible if there is retrieval interference for attachment (Lewis & Vasishth, 2005), but future studies would be required to test this conjecture.

The present results suggest that category information is well preserved across different dependency lengths. The finding is compatible with the idea that basic category information could be either maintained or retrieved in an interference-robust manner, but it does not address whether other contents of a filler phrase are maintained. Therefore, in Experiment 2 we make the measure of dependency formation contingent upon semantic features of the filler, in order to test whether those are preserved across different lengths.

## EXPERIMENT 2: SEMANTIC ANOMALY DETECTION

### Rationale

In the second experiment, we used a different measure of active dependency formation: the semantic anomaly effect (Garnsey et al., 1989; Traxler & Pickering, 1996). Example 9 illustrates

a semantic anomaly contrast. If the filler is initially treated as the direct object of the verb “erect”, then the initial interpretation describes a plausible situation in 9(a), but not in 9(b).

- (9) (a) The **monument which** the pilgrim erected \_\_\_\_ to appease the gods ...  
 (b) The **gods which** the pilgrim erected the monument to appease \_\_\_\_ ...

For this reason, a reading time slowdown is expected at the verb or shortly thereafter.

We created contrasts like Example 9 for three different dependency lengths: a short single-clause dependency, a longer single-clause dependency, and a longer biclausal dependency. If the semantic details of the filler are maintained in the focus of attention to participate in active dependency formation across the different dependency lengths, then we expect to observe a semantic anomaly effect either at the critical verb or shortly thereafter.

## Method

### Materials and procedure

Participants were 36 native speakers of English from the University of Maryland community who received \$10 for taking part in the experiment.

This experiment crossed the factors filler plausibility and length in a  $2 \times 3$  factorial design. Filler category was either “plausible” or “implausible”. Length was “short”, “+PP”, or “+CP”, identical to Experiment 1. Table 4 illustrates a full materials set.

In this design, the verb is the critical region because it is where the semantic fit of the filler can first be detected. However, because such effects are prone to spillover in self-paced reading tasks, the direct object region was made 4 words in length. Consequently, bottom-up evidence for a missing constituent would also be unavailable for some time after the verb. A reading time slowdown in the implausible conditions either at the verb or in the 4-word direct object region could be considered an effect of active dependency construction.

Table 4. Sample materials set from Experiment 2

Length	Filler plausibility	Sentence
Short	Plausible	The <b>acorns</b> which the squirrels quickly <u>crammed</u> their small puffy cheeks with ____ before scurrying out of the park had fallen from a sickly oak.
	Implausible	The <b>cats</b> which the squirrels quickly <u>crammed</u> their small puffy cheeks with ____ before scurrying out of the park were basking in the sun.
+PP	Plausible	The <b>acorns</b> which the squirrels with the bushy black tails quickly <u>crammed</u> their small puffy cheeks with ____ before scurrying out of the park had fallen from a sickly oak.
	Implausible	The <b>cats</b> which the squirrels with the bushy black tails quickly <u>crammed</u> their small puffy cheeks with ____ before scurrying out of the park were basking in the sun.
+CP	Plausible	The <b>acorns</b> which the experienced naturalist concluded that the squirrels quickly <u>crammed</u> their small puffy cheeks with ____ before scurrying out of the park had fallen from a sickly oak.
	Implausible	The <b>cats</b> which the experienced naturalist concluded that the squirrels quickly <u>crammed</u> their small puffy cheeks with ____ before scurrying out of the park were basking in the sun.

Note: The plausibility manipulation is indicated with bold face, the critical region with double underlining, and the gap position with an underscore. PP = prepositional phrase; CP = complementizer phrase.

The verbs and fillers chosen in this experiment were the same as those in Wagers and Phillips (2009). The verbs were spray-load-type alternating locative verbs (Fraser, 1971; Rappaport & Levin, 1986), which take two internal arguments: the figure and the ground arguments, arguments that correspond to a moving object and a location, respectively. For example, in the sentence “The squirrels *crammed* their cheeks with acorns”, the argument *their cheeks* is the ground, and *acorns* is the figure. All verb phrases in the experiment allow alternation between ground-figure and figure-ground argument orderings (e.g., “crammed acorns into their cheeks”), but the ground-figure ordering of arguments was used in all experimental sentences. Crucially the verb-filler combinations were normed so that the implausible fillers were equally implausible as either figure or ground (see Wagers & Phillips, 2009, for norming details). Because of this, comprehenders could not use the plausibility of the filler as a particular argument of the verb to guess its grammatical role. Thus by factoring out cues to the filler’s grammatical role from the verb, we

could be confident about attributing plausibility effects in the reading times before the gap to active dependency completion processes.

Moreover, this norming responds to the observations of Boland et al. (1995) and Pickering and Traxler (2001) that available alternative argument positions could mute a semantic anomaly.<sup>5</sup> To satisfy the double constraint of anomaly as either figure or ground argument, it proved to be practically necessary to make the anomaly one that depended on real-world knowledge, rather than selectional restrictions. Partially as a consequence of this constraint, we were able to avoid a confound with the animacy of the filler. The processing complexity of object relative clauses (RCs) is sensitive to the animacy of both the filler and the relative clause subject. In particular, when the filler and RC subject positively match in animacy, the RC is more difficult to comprehend (Traxler, Morris, & Seely, 2002). In 18 of our item sets, both the plausible and implausible fillers were inanimate; in five sets, the implausible filler was animate, and the plausible filler was animate; in one set, the implausible filler was inanimate, and the plausible filler

<sup>5</sup>Although the design of the current study required that the verbs be distinct from the verbs in Experiment 1, there are several similarities. The verbs in both Experiment 1 and Experiment 2 allow two internal arguments (theme/benefactive in Experiment 1; figure/ground in Experiment 2). Verbs in both experiments allow alternative orderings of the arguments. Finally verbs in both experiments only require one of the arguments to be realized syntactically. Thus the relative complexity of the two classes is plausibly equivalent.

Table 5. Experiment 2: Comprehension accuracy

Filler category	Length			Mean
	Short	+PP	+CP	
Plausible	91 (2)	95 (2)	89 (3)	92
Implausible	90 (3)	94 (2)	88 (3)	90
Mean	91	94	89	91

Note: Accuracy expressed as average percentage correct over participants, with row, column, and grand means. Standard error of the cell means across subjects is reported in parentheses. PP = prepositional phrase; CP = complementizer phrase.  $N=35$ .

was animate. In order to test for possible effects of animacy, we performed analyses identical to the reported analyses except that they excluded the sets involving extraction of an animate argument; these did not reveal qualitatively different results, and patterns of significance were preserved. Therefore we only report analyses with all items included. An analysis without animate extractions can be replicated from the materials provided on the authors' web site.

Analysis proceeded identically to the analysis in Experiment 1. Filler conditions were sum coded for filler category (implausible = +1). Helmert contrasts for length were identical to the length contrasts in Experiment 1. One participant was excluded from analysis due to exceptionally low accuracy (63%; greater than 2 standard deviations from the mean, expressed in logits).

## Results

### Comprehension accuracy

Mean comprehension accuracy was 91%. Comprehension accuracy by condition is reported in Table 5. The only significant difference was between PP-length conditions and CP-length conditions, the former being answered more accurately ( $0.79 \pm 0.33$ ,  $t = 2.4$ ,  $p < .05$ ).

