RESEARCH

Minding the gap?: Mechanisms underlying resumption in English

Dustin A. Chacón

University of Minnesota, Minneapolis, MN, US

dustin@umn.edu

In processing filler-gap dependencies, comprehenders quickly postulate gaps in syntactically licensed positions, but not in syntactic islands. This suggests that comprehenders can accurately use syntactic constraints to guide processing. However, resumptive pronouns appear to challenge this generalization. Resumption is ungrammatical in English. Nevertheless, they appear to immediately allow resolution of a filler dependency in syntactic islands (Hofmeister & Norcliffe 2013). I resolve this tension by arguing that pronouns are analyzed as resumptive when typical filler-gap dependency processing fails. I argue that processing a filler-gap dependency requires anticipatorily building a gapped structure. However, as further linguistic material is processed, this representation degrades in memory. Resumption facilitates processing by triggering a reference dependency, which allows the comprehender to recover a coherent interpretation of the sentence. This predicts that the accessibility of filler NP as a referent for a pronoun, length, and processing difficulty all contribute to the acceptability of resumption. I present the results of four acceptability judgment studies that support this claim. I also introduce a novel experimental paradigm, in which participants' working memory capacity is taxed while processing a sentence. This increase in processing strain decreases sensitivity to ungrammatical filler dependencies. I argue that this partially explains the acceptability of resumption in syntactic island contexts, which are likely resource-intensive.

1 Introduction

To interpret a sentence like (1), the filler (*the case*) must be associated with the later gap (represented as an underscore). The gap allows the comprehender to recover the grammatical function of the filler, identify its thematic role in the event, and integrate it into the conceptual representation of the sentence.

(1)   Dale solved **the case** that Chet had investigated __.

Many studies have demonstrated that the gap site is constructed shortly after the filler is identified, a process called active dependency formation (Fodor 1978; Crain & Fodor 1985; Stowe 1986; Traxler & Pickering 1996; Kaan et al. 2000; Aoshima et al. 2004; Phillips et al. 2005; Omaki et al. 2015; Chacón et al. 2016, see Pablos 2008 for review). For instance, Stowe (1986) found that reading times increased at the object *us* in (2) compared to controls. This finding suggests that the comprehender initially interpreted the filler *who* as the direct object of the verb *bring*. However, upon detecting the actual direct object *us*, she must reject her previously constructed analysis to accommodate the bottom-up input, and project another, later gap.

(2)   My brother wanted to know **who** Ruth would bring **us** home to ___ at Christmas.

Importantly, active gap formation appears to be suppressed in syntactic islands (Stowe 1986; Traxler & Pickering 1996; Phillips 2006; Wagers & Phillips 2009). For instance, Stowe (1986) found that there is no increased reading times at *Greg's older brother* in (3a), suggesting that the comprehender had not initially committed to a gap in this position. Relatedly, a filler-gap dependency resolving into this position would be a complex NP island violation, shown in (3b). I take these findings to suggest that there are grammatical constraints against filler-gap dependencies crossing into syntactic island configurations, and the comprehender uses these constraints to guide active dependency formation (Phillips 2006; Wagers & Phillips 2009; Yoshida et al. 2014, see Hofmeister & Sag 2010 and Phillips 2013 for discussion).

(3)   a.   The teacher asked **what** [$_{\text{CNPC-Island}}$ the silly story about **Greg's older brother**] was supposed to mean ___.
        b.   The teacher asked **what** [$_{\text{CNPC-Island}}$ the silly story about ___] was supposed to mean something.

Occasionally, fillers associate with a resumptive pronoun instead of a gap. For instance, in (4a), *a zoo* heads a relativization dependency that does not associate with a gap. The filler *a zoo* is understood as the object of the preposition *to*, i.e., the pronoun *it* functions as the resolution site for the filler dependency. Importantly, resumptive dependencies frequently appear in syntactic islands. For instance, replacing the resumptive pronoun with a gap is unacceptable in (4b) (Example sentences taken from the podcast *Cool Games Inc.*, Episode 51: 'The Prestige Goose').

(4)   a.   St. Louis has **a zoo** that [$_{\text{Adjunct-Island}}$ the first time I went to **it**], there's like, an otter exhibit.
        b.   *St. Louis has **a zoo** that [$_{\text{Adjunct-Island}}$ the first time I went to ___], there's like, an otter exhibit.

Traditionally, resumptive pronouns are described as syntactic mechanisms for repairing sland violations (Ross 1967). However, subsequent work suggests that, in English, resumptive pronouns are perceived to be ungrammatical. Resumptive pronouns are used because they can facilitate processing, particularly in sentence production (Kroch 1981; Chao & Sells 1983; Creswell 2002; Ferreira & Swets 2005; Heestand et al. 2011; Asudeh 2012; Keffala 2013; Beltrama & Xiang 2016; see also Ackerman et al. 2018). Resumptive pronouns do not display the same grammatical characteristics as grammatical filler-gap dependencies (Chao & Sells 1983), and they are assigned low ratings in off-line acceptability judgment tasks (Alexopoulou & Keller 2007; Heestand et al. 2011). This evidence has been used to argue against the traditional analysis of resumption as a repair mechanism for an island violation.

In a self-paced reading study, Hofmeister & Norcliffe (2013) found that reading times were reduced shortly after a resumptive pronoun was encountered, but only if the filler-resumptive dependency is sufficiently long. Similarly, Alexopoulou & Keller (2007) showed resumptive dependencies were rated on par with filler-gap dependencies for particularly long sentences. Thus, although resumptive pronouns may be ungrammatical, the presence of a resumptive pronoun may immediately facilitate processing, particularly in configurations that may be already difficult to process.

The finding that resumptive dependencies appear to be constructed on-line challenges the generalization that filler dependency processing is guided by grammatical constraints. First, it immediately raises the question of why comprehenders do not search

for resumptive pronouns in syntactic islands generally. For instance, as described above, active gap formation appears to be suppressed in syntactic islands. However, if comprehenders may actively search for resumptive pronouns in syntactic islands, then it is unclear why active dependency formation processes should be island-sensitive, since resumptive pronouns commonly appear in these configurations.

This is further underscored by "McCloskey's Generalization", which observes that resumptive pronouns are morphologically identically to anaphoric pronouns (McCloskey 2006). In processing a sentence like (5), the comprehender expects a grammatically licensed gap for the filler *the case* to bind. Upon encountering the pronoun *it*, she must determine whether this pronoun *it* is anaphoric, i.e., refers to another entity in the discourse, or resumptive, i.e., serves as the resolution site for the filler dependency. If resumption in English is ungrammatical, and if the comprehender applies grammatical constraints to avoid ungrammatical interpretations, then *it* should always be interpreted as anaphoric. However, the findings by Hofmeister & Norcliffe (2013) suggest that this prediction is incorrect for longer dependencies.

(5)   Dale solved **the case** [$_{\text{RelCl-Island}}$ that **it** …]

Previous approaches have suggested that resumption facilitates the processing of filler dependencies by helping the comprehender identify the intended resolution site in complex structures (Keenan & Comrie 1977; Erteschik-Shir 1992; Ariel 1999; Hawkins 1999; 2004). For instance, Hawkins (1999; 2004) proposed that resumptive pronouns allow the argument structure of a predicate to be processed without needing to identify a gap, and thus without having to retrieve the filler phrase. By contrast, Ariel (1999) proposed that the comprehender searches for gaps while processing filler dependencies with highly accessible antecedents, but searches for pronouns while processing filler dependencies with less accessible antecedents. However, it is unclear how to reconcile this intuition with the generalizations about the processing of filler-gap dependencies and pronouns. Pronouns immediately trigger a search in memory for an antecedent (Arnold et al. 2000; Badecker & Staub 2002; Runner et al. 2006). Similarly, gaps require retrieval of their filler phrase (McElree et al. 2003). Resumptive pronouns therefore do not mitigate the need to retrieve an antecedent. Arguably, identifying the antecedent of a gap is more straightforward than identifying the antecedent of a pronoun. A gap's antecedent is typically the immediately preceding filler, whereas pronouns may be interpreted as anaphoric or resumptive. Thus, resumptive pronouns are at best an ambiguous cue for signaling the intended antecedent of a dependency. In contexts where the filler NP is not highly accessible, it is unclear why a pronoun would be a better marker of intended resolution site, for this reason. Put differently, a resumptive pronoun should only facilitate processing a filler dependency if the comprehender already knew that the pronoun was intended to be resumptive. Finally, as stated above, if filler dependency processing accurately obeys grammatical constraints, ungrammatical resumptive dependencies shouldn't be entertained on-line in the first place.

To resolve this tension, Chacón (2015) argued that resumption in English follows from an interaction between pronoun interpretation and active gap formation. On his proposal, active dependency formation is an active search for a structure that licenses a thematic role for the filler. This search is grammatically constrained, i.e., gaps are not projected into island contexts. Upon encountering a pronoun, however, the comprehender launches a retrieval for an antecedent. If the filler phrase is selected as the antecedent of the pronoun, then the thematic role of the filler is identified with that of the pronoun. This allows the filler to be integrated into the meaning of the sentence. Thus, the comprehender may

resolve the dependency without actively searching for a resolution site in an island context. Instead, resumption is "pronoun-driven", i.e., depends on the retrieval mechanics of pronoun processing, which is not sensitive to island constraints (cf. Erteschik-Shir 1992).

In a sentence completion task, Chacón (2015) asked participants to complete sentences that contained an unresolved filler-gap dependency and a pronoun that could refer to the filler NP. He found that coreference between the filler NP and the pronoun resulted in fewer completions that contained a gap. For instance, participants were less likely to complete the sentence in (6) with a gap (e.g., *offend __*), and instead favored completions that had no gaps (*offend somebody*).

(6)   The bridesmaid speculated **which groomsman** [_Subject-Island_ the speech that **he** prepared] would …

He interpreted this as implying that the search for a gap was abandoned if the pronoun referred to the filler. Importantly, this effect was modulated by the availability of other potential antecedents for the pronoun. He found that the addition of a masculine referent, such as replacing the *the bridesmaid* with *the groom* in (6), resulted in higher completions containing gaps. This suggests that resumption was partially gated by availability of coreference.

The assumption that comprehenders are capable of abandoning an expected gap conflicts with findings on the processing of multiple-gap constructions, such Across-The-Board (ATB) configurations. Grammatical principles require that if a filler binds a gap in one conjunct, it must also bind a gap in all subsequent conjuncts (Ross 1967), as demonstrated in (7).

(7)   a.   The man that Dale arrested __ and Harry interrogated __.
        b.   *The man that Dale arrested Ben and Harry interrogated __.
        c.   *The man that Dale arrested __ and Harry interrogated Ben.

Wagers & Phillips (2009) and Parker (2017) investigated the processing of ATB configurations, and found that active dependency formation processes persisted even after the first gap was identified. For instance, in sentences like (8), Wagers & Phillips (2009) found increased reading times at the verb *sipping*. This suggests that a gap had been postulated in the second conjunct, even after interpreting *cheeses* as the object of *discussing*. This suggests that integrating a filler phrase into the semantic/conceptual representation is not sufficient for abandoning active gap search. Put differently, comprehenders persist in searching for grammatically-required gaps even after assigning a thematic role to a filler phrase, contrary to the proposal by Chacón (2015).

