Hyper-Active Gap Filling: Pre-verbal Object Gap Creation in English Filler-Gap Dependency Processing

Akira Omaki¹, Ellen Lau², Imogen Davidson White³, Myles Louis Dakan⁴, Colin Phillips²

¹Johns Hopkins University, ²University of Maryland, ³University College London, ⁴Northeastern University

Address for correspondence:
Akira Omaki
Department of Cognitive Science
Room 237 Krieger Hall, Johns Hopkins University
3400 North Charles Street
Baltimore, MD 21218
Email: omaki@cogsci.jhu.edu
Author Notes

We thank Michael Shvartsman and Erika Hussey for technical assistance in conducting the experiments, and Amanda Glasser for her assistance with the corpus analysis. This work was supported in part by a Wylie Dissertation Fellowship to AO, NIH F32-HD063221-02 to EFL, NSF BCS-0954651 to AO and CP, NSF BCS-0848554 to CP, and NSF DGE-0801465 to the University of Maryland.
Abstract

Much work has demonstrated that speakers of verb-final languages are able to construct rich syntactic representations in advance of verb information. This may reflect general architectural properties of the language processor, or it may only reflect a language-specific adaptation to the demands of verb-finality. The present study addresses this issue by examining whether speakers of a verb-medial language (English) wait to consult verb information before constructing filler-gap dependencies, where internal arguments are fronted and hence precede the verb. This configuration makes it possible to investigate whether the parser actively makes representational commitments on the gap position before verb information becomes available. A key prediction of the view that rich pre-verbal structure-building is a general architectural property is that speakers of verb-medial languages should predictively construct dependencies in advance of verb information, and therefore that disruption should be observed when the verb has properties that are incompatible with the predicted structure. In three reading experiments (self-paced and eye-tracking) that manipulated verb transitivity we found evidence for reading disruption when the verb was intransitive, although no such reading difficulty was observed when the critical verb was embedded inside a syntactic island structure, in which filler-gap dependency completion is blocked by linguistic constraints. These results are consistent with the hypothesis that in English, as in verb-final languages, information from preverbal NPs is sufficient to trigger active gap creation without having access to verb transitivity information.
HYPER-ACTIVE GAP FILLING: PRE-VERBAL OBJECT GAP CREATION IN ENGLISH
FILLER-GAP DEPENDENCY PROCESSING

A leading goal of sentence processing research is to understand how the parser adapts to a multitude of linguistic differences across languages to enable successful comprehension. In this regard, comparisons of verb-medial and verb-final languages have provided a valuable source of evidence (Inoue & Fodor, 1995; Mazuka & Lust, 1990). The head of a verb phrase (VP) contains rich information such as subcategorization and thematic role information that is critical for constructing structural analyses and interpretations of the phrase (e.g., Chomsky, 1965; Grimshaw, 1990; Pollard & Sag, 1994), and much experimental evidence shows that the verb is a valuable source of information for parsing (e.g., Blodgett & Boland, 2004; Boland, Tanenhaus, & Garnsey, 1990; Ford, Bresnan, & Kaplan, 1982; Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Mauner & Koenig, 2000; MacDonald, Pearlmutter, & Seidenberg, 1994; Snedeker & Trueswell, 2004; Spivey-Knowlton & Sedivy, 1995; Tanenhaus & Carlson, 1989; Traxler, Pickering, & McElree, 2002). The importance of the information from the verb head has engendered theoretical claims that structure building processes do not even start until the parser encounters the head of a structural unit to be constructed, even in verb-final languages where this would be significantly delayed (Pritchett, 1992; Abney, 1989).

However, subsequent empirical research on verb-final languages like Japanese or German has generated strong evidence against such head-driven parsing theories in their strongest form, demonstrating that the parser uses various morphological and syntactic cues to incrementally build structures and interpretations in verb-final languages (Bader & Lasser, 1994; Kamide, Altmann, & Haywood, 2003; Kamide & Mitchell, 1999; Koh, 1997; Konieczny, 2000; Yoshida, 2006; Aoshima, Yoshida, and Phillips, 2009; Bornkessel, Schlesewsky, & Friederici, 2022).
Thus, although verb information strongly influences parsing decisions when available, speakers of verb-final languages often begin building syntactic and semantic structure in advance of the verb.

These findings raise the question of whether pre-verbal structure building reflects a language-specific adaptation to the processing demands of verb-finality, or rather a property of a general parsing architecture that speakers of all languages use. For example, consider less frequent cases in verb-medial languages where multiple arguments precede the verb. A classic example of this comes from processing of ‘filler-gap’ dependencies as illustrated by the relative clause construction shown in (1), where the object noun phrase (NP) the city (called the filler) is dislocated from the post-verbal thematic position (called the gap\(^1\)), and the parser needs to associate the filler and the gap in order to assign a thematic interpretation.

(1) The city that the author visited ____ was named for an explorer.

It has been reported that speakers of verb-final languages complete filler-gap dependencies in advance of verb information, associating the filler with the earliest structural position where a thematic role could be assigned (Aoshima, Phillips, & Weinberg, 2004; Nakano, Felser, & Clahsen, 2002). The current study examines whether this may also be the case in a verb-medial language like English. If positing a gap in advance of the verb is a language-general parsing procedure rather than an adaptation specific to verb-final languages, then we predict that English speakers should posit a gap irrespective of whether the verb ultimately licenses a direct object

\(^1\) In this paper we use the ‘gap-filling’ terminology in a theory-neutral way, as is typical in the psycholinguistic literature. This terminology should not be taken as indicating a commitment to representations that include gaps or traces; all of the processing theories we discuss here could be specified in terms of representations that do not include empty categories.
gap position, and that signs of reading disruption should be observed in cases where the verb does not accommodate a direct object.

We report the results of three on-line reading experiments in English that tested this prediction by examining the effect of verb transitivity on reading times in filler-gap configurations. The results are consistent with the hypothesis that the parser creates the object gap in advance of the verb in English, specifically as soon as a VP constituent can be anticipated based on the presence of the filler and the subject NP, as well as grammatical knowledge that a VP can potentially contain an object NP position. These results suggest that the filler-gap dependency completion procedure may be uniform across languages, regardless of differences in verb positions. We argue that this finding is consistent with the view that the parser predictively constructs rich representations at the earliest possible moment in advance of critical bottom-up evidence.

Background on Active Filler-Gap Dependency Processing

Past research on filler-gap dependency processing has established that the parser postulates a gap before there is sufficient bottom-up evidence to confirm that analysis (Active gap filling: Fodor, 1978; Crain & Fodor, 1985; Stowe, 1986; Frazier & Flores D’Arcais, 1989). For example, Stowe (1986) observed the so-called Filled gap effect in (2), i.e., slower reading times at the direct object position *us* in the wh-fronting condition (2a) than in a control condition that did not involve wh-fronting (2b). This pattern of reading times suggests that the parser had already posited a gap following the transitive verb, before checking whether the direct object position was occupied.
(2)  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.

Converging evidence comes from an eye-tracking experiment by Traxler and Pickering (1996), who manipulated the thematic fit between the filler and the potential verb host, as in (3).

(3)  We like the city / book that the author wrote unceasingly and with great dedication about ____ while waiting for a contract.

Traxler and Pickering found a plausibility mismatch effect at the critical verb in (3), i.e., the first fixation time at the optionally transitive verb wrote increased when the filler was an implausible object of the verb (i.e., the city), compared to when the filler was a plausible object of the verb (i.e., the book). This suggests that at least as early as the verb position, the parser postulates a gap and analyzes the filler as the object of the verb. In fact, there is ample time course evidence for active object gap creation in many languages, using a variety of dependent measures such as reading time and gaze duration measures (Crain & Fodor, 1985; de Vincenzi, 1991; Frazier, 1987; Frazier & Clifton, 1989; Phillips, 2006; Pickering & Traxler, 2001, 2003; Wagers & Phillips, 2009), cross-modal priming (Nicol, 1993; Nicol & Swinney, 1989), visual world eye-tracking (Sussman & Sedivy, 2003) as well as event-related potentials (Featherston, Gross, Münte, & Clahsen, 2000; Felser, Münte, & Clahsen, 2003; Garnsey, Tanenhaus, & Chapman, 1989; Gouvea, Phillips, Kazanina, & Poeppel, 2010; Kaan, Harris, Gibson, & Holcomb, 2000; Phillips, Kazanina, & Abada, 2005).

The work summarized above may suggest that filler-gap dependency completion is
triggered only after the parser gains access to the verb and confirms the transitivity and thematic properties of the verb. However, research on subject gap creation in English as well as studies that investigated object gap creation in verb-final languages have presented evidence that active gap creation does not depend on verb information. For example, Lee (2004) used sentences like (4) to reveal a filled gap effect in the subject NP position.

(4)    a. That is the laboratory which, on two different occasions, Irene used a courier to deliver the samples to ___.

    b. That is the laboratory to which, on two different occasions, Irene used a courier to deliver the samples ___.

Here, the content of the wh-filler is manipulated in such a way that the wh-filler can plausibly be a subject (4a) or not (4b). The results showed a longer reading time at the subject NP Irene in (4a) than in (4b), suggesting that the parser had postulated a subject gap before encountering the actual subject NP. Although this interpretation has been challenged (Staub, 2010), it would in any case not be surprising that the parser actively creates a subject gap without having access to verb information, given that a subject is present in any sentence, regardless of verb properties. In this sense, if verb information were to play a role in the parser’s attempt to create a gap, the critical empirical evidence should come from object gap creation, where the presence or absence of an object gap position relies on properties of the verb.

Evidence for pre-verbal object gap creation has been reported for verb-final languages like Japanese in which the object gap position linearly precedes the verb. For example, Aoshima and colleagues examined processing of long-distance scrambling sentences in which an
embedded dative object NP was dislocated to sentence initial position, and found a filled gap effect at a pre-verbal dative object position (Aoshima et al., 2004). Using similar sentences, Nakano and colleagues reported evidence for an antecedent priming effect for the scrambled NP at a pre-verbal gap position (Nakano et al, 2002). These data indicate that the parser can in principle complete filler-gap dependencies before accessing verb information.

In verb-medial languages, no such evidence for pre-verbal object gap creation has been reported to date. This may reflect a real difference between languages in processing strategy, and pre-verbal object gap creation in verb-final languages may reflect the parser’s adaptation to the demands of processing these languages. Maintaining the filler in memory while it is structurally unintegrated or when it has not received a thematic interpretation has been argued to impose a burden on working memory (Gibson, 1998; Gordon, Hendrick, & Levine, 2002, Haarmann & Cameron, 2005; King & Just, 1991). Alternatively, the parser may be architecturally constrained to assign a thematic interpretation to the filler as soon as possible (Aoshima, Phillips, & Weinberg, 2004; Pickering & Barry, 1991). On this view, the parser should prioritize integrating the filler into the first grammatically permissible structural position that can potentially receive a thematic role. Given that filler-gap dependencies are potentially unbounded, waiting for the verb before constructing the ultimate object gap position could impose a large processing burden on speakers of verb-final languages.

In verb-medial languages like English, verbs become available relatively earlier in the sentence, such that the average working memory cost of waiting for the verb would be less than in verb-final languages. The advantage of waiting for the verb information is that the parser can reduce the likelihood of making risky commitments, because the verb may turn out to be intransitive and disallow an object NP analysis for the filler. In English, therefore, the parser may
create an object gap position only after the verb is confirmed to be transitive. This still constitutes active gap filling, in the respect that the ultimate gap position may turn out to be somewhere later than the object position in the sentence (as in the sentences that are used to illustrate filled gap effect in (2)). Let us call this a *conservative active gap filling* mechanism, since the bottom-up information from the verb still plays a critical role in the parser’s decision on whether to postulate an object gap or not. This view of active gap filling is rather standard for explaining filler-gap dependency completion in verb medial languages like English. For example, McElree and colleagues have argued that the dependency completion process is triggered when the parser accesses information from the verb and initiates the retrieval process for the filler that is stored in working memory (McElree & Griffith, 1998; McElree, Foraker, & Dyer, 2003; see also Lewis & Vasishth, 2005; Pickering & Barry, 1991).