### Reading times

Reading time results are reported in Figure 2. Coefficient tables from the linear mixed-effects models of the four regions of analysis are given in

Table 6. Words 12–16, for which a slowdown due to implausibility would count as evidence for active dependency completion, are annotated as “active filling” regions and are analysed separately as the verb and the direct object regions. Words 18–20, for which a slowdown due to implausibility would show that participants recognized the direct evidence for the gap location, are annotated as “gap-driven” regions. Word 11 was analysed as the precritical adverb region, and there were no effects in that region.

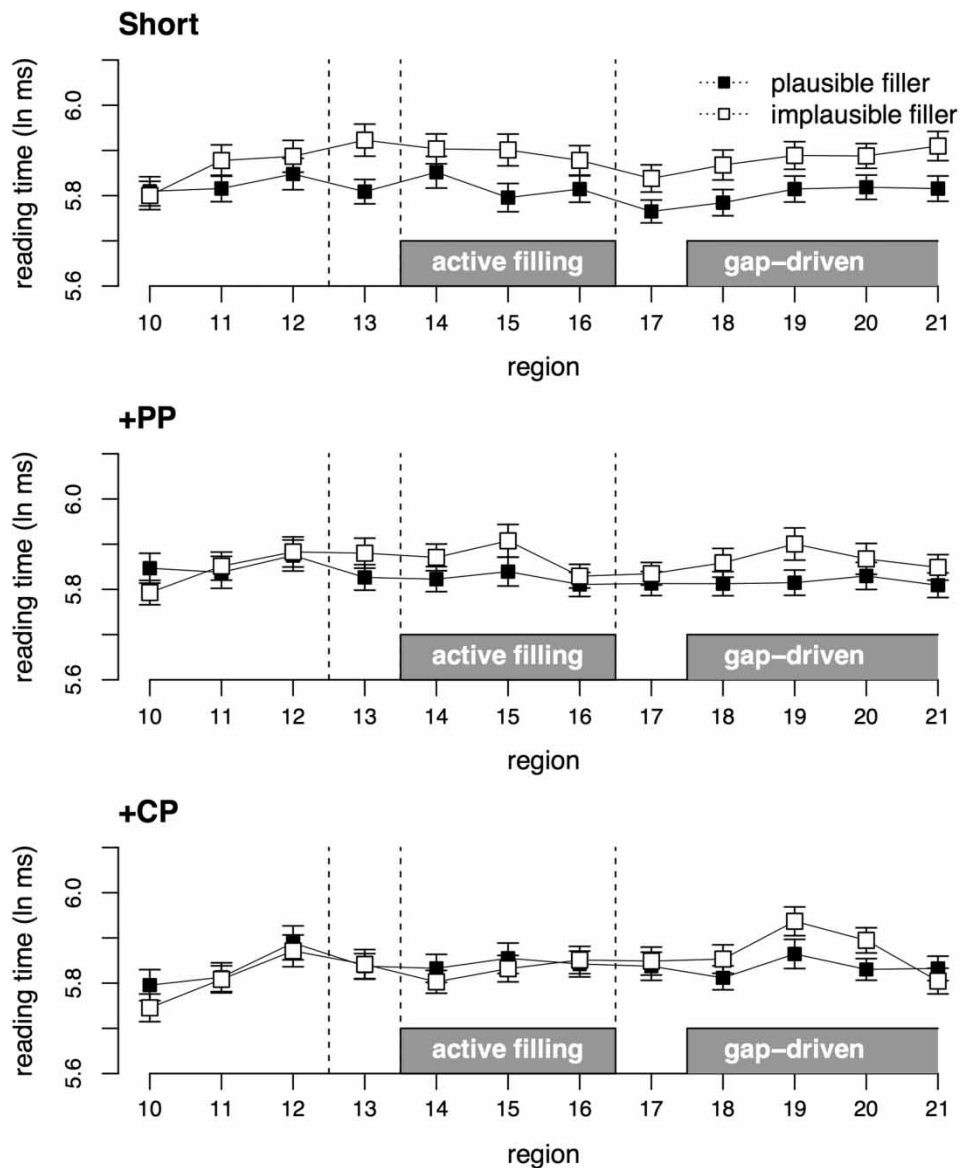
Active filling effects were most robustly found in the short length conditions. In the active filling regions (verb and direct object), there was a slowdown for the implausible filler conditions, indicating sensitivity to the filler's fit as an argument of the critical verb. In the verb region, this slowdown was only present in short conditions. In the direct object region, it was present as a simple effect, but was smaller for the +PP/+CP-length conditions overall. A pairwise test of the filler plausibility contrast, restricted to just +CP-length conditions, showed that it was absent for that length ( $t = -0.41$ ,  $p = .70$ ).

However, in the gap-driven regions (Words 18–20), both the +PP- and +CP-length conditions showed clear evidence of sensitivity to the plausibility manipulation. Here the opposite pattern was observed compared to the active region: The effect of plausibility was greatest in this region for +CP-length conditions.

Overall, this pattern of results contrasts with the filled-gap effect in Experiment 1, which was detected for all length conditions. It replicates the difference observed by Wagers and Phillips (2009) for short versus +PP-length dependencies, and it extends the finding to +CP-length dependencies.

## Discussion

The goal of this study was to test whether the semantic anomaly index of active dependency formation is observed at long dependencies. At all dependency lengths we found that comprehenders were ultimately sensitive to the semantic fit of the filler with the verb that hosted its gap. However, the onset of that sensitivity, as reflected in slower reading times



**Figure 2.** Experiment 2 reading times. Each panel corresponds to a different length condition. PP = prepositional phrase; CP = complementizer phrase. Closed symbols represent plausible fillers, open symbols implausible fillers. Reading times are reported in natural log ms (ln ms). Error bars indicate standard error of the cell means.

Sample sentence:

[... x ...]<sub>10</sub> quickly<sub>11</sub> crammed<sub>12</sub> their<sub>13</sub> small<sub>14</sub> puffy<sub>15</sub> cheeks<sub>16</sub> with<sub>17</sub> \_\_\_\_ before<sub>18</sub> scurrying<sub>19</sub> out<sub>20</sub> [of the park ...]<sub>>20</sub>

for implausible verb–filler combinations, varied with dependency length. For long biclausal dependencies, comprehenders only showed sensitivity to plausibility once there was direct evidence in the

input for the location of the missing constituent. This contrasted strikingly with short monoclausal dependencies, where sensitivity arose immediately at the verb. Monoclausal dependencies that were