(8)   **The cheeses** which the gourmets were energetically discussing __ or slowly sipping __ during the banquet were rare imports from Italy.

In this paper, I elaborate on this proposal, by arguing that the conditions of resumptive dependency formation partially follow from an interaction between anaphoric processes and disrupted active gap formation processes. I submit that shortly after encountering a filler, the comprehender builds a representation of a structure that contains a gap, e.g., a VP dominating a trace. This representation must be maintained in working memory, until the filler and the gap are associated. Thus, in my view, active gap formation depends in part in maintaining an active representation of upcoming structure that hosts the gap site, and incrementally determining whether this prediction has been satisfied.

In many models, representations that are stored in memory are vulnerable to being lost. Shunting new information into the focus of attention may displace representations that are stored in working memory, or may otherwise lead to interference effects that result in difficulty of maintaining a information in working memory (Lewis 1996; Van Dyke & Lewis 2003; Gordon et al. 2006; see Lewis et al. 2006 and Jäger et al. 2017 for overviews). Similarly, representations may decay or decrease in activation over time (Kempen & Vosse 1989; Lewis & Vasishth 2005). If processing filler-gap dependencies requires maintaining a representation of upcoming structure in memory, then this prediction may degrade or suffer from interference in high processing demand contexts. Thus, they are predicted to be less sensitive to unresolved filler dependencies. Consequently, the filler NP now plays no role in the semantic or conceptual representation of the sentence. If the comprehender encounters a pronoun that refers to the filler NP, then the filler NP may now be related to interpretation of the sentence, which may increase perception of acceptability. This is sketched in Figure 1.

The first major prediction is that increasing demand on working memory overall should reduce sensitivity to the ungrammaticality of an unresolved filler dependency. Upon encountering a filler, the comprehender generates a predicted structure containing a gap, which is stored in memory. If the bottom-up input mismatches this prediction, then the sentence is perceived to be unacceptable. However, if this representation sufficiently degrades, then the comprehender will be unable to detect that the sentence is ill-formed.

This component of my proposal shares similarities with "forgetting effects" in the processing of multiple center-embedded structures (Frazier 1985; Gibson & Thomas 1999; Vasisth et al. 2010). Ungrammatical sentences with multiple center-embeddings and too few predicates may be perceived to be more acceptable than sentences in which each dependency is grammatically resolved (e.g., *the mouse_{N1} the cat_{N2} the dog_{N3} bit_{V3} chased_{V2} ate_{V1} the cheese* may be perceived to be better than *the mouse_{N1} the cat_{N2} the dog_{N3} bit_{V3} ate_{V1} the cheese*). Gibson & Thomas (1999) proposes that this is because comprehenders forget one of the predicted VPs while processing the intervening complex structures. Since the prediction for one of the VPs is lost, the unacceptability of the sentence goes undetected. In the same vein, I argue that comprehenders are more likely to accept an unresolved filler dependency if there is complex linguistic material intervening between the filler and the projected resolution site.

The second major prediction is that, in contexts in which the predicted gap has decayed, a "pronoun-driven" coreference relation between the filler NP and pronoun aids coherence. When the predicted gapped structure decays, there is no syntactically licensed mechanism for relating the filler NP to the meaning of the sentence. However, if the comprehender selects the filler NP as the pronoun's antecedent, then the interpretation of the sentence is recovered. Thus, acceptability should increase when coreference between

	Syntax	Semantics
Maintained Prediction:	*which butler_i …[Dale [_{VP} saw him_j]]* predicted: …([_{VP} V __])	?(?,Butler) Saw(Dale, ?)
Degraded Prediction:	*which butler_i …[Dale [_{VP} saw him_i]]* predicted: …([_{VP} V __])	Saw(Dale, Butler)

Figure 1: Sketch of the proposal. If the comprehender maintains a prediction of upcoming gapped structure, then mismatch between bottom-up input and predicted structure is a cue that the sentence is ill-formed. If the predicted structure degrades in memory, then the ungrammaticality goes undetected. This allows coreference between the filler and pronoun, which enables a coherent semantic interpretation.

the filler NP and the pronoun is favored, but only when sensitivity to the syntactic ill-formedness of the dependency had already diminished. I argue that these two factors independently conspire to increase acceptability of a filler associating with a pronoun in syntactically complex configurations, which occasionally results in the sentence being perceived as acceptable.

To clarify what I mean by coherence, examine (9a). In this sentence, the filler does not bind a gap, resulting in an unresolved filler dependency. Moreover, in (9a), the filler *the butler* has no apparent relation to the sentence, which describes a situation of the maid's friend liking kids. Thus, the sentence is both ungrammatical and incoherent. However, if there is an anaphoric relation between the filler and the pronoun, as in (9b), then the antecedent filler can be integrated into the interpretation of the sentence, i.e., the babysitter's friend is the one that likes kids. Although this sentence is ungrammatical, the ability to relate the filler to the rest of the sentence may result in enhanced comprehensibility (Asudeh 2012; Keffala 2013; Beltrama & Xiang 2016).

(9)   a.   **No Gap, No Resumptive Pronoun**:
               The maid said that this is the butler that her friend really likes kids.

        b.   **No Gap, Resumptive Pronoun**:
               The butler said that this is the babysitter that her friend really likes kids.

I share with previous approaches the claim that processing difficulty is a crucial factor to the acceptability of resumption in English. However, I depart from previous approaches in two crucial ways. First, I do not characterize resumption as a mechanism for salvaging a dependency in syntactic islands/complex structures. Instead, difficult-to-process structures interfere with the memory representation of the predicted resolution site for the filler. This causes the comprehender to be less sensitive to unresolved dependencies generally. Put differently, resumption does not aid difficult-to-process structures. Instead, difficult-to-process structures enable resumption, in part by disrupting the comprehender's ability to discharge the filler dependency in a syntactically licensed way.

Secondly, this proposal makes no reference to islandhood as such. The significant majority of work on resumption in theoretical syntax (Ross 1967; Boeckx 2003) and experimental syntax (McDaniel & Cowart 1999; McKee & McDaniel 2001; Omaki & Nakao 2010; Heestand et al. 2011; Han et al. 2012; Keffala 2013; Ackerman et al. 2018) specifically addresses the interaction between syntactic islands and resumption. On my proposal, islands are only relevant insofar as they tax memory resources (Kluender & Kutas 1993; Kluender 1998; Hofmeister & Sag 2010, but see Sprouse et al. 2012), contributing to the decay of a representation of a gapped structure. By the same token, if memory representations decay over time, then length also is expected to facilitate resumption (Alexopoulou & Keller 2007; Hofmeister & Norcliffe 2013). This raises a number of issues about the distribution of resumption in English. Resumptive pronouns are more frequently produced in some structures than others (Prince 1990). Additionally, although resumption in English is generally disfavored, the acceptability of resumptive pronouns in English varies as a function of their syntactic position, dependency length, and island type (McDaniel & Cowart 1999; McKee & McDaniel 2001; Alexopoulou & Keller 2007; Han et al. 2012; Keffala 2013; Morgan & Wagers 2018). If my approach is generally right, then it remains to work out how syntactic islands differentially induce processing difficulty, such that resumption is better in some contexts than in others.

Finally, a note on cross-language variation. This paper focuses on English, in which resumption is generally ungrammatical. This contrasts with languages like Irish,

Hebrew, and Arabic, in which resumption is grammatically licensed (Shlonsky 1992; Aoun et al. 2001; McCloskey 2002; 2006). The mechanistic account that I propose is unlikely to extend to languages with grammaticized resumption. If filler dependency processing involves active prediction of a grammatical resolution site, then in principle the comprehender could actively construct a resumptive dependency in syntactic island contexts. In fact, Keshev & Meltzer-Asscher (2017) argue that this is observed in Hebrew. They observed evidence of active dependency formation in strong islands that admit resumption, but not in islands that do not tolerate resumption (see also Farby et al. 2010). However, McCloskey (2017) suggests that there may be a stronger preference for resumption in syntactically complex configurations in Irish, even outside of syntactic islands. This may suggest some processing factor even in these languages. Thus, I set aside the questions of cross-language and cross-construction variation for future work.

In this paper, I test three predictions of this analysis. First, I examine the influence that length has on acceptability of resumption. Longer dependencies should favor resumption, as previously shown (Alexopoulou & Keller 2007; Hofmeister & Norcliffe 2013). Secondly, I examine the influence that syntactic prominence of the filler phrase has on resumption. By hypothesis, syntactically prominent antecedents are more likely to license resumption than less prominent antecedents, if resumption is "pronoun-driven" (Ariel 1999; Chacón 2015). Finally, I examine the influence of memory strain on acceptance of gapless filler dependency resolutions. On my proposal, increased memory strain should result in a greater perception of acceptance for unresolved filler-gap dependencies overall. Greater strain on memory systems may raise the acceptability of resumption above a critical threshhold, such that these constructions are perceived as fully grammatical. In other words, strain on memory resources may render the ungrammaticality of these sentences "subliminal", in the sense of Almeida (2014).

I present on the results of four judgment studies. Experiments 1 and 2 were speeded acceptability judgment tasks, and Experiments 3 and 4 used a new experimental paradigm. In these experiments, participants judged sentences while simultaneously maintaining a list of words in memory. In the critical stimuli, filler dependencies did not resolve with a predicate containing a gap (a *gapless* predicate). Additionally, there was a pronoun in the sentence. The range of antecedents of this pronoun was manipulated by changing the stereotypical genders of the NPs in the sentence. I predicted that the gapless sentences would be rated much lower than gapped sentences, on average. However, as the demand on working memory increases, the sensitivity to the ungrammaticality of gapless dependencies should decrease, i.e., ratings should improve for gapless predicates. Secondly, coreference between the pronoun and the filler NP should improve ratings even more when the comprehender is placed under processing strain, due to the loss of coherence of the sentence. When the gap is unlikely to have degraded, then gapless sentences are expected to be assigned lower ratings with minimal effect of the pronoun's interpretation, conversely.

In Experiment 1, I show that participants robustly prefer gapped resolutions for filler dependencies. In Experiment 2, I show that increased length between the filler and the embedded predicate favors resumption. In Experiment 3 and 4, I show that the addition of a working memory task overall raises the acceptability of gapless resolutions, although the relative patterns in Experiments 1 and 2 are replicated in Experiments 3 and 4. In a meta-analysis of these four studies, I then suggest that the pattern of results within and between experiments are consistent with the proposal here, and show that memory strain crucially diminishes sensitivity to gapless resolutions, which may lead to a preference for resumption.

2 Experiment 1

2.1 Rationale

There were two goals for Experiment 1. The first goal was to demonstrate that gapped predicates are strongly preferred to gapless ones if a sentence contains a filler dependency. The second goal was to determine whether the availability of resumption would improve ratings for gapless predicates. In Experiment 1, I purposefully constructed materials that contained short filler dependencies, which should be unlikely to decay in memory. Thus, I predicted that participants would overwhelmingly reject fillers with gapless predicates, i.e., there should be no detectable effect of resumption in Experiment 1. In Experiments 2–4, I changed aspects of this design, which resulted in increased acceptance of resumption.