On the other hand, pre-verbal object gap creation in verb-final languages may reflect a language-general property of the processing architecture, although evidence for such mechanisms may be simply more difficult to obtain in verb-medial languages. In the English filler-gap case, for example, in any parser that adopts some form of left-corner strategy, the presence of the subject NP allows the parser to predict the presence of a VP (Abney & Johnson, 1991; Crocker, 1996; Gibson, 1991; Kimball, 1975; Resnik, 1992; Shieber & Johnson, 1993, Stabler, 1994). Given that a VP can contain an object NP position, the parser could project a VP with an object NP slot and assign the filler to this object position before confirming whether the upcoming verb is a transitive verb or not. Let us call this a *hyper-active gap filling* mechanism, because this involves a more risky predictive structure building process than is standardly assumed for active object gap creation in English. Filler retrieval and structural integration is still integral to the hyper-active gap filling mechanism, but the crucial difference is in what
information triggers the retrieval and integration, and consequently, at what point in the sentence this process is executed.

It is important to note that either of these two active gap filling mechanisms is compatible with the existing data on active object gap creation reviewed above. A filled gap effect only indicates that the gap had been created before the actual object NP is processed, but given that both hyper-active gap filling and conservative active gap filling mechanisms assume that object NP gap creation happens before or on the verb, this result is predicted by both accounts. A plausibility mismatch effect indicates that when the verb is potentially transitive, then the semantic fit between the filler and the verb is immediately assessed. This is also predicted by both accounts. The assessment of the semantic relation between the filler and the verb requires the parser to access the content of the verb, by which point the object gap position should have been created on either account. Thus, neither paradigm allows us to tease apart the two hypotheses on what kind of information is sufficient for triggering object gap creation.

In the current study we test the predictions of two hypothesized mechanisms for active object gap creation processes. If English speakers construct the gap site before encountering the verb, just like speakers of verb-final languages, then English speakers risk the possibility that the verb transitivity information might not ultimately license this structure. Therefore, disruption should be observed in filler-gap configurations when the verb turns out to be intransitive, relative to transitive verbs (e.g., The party that the student arrived/planned...). According to the conservative active gap filling mechanism outlined above, the parser waits for a transitive verb before postulating the corresponding gap structure. If this is the mechanism used by English speakers, one should not expect to see disruption at an intransitive verb, since no gap that would require a transitive verb would have been posited in advance of the verb.
Two previous studies are relevant to the two hypotheses about active object gap creation in English. Previous work by Pickering and Traxler (2003) examined the effect of subcategorization frequency in optionally transitive verbs (e.g., *Those are the lines/props that the author spoke [about]...*). It was found that readers did not take differences in subcategorization frequency into account in deciding where to posit a gap, as there was a strong preference to posit a gap in the verb object position (NP complement) even with verbs that more frequently take a PP complement. The absence of a subcategorization preference in active object gap creation could be taken to indicate that verb information is not relevant for object gap creation processes. However, all of the verbs in Pickering and Traxler’s study could grammatically accommodate an NP complement, and that the parser may therefore have relied on the transitivity information of the verb to create an object gap. Therefore, this finding does not distinguish the predictions of the two proposed mechanisms for active object gap creation.

To our knowledge, the only previous test of these two active object gap creation hypotheses is in Experiment 3 of Staub (2007). The test sentences in this experiment (5a-d) manipulated the transitivity of the verb (*called* vs. *arrived*) and sentence structure (relative clause with a gap vs. simple declarative with no gap). The filler was manipulated to be an implausible object of the transitive verb (*gadget-called*). Under the hyper-active gap filling hypothesis, the parser in effect predicts the presence of a transitive verb, and therefore the reading processes in the gap conditions should be disrupted when the verb turns out to be intransitive, and processing should also be disrupted when the verb is transitive because of the plausibility mismatch effect. On the other hand, the conservative active gap filling mechanism postulates a gap only after checking whether the verb is capable of hosting an object NP, and therefore reading disruption is predicted only in the transitive gap condition due to the plausibility mismatch effect.
(5)  a. The gadget that the manager *called* occasionally about …  
    b. The manager *called* occasionally about the gadget …  
    c. The party that the student *arrived* promptly for …  
    d. The student *arrived* promptly for the party …

Staub (2007) found longer first-fixation durations in the transitive gap condition (5a) than in the transitive no-gap condition (5b), but no such difference was observed between the intransitive gap and no-gap conditions (5c) and (5d). This pattern of data supports the prediction of the conservative active gap filling hypothesis, suggesting that the parser does not create an object gap until it checks the transitivity information of the verb. One concern about this design, however, is whether the no-gap condition was truly a neutral baseline against which a transitivity mismatch could be measured, as the gap and no-gap conditions differed substantially in both the linear and structural position of the verb. As Staub (2007) points out, one piece of data suggesting that the control may not have been completely neutral is the fact that reading times on the intransitives were numerically (but non-significantly) shorter in the gap condition than in the no-gap condition. One notable difference between the conditions is that the gap conditions (5a) and (5c) contain an extra NP (i.e., the head of the relative clause) prior to the critical verb region in comparison to the no-gap conditions (5b) and (5d), leading to a difference in the amount of contextual information available prior to the verb. Increased contextual information can facilitate processing for subsequent lexical items (Kutas & Federmeier, 2000; Stanovich & West, 1983; Van Petten & Kutas, 1990), and for this reason, lexical access for the intransitive verb in the gap condition may have become faster and masked the potential reading time slowdown associated
with the structural manipulation. In an attempt to provide a better test of the predictions of the hyper-active and conservative active gap filling accounts, the current study used relative clause islands as a control condition, which allowed the target sentences to more closely match in informational content and word position.

EXPERIMENT 1

Experiment 1 was a self-paced reading study that was designed to test the predictions of the hyper-active and conservative active gap filling hypotheses, while addressing methodological concerns about previous work. We employed the transitivity mismatch paradigm used in Staub (2007) in order to test whether a verb transitivity manipulation affects reading time at the verb. Critically, in the baseline conditions the critical verb was embedded inside a relative clause structure, a syntactic ‘island’ domain that prohibits filler-gap dependency formation (Ross, 1967; for a review, see Szabolcsi & den Dikken, 2003). A sample set of stimuli is shown in (6).

(6) a. Transitive, Non-island

The city that the author wrote regularly about was named for an explorer.

b. Transitive, Island

The city that the author who wrote regularly saw was named for an explorer.

c. Intransitive, Non-island

The city that the author chatted regularly about was named for an explorer.

d. Intransitive, Island

The city that the author who chatted regularly saw was named for an explorer.
A number of previous studies have shown that the parser respects island constraints in real-time syntactic processing, such that it avoids actively constructing filler-gap dependencies that span syntactic island boundaries (Kluender & Kutas, 1993; McElree & Griffith, 1998; McKinnon & Osterhout, 1996; Stowe, 1986; Traxler & Pickering, 1996; Wagers & Phillips, 2009; Yoshida, 2006). The relative clause island condition thus provided a baseline measure of reading times for the critical transitive and intransitive verbs, independent of processes of filler-gap dependency completion. The use of island configurations allowed us to address the methodological concerns with previous work. First, this design allowed us to retain the filler and the gap surrounding the island domain, such that the same amount of contextual information from the lexical items was present in advance of the critical verb region across the four conditions. Second, the critical region was closely matched across conditions (word 6 in the non-island conditions, word 7 in the island conditions), and it was also placed away from the early portion of the sentence.

Furthermore, following Staub’s design, we selected transitive verbs that are implausible hosts for the filler. Under this design, the hyper-active gap filling hypothesis predicted a reading time slowdown in both the non-island transitive (6a) and the non-island intransitive (6c) conditions relative to their baseline conditions (6b) and (6d), but for a different reason in the two cases. In the transitive condition, the slowdown would reflect a plausibility mismatch effect triggered by the semantic misfit between the filler and the verb. In the intransitive condition, the slowdown would result from a transitivity mismatch effect due to the mismatch between the expected subcategorization property of the verb (i.e., transitive) and the actual subcategorization property of the verb. On the other hand, the conservative active gap filling hypothesis predicted an interaction. A reading time contrast should be observed between the non-island transitive
condition (6a) and the island transitive condition (6b) due to the plausibility mismatch effect, but no corresponding contrast should be observed between the two intransitive conditions (6c) and (6d), given that the parser should not actively create an object gap in either condition. Note that the lexical difference in the critical verb region across conditions was not problematic, since the critical contrast was between non-island and island conditions within each verb type.

Method

Participants

We recruited 32 native speakers of American English from the University of Maryland community. They all received a course credit or were paid $10 for their participation and were naïve to the purpose of the experiment.

Materials

We used 28 sets of four sentences like (6a) to (6d), which are all listed in Appendix A. The transitive non-island and island conditions were taken from the implausible semantic fit conditions in Omaki and Schulz (2011), who used a modified version of the plausibility manipulation materials from Traxler and Pickering (1996). Omaki and Schulz replicated Traxler and Pickering’s plausibility mismatch effect with native and non-native speakers alike, confirming that the semantic fit between the filler and the verb affects the reading time for the verb when the verb is in a gap filling (i.e., non-island) environment, but not when the verb is inside a relative clause island. Critically, it was also found that the implausible verb-filler combination in a non-island environment (e.g., city-wrote) led to a significant slow down at the verb compared to its island counterpart with the same implausible verb-filler combination. Thus, even though the current experiment did not include a plausible counterpart of the implausible
transitive verb condition, we could be confident that a reading time contrast between the transitive non-island and island conditions results from the semantic misfit between the filler and the verb. In other words, the finding in Omaki and Schulz’s study supports the notion that island conditions in general can be used as baseline conditions for a reading disruption associated with active object gap creation. The intransitive conditions were modeled after the transitive conditions by replacing the optionally transitive verb with unergative or unaccusative intransitive verbs (Levin & Rappaport Hovav, 1995).

The non-island and island conditions differed in the number of relative clauses: The non-island condition had only one relative clause (the city that the author wrote/chatted regularly about) such that the object position of the verb wrote/chatted was the first potential gap position after the embedded subject was encountered, whereas in the island conditions the critical verb was embedded inside another relative clause the author who wrote/chatted regularly, such that linearly this was still the first verb but grammatically the filler should not be accessible to the verb due to the relative clause island constraint. Thus, the first verb served as the critical region for testing the plausibility and transitivity mismatch effects. All the transitive verbs were optionally transitive, such that the sentences in the island conditions were all ultimately grammatical. The subcategorization frequency of the optionally transitive verbs was not controlled, since Pickering and Traxler (2003) have demonstrated that plausibility mismatch effects are attested for optionally transitive verbs regardless of subcategorization frequency. In all four conditions the same adverb immediately followed the verb, making it possible to observe potential spill-over effects. The 28 sentence sets were counter-balanced across four lists so that each participant saw only one version of the target items and consequently read 7 tokens of each condition. In addition, 72 fillers of similar length and complexity were constructed and added to
each list.