**Table 6.** *Experiment 2: Mixed-effect model coefficient tables*

Region and coefficient names	Residual trimmed dataset				Untrimmed dataset			
	Estimate	SE	t	p <sub>MCMC</sub>	Estimate	SE	t	p <sub>MCMC</sub>
Adverb								
Intercept	5.80	0.04	140	<.001	5.90	0.055	110	<.001
Filler	0.006	0.008	0.77	.44	−0.003	0.011	−0.29	.34
Len.long:								
Sht – mean(CP,PP)	0.001	0.016	0.06	.95	−0.029	0.027	−1.1	.40
Len.cls: PP – CP	0.018	0.019	0.95	.34	−0.026	0.023	−1.1	.19
Filler:len.long	0.011	0.02	0.71	.48	−0.007	0.027	−0.03	.20
Filler:len.cls	−0.005	0.01	−0.26	.79	0.016	0.023	0.69	.75
Active filling verb								
Intercept	5.84	0.043	135	<.001	5.88	0.046	128	<.001
Filler	−0.004	0.008	−0.51	.62	0.005	0.011	0.44	.67
Len.long:								
Sht – mean(CP,PP)	−0.018	0.018	−1.0	.33	−0.012	0.024	−0.51	.61
Len.cls: PP – CP	−0.0002	0.020	0.01	.98	−0.001	0.027	−0.05	.96
Filler:len.long	0.043	0.018	2.4	.02	0.022	0.023	0.93	.35
Filler:len.cls	0.003	0.020	0.16	.87	0.013	0.027	0.47	.67
Active filling object NP								
Intercept	5.81	0.035	160	<.001	5.85	0.04	160	<.001
Filler	0.01	0.004	3.4	<.001	0.02	0.005	3.9	<.001
Len.long:								
Sht – mean(CP,PP)	0.01	0.001	1.3	.18	0.02	0.01	1.6	.12
Len.cls: PP – CP	0.01	0.009	1.4	.16	0.01	0.01	0.90	.37
Filler:len.long	0.02	0.008	2.5	.01	0.03	0.01	2.9	<.005
Filler:len.cls	0.01	0.009	1.3	.20	0.03	0.01	2.3	.02
Gap-driven								
Intercept	5.82	0.034	170	<.001	5.85	0.035	170	<.001
Filler	0.02	0.004	4.6	<.001	0.03	0.006	5.5	<.001
Len.long:								
Sht – mean(CP,PP)	−0.005	0.009	−0.52	.60	−0.012	0.012	−1.0	.31
Len.cls: PP – CP	−0.019	0.010	−1.7	.09	−0.017	0.014	−1.2	.23
Filler:len.long	0.01	0.009	1.1	.29	0.008	0.012	0.64	.53
Filler:len.cls	−0.02	0.011	−1.9	.06	−0.002	0.014	−0.12	.91

*Note:* Left: models on residual-trimmed data. Right: models on untrimmed data. PP = prepositional phrase; CP = complementizer phrase; NP = noun phrase; len = length; cls = clause; sht = short; MCMC = Markov Chain Monte Carlo. In filler contrasts, implausible fillers are positive (+1); in length.long contrasts, short conditions are +2/3; in length.clause contrasts, +PP conditions are (+1/2). *P*-values are reported rounded to 2 decimal places, except where they are less than .005. Dark shading highlights coefficients, excluding the intercept, with a *p*-value less than .05; light shading for  $.05 \leq p < .10$ . *SE* = standard error.

lengthened by modifying the subject with a PP showed weaker sensitivity to the semantic fit of the filler prior to the gap region. However, the RT difference in the active region was reliably smaller for +PP-length dependencies than for the short monoclausal dependencies. This contrast between anomaly detection in short and +PP-length

dependencies replicates the previous finding by Wagers and Phillips (2009; Experiment 3).

This pattern of results can be interpreted in one of two ways: Either (a) active dependency formation ceased by the time the most deeply embedded verb was reached in a biclausal dependency, or (b) the lexicosemantic features of the

filler required to elicit implausibility effects did not participate in the active part of dependency construction for longer dependencies. The results of Experiment 1 make Interpretation (a) unlikely. In that experiment, the filled-gap effect was present in monoclausal and biclausal dependencies alike. Therefore, some version of Interpretation (b) is likely to be correct. It is important to observe, however, that it is not the case that longer dependencies are simply more taxing for comprehenders and that the effect of semantic fit declines as dependency length increases. In both of the long dependency conditions, the effect of semantic fit was reliable in the postgap region and not different from that in short dependency conditions. Indeed, the most robust postgap effect was found for +CP-length dependencies, which showed no evidence of a plausibility effect during the active filling region. Therefore the lexicosemantic features of the filler are not forgotten. Rather they specifically become inaccessible or ineffective for active dependency formation.

In the context of the results of Experiment 1, we propose that long-distance dependency formation reflects the integration of information in two states. The heart of active dependency formation is the act of predictive structure building: Dependencies are projected forward in time by the comprehender, in advance of direct evidence for the gap location. However, we propose that the extent to which the dependency is initially elaborated or evaluated is limited by the information about the filler that is carried forward in time. In other words, the comprehender maintains what she can about the filler in an active representational state, and this is what guides initial dependency construction. This “active” state can be identified with the focus of attention or the capacity of concurrent processing (Broadbent, 1958; Jonides et al., 2008; McElree, 2006).

Since this capacity is limited, a complete representation of the filler cannot be maintained. When other processing events intervene, some details of the filler may be displaced. Therefore the initial dependency might be thought of as a provisional representation, the details of which are filled out by retrieving the complete representation

of the filler from memory, the process normally identified as reactivation (Nicol & Swinney, 1989).

The results of Experiments 1 and 2 together suggest that the category identity of the filler is maintained across all dependency lengths, whereas its semantic details are not. This has both architectural and functional advantages. From an architectural standpoint, the information that encodes category identity is more compact than the information about semantic features. This may be because the category identity is a coarser, more general representation for which there are only a few possibilities. In contrast, individual semantic features (e.g., +ANIMATE, +CONCRETE, +MASS, +LIQUID, etc.) are finer grained, and there are many more possibilities. From a functional standpoint, category information is grammatically more useful in evaluating the well-formedness of a dependency. A sentence may express an unusual or implausible state of affairs and yet be grammatical.

We develop this account in greater detail in the General Discussion, but immediately it raises the question of what the appropriate level of “coarseness” is for maintenance. In Experiment 3, we examine another measure of active dependency formation, in which the information that must be maintained about the filler is not as general as category identity, but is still more constrained than the semantic features.

## EXPERIMENT 3: SUBCATEGORIZATION MATCH

### Rationale

In the third experiment, we devised another measure of active dependency formation by manipulating subcategorization match. It is based upon the fact that arguments bearing particular thematic roles must occur as the complement of specific prepositions. For example, the verb “inherit” requires a source argument. In English, source arguments are the complement of the preposition “from” (Example 10a). In contrast, verbs like “entrust” require a goal argument. Goal arguments typically appear as the complement of the preposition “to” (Example 10b).

- (10) (a) The orphan inherited the prayer book from/\*to his uncle.  
(b) The monk entrusted the prayer book \*from/to his novice (where the asterisk denotes ungrammaticality).

If we displace the source/goal argument, the restriction on preposition identity still holds.

- (11) (a) The uncle from/\*to whom the orphan inherited the prayer book.  
(b) The novice \*from/to whom the monk entrusted the prayer book.

If comprehenders are sensitive not just to the category identity of a displaced PP, but also to the particular preposition that heads the PP, then there should be a reading time slowdown when the displaced preposition is not appropriate for the kind of semantic argument the verb requires.

The subcategorization mismatch manipulation was tested at two different dependency lengths: short monoclausal dependencies and PP-extended monoclausal dependencies. We chose not to include a CP-extended dependency in this experiment since most of the embedding verbs used in Experiments 1–2 allow PP arguments headed by *to* and *from* (“conclude from experience,” “say to a friend”, etc.). This would allow a fronted PP-filler to be prematurely interpreted in the main clause, potentially obscuring the effects of interest at the critical verb.