Previous work on resumption in English (McDaniel & Cowart 1999; Mc-Kee & McDaniel 2001; Alexopoulou & Keller 2007; Omaki & Nakao 2010; Heestand et al. 2011; Han et al. 2012; Beltrama & Xiang 2016; Ackerman et al. 2018) primarily compared resumptive pronouns in syntactic islands with gaps in the same position in order to assess the ability of resumption to ameliorate syntactic island violations. The results from previous studies have demonstrated a complex pattern of judgments. Overall, however, resumptive pronouns in syntactic islands appear to be rejected at similar rates to gaps in syntactic islands, if not more often.

For instance, in an oral judgment task, McKee & McDaniel (2001) found that resumptive pronouns were largely rejected in the subject, direct object, and prepositional object positions. By contrast, resumptive pronouns were accepted at rates greater than 60% in genitive object positions, and near ceiling in subject positions in *wh*-islands and tensed adjuncts. McDaniel & Cowart (1999) found that gaps and resumptive pronouns were rated equally low in subject islands, but found a preference for resumptive pronouns in the subject position of a *wh*-island, which they attributed to amelioration of the Empty Category Principle (ECP)/*that*-trace constraint (Chomsky 1981) (*That's the girl that I wonder when {she/__} met you*). This finding was also replicated by Omaki & Nakao (2010).

In contrast to these findings, Alexopoulou & Keller (2007) found that resumptive pronouns were often rated worse than gaps outside of island contexts. With further embeddings, the ratings between gaps and resumptive pronouns approached convergence. In syntactic island contexts, ratings for both gaps and resumptive pronouns were equally low, regardless of the number of embeddings. A similar profile was found by Heestand et al. (2011). Their results showed that resumption was dispreferred to gaps outside of syntactic islands, and both were equally low inside syntactic islands.

In contrast to these findings, Han et al. (2012) and Keffala (2013) found a relatively consistent profile of low judgments for resumptive pronouns in a variety of syntactic configurations. However, they found that the unacceptability of gaps varied more dramatically in the same structures. Keffala (2013) suggested that this may underlie the apparent selective acceptability of resumption across constructions. On her account, moderately unacceptable resumptive dependencies may be preferable to strongly unacceptable island violations, but dispreferred compared to mildly unacceptable island violations.

Finally, Beltrama & Xiang (2016) found that resumption improved ratings for sentences when participants were asked to judge the *comprehensibility* of the sentence as opposed to its *acceptability*. These findings may suggest that resumption facilitates comprehension without being grammatical. However, in a forced-choice task, Ackerman et al. (2018) found that participants were more likely to prefer resumption over gaps. They argued that previous failures to find an amelioration effect of resumption was a methodological

artifact of comparison across trials in experiments with randomized stimulus presentation. When the two alternatives were significantly salient, a preference for resumption emerged.

Thus, although the pattern of acceptance rates for resumption is mixed, resumption appears to generally be disfavored in English. In all these cases, the crucial comparisons were between a gap and a resumptive pronoun in the same syntactic position. From my perspective, the more crucial question is whether the comprehender prefers resolving a filler dependency with a pronoun over the possibility of a later, well-formed gap, and how acceptability of resumption compares to the baseline of a filler dependency being unresolved entirely. For this reason, the crucial manipulation in Experiments 1–4 was the presence of a gap after a (potentially) resumptive pronoun, and the possible interpretations of this pronoun. If resumption in English is ungrammatical, then it was predicted that gapped sentences would be overwhelmingly preferred to gapless sentences, regardless of whether a resumptive pronoun was present. However, if resumption is grammatical, then I predicted that comprehenders may be more likely to accept the pronoun as a suitable resolution site for the dependency if the pronoun and filler NP corefer.

2.2 Participants

There were 53 participants recruited for Experiment 1 from Amazon's Mechanical Turk platform (http://www.mturk.com). They must have been from the United States, and have completed over 500 HITs (Amazon Mechanical Turk tasks) with 95% approval rating or higher. All participants self-identified as native English-speakers, and the mean age of participants was 38.

2.3 Materials

For Experiment 1, I designed 36 sets of target stimuli, and 32 filler items. The fillers were 50% grammatical. Most filler sentences contained two clauses, and most contained some kind of long-distance dependency, such as a cleft, pseudo-cleft, or relativization. Ungrammatical sentences had a variety of errors, but mostly used subcategorization errors, or local morphosyntactic errors such as agreement or case mismatches. Some filler items had multiple errors.

For the target stimuli, I manipulated $\pm$ Gap and Pronoun Reference (Ambiguous, Filler, Subject), yielding a 2 × 3 design. All target sentences contained a cleft (focused relativization) dependency and a pronoun. The gender of the pronoun was counterbalanced across all conditions across items. The pronoun was always a possessor in the subject of the embedded clause. This position was chosen because extraction from genitives is not allowed in English, and English-speakers readily produce resumptive pronouns in genitive NP positions (Zukowski & Larsen 2004). Furthermore, I assumed that processing an NP with a pronominal possessor is unlikely to introduce any significant processing costs. The findings from Chacón (2015) and previous pilot studies found that other sentence structures produced substantially noisy results, which I attribute to the additional complexity associated with the the structures of these sentences. Thus, I chose to use a simpler design to minimize the effect of syntactic complexity.

For the $\pm$ Gap factor, I manipulated whether the end of the sentence contained a gap for the open filler dependency (+ Gap) or not (–Gap). In the + Gap conditions, I used verbs that I judged to have a strong bias for taking an NP object, and that could plausibly take the filler as an argument. For the –Gap conditions, I used verbs that had a strong bias against taking an NP object, or transitive constructions that had NP objects. This was to ensure that comprehenders could quickly detect that the filler

was not an argument of the critical verb upon entering the VP, at least if they still maintained a prediction for a gapped structure. All target materials were matched in length.

Each target stimulus contained two NPs before the critical pronoun: the subject NP of the main clause, and the filler NP. Both NPs could be potential antecedents for the critical pronoun. I manipulated the stereotypical genders of these NPs, using the strongly masculine and feminine stereotyped nouns from the norming study in Kennison & Trofe (2003). In the Ambiguous conditions, both the subject NP of the sentence and the filler NP matched in gender with the pronoun. In the Filler conditions, only the filler NP matched in stereotypical gender with pronoun, and in the Subject conditions, only the subject matched. The items were designed to keep both interpretations relatively plausible and salient. A sample set of stimuli are given in (10), and the full set of stimuli for Experiments 1–4, including fillers, are given in the Appendix. In the Ambiguous and the Filler conditions, the pronoun was potentially a resumptive pronoun, since reference between the filler and the pronoun was licensed by the stereotypical genders of the NPs. However, in the Subject condition, the pronoun was unlikely to refer to the filler, thus preventing a resumptive analysis, and strongly favoring an analysis in which the pronoun was anaphoric to the non-filler subject NP.

(10)   a.   **+ Gap, Ambiguous**:
              The maid$_{[F]}$ said that this is the babysitter$_{[F]}$ that **her**$_{[F]}$ friend really highly recommended ___.

        b.   **+ Gap, Filler**:
              The butler$_{[M]}$ said that this is the babysitter$_{[F]}$ that **her**$_{[F]}$ friend really highly recommended __.

        c.   **+ Gap, Subject**:
              The maid$_{[F]}$ said that this is the butler$_{[M]}$ that **her**$_{[F]}$ friend really highly recommended __.

        d.   **–Gap, Ambiguous**:
              The maid$_{[F]}$ said that this is the babysitter$_{[F]}$ that **her**$_{[F]}$ friend really liked kids.

        e.   **–Gap, Filler**:
              The butler$_{[M]}$ said that this is the babysitter$_{[F]}$ that **her**$_{[F]}$ friend really liked kids.

        f.   **–Gap, Subject**:
              The maid$_{[F]}$ said that this is the butler$_{[M]}$ that **her**$_{[F]}$ friend really liked kids.

I predicted that there should be an effect of ± Gap, with + Gap conditions strongly preferred to –Gap conditions. This is because I understood resumption to be ungrammatical in English, and thus the addition of a pronoun would likely not reduce the "penalty" for gapless sentences. However, if comprehenders could resolve the filler dependency with the pronoun, then acceptance rates should increased for the –Gap, Ambiguous and –Gap, Filler sentences compared to the –Gap, Subject sentences. If resumptive dependencies are only constructed in contexts of high memory demand, then I did not expect the pronoun's interpretation to have much effect in Experiment 1. This is because the sentences were specifically designed to not include structures that introduced exceptional processing difficulty, even though they were contained in a syntactic island. In other words, I predicted no interaction effects between ± Gap and Pronoun, given that processing demands were designed to be minimal in Experiment 1.

2.4 Methods

For Experiment 1, I used a speeded acceptability judgment task. I chose this task because it has been used to reveal "grammatical illusions", or sentences that are ungrammatical but are momentarily perceived to be acceptable (e.g., Wagers et al. 2009), and because it places the participant under time pressure in processing the sentence. Additionally, participants must choose a rating based on whatever has been stored in memory, since they are not able to look back at the sentence and reassess its acceptability.

Participants were recruited from Amazon Mechanical Turk. Upon accepting the HIT (Human Interaction Task) on the Mechanical Turk platform, participants were instructed to navigate to the experiment hosted on the IbexFarm (Drummond 2018). Here, participants read the consent form and instructions, and were given the opportunity to volunteer their age, location, native language, and any additional second languages. After reading the consent form and instructions, participants judged five practice trials, which featured sentences that were clearly acceptable or clearly unacceptable. Afterwards, the main experiment began.

Sentences were displayed word-by-word centered on the screen, using a rapid serial visual presentation (RSVP) design. Each word was displayed for 300 ms. Each sentence was preceded with a fixation cross and followed by a period, both of which were displayed for 500 ms. After the sentence ended, participants were asked whether the sentence was acceptable. They entered their answer either by clicking on a "Yes" or "No" button displayed on the screen, or by using the 1 and 2 keys. All experimental materials were presented in a randomized order, distributed in six separate lists in a 2 × 3 design.

After completing the task, participants were given a code that they entered into Mechanical Turk. They were compensated $2.00 for participation, and it took approximately 20 minutes to complete the task.

2.5 Results

For the filler items, I constructed a logit mixed-effects model using Grammaticality as the factor, and random effects for participant and item.[1] Grammaticality was significant ($\beta = 3.96$, SE $= 0.34$, $z = 11.61$, $p < 0.01$). The mean proportion of acceptance for Grammatical sentences was $88.1 \pm 1.1\%$, and the mean proportion of acceptance for Ungrammatical sentences was $19.1 \pm 1.3\%$. Thus, participants overwhelmingly accepted grammatical sentences, and rejected ungrammatical sentences.