Procedure

The self-paced reading task was implemented on the Linger software developed by Doug Rohde (http://tedlab.mit.edu/~dr/Linger/). We used a word-by-word, non-cumulative moving window presentation (Just, Carpenter, & Woolley, 1982). In this design, each sentence initially appears as a series of dashes, and these dashes are replaced by a word from left to right every time the participant presses the space bar. In order to ensure that the participants were paying attention while reading the sentences, all sentences were followed by yes-no comprehension questions, and feedback was provided if the questions were answered incorrectly. Comprehension questions never addressed the critical filler-gap portion of the sentence. At the beginning of the experiment, participants were instructed to read at a natural pace and to answer the questions as accurately as possible. Seven practice items preceded the self-paced reading experiment, and the order of presentation was randomized for each participant. The experiment took approximately 30 minutes.

Analysis

The data from two items were excluded from analyses due to coding errors, and only trials in which the comprehension question was answered accurately were included in the analysis. We also analyzed the data without excluding the trials based on comprehension accuracy, but the overall pattern of results did not change. Self-paced reading times for the target sentences were examined for each successive region, although the words after the auxiliary was were combined into a single region because these lay beyond the critical regions and were unlikely to show effects relevant for the critical manipulation. Reading time data that exceeded three standard deviations from the group mean at each region and in each condition were
excluded, affecting 1.7% of the data. The participant mean (F1) and item mean (F2) of the remaining reading time data for each region were submitted to a repeated measures $2 \times 2$ ANOVA with the factors structure type (non-island vs. island) and verb type (transitive vs. intransitive). In the critical regions, planned comparisons were conducted to test for systematic differences between non-island and island conditions within each verb type.

**Results**

*Comprehension accuracy.* The mean comprehension question accuracy for experimental items across participants and items was 93.0%. For the non-island conditions, the transitive items were answered with an accuracy of 93.7% ($SE = 1.9$), and the intransitive items with an accuracy of 94.6% ($SE = 1.4$). For the island conditions, the transitive items were answered with an accuracy of 91.5% ($SE = 1.7$), and the intransitive items with an accuracy of 92.0% ($SE = 2.2$). The mean accuracy did not differ reliably across conditions, although the fact that the mean accuracy for island conditions was numerically lower may reflect the complexity difference between non-island and island conditions.

*Reading time data.* Regions were defined as in the following example: *The$_1$ city$_2$ that$_3$ the$_4$ author$_5$ (who)$_6$ wrote$_7$ regularly$_8$ about/(saw)$_9$ was$_{10}$ [named for an explorer]$_{11}$. The critical regions where a potential plausibility or transitivity mismatch effect was expected consist of Region 7 (i.e., the verb *wrote/chatted*) and the following Region 8 (i.e., the adverb *regularly*), in which spill-over effects could be observed. Regions 1 through 6 were predicted to show no difference across conditions, since they were lexically matched. Regions 9 through 11 could reveal reading time differences after the filler-gap dependency is completed (Region 9 hosts the true gap site), and with a possible additional difference in the island conditions, due to the complexity associated with the extra relative clause in these conditions.
The region-by-region mean reading time for the transitive conditions is presented in Figure 1, and the mean region-by-region reading time for the intransitive conditions is presented in Figure 2.

![Graph showing mean reading time](image)

**Figure 1.** Mean reading time (ms) for the transitive non-island and island conditions. Error bars indicate standard error of the mean.

Sample sentence (words in parentheses appear only in island conditions; words in brackets represent one region):

The\textsubscript{1} city\textsubscript{2} that\textsubscript{3} the\textsubscript{4} author\textsubscript{5} (who)\textsubscript{6} wrote\textsubscript{7} regularly\textsubscript{8} about/(saw)\textsubscript{9} was\textsubscript{10} [named for an explorer]\textsubscript{11}. 
Figure 2. Mean reading time (ms) for the intransitive non-island and island conditions. Error bars indicate standard error of the mean.

Sample sentence (words in parentheses appear only in island conditions; words in brackets represent one region):

The\textsubscript{1} city\textsubscript{2} that\textsubscript{3} the\textsubscript{4} author\textsubscript{5} (who)\textsubscript{6} chatted\textsubscript{7} regularly\textsubscript{8} about/(saw)\textsubscript{9} was\textsubscript{10} [named for an explorer]\textsubscript{11}.

The statistical analysis of reading time data revealed some spurious effects in non-critical regions (Regions 3 and 4), but crucially the predicted effects in the critical regions were larger than those effects.

In the non-critical Regions 1 to 6, there were no significant differences in Regions 1, 2, 5 and 6 ($F$s < 1). In Region 3 there was a main effect of verb type, $F1(1, 31) = 10.30$, MSE = 9671, $p < .005$, $F2(1, 25) = 6.63$, MSE = 13687, $p < .05$, due to slower reading times in the transitive conditions than in the intransitive conditions (381 ms vs. 358 ms). Region 4
showed a main effect of the island manipulation in the participant analysis only, $F1(1, 31) = 5.20$, MSE = 5391, $p < .05$, $F2(1, 25) = 1.79$, MSE = 2166, $p > .1$, due to slower reading times in the island conditions (352 ms vs. 365 ms). Since these regions were lexically matched across conditions, we conclude that these must be spurious effects. But given that the effects were small and occurred well ahead of the critical regions, these unexpected effects were unlikely to have impacted the observations in the critical regions.

At the critical verb in Region 7 there were no significant differences ($F$s < 1). The following spill-over region (Region 8) revealed no main effect of verb type ($F$s < 1), but there was a main effect of structure type, $F1(1, 31) = 16.90$, MSE = 287354, $p < .0005$, $F2(1, 25) = 14.79$, MSE = 230119, $p < .005$, reflecting the fact that the non-island conditions produced significantly slower reading times than the island conditions (529 ms vs. 435 ms). There was no significant interaction of verb type and structure type ($F$s < 1). A pair-wise comparison revealed that the reading times in the non-island condition were significantly slower than the reading times in the island condition for transitive sentences (529 ms vs. 441 ms), $t1(31) = 3.15, p < .005, t2(25) = 2.54, p < .05$, as well as for intransitive sentences (542 vs. 441 ms), $t1(31) = 4.05, p < .0005, t2(25) = 3.64, p < .005$.

Region 9 consisted of a second verb in the island conditions and a preposition in the non-island conditions. We observed a main effect of structure type in Region 9, $F1(1, 31) = 13.53$, MSE = 150149, $p < .005$, $F2(1, 25) = 13.61$, MSE = 108850, $p < .005$, as well as in Region 10, $F1(1, 31) = 12.75$, MSE = 64094, $p < .005$, $F2(1, 25) = 5.78$, MSE = 48875, $p < .05$, in these cases due to slower reading times in the island conditions (Region 9: 519 ms vs. 451 ms, Region 10: 451 ms vs. 406 ms). Region 11 revealed no significant differences ($F$s < 1).
Discussion

In Experiment 1, we tested the predictions of two hypotheses about active object gap creation. The hyper-active gap filling hypothesis predicted the presence of reading disruption at intransitive verbs, because encountering an intransitive verb in a filler-gap context would be incompatible with the object gap structure constructed earlier. On the other hand, the conservative active gap filling hypothesis predicted no such reading disruption, because the parser should first consult the transitivity information of the verb to decide whether to posit an object gap or not. As a baseline for estimating the degree of disruption at the verb due to plausibility mismatch and transitivity mismatch, we used relative clause island constructions, which block the association of the filler with the critical verb. In the region following the verb, we observed slower reading times for both plausibility-mismatched transitive verbs and intransitive verbs in non-island conditions than in corresponding island conditions. Previous work has shown that these items with plausibility mismatched transitive verbs in a non-island environment elicit longer reading times than their plausible non-island or plausible/implausible island counterparts (Omaki & Schulz, 2011), and here we replicated the finding of a slowdown after the optionally transitive verb in the implausible non-island condition relative to the implausible island condition. This slowdown can be interpreted as the result of active association of the filler with the transitive verb, which in these stimuli resulted in a verb-object plausibility mismatch. In the island condition, the verb was inaccessible as a potential gap position and thus this implausible verb-object combination could not be formed. The slowdown observed in the intransitive non-island condition relative to the intransitive island condition can be interpreted as a transitivity mismatch. This suggests that the parser does not wait for bottom-up evidence from the verb that the verb can syntactically license a gap, but rather attempts to construct the
dependency before this information is available. This slowdown cannot reflect the cost of maintaining the filler in working memory, because a filler is also being maintained at this position in the baseline island condition.

In Regions 9 and 10, the island conditions were read more slowly for both levels of verb type. Region 9 corresponds to the word that licensed the true gap site across all conditions, and hence this slowdown could reflect a difference in the so-called integration cost (Gibson, 1998; 2000) between non-island and island conditions. Previous work on filler-gap dependency processing has demonstrated that increased complexity and length differences result in increased processing difficulties at the gap site, as measured by reading time (Gibson & Warren, 2004) and reduced accuracy in speeded acceptability judgment tasks (McElree et al., 2003).

Note that it is unlikely that the reading time contrast between non-island and island conditions in Region 8 is related to the overall complexity of the constructions used in our stimuli, given that on all accounts that we are aware of, island domains have been argued to be syntactically more complex and more taxing for working memory resources (Deane, 1991; Kluender, 1999, 2004; Kluender & Kutas, 1993; Hofmeister & Sag, 2010). The fact that the putatively less complex non-island conditions were read more slowly allows us to attribute the slowdown to processes that uniquely occur in the non-island conditions, namely filler-verb association.

In summary, the presence of both a plausibility mismatch effect and a transitivity mismatch effect lends support to the hyper-active gap filling hypothesis, and argues against a conservative active gap filling hypothesis under which transitivity information is consulted before attempting to create an object gap. This finding directly contrasts with that of Staub (2007), who did not find evidence for a transitivity mismatch effect. One possible reason for this
discrepancy is the difference in the baseline conditions used in the two studies. As noted above, the lack of the filler NP in Staub’s baseline condition may have unexpectedly created a difference in the amount of helpful contextual information in the two conditions and hence may have complicated the reading time prediction for the critical verb region. Moreover, there was a large difference in the word position of the critical verb across conditions, and this may have additionally complicated the reading time patterns. In the current study, both of these complicating factors were controlled. Both the target and baseline conditions contained the same filler (and hence the same contextual information) in advance of the critical verb region, and the word position of the critical verb was closely matched across conditions. Given that our study used a verb transitivity manipulation design similar to Staub (2007), the removal of these complicating factors seems to be the most likely explanation for the presence of verb transitivity mismatch effect.

However, there are other methodological differences between our Experiment 1 and Staub (2007) that could account for the difference in the findings. First, our intransitive materials consisted of two types of intransitive verbs: We mainly used unergative verbs which only take a semantic agent as an argument, but we also used unaccusative intransitive verbs that only take a theme/experiencer as an argument (Perlmutter, 1978; Levin & Rappaport-Hovav, 1995). On the other hand, Staub’s intransitive condition used only unaccusative intransitive verbs. Both types of intransitive verbs are generally incompatible with an overt direct object NP, but in some restricted contexts unergative verbs are capable of hosting an NP object (e.g., “laugh a big laugh”; see Keyser & Roeper, 1984). It is possible that this special property of unergative verbs may have led the parser to treat it in the same way as transitive verbs in our experiment, whereas unaccusative intransitive verbs admit no such exceptions.
Second, Staub (2007) used an eye-tracking during reading method and reported evidence for active gap creation on the transitive verb region, whereas in our self-paced reading experiment, evidence for reading disruption for transitive and intransitive verbs (i.e., the slowdown in non-island conditions compared to island conditions) was not observed until the spill-over adverb region. Spill-over effects are common in self-paced reading experiments and it is thus common to attribute spill-over effects to processes triggered in a preceding region, but in our experiment there is an alternative explanation for the effect in the adverb region that would not require hyper-active gap filling. For the intransitive condition, the slowdown in the adverb region could indicate that the parser had expected the presence of a preposition, which would allow structural integration of the filler. Under this alternative account, the slowdown is not due to a transitivity mismatch on the verb, but rather to a word category expectation mismatch in the adverb region that was triggered by the verb itself. This account is entirely consistent with the conservative active gap filling hypothesis, since the parser’s expectation regarding filler-gap dependency completion is based on the information from the verb, and this may be the reason why the eye-tracking during reading study by Staub (2007) demonstrated reading disruption for transitive verbs at the verb region but not for intransitive verbs (for another eye-tracking demonstration of plausibility mismatch effect directly on the transitive verb, see Traxler & Pickering, 1996). On the other hand, an eye-tracking during reading method generally provides better temporal precision than the self-paced reading method used in our study (Rayner, 1998; Rayner & Pollatsek, 2006). Thus, if we use an eye-tracking during reading method while maintaining our structural manipulations, we may find evidence for a transitivity mismatch effect on the verb region as well, which would lend support to our interpretation of the self-paced reading results as evidence for the hyper-active gap filling hypothesis.
In sum, results from Experiment 1 are consistent with the predictions of the hyper-active gap filling hypothesis, contrary to the results reported in Staub (2007), possibly due to the more closely matched baseline conditions in our experiment. However, there were two methodological differences between our study and Staub’s study that call for caution in this conclusion, namely the broader range of intransitive verbs used in our study, and also the fact that we used a dependent measure that does not provide as good temporal precision as the eye-tracking during reading method. Experiment 2 was designed to address both of these concerns.