Method

Materials and procedure

Participants were 54 native speakers of English who were recruited via Amazon Mechanical Turk (we attempted to recruit as many participants as possible over a 2-day period). They received \$4 for taking part in the experiment. Participants completed the study using the online experiment platform Ibex (Drummond, 2010), which allows a self-paced reading experiment to be deployed in the browser. An earlier study of the same design, conducted in the lab with 18 native speakers, yielded the same results; this allays some potential concerns about the robustness of real-time psycholinguistic data derived from a web-based participant sample.

This experiment crossed the factors filler match and length in a 2 × 2 factorial design. The filler match factor contrasted “match” and “mismatch” conditions, corresponding to the subcategorization match between the verb and the PP. The length factor contrasted “short” and “long” conditions. As in Experiment 1, the short length conditions separated the filler and the verb by a two-word subject and an adverb. The long conditions attached a five-word PP modifier to the subject. Table 7 illustrates a full materials set. Items were balanced so that the matching preposition was *to* in 12 item sets and *from* in 12 item sets. Materials were distributed according to a Latin

Table 7. Sample material set for Experiment 3

Length	Match	Sentence
Short	Match	Item frame: “The courier ... unfortunately wrecked his bike in traffic.” ... <b>to whom</b> the secretary warily <u>entrusted</u> the confidential business correspondence ____ after some hesitation ...
	Mismatch	... <b>from whom</b> the secretary warily <u>entrusted</u> the confidential business correspondence ____ after some hesitation ...
Long	Match	... <b>to whom</b> the secretary for the high-powered defense attorney warily <u>entrusted</u> the confidential business correspondence ____ after some hesitation ...
	Mismatch	... <b>from whom</b> the secretary for the high-powered defense attorney warily <u>entrusted</u> the confidential business correspondence ____ after some hesitation ...

Note: The match manipulation is indicated with boldface, the start of the critical region with underlining, and the gap position with an underscore.

**Table 8.** *Experiment 3: Comprehension accuracy*

Filler subcategorization match	Length		Mean
	Short	Long	
Match	85 (2)	85 (2)	85
Mismatch	86 (2)	80 (2)	83
Mean	85	82	84

*Note:* Accuracy expressed as average percentage correct over participants, with row, column, and grand means. Standard error of the cell means across subjects is reported in parentheses.  $N = 54$ .

square across four lists so that participants each read six sentences per condition.

The analysis details were as described in Experiment 1. Regions of interest were structurally aligned so that the VP words common to all conditions in an item set corresponded to the same regions across conditions. Because there were only two levels of the length factor, it was sum coded in this study, with long (+PP) represented by the positive coefficient.

## Results

### *Comprehension accuracy*

Comprehension accuracy is reported in Table 8. Overall accuracy was 84%. There was a 3% decrement in accuracy associated with long dependencies and a 2% decrement for verbs with mismatching PP arguments. However none of these effects, nor their interaction, achieved significance.

### *Reading times*

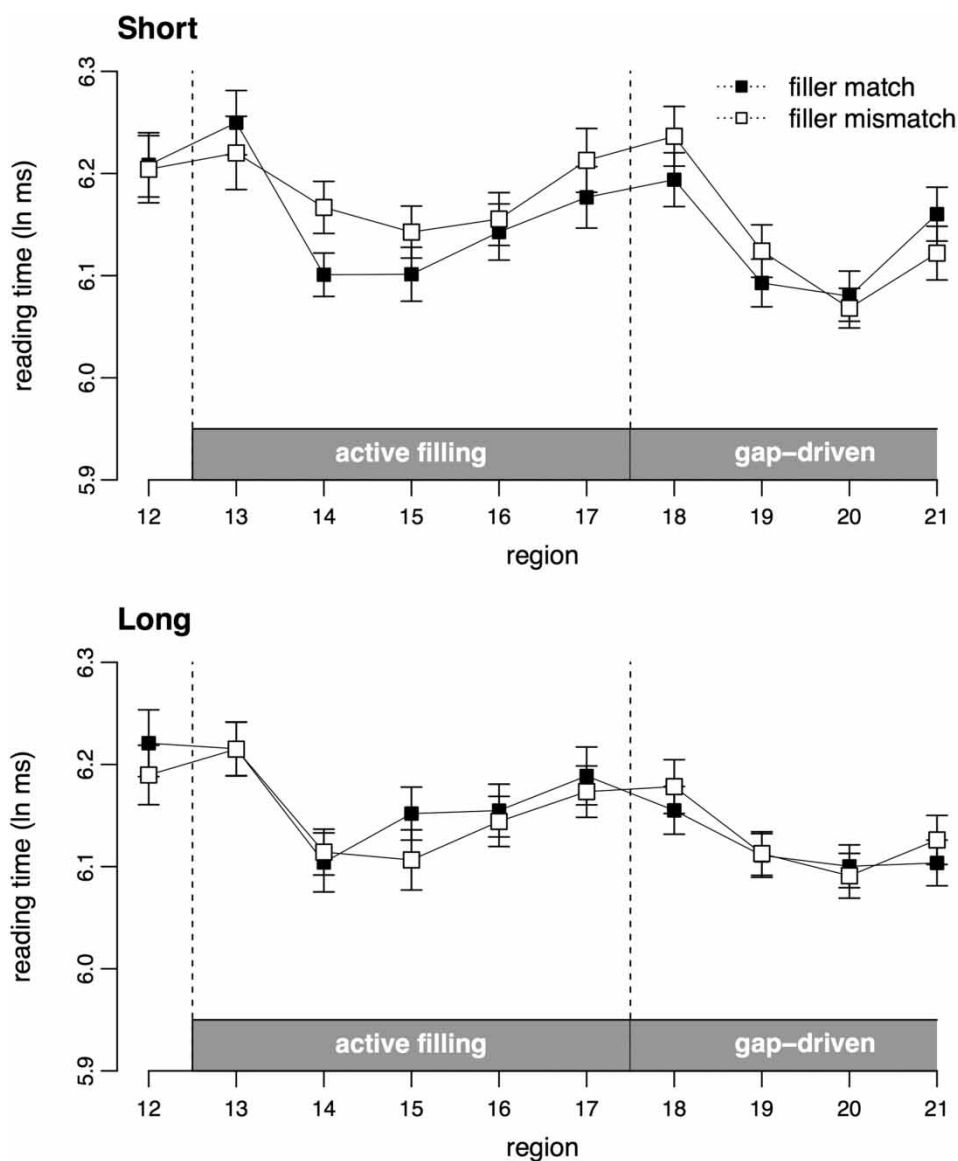
Reading time results are reported in Figure 3. The main finding from the reading time data is that sensitivity to the verb–PP match was observed principally only in the short dependency conditions. As in Experiment 2, we report four regions of analysis: the adverb, two “active filling” regions (the verb and the direct object NP), and the three following words (“gap-driven”). Model coefficients are given in Table 9. In the adverb and verb regions there were no effects.

In the object NP region, there was an anomaly effect: Filler mismatch conditions were read more slowly than filler match conditions. However, there was also an interaction with length, such that the contrast was almost entirely abolished for long conditions. This is apparent in the untrimmed data as well, as Figure 3 makes clear. In the gap region, the anomaly effect of filler mismatch approached marginal significance, as did the interaction with length; but neither effect was reliable at  $\alpha = .05$ .