For the target items, I constructed a logit mixed-effects models using the `lme4` package in R (R Development Core Team 2008; Bates et al. 2015). I used a maximal random effects structure (Barr et al. 2013). The model included ±Gap, Pronoun, and their interaction term as factors. The model also included random slopes for ±Gap, Pronoun, and their interaction for participants and for items.[2] Because the factor Pronoun had three levels without an obvious choice for the reference level, I sum-coded these factors around 0. This resulted in two separate coefficients for the factor Pronoun. The ±Gap factor was fit with +Gap coded as 1 and –Gap coded as –1. The Pronoun factor was fit with two coefficients, with Ambiguous fit as 1 and 0 for the two coefficients respectively, Filler as 0 and 1, and Subject as –1 and –1. This complicates the interpretation of the mixed-effects model, but planned pairwise comparisons between each of the three Pronoun levels were compared within each of the ±Gap levels. The key comparison was between the –Gap, Filler resumptive condition compared to –Gap, Subject anaphoric condition. Pairwise

---

[1] The structure of this model using `lme4` syntax is: Response ~ Grammatical + (1|Participant) + (1|Item).
[2] The structure of this model using `lme4` syntax is: Response ~ Gap * Pronoun + (1 + Gap*Pronoun|Participant) + (1 + Gap*Pronoun|Item).

comparisons used the Tukey honestly significant difference correction for multiple comparisons. The mean acceptance rates by condition are given in Figure 2. The results of the logit mixed effects model fit to the acceptability data are given in Table 1, and the results of the pairwise comparisons are given in Table 2.

Most strikingly, the +Gap conditions were overwhelmingly accepted, and the –Gap conditions were overwhelmingly rejected. This was predicted, because I assumed that resumption is ungrammatical in English, i.e., participants would only accept sentences in which the filler dependency resolves with a gap. Participants did not appear to treat the pronoun in the subject NP as a suitable resolution site for the cleft dependency.

Performance on Experiment 1 could also be understood as a signal detection task, i.e., correctly distinguishing grammatical sentences from ungrammatical sentences. For this reason, I computed the sensitivity index, or d′, to assess accuracy. D′-scores are 'a measure of a participant's ability to discriminate between the signal, while taking into account participant response biases (Macmillan & Creelman 2005). To calculate d′-scores, I treated acceptances of +Gap target items and Grammatical filler items as 'hits', and rejections of –Gap target items and Ungrammatical filler items as 'correct rejections'. The average d′-scores across subjects for all items in Experiment 1 was 2.47 ± 0.14, indicating moderately high ability to discriminate between grammatical and ungrammatical sentences. In the filler items specifically, the average d′-score was 2.41 ± 0.14. Across all target items, the average d′-score was 3.25 ± 0.24, suggesting a robust ability to distinguish grammatical +Gap target items from ungrammatical –Gap target items.

2.6 Discussion

The goal of Experiment 1 was to determine whether the availability of resumption improved the acceptability of gapless sentences. The results from Experiment 1 found no such improvement. Previous work demonstrated that sentences with resumptive dependencies often were assigned low ratings in off-line judgments (Alexopoulou & Keller 2007; Omaki & Nakao 2010; Heestand et al. 2011; Han et al. 2012; Keffala 2013; Beltrama & Xiang 2016), which is consistent with the proposal that resumption is not grammatically available in English. However, these findings compared gaps with resumptive pronouns in the same syntactic positions. In this study, I compared resumptive dependencies
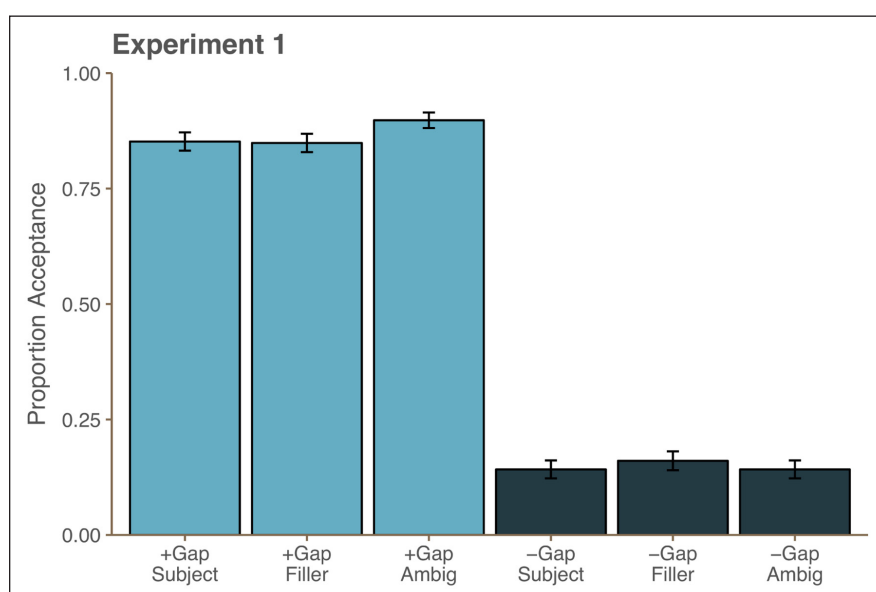

Figure 2: Mean acceptance rates by condition from Experiment 1, with error bars representing one standard error from the mean.

Table 1: Results of logit mixed-effects model fit to the acceptability data for the target items in Experiment 1. Starred (*) rows correspond to *p*-values below 0.05.

	β	SE	*z*	*p*	
Intercept	−0.18	0.21	−0.86	0.39	
Gap	2.69	0.23	11.73	<0.01	*
Pronoun:1	0.13	0.21	0.62	0.54	
Pronoun:2	−0.01	0.20	−0.06	0.95	
Gap × Pronoun:1	0.25	0.21	1.23	0.22	
Gap × Pronoun:2	−0.28	0.20	−1.41	0.16	

Table 2: Results of pairwise comparisons for the model fit to the acceptability judgments of Experiment 1. Comparisons were made between the three levels for the Pronoun factor nested within each level of the ±Gap factor.

		β	SE	*z*-ratio	*p*
+Gap	Ambiguous – Filler	0.68	0.47	1.43	0.32
	Ambiguous – Subject	0.47	0.47	1.01	0.57
	Filler – Subject	−0.20	0.46	−0.44	0.90
−Gap	Ambiguous – Filler	−0.39	0.51	−0.76	0.73
	Ambiguous – Subject	0.02	0.57	0.03	1.00
	Filler – Subject	0.41	0.55	0.75	0.74

with filler-gap dependencies, and anaphoric interpretation of pronouns with resumptive interpretations of pronouns, which is a novel comparison. The results from Experiment 1 further demonstrated that comprehenders were reluctant to interpret a pronoun as a resolution site for short filler dependencies. To my knowledge, this is also the first demonstration of the bias against resumptive dependencies using speeded acceptability judgment tasks. This is a novel finding, because it demonstrates that placing a participant under time pressure is not sufficient for increasing the acceptability of resumption.

One potentially surprising finding was that there was no detectable weak cross-over effect in Experiment 1 (Postal 1971; Wasow 1972). Weak cross-over effects are the penalty in acceptability associated with a filler NP that binds a pronoun and a gap, as in (11). If weak crossover were reliably detectable in judgments, then I would have expected lower ratings for the + Gap, Filler sentences compared to the + Gap, Subject sentences. This is because + Gap, Filler sentences contained a filler NP that was bound both the pronoun and the later gap.

(11)   *Who$_i$ did his$_i$ mother love __?

In a real-time processing study, Kush et al. (2017) found that the head of a an unresolved filler-gap dependency was considered as a suitable antecedent for a pronoun, even though this interpretation resulted in a weak cross-over violation. However, this interpretation was rejected in off-line measures. These findings suggested that the weak cross-over violations are not immediately detected in real-time processing. Because speeded acceptability judgment tasks are intended to give a measure of judgments earlier in processing, the failure to find a weak cross-over effect may be interpreted as replicating this result. However, in Experiment 3 and in the overall meta-analysis in Section 6, there was evidence of the weak cross-over effect.

In summary, the results of Experiment 1 failed to demonstrate any facilitatory effect for resumption, even under time pressure. However, this is unsurprising. I purposefully designed these materials to have shorter filler dependencies, and I purposefully selected a syntactic island that I supposed would not induce significant strain on the participants' memory. In Experiment 1, the filler NP and the pronoun were only separated by one word, the complementizer *that*. Additionally, the non-filler NP is the main clause subject, a syntactically prominent position. This makes it a highly accessible antecedent for the pronoun (Gundel et al. 1993; Grosz et al. 1995; Arnold 2010; Arnold & Lao 2015), which is a determining factor for resumption in English (Erteschik-Shir 1992; Ariel 1999; Chacón 2015). In other words, it may have been difficult for the pronoun to access the filler NP as its antecedent, in favor of the subject NP.

In Experiment 2, the position of the cleft filler NP and the non-filler subject NP were changed, such that the cleft was in the main clause and the non-filler subject NP was in the embedded clause. This increased the length of the cleft dependency, and also placed the filler NP as a prominent argument in the main clause, making it a more accessible antecedent for the pronoun. These changes were predicted to make the resumptive analysis more likely.

3 Experiment 2

As in Experiment 1, the goal of Experiment 2 was to determine whether the availability of resumption improved the acceptability of gapless sentences. In Experiment 1, there was no evidence that resumption improved acceptability. However, the structure of the materials in Experiment 1 discouraged resumption, due to the length of the filler dependency and the syntactic positions of the possible NP antecedents for the pronoun.

In Experiment 2, the order of the two NPs was reversed, such that the clefted filler NP occurred in the main clause, and the non-filer subject NP occurred in the embedded clause. I predicted that the reverse order would make the filler NP more syntactically prominent. This is important, because pronouns more easily select syntactically prominent antecedents (Gundel et al. 1993; Grosz et al. 1995; Arnold 2010). Thus, reordering the relevant NPs in the sentence should favor a resumptive interpretation more easily. Moreover, the reversed order increased the length of the cleft dependency, which is also observed to favor resumption.

3.1 Participants

There were 60 participants recruited for Experiment 2 from Amazon's Mechanical Turk platform, using the same inclusion criteria as in Experiment 1. All participants self-identified as native English-speakers, and the mean age of participants was 36.

3.2 Materials

The materials in Experiment 2 were similar to the those in Experiment 1, except that the filler NP and the non-filler subject NP were re-ordered, such that the filler NP appeared first in the main clause, and the non-filler subject NP appeared second in the embedded clause. This change made the filler NP more syntactically prominent, and increased the length of the cleft dependency. The materials for Experiment 2 are exemplified in (12).

(12) a. **+Gap, Ambiguous**:
          This is the babysitter$_{[F]}$ that the maid$_{[F]}$ said that **her**$_{[F]}$ friend really highly recommended __.
      b. **+Gap, Filler**:
          This is the babysitter$_{[F]}$ that the butler$_{[M]}$ said that **her**$_{[F]}$ friend really highly recommended __.

   c.   **+ Gap, Subject**:
This is the butler[M] that the maid[F] said that **her**[F] friend really highly recommended __.

   d.   **–Gap, Ambiguous**:
This is the babysitter[F] that the maid[F] said that **her**[F] friend really liked kids.

   e.   **–Gap, Filler**:
This is the babysitter[F] that the butler[M] said that **her**[F] friend really liked kids.

   f.   **–Gap, Subject**:
This is the butler[M] that the maid[F] said that **her**[F] friend really liked kids.

3.3 Methods

The methods in Experiment 2 were the same as in Experiment 1. Participants were compensated $2.00 for participation, and took approximately 20 minutes to complete the task.

3.4 Results

For the filler items, I constructed a logit mixed-effects model using Grammaticality as the factor, and random effects for subject and item, as in Experiment 1. Grammaticality was significant ($\beta = 3.50$, SE $= 0.34$, $z = 10.38$, $p < 0.01$). The mean proportion of acceptance for Grammatical filler items was $88.3 \pm 1.0\%$, and the mean proportion of acceptance for Ungrammatical filler items was $27.9 \pm 1.4\%$. As before, participants overwhelmingly accepted Grammatical filler items and rejected Ungrammatical ones.