EXPERIMENT 2

The goal of Experiment 2 was to address the methodological concerns raised in Experiment 1 in testing the predictions of the hyper-active gap filling and conservative active gap filling hypotheses. In order to make our experimental methodology as close as possible to the study by Staub (2007), we made two methodological changes from Experiment 1. First, we constructed new sets of stimuli that used only the unaccusative intransitive verbs that were used in Staub (2007). Given that unaccusative intransitive verbs are syntactically incapable of hosting an overt direct object NP, this class of intransitive verbs provides a stronger test of the transitivity mismatch effect. Second, we used an eye-tracking during reading method instead of the self-paced reading task used in Experiment 1. Eye-tracking provides better sensitivity to the temporal dynamics of reading processes, and in the domain of filler-gap processing, time course evidence for gap filling such as plausibility mismatch effects have been attested in the first fixation duration on the verb region itself (Traxler & Pickering, 1996). Given that the hyper-active gap filling hypothesis concerns structure building processes that are predicted to be contradicted when readers encounter a specific verb type, it is critical that we use a dependent
measure that can elicit a mismatch response on the verb region itself.

Method

Participants

We recruited 44 native speakers of American English from the University of Maryland community. All had normal or corrected-to-normal vision, and were naïve to the purpose of the experiment. They received course credit or were paid $10 for their participation, which lasted around 40 minutes.

Materials

We used 24 sets of four sentences like (7), which are all listed in Appendix B.

(7)  

a. Intransitive, non-island

The opponent that the veteran tennis player prevailed skillfully over during the tournament was very gracious.

b. Intransitive, island

The opponent that the veteran tennis player who prevailed skillfully over beat during the tournament was very gracious.

c. Transitive, non-island

The opponent that the veteran tennis player played skillfully with during the tournament was very gracious.

d. Transitive, island

The opponent that the veteran tennis player who played skillfully with beat during the tournament was very gracious.
This experiment used the same transitivity mismatch logic as Experiment 1 and manipulated the verb transitivity type (intransitive vs. transitive). However, in this experiment the semantic fit between the filler and the transitive verb was always plausible, such that no reading disruption was expected at the transitive verb in the non-island condition. As in Experiment 1 we manipulated structure type (non-island vs. island), using conditions with relative clause island structures as baseline conditions. Relative clause islands provide an effective baseline, since they include the same filler NP and other lexical material as the non-island condition, while preventing dependency completion at the critical verb. As in Experiment 1, the transitive verbs were optionally transitive and the true gap position occurred outside the island domain, allowing the sentence to continue grammatically. The 24 sentence sets were counter-balanced across four lists, such that each participant saw only one version of each of the target sentences. Since only 12 intransitive verbs were used, participants saw each intransitive verb twice across the course of the experiment, once in an island context and once in a non-island context. The target sentences were combined with 108 fillers of similar length and complexity.

The two hypotheses on active gap creation mechanisms predict different reading time patterns for the transitive and intransitive conditions. The hyper-active gap filling hypothesis states that the parser can use pre-verbal information to create an object gap before encountering the verb, and hence reading disruption is predicted in the intransitive non-island condition in comparison to the corresponding island condition. No such reading disruption is expected in the transitive non-island condition because the transitive verb expectation is satisfied, and hence no reading time contrasts should be observed between the two transitive conditions. On the other hand, the conservative active gap filling hypothesis states that the object gap creation process
occurs only after checking whether the verb can host an object NP. In other words, in the transitive non-island condition gap creation should occur, but in the intransitive non-island condition no gap should be created and hence there should be no disruption reflecting the need for reanalysis. Thus, this hypothesis predicts no reading time contrast between the two transitive conditions or between the two intransitive conditions.

Procedure

An SR Research (Mississauga, Ontario, Canada) Eyelink 1000 eye-tracker was used to record eye movements. The participant’s head was stabilized by a chin rest and a forehead rest. The position of the right eye only was monitored at a sampling rate of 1000 Hz. The eye-tracker display allowed a maximum of 100 characters per line. Some filler sentences were displayed on two lines, but all target sentences were displayed on one line. Stimuli were displayed on a 17-inch monitor, and participants were seated 60 cm from the computer screen.

Before the experiment started, participants were seated in front of the eye-tracker and received instructions for the experiment. A calibration routine was performed at the beginning of the experiment, and the experimenter monitored the calibration accuracy throughout the experiment, recalibrating when necessary. The experiment started with written instruction on the display and four practice trials. At the beginning of each trial, a black square was displayed on the left side of the monitor, which corresponded to the location of the beginning of the sentence. The text was displayed after the participant successfully fixated on the square. After reading each sentence, the participant pressed a button to remove the sentence display. Each sentence was followed by a yes-no comprehension question, and the participant answered the comprehension question by pressing a left or right button. Comprehension questions never addressed the critical filler-gap portion of the sentence. The entire experiment lasted approximately 40 minutes.
Data analysis

Data from six participants were removed due to calibration errors. Trials in which participants answered the comprehension question incorrectly were removed from the eye movement analyses. For the remaining data an automatic procedure pooled short contiguous fixations. The procedure incorporated fixations of less than 80 ms into larger fixations when they occurred within one character of each other and deleted any remaining fixations of less than 80 ms, because little information can be extracted during such short fixations (Rayner & Pollatsek, 1989). Unusually long fixations greater than 800 ms were also removed, because they usually reflect tracker losses or other anomalous events. This procedure resulted in the exclusion of 2.63% of all fixations.

For the purpose of analysis of the eye movement data, the sentences were divided into the following regions.

(8)  

a. Non-island conditions

The opponent that/ the veteran tennis player/ {prevailed | played}/ skillfully/

{over | with} during the tournament was very gracious.

b. Island conditions

The opponent that/ the veteran tennis player who/ {prevailed| played}/ skillfully/

had visited during the tournament was very gracious.

We report eye movement data in the following three regions: a) the pre-verb region (the veteran tennis player in (8a), the veteran tennis player who in (8b)), in order to ensure that there were no unexpected reading behavior differences that might compromise the interpretation of the data
from the critical region, b) the verb region, which is the critical region in which potential transitivity mismatch effects might be observed, and c) the post-verb region, which corresponds to the post-verbal adverb and could be used to probe for potential spill-over effects. The data in the remaining regions are not reported, because reading times at these regions are not critical for distinguishing the competing hypotheses, and after the post-verb region, the lexical items were not held constant across conditions and therefore any observed differences would be difficult to interpret. Each region started with the space before the first word in the region, and ended on the last character of the last word in the region.

Note that the island conditions contained one extra word, i.e., the relative pronoun (e.g., who), which could have affected reading times in the pre-verb region as well as in regression measures for subsequent regions. Furthermore, the length of the critical verbs in the transitive and intransitive conditions was not controlled, as the critical comparison was between non-island and island conditions within each level of the verb type factor. In fact, the mean length of the verbs was longer in the intransitive condition (7.58 characters) than in the transitive condition (6.33 characters), \( t(23) = 2.63, p < .05 \). Thus, the critical observation that is relevant for testing the current hypotheses comes from the structure type manipulation within each verb type, as well as from the interaction of structure type and verb type, rather than from a main effect of verb type, which may simply reflect the different properties of the lexical items across conditions.

Following the data analysis procedures used in Staub (2007), four reading time measures were computed for the three regions of interests: \textit{first fixation duration}, \textit{first pass time}, \textit{regression path time}, and \textit{percent regressions} (Rayner, 1998; Rayner & Pollatsek, 2006; Staub & Rayner, 2007). First fixation duration is the duration of the very first fixation in a region, regardless of whether there is a single word or multiple words in that region. This measure is
often used as an index of lexical difficulty (e.g., Reichle, Rayner, & Pollatsek, 2003) but is also informative about the earliest syntactic processes that immediately follow lexical access (e.g., Frazier & Rayner, 1982; Sturt, 2003). In fact, previous eye-tracking studies on filler-gap dependency completion have shown that the plausibility mismatch effect can be observed in the first fixation duration on the verb (Staub, 2007; Traxler & Pickering, 1996), and it is thus reasonable to expect a transitivity mismatch effect in this measure in the current study.

The first-pass reading time is calculated by summing the fixations in a region between the time when the eye-gaze first enters the region from the left and the time when the eye-gaze exits the region either to the left or the right. First-pass reading times also index early lexical and syntactic processes associated with a region, but given that they consist of multiple fixations on the same region, they may also reflect slightly later processes than the first fixation measure.

Regression path times are the sum of fixations from the time when the eye-gaze first enters a region from the left to the time when the eye-gaze exits the region to the right. Regression path time is identical to first-pass reading time if the eye-gaze first exits the region to the right, but if the eye-gaze exits the region to the left, then regression path times are longer than the first-pass time as they include all fixations in previous regions as well as re-fixations on the region before exiting the region to the right. Thus, regression path times are likely to reflect slightly later processes, such as integration of the critical region with the preceding context. The percent regressions indicate the probability that a reader made a regressive eye movement after fixating a given region. This measure includes only regressions made during the reader’s first pass through the region, and does not include regression made after re-fixating the region.

These eye movement measures from the pre-verb, verb, and post-verb regions were submitted to statistical analyses. The first fixation, first pass and regression path time data were
submitted to a 2 × 2 repeated measures ANOVA with verb type and structure type as within-
participants factors. For the percent regression measures, for reasons that the categorical
dependent measures such as proportion measures violate assumptions of ANOVA, we used
mixed-effects logistic regression, with verb type, structure type and their interaction as fixed
effects (See Jaeger, 2008). The fixed effects were coded using simple difference sum coding. We
tested random intercepts for participants and items, as well as random slopes for the
experimental fixed effects by participants and by items. When there was a main effect of
structure type or a significant interaction of verb type and structure type in the critical verb
region, a planned comparison was conducted to test for systematic differences between the island
and non-island conditions within each verb type.

Results

Mean comprehension accuracy for the experimental items was 91.9% across the four
conditions, and did not differ across the four conditions. Table 1 presents the participant means
on each measure for each region as well as the standard errors of the participant means.