## Discussion

This experiment provided evidence that comprehenders were sensitive online to the subcategorization restriction that verbs place on a displaced PP-argument, but only when the distance between the displaced PP and the verb was short. When more than a two-word subject intervened between the PP filler and the verb, there was no indication in the reading times that participants detected a mismatch. This finding suggests that the information about the identity of the preposition is not well preserved over a relatively long dependency. Only in the short conditions was there evidence that a mismatching PP was rapidly detected, based on reading times at the verb and in the following region. The presence of a slowdown in this region indicates that comprehenders evaluated the displaced PP as a potential argument of the verb.

An important difference between this experiment and Experiment 2 should be highlighted. In Experiment 2, the semantic anomaly led to a slowdown in the active filling regions in the short dependency conditions, but not in the longer dependency conditions. Nevertheless, the long dependency conditions did eventually show an effect of the semantic anomaly, in the postgap regions. In contrast, in the present experiment there was no statistically significant slowdown in the long dependency conditions. It could be that there was a high level of variability in the timing and strength of the mismatch effect at the level of individual trials or participants, with the consequence that we were not able to observe



**Figure 3.** Experiment 3 reading times. Each panel corresponds to a different length condition. Closed symbols represent subcategorization matching fillers, open symbols mismatching fillers. Reading times are reported in natural log ms (ln ms). Error bars indicate standard error of the cell means.

Sample sentence:

*The courier to/from whom the secretary [for the high-powered defense attorney]... warily<sub>12</sub> entrusted<sub>13</sub> the<sub>14</sub> confidential<sub>15</sub> business<sub>16</sub> correspondence<sub>17</sub> after<sub>18</sub> some<sub>19</sub> hesitation<sub>20</sub>...*

a punctate mismatch effect in the average reading times. Furthermore, the absence of a reliable effect even in the gap region raises the concern

that our study lacked the appropriate power to detect it. However, in a pilot study ( $n = 18$ ), we did replicate the exact pattern, with a different

Table 9. Experiment 3: Mixed-effect model coefficient tables

Region and coefficient names	Residual trimmed dataset				Untrimmed dataset			
	Estimate	SE	t	p <sub>MCMC</sub>	Estimate	SE	t	p <sub>MCMC</sub>
Adverb								
Intercept	6.16	0.053	120	<.001	6.21	0.060	100	<.001
Filler	−0.002	0.007	−0.23	.81	−0.009	0.010	−0.83	.42
Length	0.009	0.007	1.3	.19	−0.001	0.010	−0.10	.93
Filler:length	−0.002	0.007	−0.35	.64	−0.005	0.010	−0.47	.60
Active filling verb								
Intercept	6.19	0.046	130	<.001	6.23	0.050	120	<.001
Filler	0.002	0.008	0.24	.80	−0.007	0.012	−0.65	.51
Length	0.001	0.008	0.18	.84	−0.011	0.012	−0.95	.36
Filler:length	0.001	0.008	0.12	.90	0.009	0.012	0.75	.48
Active filling object NP								
Intercept	6.11	0.039	160	<.001	6.15	0.05	130	<.001
Filler	0.008	0.003	2.6	.01	0.005	0.005	1.1	.29
Length	0.003	0.003	0.95	.36	−0.004	0.005	−0.73	.47
Filler:length	−0.012	0.003	−3.8	<.001	−0.014	0.005	−2.8	.005
Gap-driven								
Intercept	6.09	0.036	160	<.001	6.13	0.040	150	<.001
Filler	0.006	0.003	1.6	.10	0.007	0.005	1.4	.17
Length	0.0001	0.003	0.03	.98	−0.004	0.005	−0.85	.40
Filler:length	−0.005	0.003	−1.6	.13	−0.005	0.005	−0.91	.36

Note: Left: models on residual-trimmed data. Right: models on untrimmed data. NP = noun phrase; MCMC = Markov Chain Monte Carlo. In FILLER contrasts, mismatch conditions are positive (+1); in LENGTH contrasts, long conditions are +1. *P*-values are reported rounded to 2 decimal places, except where they are less than .005. Dark shading highlights coefficients, excluding the intercept, with a *p*-value less than .05. *SE* = standard error.

sample: college students in the lab. We supply those data in our supplementary materials.<sup>6</sup>

Alternatively, it could be that the mismatch was never detected, unless it was detected actively. In a follow-up grammaticality study with rapid serial visual presentation and speeded judgement, we presented 32 participants with the target sentences from Experiment 3 (the composition of the fillers was different). In the short dependency conditions, participants accepted grammatical sentences in 83% of trials, but they also only rejected the ungrammatical sentences in 29% of trials. In the long dependency conditions, participants accepted grammatical sentences in 67% of trials and rejected

ungrammatical sentences in 43% of trials. In a mixed-effects logistic regression over percentage correct, there were only main effects of match ( $-0.71 \pm 0.49$ ,  $p < .001$ ) and length ( $-0.95 \pm 0.49$ ,  $p < .001$ ), and no interaction. What is most notable is that overall sensitivity to this violation was low for both lengths. We computed *d*-prime, a measure of sensitivity that corrects for response bias by taking the difference between the hit rate and the false-alarm rate (expressed as a *z*-score; MacMillan & Creelman, 2004). In our case, the hit rate corresponded to the endorsement rate for grammatical sentences, and the false-alarm rate the endorsement rate for ungrammatical sentences.

<sup>6</sup>If we took, as an expectation of the effect size, the average difference between filler match conditions in short dependencies, excluding the first word in the direct object region (see Figure 3), we would have a small positive difference corresponding to a within-subjects corrected Cohen's *d* of 0.33. For a one-sided *t*-test and a repeated measures design, a sample size of 58 would be required to achieve a power of .80 ( $1 - \beta$ ; G\*Power 3, Faul, Erdfelder, Lang, & Buchner, 2007).

In short dependencies, the resulting  $d$ -prime was 0.40, and for long dependencies it was 0.26. Participants therefore seemed to mostly endorse short sentences as grammatical, regardless of match, and to reduce endorsement for long sentences overall. Sensitivity to the mismatch remained very small in both cases. This pattern corroborates the idea that individuals do not retain sensitivity to the lexical identity of the preposition in these sentences for a long time, even in short sentences. A possible reason is that the verb is so semantically constraining about the possible role of the PP that only information about the NP expressed inside the PP is useful to retain. We discuss this possibility in greater detail in the General Discussion.

## GENERAL DISCUSSION

### Summary of results

In three experiments we tested the sensitivity of different measures of active dependency formation to extended dependency lengths. Characterizing the difference between early, active dependency completion and later nonactive/gap-driven reactivation of the filler is potentially a window onto how the comprehender manages memory resources in the course of comprehending an unbounded dependency. In order to distinguish active and nonactive phases of dependency completion, longer dependencies become a crucial testing ground. However, there have been surprisingly few studies examining longer dependencies. In Experiments 1 and 2, we tested the filled-gap effect and the plausibility effect, respectively, both of which are commonly used in studies on processing filler-gap dependencies (e.g., Garnsey et al., 1989; Lee, 2004; Stowe, 1986; Traxler & Pickering, 1996). In Experiment 3, we tested the effect of subcategorization mismatch. These measures differ in terms of the kind of the information needed about the filler phrase to distinguish the normal and (sometimes temporarily) anomalous sentences in each study: The filled-gap effect is sensitive to the maximal category of the filler (e.g., “which

chemicals” vs. “for which chemicals”); the subcategorization effect is sensitive to the identity of a PP filler’s closed-class head (e.g., “to which courier” vs. “for which courier”); and the plausibility effect is sensitive to particular semantic features of an NP filler’s open-class head (e.g., “acorns” vs. “cats”).