   Due to an experimenter error, 40 observations out of 2160 target items had to be discarded. After excluding these observations, I constructed a logit mixed-effects model using the same procedure as in Experiment 1. The mean acceptance rates by condition are given in Figure 3. The results of the logit mixed effects model are given in Table 3, and the results of the pairwise comparisons are given in Table 4.
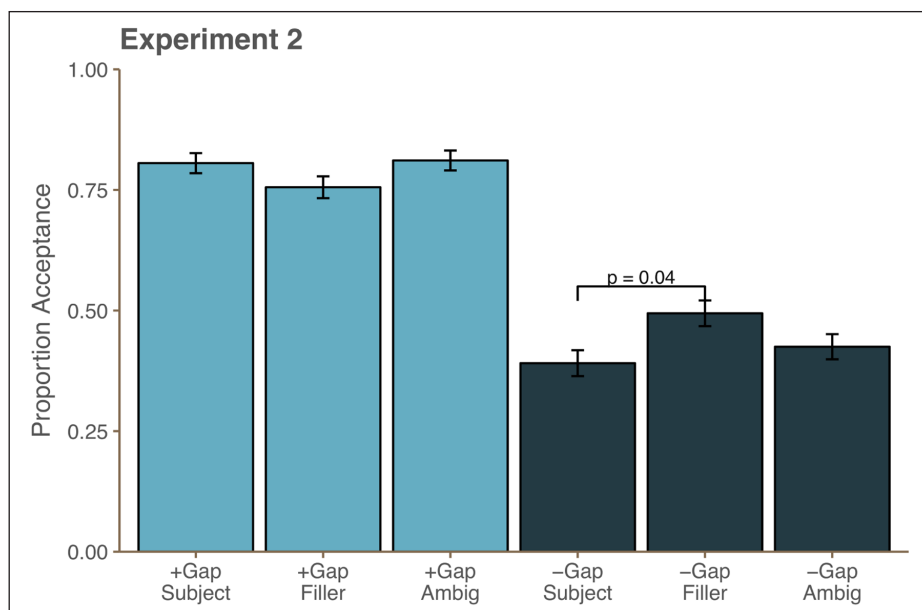

Figure 3: Mean acceptance rates by condition from Experiment 2, with error bars representing one standard error from the mean. Reported *p*-value corresponds to results from the pairwise comparisons.

Table 3: Results of logit mixed-effects model fit to the acceptability judgments for the target items in Experiment 2. Starred (*) rows correspond to *p*-values below 0.05.

	β	SE	z	p	
Intercept	0.66	0.18	3.62	<0.01	*
Gap	1.09	0.13	8.46	<0.01	*
Pronoun:1	0.13	0.11	1.23	0.22	
Pronoun:2	0.08	0.10	0.09	0.93	
Gap × Pronoun:1	0.18	0.11	1.68	0.10	
Gap × Pronoun:2	−0.33	0.10	−3.23	<0.01	*

Table 4: Results of pairwise comparisons for the model fit to the acceptability judgments from Experiment 2. Comparisons were made between the three levels for the Reference factor nested within each level of the ± Gap factor. Starred (*) rows correspond to *p*-values below 0.05.

		β	SE	z-ratio	p	
+Gap	Ambiguous − Filler	0.65	0.28	2.31	0.06	
	Ambiguous − Subject	0.28	0.30	0.92	0.63	
	Filler − Subject	−0.37	0.26	−1.42	0.33	
−Gap	Ambiguous − Filler	−0.37	0.21	−1.7	0.18	
	Ambiguous − Subject	0.23	0.23	1.00	0.58	
	Filler − Subject	0.60	0.25	2.44	0.04	*

As in Experiment 1, the most striking result is that the + Gap sentences were accepted significantly more often than the –Gap sentences. However, unlike in Experiment 1, there was a significant interaction between ± Gap and Pronoun, which was reflected in the increased acceptance rates for the –Gap, Filler condition compared to the –Gap, Subject condition. Thus, in Experiment 2, there appeared be a facilitation effect of resumption, which was not observed in Experiment 1.

The average d′-score across participants and all items for Experiment 2 overall was 1.61 ± 0.11. For filler items, the average d′-score was 2.27 ± 0.14. For target items, the mean score was 1.40 ± 0.15. This suggests that participants were moderately capable in discriminating between grammatical and ungrammatical items.

3.5 Discussion

The goal of Experiment 2 was to determine whether resumption would improve sentences with longer cleft dependencies than the sentences used in Experiment 1. In Experiment 1, there was no evidence that reference between the filler NP and pronoun improved acceptability, suggesting that comprehenders did not entertain a resumptive analysis. In Experiment 2, coreference between the filler NP and pronoun was facilitated, because the filler NP was in a syntactically more prominent position, which increased the likelihood of coreference between the pronoun and the filler (Gundel et al. 1993; Grosz et al. 1995; Arnold 2010). Additionally, the dependency was longer, which is independently observed to increase the acceptability of resumption (Alexopoulou & Keller 2007; Hofmeister & Norcliffe 2013). Thus, the design of Experiment 2 favored the acceptability of resumption.

In Experiments 3 and 4, I added the word list recall task to the speeded acceptability judgment task. This was designed to strain working memory resources while participants processed and judged a sentence. On the analysis that I propose, increased strain on memory sources was predicted to increase the acceptability of –Gap target items overall.

This is because the additional strain on memory should result in quicker decay of the prediction of a gapped structure in memory. This then should result in diminished ability to determine whether a filler dependency was resolved in a syntactically licensed way, i.e., with a gap.

4 Experiment 3

4.1 Rationale

The goal of Experiment 3 was to determine whether the acceptability of resumption increased with an increased strain on memory. As described earlier, many previous accounts of resumption have proposed that processing difficulty and syntactic complexity partially determine the distribution of resumption (Keenan & Comrie 1977; Ariel 1999; Hawkins 1999; 2004). Resumptive pronouns typically are used in syntactic island contexts (Ross 1967; Boeckx 2003; Ferreira & Swets 2005), which are often characterized as particularly syntactically complex (Kluender & Kutas 1993; Kluender 1998; 2004; Hofmeister & Sag 2010). Additionally, length affects the acceptability of resumption in English (Alexopoulou & Keller 2007; Hofmeister & Norcliffe 2013), which itself may be due to the increased processing difficulty of longer dependencies (Kluender & Kutas 1993; Hawkins 1999; 2004; Gibson 2000; Hofmeister & Sag 2010). On my account, syntactic processing difficulty favors resumption because it leads to difficulty in maintaining a predicted structure in memory. Syntactic island configurations and length are possible sources of this difficulty, but other kinds of demands on memory may also affect the representation.

To test this, I introduced a word list recall task in Experiments 3 and 4. Participants were asked to memorize a list of words before judging a sentence as acceptable or unacceptable, and then respond to a probe word after judging the sentence. Crucially, the materials in Experiment 3 were identical to those in Experiment 1. On my proposal, the addition of this task should lead to a degraded representation of the gapped structure in memory, resulting in less sensitivity to the ungrammaticality of an unresolved filler-gap dependency.

4.2 Participants

There were 60 participants recruited for Experiment 3 from Amazon's Mechanical Turk platform (http://www.mturk.com), using the same inclusion criteria as in Experiments 1 and 2. All participants self-identified as native English-speakers, and the mean age of participants was 35.

4.3 Materials

The sentences that participants were asked to judge were the same as in Experiment 1. For the word list recall task, there were three short nouns not mentioned in the stimulus sentence that preceded it. These nouns were randomly selected from a dictionary, and then hand-selected to avoid repetition between the word-list and the target sentence. I also avoiding using nouns that had clear semantic relations to the sentence or to each other. Most nouns were monomorphemic.

4.4 Methods

The methods for Experiment 2 were the same as in Experiment 1, with the exception of the word list recall task. Before each trial, participants saw the list of words displayed in the center of the screen for 1000 ms. Afterwards, the sentence to be judged was automatically displayed in an RSVP design as in Experiments 1 and 2. Then, participants were asked to judge the sentence as acceptable or unacceptable. After judging it, participants were asked whether a probe word was in the initial list of words that they were asked to

memorize. On half of the trials, the probe word was in the word list before the judgment phase. Participants were given as much time as necessary to provide a judgment for the sentence and for responding to the probe word. This process is illustrated in Figure 4.

4.5 Results

In this section, I first describe performance on the word list recall task, and then performance on the acceptability judgment task. For the acceptability results, I first conducted analysis on the raw results. Afterwards, I reanalyzed the dataset, excluding trials with incorrect responses on the word list recall task. I decided to do this, because participants may have primarily attended to one of the two tasks. Primarily attending to the judgment task may produce results more similar to those in Experiment 1, but with low accuracy rates on the word list recall task. Conversely, attending to the word list recall task may produce noisier performance on the acceptability judgment task, or it may amplify the effect of memory strain on the acceptability judgments. Because I am interested in the effect of memory strain on judgments in Experiment 3, I chose to err on the side of excluding trials in which participants were not attending to the word list recall task. For transparency, I report on both results, and highlight any crucial differences.

For probe word recall accuracy, I report on both percentage of correct trials and average d'-scores across participants. Overall, probe word recall accuracy was $80.4 \pm 0.4\%$. The average d'-score across all items and participants was 1.92 ± 0.10, suggesting moderate accuracy in the recall task in Experiment 3. For the filler items, probe word recall accuracy was $80.0 \pm 0.06\%$. For Grammatical filler items, recall accuracy was $81.1 \pm 0.9\%$, and for Ungrammatical filler items, recall accuracy was $78.9 \pm 0.9\%$. In d'-scores, recall accuracy for filler items was 2.2 ± 0.17. For target items, the mean accuracy was $80.8 \pm 0.06\%$. Mean accuracy by condition for target items is broken down in Table 5. A logit mixed effects model with $\pm$ Gap, Pronoun, and their interaction term as factors, and with participant and item as random effects,[3] revealed higher recall accuracy for the Subject conditions ($\beta = 0.47$, $SE = 0.14$, $z = 3.30$, $p < 0.01$), but no other effects were found (all $ps > 0.05$). The average d'-score on the word list recall task for target items was 2.46 ± 0.26. This suggests that participants attended to the word list recall task with moderately high accuracy.
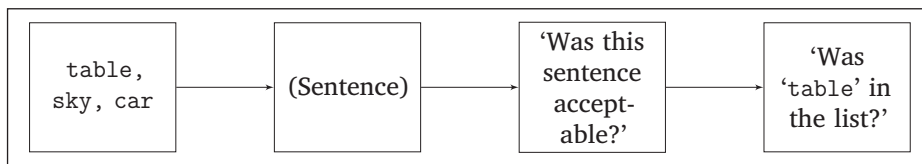

Figure 4: Method for Experiment 3. First, the word list was displayed. Then, the speeded acceptability judgment task was conducted. Afterwards, participants were asked whether a probe word was contained in the initial list.

Table 5: Mean accuracy and standard error on word list recall task by condition for Experiment 3.