Table 1
Experiment 2 Participant Mean Reading Times in Milliseconds (Standard Error) and
Percent Regressions

<table>
<thead>
<tr>
<th>Measure</th>
<th>Pre-verb region</th>
<th>Verb region</th>
<th>Post-verb region</th>
</tr>
</thead>
<tbody>
<tr>
<td>First fixation duration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intransitive, non-island</td>
<td>226 (7)</td>
<td>299 (10)</td>
<td>271 (9)</td>
</tr>
<tr>
<td>intransitive, island</td>
<td>222 (6)</td>
<td>270 (8)</td>
<td>259 (8)</td>
</tr>
<tr>
<td>transitive, non-island</td>
<td>229 (8)</td>
<td>277 (8)</td>
<td>268 (11)</td>
</tr>
<tr>
<td>transitive, island</td>
<td>236 (8)</td>
<td>266 (8)</td>
<td>258 (9)</td>
</tr>
<tr>
<td>First-pass time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intransitive, non-island</td>
<td>349 (19)</td>
<td>379 (13)</td>
<td>340 (15)</td>
</tr>
<tr>
<td>intransitive, island</td>
<td>461 (21)</td>
<td>345 (20)</td>
<td>330 (21)</td>
</tr>
<tr>
<td>transitive, non-island</td>
<td>367 (22)</td>
<td>319 (11)</td>
<td>308 (14)</td>
</tr>
<tr>
<td>transitive, island</td>
<td>468 (29)</td>
<td>316 (14)</td>
<td>321 (16)</td>
</tr>
<tr>
<td>Regression path time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intransitive, non-island</td>
<td>538 (43)</td>
<td>528 (38)</td>
<td>545 (40)</td>
</tr>
<tr>
<td>intransitive, island</td>
<td>762 (48)</td>
<td>527 (54)</td>
<td>497 (43)</td>
</tr>
</tbody>
</table>
In the pre-verb region, the first fixation duration measure showed no significant differences ($F$s < 2.5, $p$ > .13). The first pass reading time measure showed a main effect of the structure type manipulation, $F1(1, 37) = 31.39$, MSE = 426763, $p$ < .0005, $F2(1, 23) = 43.88$, MSE = 245985, $p$ < .0005, due to slower reading times in the island conditions (358 ms vs. 464 ms), and there was no main effect of verb type or interaction of the two factors ($F$s < 1).

Similarly, in the regression path duration measure, we found a main effect of structure type, $F1(1, 37) = 36.67$, MSE = 1518807, $p$ < .0005, $F2(1, 23) = 29.08$, MSE = 812343, $p$ < .0005, due to slower reading times in the island conditions (534 ms vs. 734 ms), and there was no main effect of verb type or interaction of the two factors ($F$s < 2.5, $p$ > .12). This structure type effect was expected, since the island conditions contained an extra relative pronoun. In the percent regressions measure, there were no significant effects of verb type or structure type ($\beta$ < .09, $p$ > .57), but there was a marginally significant interaction ($\beta$ = .55, $SE$ = .32, $p$ = .085). However, a planned comparison revealed that no reliable structure effect was present in either intransitive conditions ($\beta$ = .34, $p$ = .13) or transitive conditions ($\beta$ = .22, $p$ = .34).

Turning to the critical verb region, in the first fixation duration there was a main effect of structure type in both the participants and the items analyses, $F1(1, 37) = 5.17$, MSE = 15614, $p$ < .05, $F2(1, 23) = 5.09$, MSE = 8395, $p$ < .05, as well as a marginal effect of verb type in the participants analysis, $F1(1, 37) = 3.77$, MSE = 6641, $p$ = .06, $F2(1, 23) = 2.20$, MSE = 4215, $p$ > .1, with no significant interaction between the factors ($F$s < 1.7, $p$ > .2). A pair-wise
comparison was conducted to test whether the structure type manipulation affected the first fixation duration for both verb types. This comparison revealed that the reading time in the non-island intransitive condition was significantly longer than in the island intransitive condition (299 ms vs. 270 ms), $t_1(37) = 2.31, p < .05$, $t_2(23) = 2.49, p < .05$, but that the difference in first fixation duration between the transitive non-island and island conditions did not reach significance (277 ms vs. 266 ms; $ts < 1.1$).

On the first pass measure in the verb region, we observed a main effect of verb type, $F_1(1, 37) = 13.91$, $MSE = 73856$, $p < .005$, $F_2(1, 23) = 10.73$, $MSE = 45071$, $p < .005$, due to slower reading times in the intransitive conditions than in the transitive conditions (362 ms vs. 318 ms), and there was a significant interaction of the two factors in the items analysis, $F_1(1, 37) = 1.81$, $MSE = 8736$, $p > .1$, $F_2(1, 23) = 4.37$, $MSE = 11062$, $p < .05$. A pair-wise comparison revealed that the reading time in the non-island intransitive condition was significantly longer than in the island intransitive condition in the items analysis, though the effect was marginal in the participants analysis (379 ms vs. 345 ms), $t_1(37) = 1.7, p = .097$, $t_2(23) = 2.26, p < .05$. However, the difference in first pass between the transitive non-island and island conditions did not reach significance (319 ms vs. 315 ms; $ts < 1$).

Because the regression path duration measure reflects differences in the probability of regressing from this region, we discuss the percent regressions results on the verb region first. The percent regressions results demonstrated a significantly greater proportion of regressions in the island conditions relative to the non-island conditions ($\beta = .926$, $SE = .398$, $p < .0005$), and there were no other significant effects ($p > .26$). A model containing planned comparisons revealed that the structure type effect was present for both verb types: fewer regressive eye movements were made in the intransitive non-island condition than in the intransitive island
condition (14.4% vs. 24.0%; $\beta = .703$, $SE = .271$, $p < .001$), and the same pattern was observed in the corresponding transitive conditions, (11.7% vs. 28.4%; $\beta = 1.149$, $SE = .292$, $p < .0005$).

The larger proportion of percent regression in island conditions most likely reflects the structural complexity of the island conditions.

In the analysis of the regression path time in the verb region, there was a marginally significant effect of structure type, $F1(1, 37) = 3.30$, MSE = 169918, $p = .078$; $F2(1, 23) = 3.63$, MSE = 80697, $p = .07$, and a main effect of verb type that was significant in the participants analysis but only marginal in the items analysis, $F1(1, 37) = 4.26$, MSE = 209916, $p < .05$; $F2(1, 23) = 3.70$, MSE = 83712, $p = .067$. There was also a significant interaction of structure type and verb type in the items analysis though the effect was marginal in the participants analysis, $F1(1, 37) = 2.87$, MSE = 170952, $p = .099$, $F2(1, 23) = 5.41$, MSE = 143218, $p < .05$. A pair-wise comparison revealed that the structure type effect was present for transitive verb sentences but not for intransitive verb sentences: The regression path time in the transitive non-island condition was significantly faster than in the transitive island condition (386 ms vs. 520 ms), $t1(37) = 2.77$, $p < .01$, $t2(23) = 2.92$, $p < .01$, whereas there was no difference between the intransitive non-island and island conditions (528 ms vs. 527 ms; $ts < 1$). This interaction was rather unexpected, but it receives a straightforward explanation once we consider the fact that the regression path time reflects two different underlying processes: the first pass time and time spent regressing to earlier regions (for discussion, see Staub & Clifton, 2006). In the percent regressions measure, the two island conditions had high percentages of regressions (24.0% and 28.4%), and this is reflected in the large regression path time in these conditions. In the non-island conditions, the intransitive non-island condition had a higher percentage of regressions than the transitive non-island condition (14.4% vs. 11.7%), and moreover, the intransitive non-island condition had a
larger first pass time than the transitive non-island condition (379 ms vs. 319 ms.). These data indicate that in reality the regression path time merely reflects reading time increase due to large processing demand in the two island conditions, and the significant interaction should not be taken as directly relevant to the experimental hypotheses.

In the post-verb region, there was no significant effect of verb type or structure type, and no interaction of the two factors in any of the dependent measures (all $F$s < 2.6, $p > .11$). The analysis of percent regression revealed no significant effects ($\beta < .291$, $p > .32$).

**Discussion**

The results of this experiment can be summarized as follows. First, the first fixation duration for intransitive verbs in a structure that would allow a gap (non-island condition) was significantly longer than when the same verb appeared within an island configuration. This effect was not observed when the critical verb was transitive. The fact that there was a reading disruption for intransitive verbs but not for transitive verbs is consistent with the prediction of the hyper-active gap filling hypothesis: if the parser creates an object gap and integrates the filler into the object position before having access to verb transitivity information, reading disruption in the non-island intransitive condition should result from the mismatch between the predicted transitivity and actual transitivity of the verb. A similar pattern was observed in first pass times, although differences here were less robust.

Under the proposed hyper-active gap mechanism, first fixation duration is the most appropriate dependent measure for attesting the transitivity mismatch effect, because this mechanism effectively predicts the presence of a transitive verb. Therefore, the evaluation of this prediction should be possible as soon as the parser gains access to the lexical content of the verb,
and this effect can be reasonably expected in the first fixation duration measure, which has generally been shown to reflect the earliest phase of lexical access (Rayner, 1998; Staub & Rayner, 2007). Thus, the intransitive vs. transitive contrast in first fixation duration provides support for the hyper-active gap filling hypothesis, contrary to the conclusion reached in Staub (2007).

Regression path times at the verb region were much shorter for the transitive non-island condition than the other three conditions. As discussed in the results section, this is due to a combination of the higher percentage of regressions in the island conditions and the longer first pass time in the intransitive non-island condition. Although speculative, one possible interpretation of the larger percentage of regressions in the island conditions is that island conditions contain an extra word (i.e., the relative pronoun who) and also possibly incur greater complexity (Deane, 1991; Kluender & Kutas, 1993). Therefore, the fact that the non-island transitive was faster than its corresponding island condition whereas the non-island intransitive was not faster than its corresponding island condition may in fact reflect a relative slowdown in the non-island intransitive condition due to a transitivity mismatch effect. However, more studies with additional conditions would be needed to confirm this interpretation of the regression path results.

Although the first fixation data lend support to the hyper-active gap filling hypothesis, there are two reasons to treat these results with caution. First, although the pairwise comparisons within transitive and intransitive conditions demonstrated the pattern predicted by the hypothesis, the interaction between structure and verb type was not significant. In fact, the experiment design was constructed in such a way that the hyper-active gap hypothesis does not necessarily predict an interaction of structure and verb types: Given that the lexical item in the critical region differs
across verb types, the empirical support for the hyper-active gap filling account is to show a
structure type effect in the intransitive conditions but not in the transitive conditions. However,
the lack of significant interaction may invite skepticism about the robustness of our finding.
Second, anonymous reviewers pointed out that some of the verbs used in our intransitive
conditions seem to be able to occur in a transitive frame for some speakers (e.g., *departed the
city*, *vanished the coin*). While this concern does not straightforwardly account for the transitivity
mismatch effect we observed in Experiment 2, this may also raise questions about the
generalizability of our finding. In order to further test the robustness of the hyper-active gap
filling account, we conducted an additional eye-tracking experiment with more strictly
intransitive verbs.

EXPERIMENT 3

The goal of Experiment 3 was to replicate the finding in Experiment 2 with improved
materials. Specifically, we conducted a corpus search to identify strictly intransitive unaccusative
verbs that are used in a transitive frame at a near-zero frequency, and replaced some of the
materials from Experiment 2.