The filled-gap effect tested in Experiment 1 was apparent at all dependency lengths. The logic of the filled-gap manipulation was as follows: If the parser posited a direct object gap for NP fillers, but not for PP fillers, then there should be longer reading times in the direct object region for NP conditions than for PP conditions, probably reflecting reanalysis that was required in the NP-filler sentences. The fact that a filled-gap effect was found at all dependency lengths suggests that the parser actively completed the filler-gap dependency in all of those contexts and that it specifically projected direct object gaps for NP fillers.

In contrast to the filled-gap effect, the plausibility effect in Experiment 2 showed more variability across conditions. The logic of the plausibility effect was as follows: If the parser immediately analysed the filler as an argument of the critical verb, then fillers that were semantically anomalous arguments should lead to increased reading times at the verb or in the direct object region (the “active filling” regions). In the active filling regions, the plausibility effect was strong in short dependency conditions but completely absent in long biclausal dependency conditions. In the long, PP-extended dependency conditions, the plausibility effect was present, but weak. Of particular importance is the fact that the long conditions that showed weak or no plausibility effects in the active filling regions nevertheless showed a clear plausibility effect in the postgap regions. Comprehenders therefore did not fail to notice the semantically anomalous arguments, but they did not detect the semantic anomaly before reaching the gap. We observed the same pattern of results in an earlier experiment (Wagers & Phillips, 2009, Experiment 3), which only included the +PP-lengthened condition.

Finally, the subcategorization mismatch effect in Experiment 3 showed a profile similar to the plausibility effect in Experiment 2. The logic of this index was as follows: If the parser immediately analysed

the PP filler as an argument of the critical verb, then fillers whose prepositional head did not match the requirements of the verb should lead to increased reading times at the verb or in the direct object region that preceded the gap. However, the mismatch effect was only apparent in the short dependency conditions. As in Experiments 1 and 2, the active dependency formation effect in short dependencies was large and occurred early in the regions of interest. Paralleling the findings of Experiments 1 and 2, the mismatch effect was attenuated in a long, PP-extended dependency. In fact, in this study the effect in the PP-extended conditions was completely absent in active regions or after the gap. It is therefore unclear how robustly comprehenders recovered sensitivity to the identity of the filler's prepositional head in long dependencies. A follow-up speeded grammaticality test indicated that end-of-sentence judgements were not very accurate; however, there was no interaction with length.

### Memory and control processes in active dependency construction

We conclude that filler-gap dependencies are actively constructed regardless of dependency length. This can explain the pattern of results observed in Experiment 1, and by Frazier and Clifton (1989): Even long, biclausal filler-gap dependencies are formed actively. What must be accounted for, then, is why longer dependencies failed to show any active effects involving properties of the filler phrase that were more specific than syntactic category. The likely explanation, we argue, is that certain information is immediately accessible to the processor to make parsing decisions, while other information must be retrieved from passive memory once the dependency is formed. Findings from a variety of cognitive tasks indicate that memory is partitioned into a highly capacity-limited focal state (Cowan, 2001; Garavan, 1998; Jonides et al., 2008; McElree, 2006; Verhaegen, Cerella, & Basak, 2004) and a virtually unlimited nonfocal state. Information in the focal state can be incorporated into ongoing processing directly. However, information that has been displaced

from this state must be reincorporated by retrieval operations, a finding supported by speed-accuracy trade-off time-course studies of list memory (McElree, 2006; McElree & Doshier, 1989) and, more recently, studies of subject-verb dependencies in language comprehension (McElree et al., 2003).

Interpreted in view of this architecture, our data suggest that the comprehension system completes filler-gap dependencies based on two complementary processes: a "leading" process, which projects structure ahead of the input based on a limited amount of information in the focus of attention; and a "lagging" process, which involves the retrieval or reactivation of the full filler phrase. The filled-gap effect's relative insensitivity to length derives from the fact that active dependency formation can be driven by a minimal amount of information—that is, the information that a filler of category XP exists, and that this small amount of information can survive during the course of the sentence, even in the face of unrelated processing events. The filled-gap effect depends upon the minimal amount of information necessary to drive active filling, whereas the other measures of active gap creation depend on maintenance of fuller syntactic and semantic details. In Experiments 2 and 3, when the filler-gap dependency was initially formed in long dependencies, lexically specific information about the filler phrase was not available that would lead the comprehender to notice any anomalies. Once the comprehender retrieved the filler, however, the acceptability of the dependency could be evaluated.

The core idea that our data support is that certain features of the representation are privileged in anticipation of how future information should be integrated. This can be achieved in a variety of models. For example, the proposed interplay of predictive, leading processes and retrospective, lagging processes can be combined in a straightforward hybridization of the two leading mechanisms of filler-gap processing: a maintenance mechanism, as in Wanner and Maratsos' (1978) HOLD CELL hypothesis (and Frazier's, 1987, related idea that the filler is somehow not "inert"); and a reactivation mechanism, proposed on the basis of cross-modal lexical priming and probe recognition tasks

(Bever & McElree, 1988; Nicol et al., 1994; Nicol & Swinney, 1989). This decomposition of mechanisms actually seems to be reflected rather directly by the dissociation of the plausibility effect in Experiment 2: The effect appeared on the verb for short dependencies, but downstream from the gap for longer dependencies. However, the question arises of why retrieval does not occur immediately, rather than after the position where the gap site is confirmed. One possibility is simply that other interpretive relations are being constructed simultaneously. After projecting the gap position, the processor shifts its resources to interpret the direct object. Data from cross-modal priming studies are relevant here, but are somewhat mixed (see McKoon & Ratcliff, 1994; Nicol et al., 1994). Among the studies that have found evidence for reactivation, Nicol and Swinney (1989) report that facilitation (in RTs) for a visually presented, semantic associate of the filler is found in the position immediately following the verb. On the other hand, in a synonym judgement task, McElree (2001) obtained the greatest facilitation (in accuracy) several words downstream, after which the gap had been unambiguously located. It is important to note that in both of these cases, the gap was in direct object position. In our case, the gap was in an oblique position. Further research is therefore needed to test the generality of the results.

An alternative way of differentially privileging different kinds of information in processing is through different data types. Syntactic constituents (nodes, items) might be distinguished from syntactic dependencies (arcs, associations) in terms of their mnemonic properties. This opens up several degrees of freedom. First, we could grant that certain features inhere in or are bound to arcs while others are bound to constituents. Secondly, we could distinguish between predictions for particular constituents and predictions for dependencies between constituents. Finally, different data types may have their own advantages and vulnerabilities in working memory: The constituents themselves could be susceptible to interference, the dependencies linking them could be susceptible, or both. How could this help explain our data? If comprehenders posit the link between

filler and gap predictively, this component of the representation and not access to the features of the filler item itself could drive the filled-gap effect. The specific filler constituent would need to be reactivated to interpret it and thus to generate the anomaly effect. Moreover, if a single labelled dependency link, bearing information about syntactic category, were a relatively strong or unique component of the representation, its distinctness at retrieval could render it more robust to interference (Nairne, 2006). Thus an account is possible without direct appeal to maintaining a syntactic constituent, but robust predictive activation of some information does seem to be required.

Our present data, and understanding, probably cannot choose between all these possibilities, but they do point the way for future investigation on how distinct features types interact with one another in working memory during language processing. In the next section, we expand upon the relationship between maintenance and distinctness.