	+Gap	−Gap
Ambiguous	78.1 ± 1.5%	76.9 ± 1.6%
Filler	81.1 ± 1.5%	78.9 ± 1.5%
Subject	85.6 ± 1.3%	83.9 ± 1.4%

---

[3] The structure of this model using `lme4` syntax is: Response ~ Gap * Pronoun + (1|Participant) + (1|Item).

Next, I describe the acceptance rates in the acceptability judgment task. First, as before, I constructed a logit mixed effects model to analyze the filler items, with the same structure as Experiments 1 and 2. Grammatical filler items were accepted more often than Ungrammatical filler items, in both the dataset including incorrect trials on the word list recall task ($\beta = 3.11$, SE $= 0.27$, $z = 11.41$, $p < 0.01$), and the dataset excluding these trials ($\beta = 3.26$, SE $= 0.31$, $z = 10.57$, $p < 0.01$). Before excluding these trials, the Grammatical filler items were accepted $84.3 \pm 0.8\%$ of the time, and the Ungrammatical filler items were accepted $31.3 \pm 1.1\%$ of the time. After exclusion of incorrect trials, the Grammatical filler items were accepted $85.0 \pm 0.9\%$ of the time, and the Ungrammatical filler items were accepted $31.2 \pm 0.01\%$ of the time.

For the target items, I constructed a logit mixed-effects model for both datasets with the same structure as in Experiments 1 and 2, and conducted the same pairwise comparisons. The mean acceptance rates by condition after exclusion are plotted in Figure 5. The results of the logit mixed effects model after exclusion are given in Table 6, and the results of the pairwise comparisons are given in Table 7.

As in Experiments 1 and 2, the most striking fact is the main effect of $\pm$Gap, with the $+$Gap conditions accepted significantly more often than the $-$Gap conditions. Additionally, there was a main effect of Pronoun:2. In the dataset with incorrect memory trials excluded, pairwise comparisons revealed a dispreference for coreference between a filler NP and the pronoun within the $+$Gap level. This is consistent with the weak cross-over effect described in Section 2.6, since it reflects a bias against a filler NP binding both a pronoun and a gap in the same sentence. However, no comparisons were significant within the level $-$Gap.

The model fit to the dataset that included incorrect trials on the word list recall task showed a similar profile, with a main effect of $\pm$Gap ($\beta = 2.51$, SE $= 0.28$, $z = 8.91$, $p < 0.01$), a marginal effect of Pronoun:1 ($\beta = 0.45$, SE $= 0.25$, $z = 1.81$, $p = 0.07$), a significant effect of Pronoun:2 ($\beta = -0.59$, SE $= 0.18$, $z = 3.32$, $p < 0.01$), and an interaction effect between $\pm$Gap and Pronoun:2 only observed in this dataset ($\beta = -0.54$, SE $= 0.19$, $z = 2.86$, $p < 0.01$). Pairwise comparisons in this dataset again revealed a
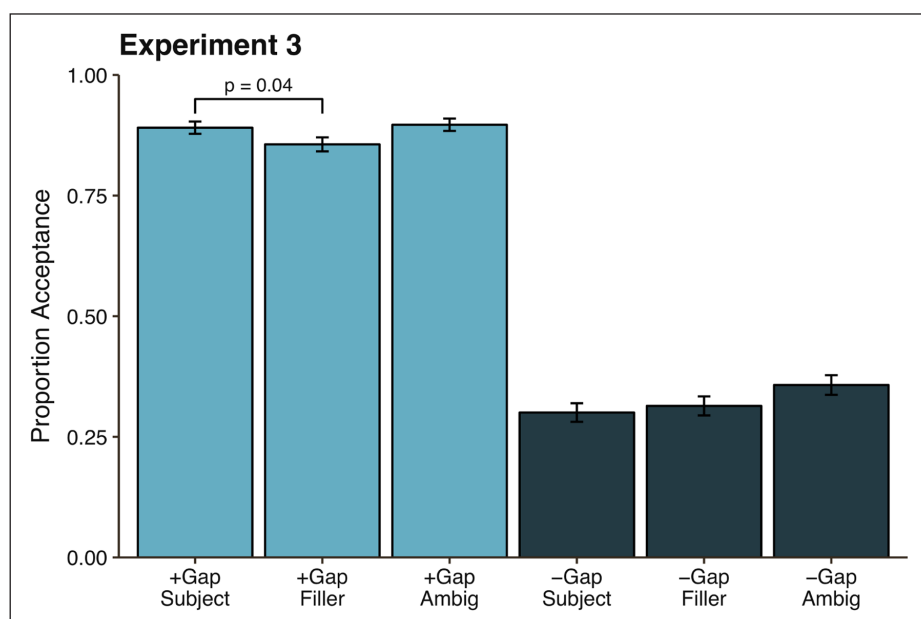

Figure 5: Mean acceptance rates by condition in Experiment 3, with error bars representing one standard error from the mean. Reported *p*-value corresponds to results from the pairwise comparisons.

Table 6: Results of logit mixed-effects model fit to the acceptability judgments for the target items with correct word list recall in Experiment 3. Starred (*) rows correspond to *p*-values below 0.05.

	β	**SE**	*z*	*p*	
Intercept	0.95	0.34	2.81	<0.01	
Gap	2.90	0.33	8.87	<0.01	*
Pronoun:1	0.24	0.33	0.74	0.46	
Pronoun:2	−0.75	0.26	−2.90	<0.01	*
Gap × Pronoun:1	−0.28	0.28	−1.02	0.31	
Gap × Pronoun:2	−0.44	0.26	−1.69	0.09	

Table 7: Results of pairwise comparisons for the model fit to the acceptability judgments for target items with correct word list recall in Experiment 3. Comparisons were made between the three levels for the Reference factor nested within each level of the ± Gap factor. Starred (*) rows correspond to *p*-values below 0.05.

		β	**SE**	*z*	*p*	
+Gap	Ambiguous – Filler	1.15	0.68	1.70	0.21	
	Ambiguous – Subject	−1.27	1.12	−1.14	0.49	
	Filler – Subject	−2.42	0.96	−2.53	0.04	*
−Gap	Ambiguous – Filler	0.84	0.53	1.57	0.26	
	Ambiguous – Subject	0.74	0.53	1.40	0.34	
	Filler – Subject	−0.11	0.56	−0.19	0.98	

dispreference for interpreting the filler NP and the pronoun as coreferential within the level +Gap ($\beta = 1.62$, SE = 0.61, z-ratio = 2.67, $p = 0.02$). Additionally, Subject was accepted more often than Filler within the level +Gap, suggesting that Filler reference is specifically dispreferred ($\beta = 1.81$, SE = 0.61, z-ratio = 2.98, $p = 0.01$). Including incorrect memory trials appears to more clearly bring out the weak cross-over effect in Experiment 3.

Before exclusion, the average d′ score across all participants and items was 1.81 ± 0.14, and after exclusion, 1.84 ± 0.13. Thus, participants were moderately successful at discriminating between grammatical and ungrammatical sentences. For the filler items, the average d′-scores across participants was 1.86 ± 0.14 before exclusion, and 1.85 ± 0.13 after exclusion. For the target items, the average d′-score was 2.46 ± 0.26 before exclusion, and 2.62 ± 0.25 after exclusion.

4.6 Discussion

Participants performed accurately on the word list recall task, while maintaining sensitivity to the distinction between grammatical and ungrammatical sentences, both in the filler items and in the target items. This demonstrated that participants were capable of maintaining the word list in memory while simultaneously processing the complex sentences. Interestingly, participants were more accurate at the word list recall task in the Subject conditions than other target items, which was unpredicted.[4] One possible explanation for this finding is that the Subject sentences were easier to process, thereby mitigating their impact on maintaining the word list in memory. In the Ambiguous and Filler conditions, coreference was possible between the filler NP and the pronoun. This may have engen-

---

[4] I thank an anonymous reviewer for pointing this out to me.

dered some processing difficulty, perhaps because participants momentarily considered the resumptive interpretation. This increased processing difficulty may have in turn resulted in degradation of the word list in memory, resulting in lower performance on the Ambiguous and Filler conditions. However, this is speculation, and this effect was not predicted.

In Experiment 1, participants overwhelmingly rejected the –Gap sentences, regardless of the pronoun's interpretation. This was likely because the materials were designed to not induce significant demand on memory resources. By contrast, it was predicted that the addition of the word list recall task would result in increased acceptability for the –Gap sentences and for resumption generally. However, there was no specific increase in judgments for the –Gap, Subject sentences in the analysis of either dataset, as I had predicted. One possibility is that dependency length is a precondition for resumption in English. Thus, although participants were placed under memory strain in Experiment 3, they did not analyze the pronoun as resumptive, due to the dependency being too short. Relatedly, length may be a precondition for resumption because decay is a significant factor to degradation of the representation in memory. In other words, the addition of the word list recall task was not sufficiently difficult enough to engender resumption. Alternatively, participants may not have considered a resumptive analysis due to relative salience of the filler NP as an antecedent. Like Experiment 1, the filler NP was not located in the main clause. Thus, the non-filler subject NP may been significantly more accessible, and therefore participants were less able to access the filler NP as an antecedent.

The account of resumption sketched in this paper also predicted that –Gap sentences may be accepted more frequently overall with the addition of the word list recall task. Qualitatively, the –Gap target item acceptance rates increased compared to Experiment 1. However, as in Experiment 2, the acceptance rates for Ungrammatical filler sentences also increased, suggesting that participants may have simply guessed more frequently in Experiment 3. I turn to this issue in Section 6.

Next, the goal of Experiment 4 was to further explore the interaction between memory strain and length. In Experiment 4, I used the same items as in Experiment 2. These items were designed to favor resumption. If the ambiguity advantage in Experiment 3 resulted from failure to maintain both possible referents, then it is predicted that a similar ambiguity advantage should surface in Experiment 4.

5 Experiment 4

As in Experiment 3, the goal of Experiment 4 was to determine whether the additional memory strain would result in higher rate of acceptance of resumptive dependencies, as reflected in higher ratings for –Gap, Subject sentences compared to –Gap, Subject sentences. In Experiment 4, I used the same materials from Experiment 2, which had proven more likely to engender a resumptive interpretation.

5.1 Participants

There were 60 participants recruited for Experiment 4 from Amazon's Mechanical Turk platform (http://www.mturk.com), using the same inclusion criteria as in Experiments 1–3. As before, all participants self-identified as native English-speakers, and the mean age of participants was 35. One participant was excluded for reliably low results on the word list recall task ($< 50\%$).

5.2 Materials

The sentences that participants were asked to judge were the same as in Experiment 2. For the word list recall task, the same three short inanimate nouns that were generated in Experiment 3 were used.

5.3 Methods

The methods for Experiment 4 were the same as in Experiment 3. Participants were compensated $2.00 for participation, and it took approximately 20 minutes to complete the task.

5.4 Results

As in Section 4.5, I first report on the word list recall results, and then the acceptability judgment results. I first conducted analysis on the raw acceptability results, and then excluding trials with incorrect memory recall.