Method

Participants

We recruited 40 native speakers of American English from the University of Maryland
community. All had normal or corrected-to-normal vision, and were naïve to the purpose of the
experiment. They received course credit or were paid $10 for their participation, which lasted
around 40 minutes.
Materials

The basic material design in this experiment was identical to that of Experiment 2, and we replaced some of the items that had an intransitive verb that could occasionally appear in a transitive frame. In order to identify strictly intransitive unaccusative verbs, we used the Corpus of Contemporary American English (Davies, 2010; available online at http://corpus.byu.edu/coca/). The corpus consists of 425 million words in American English, which include spoken and written materials that were collected from a wide range of genres between 1990 and 2011. In our search, we first identified the total occurrence of the intransitive verbs of interest, and subsequently narrowed the search to calculate how frequently the verbs were used in a transitive frame. We selected 12 intransitive verbs that were almost never used in a transitive frame; as a result, the mean frequency of transitive use for the 12 intransitive verbs was 0.27% ($SE = .16$). While there were some exceptional cases in which these verbs were used in a transitive frame, it is important to note that none of the transitive frames included a filler-gap dependency. This suggests that the transitive frames are even more unlikely in an object relative clause context of the kind used in this experiment. The 24 sentence sets as well as the frequency of transitive usage of each intransitive verb are listed in Appendix C. These 24 sentence sets were counter-balanced across four lists, such that each participant saw only one version of each of the target sentences. The materials were counterbalanced in such a way that the critical intransitive verb was never repeated within a given participant. The target sentences were combined with 108 fillers that were also used in Experiment 2.

Procedure and Data Analysis

The experimental procedure and data analysis were identical to those of Experiment 2.
Results

Mean comprehension accuracy for the experimental items was 94.3%, and did not reliably differ across the four conditions. Table 2 presents the participant means on each measure for each region as well as the standard errors of the participant means.

Table 2
Experiment 3 Participant Mean Reading Times in Milliseconds (Standard Error) and Percent Regressions

<table>
<thead>
<tr>
<th>Measure</th>
<th>Pre-verb region</th>
<th>Verb region</th>
<th>Post-verb region</th>
</tr>
</thead>
<tbody>
<tr>
<td>First fixation duration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intransitive, non-island</td>
<td>241 (7)</td>
<td>303 (10)</td>
<td>292 (10)</td>
</tr>
<tr>
<td>intransitive, island</td>
<td>239 (7)</td>
<td>279 (8)</td>
<td>278 (8)</td>
</tr>
<tr>
<td>transitive, non-island</td>
<td>238 (8)</td>
<td>283 (7)</td>
<td>272 (9)</td>
</tr>
<tr>
<td>transitive, island</td>
<td>228 (7)</td>
<td>275 (7)</td>
<td>266 (10)</td>
</tr>
<tr>
<td>First-pass time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intransitive, non-island</td>
<td>401 (20)</td>
<td>388 (13)</td>
<td>371 (19)</td>
</tr>
<tr>
<td>intransitive, island</td>
<td>506 (30)</td>
<td>376 (20)</td>
<td>344 (13)</td>
</tr>
<tr>
<td>transitive, non-island</td>
<td>382 (24)</td>
<td>339 (12)</td>
<td>355 (18)</td>
</tr>
<tr>
<td>transitive, island</td>
<td>518 (37)</td>
<td>331 (12)</td>
<td>344 (17)</td>
</tr>
<tr>
<td>Regression path time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intransitive, non-island</td>
<td>560 (27)</td>
<td>516 (41)</td>
<td>621 (45)</td>
</tr>
<tr>
<td>intransitive, island</td>
<td>712 (44)</td>
<td>628 (42)</td>
<td>612 (62)</td>
</tr>
<tr>
<td>transitive, non-island</td>
<td>545 (33)</td>
<td>446 (30)</td>
<td>452 (28)</td>
</tr>
<tr>
<td>transitive, island</td>
<td>774 (50)</td>
<td>487 (35)</td>
<td>602 (63)</td>
</tr>
<tr>
<td>Percent regressions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intransitive, non-island</td>
<td>26.0 (3.2)</td>
<td>14.4 (2.8)</td>
<td>36.0 (3.8)</td>
</tr>
<tr>
<td>intransitive, island</td>
<td>28.5 (4.6)</td>
<td>31.5 (4.1)</td>
<td>24.5 (3.6)</td>
</tr>
<tr>
<td>transitive, non-island</td>
<td>24.3 (3.5)</td>
<td>14.7 (2.8)</td>
<td>15.5 (2.8)</td>
</tr>
<tr>
<td>transitive, island</td>
<td>29.0 (4.9)</td>
<td>22.7 (3.0)</td>
<td>27.5 (3.6)</td>
</tr>
</tbody>
</table>

The pre-verb region showed a very similar pattern of results to those in Experiment 2. The first fixation duration measure showed no significant differences ($F$s < 1.4, $p > .26$). The first pass time measure showed a main effect of the structure type manipulation, $F_I(1, 39) = 31.63$, MSE = 584032, $p < .0005$, $F_2(1, 23) = 46.00$, MSE = 374153, $p < .001$, due to slower reading times in the island conditions (391 ms vs. 512ms). There was no main effect of verb type or interaction of the two factors ($F$s < 1). Similarly, in the regression path duration measure, we found a main
effect of structure type, $F1(1, 39) = 28.79$, MSE = 1450283, $p < .001$, $F2(1, 23) = 61.45$, MSE = 918299, $p < .001$, due to slower reading times in the island conditions (552 ms vs. 743 ms), and there was no main effect of verb type or interaction of the two factors ($Fs < 1.8, p > .18$). As discussed in Experiment 2, this structure type effect likely reflects the presence of an extra relative pronoun in island conditions. We observed no significant effect in the percent regressions measure ($ps > .17$).

The data in the critical verb region also replicated the results in Experiment 2 closely. The first fixation duration showed a main effect of structure type in both the participants and the items analyses, $F1(1, 39) = 5.76$, MSE = 10642, $p < .05$, $F2(1, 23) = 5.49$, MSE = 5220, $p < .05$, as well as a marginal effect of verb type in the participants analysis, $F1(1, 39) = 4.03$, MSE = 5878, $p = .051$, $F2(1, 23) = 2.23$, MSE = 4062, $p > .1$, with no significant interaction between the factors ($Fs < 1.5, p > .23$). A planned comparison was conducted to test whether the structure type manipulation affected the first fixation duration for both verb types. This comparison revealed that the reading time in the non-island intransitive condition was significantly longer than in the island intransitive condition (303 ms vs. 279 ms; $t1(39) = 2.51$, $p < .05$, $t2(23) = 2.26$, $p < .05$), but the difference in reading time between the transitive non-island and island conditions did not reach significance (283 ms vs. 275 ms; $ts < 1.01, p > .32$).

The first pass measure revealed a main effect of verb type, $F1(1, 39) = 15.63$, MSE = 88023, $p < .001$, $F2(1, 23) = 14.75$, MSE = 58066, $p < .001$, due to slower reading times in the intransitive conditions than in the transitive conditions (382 ms vs. 335 ms), but there was no main effect of structure type or interaction of the two factors ($Fs < 1$). As discussed in Experiment 2, the main effect of verb type plausibly reflects the greater length and relative infrequency of intransitive verbs.
The percent regressions results demonstrated a significantly greater proportion of regressions in the island conditions than in the non-island conditions, $\beta = .843, SE = .184, p < .0005$, and there were no other significant effects ($ps > .15$). A model containing pairwise comparisons revealed that the structure type effect was present for both verb types: fewer regressive eye movements were made in the intransitive non-island condition than in the intransitive island condition (14.4% vs. 31.5%), $\beta = 1.103, SE = .253, p < .0005$, and the same pattern was observed in the corresponding transitive conditions, (14.7% vs. 22.7%), $\beta = .58, SE = .265, p < .05$. On the regression path time measure, we found a main effect of structure type, $F(1, 39) = 4.73, MSE = 233329, p < .05; F(1, 23) = 5.69, MSE = 197512, p < .05$, due to longer regression path times in the island conditions (558 ms vs. 481 ms), as well as a main effect of verb type ($F(1, 39) = 15.84, MSE = 443803, p < .001; F(1, 23) = 8.87, MSE = 229046, p < .01$) due to longer regression path times in the intransitive conditions (572 ms vs. 467 ms). This resembles the verb type effect found in the first-pass time measure. There was no significant interaction of structure type and verb type ($Fs < 1.1, ps > .3$). As discussed in Experiment 2, the structure type effects in regression-path and percent regression measures likely reflect the greater structural complexity of the island conditions.

In the post-verb region, the first fixation duration measure showed a main effect of verb type in the participants analysis, $F(1, 39) = 5.25, MSE = 9356, p < .05, F(1, 23) = 2.15, MSE = 3962, p > .15$, due to slower reading times in the intransitive conditions than in the transitive conditions. There was no main effect of structure type or interaction of the two factors ($Fs < 2.3, ps > .15$). This likely reflects the verb type effect observed in first-pass time and regression-path time in the critical verb region. The first pass time in the post-verb region revealed no significant differences ($Fs < .135, ps > .25$).
The percent regression measure in the post-verb region revealed no main effect of structure \((p > .9)\), but there was a main effect of verb type, \(\beta = -.502, SE = .164, p < .01\), as well as a significant interaction of the two factors, \(\beta = -1.324, SE = .328, p < .001\). A pairwise comparison revealed that the structure type effect was present in intransitive conditions \((\beta = -0.643, SE = .217, p < .005)\) due to the larger probability of regressive looks in the non-island condition than in the island condition (36.0% vs. 24.5%). The structure effect was also present for transitive conditions \((\beta = -0.681, SE = .246, p < .01)\), but the effect was in the opposite direction from the intransitive conditions, as the probability of regressive looks in the non-island condition was reliably smaller than in the island condition (15.5% vs. 27.5%).

In the analysis of the regression path time in the post-verb region, we found no main effect of verb type \((Fs < 2.9, ps > .1)\), but there was a main effect of structure type in the participants analysis, \(F1(1, 39) = 4.84, MSE = 203007, p < .05, F2(1, 23) = 2.43, MSE = 113318, p > .13\), due to slower reading times in the island conditions, as well as a significant interaction of the two factors in the items analysis, \(F1(1, 39) = 3.60, MSE = 253495, p = .07, F2(1, 23) = 5.76, MSE = 195549, p < .05\). Planned comparisons revealed that reading times in the island transitive condition were significantly longer than in the non-island transitive condition (602 ms vs. 452 ms; \(t1(39) = 2.78, p < .01, t2(23) = 2.49, p < .05\), but there was no significant difference between the intransitive conditions \((t < .43, p > .67)\). This pattern resembles the regression path time in the critical region in Experiment 2; thus, we suggest that the interaction here reflects the larger increase of regression path-time in the intransitive non-island condition than in the transitive non-island condition.

Discussion
The results of Experiment 3 can be summarized as follows. When the critical verb was an intransitive verb, the first fixation on the verb region was significantly longer in the non-island condition than in the island condition. The first-pass time was reliably longer when the verb was intransitive, and the regression-path time was longer in intransitive conditions as well as in island conditions, although there was no interaction. These results show a striking similarity to results from the critical region of Experiment 2, and this leads to the following two implications. First, the fact that previous results were replicated with more strictly intransitive unaccusative verbs suggests that the findings in Experiment 2 were not due to the weaker intransitivity of a subset of the critical verbs. Second, the replication of the slow-down in the first fixation duration measure in the non-island intransitive condition across two experiments with different participants and materials strengthens the evidence that intransitive verbs elicit processing difficulties in filler-gap configurations.

GENERAL DISCUSSION

Experiments 1, 2 and 3 all demonstrated evidence for reading disruption at an intransitive verb when the verb was in a potential gap-filling environment. The reading disruption that can be attributed to a transitivity mismatch effect was observed at the same region as the region that revealed a plausibility mismatch effect (Experiment 1), and this reading disruption for an intransitive verb was observed as early as the first fixation on the intransitive verb (Experiments 2 and 3). These results lend support to the hyper-active gap filling hypothesis, which claims that in English filler-gap dependency processing object gap creation can be initiated based on pre-verbal information and can thereby lead the parser to expect a transitive verb. This is indeed what has been proposed for the filler-gap dependency processing mechanism in head-final
languages (Aoshima et al., 2004; Nakano et al., 2002), but the current work suggests that the same mechanism extends to the processing of filler-gap dependency in verb-medial languages like English as well.