### The capacity of concurrent processing and underspecification

The amount of information that can be concurrently processed is clearly limited (Broadbent, 1958), but the exact nature of these limitations for processing linguistic structure remains poorly understood. For list memory it is clear that focal attention does not apply merely to the last percept, and that chunking and task expectations play an important role in determining what is maintained in focal attention (McElree, 2006). An idea with deep roots in psycholinguistics is that a capacity bottleneck causes the chunking of an expression into its major constituents, a notion that motivated the click dislocation experiments of the 1960s (Bever, Lackner, & Kirk, 1969). More recently, in the ACT-R model of sentence processing (Lewis & Vasishth, 2005), it is one maximal projection (i.e., an XP) that can be maintained in the “active” buffer. What these ideas have in common is that the capacity bottleneck functionally carves up expressions into contiguous extents of structure. The span of concurrent processing thus

specifically limits the breadth of information that can be represented at any given moment.

An alternative view, which is consistent with our proposal for processing filler-gap dependencies, is that limits on concurrent processing exert their influence primarily on the depth of information that can be represented at any given moment. Instead of maintaining all the features of a given constituent with full precision, the comprehender may choose to discard some features from focal attention, in order to accommodate information from other constituents. For filler-gap dependencies, maintaining a fully articulated encoding of the filler phrase across the entire extent of the dependency while processing intermediate portions of the sentence may simply be impossible. But it may nonetheless be feasible to maintain a pared-down representation of the filler-phrase, with just enough information to make the most crucial parsing decision—that is, where to posit a gap. In the shortest length conditions, when active effects were observed for all contrast types, we propose that most of the features of the filler remained in focal attention and were yet to be displaced.

There are models that motivate the paring down of concurrently active features in a partially constructed representation without strictly separating maintenance and retrieval processes. For example, Oberauer and Kliegl (2006) present a formal model that traces the construct of maintenance capacity to the underlying cause of feature overwriting. In this view, items are lost from active processing because other items, with similar component features, have overwritten them. Capacity is thus a function of distinctness, just like retrievability (Nairne, 2006). This model thus achieves a broader unification of interference theory and forgetting. For language processing, it could be that certain syntactic features are simply more interference-robust than specific lexical features—that is, the syntactic category of the filler in Experiment 1 is less likely to be overwritten than the lexical features in Experiment 2. Those “robust” features would constitute part of the pared-down representation. This might seem initially implausible, since syntactic categories are fewer in number than the set of potential lexical features; thus the syntactic

category of the filler and other constituents are more likely to overlap than are their lexical features. But it is important to keep in mind that we do not know the exact lexical features responsible for anomaly detection in Experiment 2 or the syntactic features responsible for the filled gap effect in Experiment 1.

Most generally, this line of inquiry raises the question of whether we can have a theory of what kinds of information are interference-robust in language processing and what kinds are not. Distinctness is clearly a crucial part of the picture. But we conjecture that a complementary limitation derives from the binding of specific lexical material to positions in the structure, an idea supported by recent computational modelling of parsing (Van der Velde & de Kamps, 2006; for nonlinguistic information: Oberauer & Vockenberg, 2009; Treisman & Gelade, 1980). Thus we distinguish the storage of atomic features from the storage of feature/feature bindings. In our case, we might suppose that syntactic category features are bound together with the +WH filler feature (i.e., introduced by “which” in “which acorns”), but that lexical details are linked to the NP restrictor (e.g., “acorns” in “which acorns”). Multiple NPs were encountered along the dependency path, which could interrupt the binding of the lexical features to the NP restrictor via overwriting. However, no further +WH categories were encountered, effectively protecting the information bound to that feature. This view has two advantages: First, it is consistent with the idea that syntactic categories are complex and structured (Gazdar, Klein, Pullum, & Sag, 1985; Jackendoff, 1977). Secondly, it makes a strong prediction: If it is correct, experimentally introducing further (nested) filler phrases (e.g., a relative clause attached to a subject) should disrupt the filled-gap effect.

Limits on the depth of representation can be thought of as a kind of underspecification, a concept that has received renewed attention in psycholinguistics (Frisson, 2009; Sturt, Sanford, Stewart, & Dawydiak, 2004; Weinberg, 1993). By underspecifying most constituents in the syntactic context, the comprehender frees up capacity to

encode and process incoming information, but nonetheless has at the ready some amount of global context upon which to base parsing decisions. Though locality biases are evident in many comprehension processes, nonlocal dependencies can often be constructed with great accuracy (Phillips, Wagers, & Lau, 2011). Thus the ability to have ready access to information about global structure would be an advantage. Understanding the generality of this conclusion will depend on examining other species of linguistic dependency in a way that varies the predictability and distinctness of their dependent elements.

Original manuscript received 6 January 2012  
Accepted revision received 21 September 2013  
First published online 9 December 2013

## REFERENCES

- Aoshima, S., Phillips, C., & Weinberg, A. S. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language*, 51, 23–54.
- Baayen, H. (2008). *Analyzing linguistic data*. Cambridge, UK: Cambridge UP.
- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3, 12–28.
- Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-42. Retrieved from <http://CRAN.R-project.org/package=lme4>
- Bever, T. G., Lackner, J. R., & Kirk, R. (1969). The underlying structures of sentences are the primary units of immediate speech processing. *Perception & Psychophysics*, 5, 225–234.
- Bever, T. G., & McElree, B. (1988). Empty categories access their antecedents during comprehension. *Linguistic Inquiry*, 19, 35–43.
- Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language*, 34, 774–806.
- Bourdages, J. S. (1992). Parsing complex NPs in French. In H. Goodluck & M. S. Rochemond (Eds.), *Island constraints: Theory, acquisition and processing* (pp. 61–87). Dordrecht: Kluwer Broadbent 1958.
- Broadbent, D. E. (1958). *Perception and communication*. New York: Oxford UP.
- Chomsky, N. (1977). On Wh-Movement. In P. W. Culicover, T. Wasow, & A. Akmajian (Eds.), *Formal syntax* (pp. 71–132). New York: Academic Press.
- Chomsky, N. (1987). *Barriers*. Cambridge, MA: MIT Press.
- Clahsen, H., & Featherston, S. (1999). Antecedent priming at trace positions: Evidence from German scrambling. *Journal of Psycholinguistic Research*, 28, 415–437.
- Conklin, K., Koenig, J. P., & Mauner, G. (2004). The role of specificity in the lexical encoding of participants. *Brain and language*, 90, 221–230.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–185.
- Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing* (pp. 94–128). Cambridge: Cambridge University Press.
- Drummond, A. (2010). Ibexfarm (Version 0.3) [Software]. Retrieved from <https://github.com/addrummond/ibexfarm>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.
- Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, 47, 250–272.
- Fraser, B. (1971). A note on the spray-paint cases. *Linguistic Inquiry*, 2, 603–607.
- Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory*, 5, 519–560.
- Frazier, L., & Clifton, C., Jr. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, 4, 93–126.
- Frazier, L., & Flores D'Arcais, G. B. (1989). Filler-driven parsing: A study of gap filling in Dutch. *Journal of Memory of Language*, 28, 331–344.
- Frisson, S. (2009). Semantic underspecification in language processing. *Language and Linguistics Compass*, 3, 111–127.