Overall, the mean word list recall accuracy was $80.6 \pm 0.4\%$. The average d′-score across all items and participants was 2.06 ± 0.14. This suggests that there was moderately high accuracy in the recall task. For the filler items, recall accuracy was $81.0 \pm 0.7\%$ overall. For Grammatical filler items, the recall accuracy was $82.0 \pm 0.9\%$, and for Ungrammatical filler items, the recall accuracy was $80.0 \pm 0.9\%$. In d′-scores, memory recall accuracy for filler items was 2.28 ± 0.17. For target items, the mean accuracy was $80.3 \pm 0.6\%$. The mean accuracy by condition for target items is broken down in Table 8. As in Experiment 3, recall accuracy was higher in Subject target items ($\beta = 0.60$, SE $= 0.16$, $z = 3.73$, $p < 0.01$). The mean d′-score for the recall task on the target items was 2.24 ± 0.18.

The acceptability judgment data was analyzed with the same methods as in Experiments 1–3. For the filler items, Grammatical filler items were accepted more often than Ungrammatical filler items, in both the dataset including incorrect trials on the memory recall task ($\beta = 3.24$, SE $= 0.29$, $z = 11.02$, $p < 0.01$), and the dataset excluding these trials ($\beta = 3.13$, SE $= 0.28$, $z = 11.15$, $p < 0.01$). Before excluding incorrect recall trials, Grammatical filler items were accepted $88.2 \pm 0.01\%$ of the time, and the Ungrammatical filler items were accepted $32.6 \pm 0.01\%$ of the time. After exclusion of incorrect trials, the Grammatical fillers were accepted $88.6 \pm 0.8\%$ of the time, and Ungrammatical fillers were accepted $32.2 \pm 1.2\%$ of the time.

For the target items, I constructed a logit mixed-effects models for both datasets with the same structure as in Experiments 1–3. The mean acceptance rates by condition after exclusion are plotted in Figure 6. The results of the logit mixed effects model after exclusion are given in Table 9, and the results of the pairwise comparisons is given in Table 10.

As in the previous three experiments, there was a strong preference for $+$Gap target items over –Gap target items. Unlike in Experiment 3 however, there was no main effect of Pronoun. There was a significant interaction between $\pm$Gap and Pronoun, which was reflected in the increased acceptance rates for the –Gap, Filler sentences compared to the –Gap, Subject sentences. Thus, in Experiment 4, resumptive pronouns detectably facilitated the acceptability of a sentence, as in Experiment 2. Unlike Experiment 3, there was no evidence of the weak cross-over effect.

The model fit to the data that did not exclude incorrect memory recall trials had a similar pattern. There was a main effect of $\pm$Gap ($\beta = 1.32$, SE $= 0.20$, $z = 6.64$, $p < 0.01$), and an interaction effect between $\pm$Gap and Pronoun:2 ($\beta = 0.41$, SE $= 0.14$, $z = 2.94$, $p < 0.01$). Pairwise comparisons for each level of Pronoun nested within each level of $\pm$Gap did not

Table 8: Mean accuracy and standard error on word list recall task by condition for Experiment 4.

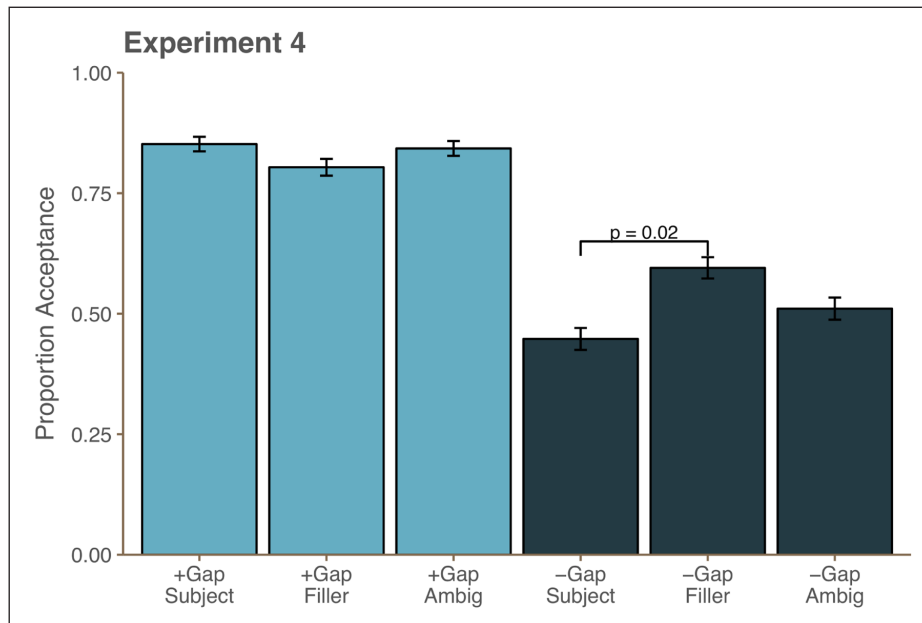|	+Gap	−Gap
Ambiguous	78.2 ± 0.02%	78.6 ± 0.02%
Filler	81.3 ± 0.02%	79.6 ± 0.02%
Subject	85.0 ± 0.01%	85.0 ± 0.01%

Figure 6: Mean acceptance rates by condition in Experiment 4, with error bars representing one standard error from the mean. Reported *p*-value corresponds to results from the pairwise comparisons.

Table 9: Results of logit mixed-effects model fit to the acceptability judgments for target items in Experiment 4. Starred (*) rows correspond to *p*-values below 0.05.

	β	SE	z	p	
Intercept	1.79	0.29	6.10	<0.01	*
Gap	1.56	0.27	5.69	<0.01	*
Pronoun:1	0.11	0.28	0.39	0.70	
Pronoun:2	0.02	0.25	0.07	0.95	
Gap × Pronoun:1	0.20	0.26	0.77	0.44	
Gap × Pronoun:2	−0.54	0.23	−2.33	0.02	*

Table 10: Results of pairwise comparisons for the model fit to acceptability data from the target items with correct word list recall from Experiment 4. Comparisons were made between the three levels for the Reference factor nested within each level of the ± Gap factor. Starred (*) rows correspond to *p*-values below 0.05.

		β	SE	z	p	
+Gap	Ambiguous − Filler	0.84	0.80	1.05	0.55	
	Ambiguous − Subject	0.10	0.84	0.12	0.99	
	Filler − Subject	−0.74	0.74	−0.99	0.58	
−Gap	Ambiguous − Filler	−0.65	0.39	−1.62	0.21	
	Ambiguous − Subject	0.37	0.45	0.82	0.69	
	Filler − Subject	1.02	0.40	2.56	0.03	*

yield any significant differences, but there was a marginal difference between the resumptive −Gap, Filler condition and the anaphoric −Gap, Subject condition ($\beta = 0.66$, SE $= 0.31$, *z*-ratio $= 2.13$, $p = 0.08$). Thus, the crucial comparison was significant in the dataset that excluded incorrect memory recall trials, but only marginally significant in the full dataset.

The average d′ across all items and participants was 1.55 ± 0.14 before exclusion incorrect recall trials, and 1.96 ± 0.16 after exclusion. The average d′-score for filler items was 2.25 ± 0.20 before exclusion, and 2.36 ± 0.22 after exclusion. The d′-scores for all target items was 1.34 ± 0.20 before exclusion, and 1.31 ± 0.22 after exclusion.

5.5 Discussion

The goal of Experiment 4 was to determine whether acceptance rates increased for sentences with a possible resumptive analysis. The general pattern of results from Experiment 2 were replicated in Experiment 4, i.e., –Gap, Filler target items were accepted more often than –Gap, Subject items. This is unsurprising, since the same materials were used across the two experiments.

Qualitatively, the results of Experiment 4 suggested a greater acceptance rates for the –Gap target items compared to Experiments 1–3. On the proposal in this paper, the length of the filler dependency and the word list recall task both strain memory resources, which results in a diminished ability to detect the ungrammatical resolution of the filler dependency. I turn to the cross-experiment comparisons in the next section to demonstrate that this is a reliable pattern across Experiments 1–4.

6 Meta-Analysis of Experiments 1–4

The proposal in this paper makes several predictions. The first prediction is that increased length between a filler NP and a pronoun should favor resumption, which has been demonstrated in previous studies (Alexopoulou & Keller 2007; Hofmeister & Norcliffe 2013). Secondly, resumption should be accepted at higher rates in contexts of high working memory demand. Finally, resumption should be more preferable in contexts in which comprehenders are less sensitive to ungrammatical filler dependency resolutions more generally. Qualitatively, these two predictions are confirmed, either in within-experiment comparisons or across-experiment comparisons. In Experiments 1 and 3, there was no detectable effect of pronoun interpretation on –Gap indicative of resumption. By contrast, in Experiments 2 and 4, there was an advantage for the resumptive –Gap, Filler target items compared to the anaphoric –Gap, Subject items. Thus, the change in materials between Experiments 1 and 3 to Experiments 2 and 4 increased the availability of resumption. These studies do not clearly demarcate the effect of referent accessibility and dependency length. However, it is likely that both contribute to the availability of resumption, and my proposal makes reference to both factors.

In this section, I demonstrate that the cross-experiment comparisons are quantitatively reliable. Moreover, I suggest that these contrasts cannot be due to overall noisier performance, corresponding to increased task difficulty. First, I conduct a meta-analysis across the acceptability ratings across the four experiments. Then, I conduct a meta-analysis of the sensitivity (d′) scores. In both cases, I show that length has a significant impact on the acceptability of resumption, and that memory load and length furthermore facilitate the acceptability of ungrammatical sentences overall.

6.1 Acceptability ratings

First, I examined the acceptance rates of the target items across Experiments 1–4. I collated the data from the four experiments, including the trials with incorrect word list recall performance in Experiment 3 and Experiment 4. I then constructed a logit mixed effects model with the acceptance rates of the target items as the dependent variable, and with ±Gap and Pronoun as within-participant factors. I also included two new between-participant factors, Memory Load and Length. Memory Load had

levels –Memory Load (Experiments 1 and 2) and +Memory Load (Experiments 3 and 4). Length had levels Short (Experiments 1 and 3) and Long (Experiments 2 and 4). I also included random slopes for ±Gap × Pronoun by participant, and random effects by item.[5] More complex models with random slopes by item did not converge. The results of mixed effects model are given in Table 11. Due to the complexity of this model, I will first discuss main effects, then significant interaction effects, then pairwise comparisons.

6.1.1 Gap and pronoun

Unsurprisingly, this meta-analysis revealed main effects of ±Gap, showing a strong preference for +Gap over –Gap. Importantly, there were also main effects of Memory Load and Length. The main effect of Memory Load signified that the addition of the word list recall task increased acceptance rates overall. Similarly, increasing the length of the filler-dependency had an overall effect of increasing acceptance rates.

Table 11: Results of logit mixed-effects model fit to the acceptability judgment data for all target items in Experiments 1-4. Starred (*) rows correspond to p-values below 0.05.