The view that object gap creation is triggered by pre-verbal information contrasts with a standard view of object gap creation in English filler-gap dependency processing that object gap creation is driven by properties of the verb (e.g., McElree et al., 2003; Pickering & Barry, 1991). In fact, the hyper-active gap filling mechanism suggests an alternative interpretation of existing data on active object gap creation. For example, a plausibility mismatch effect found in Traxler and Pickering (1996) has been taken to suggest that filler-retrieval occurs after accessing the transitivity information on the verb, and that subsequent structural integration of the filler leads to the implausible verb-object composition, which in turn results in reading time slowdown. However, under the hyper-active gap filling account, prior to the verb the reader analyzes the filler as a direct object of the upcoming verb, and given the combination of the subject NP and the hypothesized object NP, the reader may already expect a certain class of transitive verbs that would be semantically compatible with the filler noun phrase. In other words, plausibility mismatch effects could be reconsidered as a reflection of a violation of lexical expectations, which result from predictive structural analysis. Future studies are needed to examine to what extent this reinterpretation of plausibility mismatch effects are feasible.

The present study has focused on filler-gap dependency processing, but the current conclusion is consistent with a broader class of models of sentence processing that propose that the parser utilizes a variety of sources of linguistic and contextual information to predictively build structural representations (Altmann & Kamide, 1999; Gibson, 1998; Hale, 2003; Kimball, 1975; Levy, 2008). On the other hand, the present study does not reveal what kind of pre-verbal
information is critical for triggering object gap creation in advance of the verb. One possible source that was already discussed in the Introduction is the grammatical knowledge of phrase structure rules, which suggest that the upcoming VP representation can contain an object NP slot. However, it is equally feasible that the parser could use non-grammatical information in predictively positing the object gap, such as differences in relative conditional probabilities derived from the lexical and contextual information from the combination of the filler noun phrase and the subject. For example, even when a clause appears to resemble a gap structure like a relative clause, with a certain combination an adjunct gap may seem much more plausible than an object gap analysis (e.g., the day that… can continue as involving an adjunct gap as in the day that I was born, or an object gap as in the day that I have been looking forward to). Further studies are needed to investigate what kind of information contributes to such predictive object gap creation processes.

One may argue that there is an alternative interpretation of the results that still assumes that verb information plays a critical enabling role in filler-gap dependency formation in English. For example, it is possible that filler retrieval processes are automatically activated as soon as the parser accesses the category information of the verb without accessing the transitivity information of the verb. Such a procedure could be motivated by an incremental interpretation strategy that attempts to combine any N-N-V sequence into a proposition (for discussion, see e.g. Goodluck, Finney & Sedivy, 1991; Goodluck, Finney, & Ling, 1995). Under this alternative account, the transitivity mismatch effect arises because the filler that was ‘blindly’ retrieved based on the verb category information mismatches the subcategorization property of the verb that is accessed later (see van Gompel & Liversedge, 2003, for a similar proposal for a gender mismatch effect in pronoun processing).
Although our study does not completely rule out a non-predictive account, these data put important constraints on the form that such an account must take. Critically, a non-predictive account must assume that access to the contents of lexical information is ordered, such that category information is accessed earlier than the subcategorization property of the verb. However, as yet there is little evidence to support such ordered access to category vs. other contents of a verb (Farmer, Christiansen, & Monaghan, 2006 is one rare case, but see Staub, Grant, Clifton, & Rayner, 2009 for a counterargument), whereas there is an abundance of psycholinguistic and neurolinguistic research demonstrating extremely fast access to various aspects of the content of lexical items (e.g., Federmeier, Segal, Lombozo, & Kutas, 2000; Dambacher, Kliegl, Hofmann, & Jacobs, 2006; Hauk, Davis, Ford, Pulvermuller, & Marslen-Wilson 2006; Staub & Rayner, 2007; Tanenhaus, 2007; Almeida & Poeppel, submitted). On the other hand, there has been a recent surge of empirical work demonstrating that structure building processes can proceed predictively based on various types of top-down linguistic and contextual information, as discussed above (e.g., Konieczny, 2000; Kamide et al., 2003; Delong, Urbach, & Kutas, 2005; van Berkum, Brown, Zwisterlood, Kooijman, & Hagoort, 2005; Lau, Stroud, Plesch, & Phillips, 2006; Yoshida, 2006; Staub & Clifton, 2006; Yoshida, Dickey, & Sturt, in press), and the current work demonstrating extremely early object gap creation processes can be seen as another instance of such predictive structure building processes. While these other findings lead us to favor a predictive explanation, further work is needed to more firmly establish that the hyper-active gap filling hypothesis is a better account for the pattern of results observed across a variety of paradigms than this alternative category-driven approach.

The current finding may also seem to contradict findings by Boland, Tanenhaus, Garnsey, and Carlson (1995) and Pickering and Traxler (2001). These authors tested the processing of
filler-gap dependencies in sentences that contain verbs like *persuade* or *remind* that can have both an NP slot and a clausal complement slot in their argument structure, and found no evidence for reading disruption when the filler was semantically incompatible with the direct object NP slot, but compatible with the complement slot. According to the hyper-active gap filling account, encountering a *persuade*-type verb should not result in a transitivity mismatch effect since *persuade* does make available an object position, but one may wonder whether it should result in a plausibility mismatch effect when the filler is a semantically incompatible object, since an object-gap structure is hypothesized to be predictively constructed before the verb.

We can see two ways of reconciling these findings with the results presented here. First, the plausibility mismatch slowdown observed for simple transitive verbs may largely reflect the cost of reanalyzing the predicted structure to one that is compatible with the new input, which may vary depending on the argument structure of the verb. Revision may be costly in the cases where the verb is intransitive or mono-transitive and does not provide sufficient information for the parser to anticipate an alternative structural position for the filler, whereas in the *persuade/remind* cases, the revision may be less costly because the argument structure of the verb clearly indicates the presence of an upcoming clause in which the filler can be integrated. Second, the predicted filler-gap structure may be more abstract than we have indicated so far. Rather than specifically predicting an object gap when the filler and relative clause subject are encountered, the parser may simply predict an argument gap position somewhere inside the complement domain of an upcoming VP representation, such that a gap in either the NP slot or in the clausal complement slot of *persuade*-class ditransitive verbs would be consistent with the prediction. The current results are compatible with either account.

In the sentences used here, the intransitive structures are eventually resolved by the
appearance of a preposition, which provides another structural position for the filler. Although this could be recognized as a possible reanalysis even at the verb position, this position is not specifically licensed by the input until the preposition is actually encountered (in contrast with the persuade/remind cases, in which the indirect object position is available at the verb). One interesting question for further research is whether the difficulty of recovering from the simple transitive analysis is modulated by the frequency with which a particular intransitive verb co-occurs with a prepositional phrase that could host the filler. For example, many intransitive verbs can be combined with a prepositional complement to form a phrasal verb that takes the object of the preposition as an argument, e.g. listen to the music. If a particular intransitive verb occurs very frequently in a phrasal verb configuration, reanalysis to this structure in a filler-gap configuration might be less costly, even prior to the presentation of the preposition.\(^2\)

Finally, the conclusion that the same filler-gap dependency completion procedure is used across head-initial and head-final languages suggests that the parser’s structure building procedures, at least for filler-gap dependency completion, may not be qualitatively different across languages. However, this still leaves us with many open questions. For example, so far the evidence for hyper-active gap filling is reported for Japanese (Aoshima et al., 2004; Nakano et al., 2002) and English (the present paper), but this line of work obviously needs to be extended to other languages to test the robustness of the hypothesis. Moreover, predictive dependency formation processes are observed in domains other than filler-gap dependency processing (e.g., resolution of backward anaphora; Aoshima, Yoshida, & Phillips, 2009; van Gompel & Liversedge, 2003; Kazanina, Lau, Lieberman, Yoshida, & Phillips, 2007), but it is not known whether there is cross-linguistic variation in other predictive structure building processes, and if such cross-linguistic variation exists, what is its cause. These questions about the correspondence

\(^2\) We thank an anonymous reviewer for this suggestion.
between linguistic properties and psycholinguistic processes must be addressed by further cross-linguistic investigations of parsing, but we believe this is a fruitful line of investigation that could shed further light on fundamental questions about the relationship between linguistic representations and real-time processes for constructing those representations.

Conclusion

The present study tested the hypothesis that predictive structure building processes underlie filler-gap dependency completion in English. In the presence of a filler-gap dependency, intransitive verbs consistently led to reading disruption, and this pattern was replicated in self-paced reading measures as well as in eye movement measures. These findings show that English speakers do not wait to check that the verb makes an object position available, and are consistent with the hypothesis that the parser postulates an object gap at least as soon as it encounters a filler phrase and a subject NP. We suggest that the parser uses pre-verbal information to predictively create rich syntactic representations regardless of word order differences across languages.
References


integration difficulty. *Language and Cognitive Processes, 15*(2), 159-.


Cognitive Processes, 8, 573-633.


Ross, J. R. (1967). *Constraints on variables in syntax.* Unpublished doctoral dissertation, MIT,


Appendix A: Materials used in Experiment 1

1a/c. The house that the woman wrote/grinned slyly about was inspected by the board.
1b/d. The house that the woman who wrote/grinned slyly appreciated was inspected by the board.
Example comprehension question: Was it the board that inspected the house? N

2a/c. The city that the author wrote/chatted regularly about was named for an explorer.
2b/d. The city that the author who wrote/chatted regularly saw was named for an explorer.

3a/c. The injury that the victim called/sighed repeatedly about was treated improperly by doctors.
3b/d. The injury that the victim who called/sighed repeatedly saw was treated improperly by doctors.

4a/c. The jewelry that the sheriff questioned/joked sharply about was recovered after the robbery.
4b/d. The jewelry that the sheriff who questioned/joked sharply watched was recovered after the robbery.

5a/c. The billboard that the girl threw/snorted angrily at was located near the beach.
5b/d. The billboard that the girl who threw/snorted angrily saw was located near the beach.

6a/c. The wall that the boy threw/grinned mischievously at was painted fire engine red.
6b/d. The wall that the boy who threw/grinned mischievously noticed was painted fire engine red.

7a/c. The artist that the architect designed/competed passionately with was nationally well
7b/d. The artist that the architect who designed/competed passionately admired was nationally well known.

8a/c. The design that the professor lectured/sighed resignedly about was discussed in the seminar.
8b/d. The design that the professor who lectured/sighed resignedly saw was discussed in the seminar.

9a/c. The equipment that the employee phoned/frowned disapprovingly about was mentioned by the President.
9b/d. The equipment that the employee who phoned/frowned disapprovingly saw was mentioned by the President.

10a/c. The drugs that the principal threatened/frowned sternly about were discussed during the meeting.
10b/d. The drugs that the principal who threatened/frowned sternly confiscated were discussed during the meeting.

11a/c. The bridge that the tourist read/napped peacefully under was photographed by the group.
11b/d. The bridge that the tourist who read/napped peacefully missed was photographed by the group.

12a/c. The clock that the collector read/smiled fondly about was found while shopping for
antiques.

12b/d. The clock that the collector who read/smiled fondly discovered was found while shopping for antiques.

13a/c. The bacteria that the biologist instructed/struggled intensely about turned out to be highly productive.

13b/d. The bacteria that the biologist who instructed/struggled intensely studied turned out to be highly productive.

14a/c. The match that the athlete trained/struggled endlessly for was ended by the authorities.

14b/d. The match that the athlete who trained/struggled endlessly played was ended by the authorities.

15a/c. The manager that the custodian cleaned/smiled obediently for was ruined by a financial crisis.