- Garavan, H. (1998). Serial attention within working memory. *Memory & Cognition*, 26, 263–276.
- Garnsey, S. M., Tanenhaus, M. K., & Chapman, R. M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research*, 18, 51–60.
- Gazdar, G., Klein, E., Pullum, G., & Sag, I. (1985). *Generalized phrase structure grammar*. Cambridge, MA: Harvard University Press.
- Gibson, E. (1998). Syntactic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1–76.
- Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95–126). Cambridge, MA: MIT Press.
- Jackendoff, R. (1977). *X-bar syntax: A study of phrase structure*. Cambridge, MA: MIT Press.
- Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, 59, 15.1–15.32.
- Just, M. A., Carpenter, P. A., & Wooley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111(2), 228–238.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30, 580–560.
- King, J. W., & Kutas, M. (1995). Who did what and when? Using word-and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 7, 376–395.
- Kluender, R., & Kutas, M. (1993). Subjacency as a processing phenomenon. *Language and Cognitive Processes*, 8, 573–633.
- Lee, M.-W. (2004). Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research*, 33, 51–73.
- Lewis, R. L. (2000). Specifying architectures for language processing: Process, control, and memory in parsing and interpretation. In M. W. Crocker, M. Pickering, & C. Clifton (Eds.), *Architectures and mechanisms for language processing* (pp. 56–89). Cambridge, UK: Cambridge University Press.
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29, 375–419.
- Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Mahwah, NJ: Lawrence Erlbaum.
- Manzini, M. R. (1992). *Locality: A theory and some of its empirical consequences*. Cambridge, MA: MIT Press.
- McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory. *Journal of Psycholinguistic Research*, 29(2), 111–123.
- McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27, 817–835.
- McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 46, pp. 155–200). San Diego: Academic Press.
- McElree, B., & Doshier, B. A. (1989). Serial position and set size in short-term memory: Time course of recognition. *Journal of Experimental Psychology: General*, 118, 346–373.
- McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory & Language*, 48(1), 67–91.
- McElree, B., & Griffith, T. (1998). Structural and lexical constraints on filling gaps during sentence comprehension: A time-course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 432–460.
- McKinnon, R., & Osterhout, L. (1996). Constraints on movement phenomena in sentence processing: Evidence from event-related potentials. *Language and Cognition*, 11, 495–523.
- McKoon, G., & Ratcliff, R. (1994). Sentential context and on-line lexical decision tasks. *Journal of Experimental Psychology: Language, Memory and Cognition*, 20, 1239–1243.
- Nairne, J. S. (2006). Modeling distinctiveness: Implications for general memory theory. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 27–46). USA: Oxford University Press.
- Neville, H. J., Nicol, J., Barss, A., Forster, K., & Garrett, M. (1991). Syntactically-based sentence processing classes: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, 3, 151–165.
- Nicol, J. L., Fodor, J. D., & Swinney, D. (1994). Using cross-modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 20, 1229–1238.
- Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18, 5–19.
- Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, 55, 601–626.

- Oberauer, K., & Vockenberg, K. (2009). Updating of working memory: Linger-ing bindings. *The Quarterly Journal of Experimental Psychology*, 62, 967–987.
- Omaki, A., Lau, E., Davidson White, I., & Phillips, C. (2010). Hyper-active gap filling: Pre-verbal object gap creation in English filler-gap dependency processing. Submitted.
- Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition*, 33, 563–588.
- Phillips, C. (2006). The real-time status of island phenomena. *Language*, 82, 795–823.
- Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, 22, 407–428.
- Phillips, C., & Wagers, M. W. (2007). Relating structure and time in linguistics and psycholinguistics. In G. Gaskell (Ed.), *Oxford handbook of psycholinguistics* (pp. 739–756). Oxford: Oxford University Press.
- Phillips, C., Wagers, M., & Lau, E. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In J. Runner (Ed.), *Experiments at the interfaces, syntax and semantics* (Vol. 37, pp. 147–180). Bingley, UK: Emerald Publications.
- Pickering, M. J., Barton, S., & Shillcock, R. (1994). Unbounded dependencies, island constraints and processing complexity. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 199–224). London: Lawrence Erlbaum.
- Pickering, M. J., & Traxler, M. J. (2001). Strategies for processing unbounded dependencies: lexical information and verb argument assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1401–1410.
- Pickering, M. J., & Traxler, M. J. (2003). Evidence against the use of subcategorisation frequency in the processing of unbounded dependencies. *Language and Cognitive Processes*, 18, 469–503.
- R Development Core Team. (2011). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Rappaport, M., & Levin, B. (1986). *What to do with Theta Roles*. (Lexicon Project Working Papers, 11). Cambridge, MA: MIT Center for Cognitive Science.
- Rizzi, L. (1990). *Relativized minimality*. Cambridge, MA: MIT Press.
- Rohde, D. (2003). Linger (Version 2.94) [Software]. Retrieved from <http://tedlab.mit.edu/~dr/Linger/>
- Ross, J. R. (1967). *Constraints on variables in syntax*. (Doctoral dissertation). Massachusetts Institute of Technology.
- Ruchkin, D. S., Johnson, R. Jr., Canoune, H., & Ritter, W. (1990). Short-term memory storage and retention: An event-related brain potential study. *Electroencephalography and Clinical Neurophysiology*, 76, 419–439.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–1366.
- Sprouse, J., Wagers, M. W., & Phillips, C. (2012). A test of the relation between working memory capacity and syntactic island effects. *Language*, 88, 82–123.
- Staub, A. (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 550–569.
- Stowe, L. (1986). Parsing wh-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, 1, 227–246.
- Stowe, L. A., Tanenhaus, M. K., & Carlson, G. N. (1991). Filling gaps on-line: Use of lexical and semantic information in sentence processing. *Language and Speech*, 34, 319–340.
- Sturt, P., Sanford, A. J., Stewart, A., & Dawydiak, E. (2004). Linguistic focus and good-enough representations: An application of the change detection paradigm. *Psychonomic Bulletin and Review*, 11, 882–888.
- Sussman, R., & Sedivy, J. (2003). The time course of processing syntactic dependencies. *Language and Cognitive Processes*, 18, 143–163.
- Swinney, D., Ford, M., Frauenfelder, U., & Bresnan, J. (1988). *On the temporal course of gap-filling and antecedent assignment during sentence comprehension*. Unpublished manuscript.
- Tanenhaus, M. K., Stowe, L. A., & Carlson, G. (1985). *The interaction of lexical expectation and pragmatics in parsing filler-gap constructions*. Proceedings of the seventh annual cognitive science society meetings, Irvine, CA, pp. 361–365.
- Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, 47, 69–90.

- Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies. *Journal of Memory and Language*, 35, 454–475.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, 55, 157–166.
- van der Velde, F., & de Kamps, M. (2006). Neural blackboard architectures of combinatorial structures in cognition. *Behavioral and Brain Sciences*, 29, 37–70.
- Verhaegen, P., Cerella, J., & Basak, C. (2004). A working memory workout: How to expand the focus of serial attention from one to four items in 10 hours or less. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 30, 1322–1337.
- Wagers, M., & Phillips, C. (2009). Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics*, 45, 395–433.
- Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 119–161). Cambridge, MA: MIT Press.
- Weinberg, A. (1993). Parameters in the theory of sentence processing: Minimal Commitment theory goes east. *Journal of Psycholinguistic Research*, 22, 339–364.