15b/d. The manager that the custodian who cleaned/smiled obediently liked was ruined by a financial crisis.

16a/c. The recording that the instructor taught/beamed enthusiastically about was heard throughout the auditorium.

16b/d. The recording that the instructor who taught/beamed enthusiastically presented was heard throughout the auditorium.
17a/c. The princess that the clown sang/danced cheerfully with was adored by the media.
17b/d. The princess that the clown who sang/danced cheerfully admired was adored by the media.

18a/c. The accident that the governor scolded/yelled angrily about was seen on the news.
18b/d. The accident that the governor who scolded/yelled angrily witnessed was shown on the news.

19a/c. The theories that the teacher scolded/corresponded vigorously about were taught throughout the term.
19b/d. The theories that the teacher who scolded/corresponded vigorously criticized were taught throughout the term.

20a/c. The jobs that the instructor taught/sulked grudgingly about were all in food service.
20b/d. The jobs that the instructor who taught/sulked grudgingly knew were all in food service.

21a/c. The country that the man killed/prayed endlessly for was destroyed by Mongol military.
21b/d. The country that the man who killed/prayed endlessly hated was destroyed by Mongol military.

22a/c. The party that the designer dressed/laughed obnoxiously for was thought to be very important.
22b/d. The party that the designer who dressed/laughed obnoxiously enjoyed was thought to be very important.
23a/c. The poster that the manager paid/appeared unexpectedly for was sent to the office.
23b/d. The poster that the manager that paid/appeared unexpectedly saw was sent to the office.

24a/c. The magazine that the children fought/giggled persistently about could not be found anywhere.
24b/d. The magazine that the children who fought/giggled persistently read could not be found anywhere.

25a/c. The sculpture that the critics lectured/quarreled seriously about was seen in the park.
25b/d. The sculpture that the critics who lectured/quarreled seriously denounced was seen in the park.

26a/c. The money that the criminal kidnapped/chuckled cruelly for was missing during the investigation.
26b/d. The money that the criminal who kidnapped/chuckled cruelly despised was missing during the investigation.

27a/c. The game that the journalist bet/shrieked compulsively on was discussed at the pub.
27b/d. The game that the journalist who bet/shrieked compulsively watched was discussed at the pub.

28a/c. The client that the cook prepared/flirted skillfully for was disliked by the waiters.
28b/d. The client that the cook who prepared/flirted skillfully favored was disliked by the waiters.

Appendix B: Materials used in Experiment 2

1a/c. The studio that the students designed/remained peacefully in while the professors conferred was small and ugly.

1b/d. The studio that the students who designed/remained peacefully rested in while the professors conferred was small and ugly.

Example question: Was the studio ugly?

2a/c. The warehouse that the trucker phoned/departed nervously from last week was very old.

2b/d. The warehouse that the trucker who phoned/departed nervously visited last week was very old.

3a/c. The opponent that the veteran tennis player played/prevailed skillfully with/over during the game was very gracious.

3b/d. The opponent that the veteran tennis player who played/prevailed skillfully beat during the game was very gracious.

4a/c. The newsroom that the reporter called/emerged moodily from was full of problems.

4b/d. The newsroom that the reporter who called/emerged moodily oversaw was full of problems.

5a/c. The group that the speaker lectured/appeared reluctantly with at the conference was very dogmatic.
5b/d. The group that the speaker who lectured/appeared reluctantly spoke to at the conference was very dogmatic.

6a/c. The impaired plane that the pilot landed/arose wearily behind/from was a mass of twisted metal.
6b/d. The impaired plane that the pilot who landed/arose wearily had flown/crashed did not meet safety standards.

7a/c. The knight that the warrior killed/died nobly for went on to save the princess.
7b/d. The knight that the warrior who killed/died nobly admired/succumbed to went on to save the princess.

8a/c. The quarrel that the girl heard/persisted reluctantly about/in was incomprehensible and pointless.
8b/d. The quarrel that the girl who heard/persisted reluctantly resolved/won was incomprehensible and pointless.

9a/c. The building that the thief climbed/disappeared quickly over/behind during the chase was a nondescript warehouse.
9b/d. The building that the thief who climbed/disappeared quickly entered during the chase was a nondescript warehouse.

10a/c. The manager that the customer fought/erupted angrily with/at was near the register.
10b/d. The manager that the customer who fought/erupted angrily shouted at was near the register.

11a/c. The rave that the teenage boy planned/remained obligingly for/in at the underground club was hopping.

11b/d. The rave that the teenage boy who planned/remained obligingly watched at the underground club was hopping.

12 a/c. The old bed that the cat scratched/arose lazily at/in was really worn out but comfortable.

12b/d. The old bed that the cat who scratched/arose lazily sharpened her claws on was really worn out but comfortable.

13a/c. The table that the chemical burned/disappeared quietly on/from was very old and dirty.

13b/d. The table that the chemical that burned/disappeared quietly had left a stain on was very old and dirty.

14a/c. The people that the conqueror killed/prevailed fiercely for/over were [grateful to have their city back]/[upset to lose their city].

14b/d. The people that the conqueror who killed/prevailed fiercely defended were grateful to have their city back.

15a/c. The athlete that the coach taught/appeared proudly about/with before the game was nominated for a big award.
15b/d. The athlete that the coach who taught/appeared proudly trained before the game was nominated for a big award.

16a/c. The accident that the lady escaped/died mysteriously from/in last night was thoroughly investigated.
16b/d. The accident that the lady who escaped/died mysteriously had photographed last night was thoroughly investigated.

17a/c. The research that the scientist prepared/persisted determinedly for/in during his whole career was finally completed.
17b/d. The research that the scientist who prepared/persisted determinedly pursued during his whole career was finally completed.

18a/c. The gate that the limousine passed/emerged slowly through/from as it left the house was closed shortly thereafter.
18b/d. The gate that the limousine which passed/emerged slowly crossed as it left the house was closed shortly thereafter.

19a/c. The frat boy that the woman fought/erupted aggressively with/at in the department store was very obnoxious.
19b/d. The frat boy that the woman who fought/erupted aggressively disciplined in the department store was very obnoxious.
20a/c. The assistant that the magician trained/vanished skillfully with was good at her job.
20b/d. The assistant that the magician who trained/vanished skillfully complimented after the show was good at her job.

21a/c. The fugitive that the mobster hid/appeared abruptly from/with was rumored to be very dangerous.
21b/d. The fugitive that the mobster who hid/appeared abruptly feared/shot was rumored to be very dangerous.

22a/c. The airport that the ambassador left/departed rapidly for/from during the unrest was closed to most traffic.
22b/d. The airport that the ambassador who left/departed rapidly had visited during the unrest was closed to most traffic.

23a/c. The computer lab that the IT technician phoned/arrived tardily from/at was full of college kids studying for finals.
23b/d. The computer lab that the IT technician who phoned/arrived tardily despised was full of college kids studying for finals.

24a/c. The costume party that the student planned/arrived eagerly for at the fraternity house was pretty lame.
24b/d. The costume party that the student who planned/arrived eagerly attended/threw at the fraternity house was pretty lame.
Appendix C: Materials used in Experiment 3

1a/c. The concert that the teenager left/fainted suddenly from/at was later considered a turning point for the rock band.

1b/d. The concert that the teenager who left/fainted suddenly blogged about was later considered a turning point for the rock band.

Example question: Was it a jazz band that performed at the concert?

2a/c. The stage that the dancer exited/fainted abruptly from/on had long-standing ventilation problems.

2b/d. The stage that the dancer who exited/fainted abruptly fell off of had long-standing ventilation problems.

3a/c. The girls that the skilled cheerleader trained/tumbled tirelessly for were the best in the conference.

3b/d. The girls that the skilled cheerleader who trained/tumbled tirelessly encouraged were the best in the conference.

4a/c. The ladder that the child climbed/tumbled clumsily on was very old and unstable.

4b/d. The ladder that the child who climbed/tumbled clumsily broke was very old and unstable.

5a/c. The peacekeeping operation that the volunteers served/perished selflessly in was critical in preventing a civil war.

5b/d. The peacekeeping operation that the volunteers who served/perished selflessly stabilized
was critical in preventing a civil war.

6a/c. The spy that the bodyguard fought/perished bravely with failed to complete his mission.
6b/d. The spy that the bodyguard who fought/perished bravely injured failed to complete his mission.

7a/c. The military expedition that the nation planned/prospered swiftly for/from was the largest in recent history.
7b/d. The military expedition that the nation that planned/prospered swiftly launched was the largest in recent history.

8a/c. The marketing strategy that the company developed/prospered rapidly with was highly effective in transforming the industry.
8b/d. The marketing strategy that the company that developed/prospered rapidly pioneered was highly effective in transforming the industry.

9a/c. The budget proposal that the politician debated/relented momentarily about/on was an increase in funding for the arts.
9b/d. The budget proposal that the politician who debated/relented momentarily detested was an increase in funding for the arts.

10a/c. The exam that the professor prepared/relented indifferently for/on was too difficult for all of the students.
10b/d. The exam that the professor who prepared/relented indifferently underestimated was too difficult for all of the students.

11a/c. The opponent that the veteran tennis player played/prevailed skillfully with/over was very gracious after the game.

11b/d. The opponent that the veteran tennis player who played/prevailed skillfully beat was very gracious after the game.

12a/c. The people that the conqueror killed/prevailed fiercely for/over were discouraged by the lengthy war.

12b/d. The people that the conqueror who killed/prevailed fiercely defended were discouraged by the lengthy war.

13a/c. The group at the conference that the speaker lectured/appeared unwillingly with was very dogmatic.

13b/d. The group at the conference that the speaker who lectured/appeared unwillingly spoke to was very dogmatic.

14a/c. The fugitive that the mobster hid/appeared furtively from/with was rumored to be very dangerous.

14b/d. The fugitive that the mobster who hid/appeared furtively dreaded was rumored to be very dangerous.
15a/c. The impaired plane that the pilot watched/arose wearily from was no longer able to fly.
15b/d. The impaired plane that the pilot who watched/arose wearily oversaw was no longer able to fly.

16a/c. The old bed that the cat scratched/arose lazily at/from was really worn out but comfortable.
16b/d. The old bed that the cat who scratched/arose lazily rested on was really worn out but comfortable.

17a/c. The research that the scientist prepared/persisted determinedly for/in during his whole career was finally completed.
17b/d. The research that the scientist who prepared/persisted determinedly pursued during his whole career was finally completed.

18a/c. The quarrel that the girl overheard/persisted reluctantly about/in was incomprehensible and pointless.
18b/d. The quarrel that the girl who overheard/persisted reluctantly resolved was incomprehensible and pointless.

19a/c. The building that the thief climbed/disappeared quickly over/behind during the chase was a nondescript warehouse.
19b/d. The building that the thief who climbed/disappeared quickly entered during the chase was a nondescript warehouse.
20a/c. The table that the chemical burned/disappeared quietly on/from was very old and dirty.
20b/d. The table that the chemical that burned/disappeared quietly stained was very old and dirty.

21a/c. The manager that the customer fought/erupted aggressively with/at was standing near the register.
21b/d. The manager that the customer who fought/erupted aggressively berated was standing near the register.

22a/c. The intern that the supervisor lectured/erupted angrily at was very obnoxious.
22b/d. The intern that the supervisor who lectured/erupted angrily chastised was very obnoxious.

23a/c. The computer lab that the IT technician phoned/arrived tardily from/at was full of college kids studying for finals.
23b/d. The computer lab that the IT technician who phoned/arrived tardily despised was full of college kids studying for finals.

24a/c. The costume party that the student planned/arrived ahead for at the fraternity house was sparsely attended.
24b/d. The costume party that the student who planned/arrived ahead enjoyed at the fraternity house was sparsely attended.