	β	SE	z	p	
Intercept	0.39	0.15	2.53	0.01	*
Gap	1.05	0.13	7.85	<0.01	*
Pronoun:1	0.08	0.10	0.83	0.40	
Pronoun:2	0.02	0.09	0.20	0.84	
Memory Load	0.98	0.17	5.63	<0.01	*
Length	−0.53	0.22	−2.41	<0.02	*
Gap × Pronoun:1	0.13	0.10	1.24	0.22	
Gap × Pronoun:2	−0.34	0.10	−3.47	<0.01	*
Gap × Memory Load	−0.21	0.17	−1.23	0.22	
Pronoun:1 × Memory Load	0.17	0.17	0.98	0.33	
Pronoun:2 × Memory Load	−0.12	0.16	−0.74	0.46	
Gap × Length	1.81	0.21	8.43	<0.01	*
Pronoun:1 × Length	0.17	0.17	0.98	0.33	
Pronoun:2 × Length	−0.11	0.16	−0.74	0.46	
Memory Load × Length	−0.01	0.27	−0.02	0.98	
Gap × Pronoun:1 × Memory Load	0.02	0.13	0.13	0.90	
Gap × Pronoun:2 × Memory Load	−0.01	0.13	−0.01	0.93	
Gap × Pronoun:1 × Length	0.12	0.17	0.72	0.47	
Gap × Pronoun:2 × Length	0.02	0.16	0.14	0.89	
Gap × Memory Load × Length	−0.71	0.26	−2.69	<0.01	*
Pronoun:1 × Memory Load × Length	0.10	0.21	0.45	0.65	
Pronoun:2 × Memory Load × Length	−0.11	0.19	−0.55	0.58	
Gap × Pronoun:1 × Memory Load × Length	−0.24	0.21	−1.17	0.24	
Gap × Pronoun:2 × Memory Load × Length	0.10	0.20	0.52	0.61	

---

[5] The structure of this model using `lme4` syntax is: Response ~ Gap * Pronoun * Memory Load * Length + (1 + Gap * Pronoun |Participant) + (1 |Item).

6.1.2 Interaction effects

There were three significant interactions. There was an interaction between $\pm$ Gap and Pronoun, one between $\pm$ Gap and Length, and a three-way interaction between $\pm$ Gap, Memory Load, and Length.

6.1.3 Pairwise comparisons

First, I conducted pairwise comparisons of the three levels in Pronoun nested within the two levels of $\pm$ Gap. This revealed the expected pattern. Within the $+$ Gap conditions, Ambiguous items were accepted more than Filler items ($\beta = 0.68$, SE $= 0.17$, z-ratio $= 3.90$, $p < 0.01$), and Subject items were also accepted more than Filler items ($\beta = 0.48$, SE $= 0.16$, z-ratio $= 2.97$, $p = 0.01$). This reflects sensitivity to the weak cross-over constraint, because it demonstrates a bias against a filler NP binding a pronoun and a gap in the same sentence. Within –Gap conditions, Filler items were accepted more than Subject items ($\beta = 0.47$, SE $= 0.13$, z-ratio $= 3.49$, $p < 0.01$), and the other comparisons were not significant ($p > 0.10$). This demonstrates that forced coreference between a filler NP and a pronoun improves the penalty of an otherwise unresolved filler dependency, consistent with a resumptive analysis. In other words, the effect of resumption on repairing an unresolved filler dependency was observed in the meta-analysis.

Next, I discuss the interaction effect between Length and $\pm$ Gap. Pairwise comparisons revealed that Long sentences were accepted more often than Short sentences within the level –Gap ($\beta = 1.98$, SE $= 0.20$, z-ratio $= 10.01$, $p < 0.01$). Conversely, Long sentences were accepted less often than Short sentences within the level $+$ Gap ($\beta = -0.92$, SE $= 0.17$, z-ratio $= -5.50$, $p < 0.01$). In other words, longer filler dependencies increased the acceptance of gapless resolutions, and lowered the acceptance of gapped resolutions. This implies that sensitivity to whether a filler dependency resolved in a grammatically-licensed way was diminished for longer dependencies, consistent with this proposal.

Next, I examined the three-way interaction between $\pm$ Gap, Length, and Memory Load in several pairwise comparisons. First, I examine the effect of Length within Memory Load and $\pm$ Gap. Within the level –Gap. Within this level, Short, $+$ Memory Load items were assigned higher ratings than Short, –Memory Load (Experiment 1 vs. Experiment 3; $\beta = 1.89$, SE $= 0.33$, z-ratio $= 5.71$, $p < 0.01$). This suggests that the increased processing of the working memory task improved the –Gap sentences, even in shorter dependencies. Similarly, within the level $+$ Gap, the addition of the memory recall task increased acceptance rates for longer dependencies, as revealed in the pairwise comparison between Long, $+$ Memory Load and Long, -Short Memory Load (Experiment 2 vs. Experiment 4; $\beta = 1.18$, SE $= 0.25$, z-ratio $= 4.77$, $p < 0.01$).

Next, I compared the effect of Memory Load nested within the two levels of Length and $\pm$ Gap. Within the level $+$ Gap, Memory Load facilitated acceptance rates for Long sentences ($\beta = 0.77$, SE $= 0.23$, z-ratio $= 3.28$, $p < 0.01$). This shows that the additional strain induced by the word list recall task had an effect on improving acceptance rates for the gapless resolutions for longer dependencies only. However, there was no effect of Memory Load on the short dependencies within the level of $+$ Gap ($p > 0.10$). This implies that memory load is likely insufficient for entirely inducing enough working memory demand to facilitate resumption. Instead, additional working memory demand, as induced by the word list recall task, may magnify the processing demand on longer filler dependencies. The effect of the word list recall task is plotted in Figure 7.

6.2 Sensitivity scores

One possible concern about the comparison between Experiments 1–4 is that the increased acceptance rates of –Gap target items may not reflect anything specific about the processing of filler dependencies. Rather, these results may reflect noisier performance induced
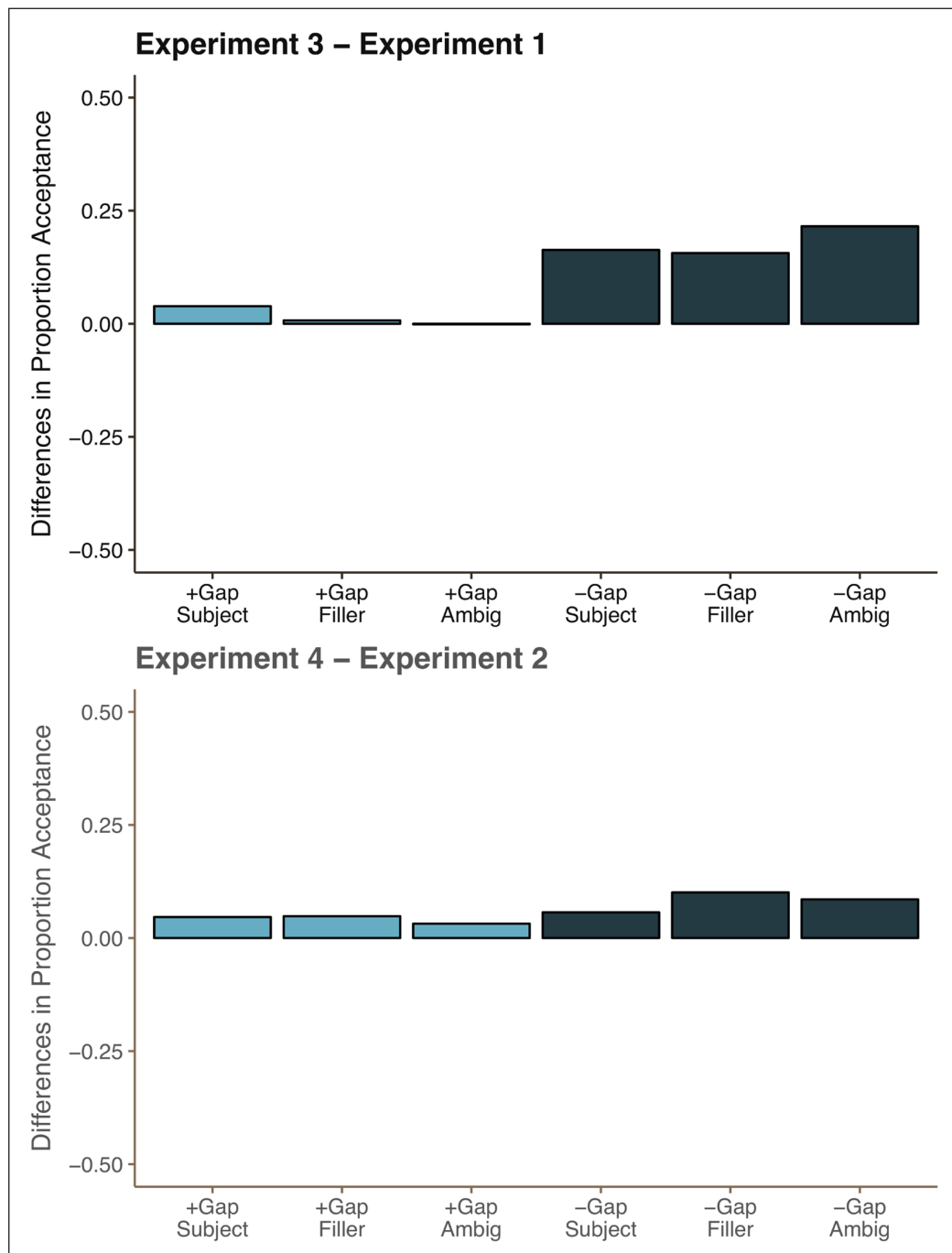
Figure 7: Effect of word list recall task on acceptance rates by condition. The y-axis corresponds to the difference between mean acceptance rates between Experiments 3 and 1 (top) and Experiments 4 and 2 (bottom).

by increased task difficulty. Increased guessing on acceptability trials should raise the acceptance rates of –Gap items, due to higher proportion of at-chance performance. The finding that +Gap, Long items were accepted less often than +Gap, Short items may also be consistent with this explanation, i.e., this may show that additional processing cost associated with longer filler dependencies may result in both –Gap and +Gap acceptance rates drawing closer to chance.

To distinguish these hypotheses, I next examine the distribution of d′-scores across participants in the four experiments. As I described in Section 2.6, the d′-score is a way of quantifying individual participants' sensitivity that takes into account individual response bias. If the cross-experiment results were due to increased task difficulty resulting in

noisier performance, then there should be a decline in d′-scores with the addition of the word list recall task and the longer filler dependencies. First, I discuss the d′-scores for target items. Then, I turn my attention to the d′-scores for filler items.

6.2.1 Target sensitivity scores

To test this, I constructed a linear mixed effects model with participants' d′-scores computed over target items as the dependent variable. I included Memory Load, Length and their interaction terms as between-participant factors, and included random effects by participant.[6] This revealed a main effect of Length, showing that participants in experiments with longer filler dependencies had decreased sensitivity to grammaticality (β = 1.85, SE = 0.30, t(225) = 6.11, p < 0.01). There was no main effect of Memory Load (β = –0.06, SE = 0.30, t(225) = –0.19, p = 85), but there was a marginal interaction effect (β = –0.74, SE = 0.43, t(225) = –1.73, p = 0.08). Pairwise comparisons revealed that d′-scores were lower in Short, + Memory Load compared to Short, –Memory Load (β = 0.79, SE = 0.31, t-ratio = 2.60, p = 0.01), but there was no difference + Memory Load and –Memory Load within the level Long (β = 0.06, SE = 0.30, t-ratio = 0.19, p = 0.85). This suggests that performance suffered with the addition of the word list recall task between Experiments 1 and 2, but not necessarily between Experiments 3 and 4.

However, this result does not distinguish between these two explanations. Increased guessing should lead to a decrease in sensitivity as the tasks become more complex across both target items and filler items. On my analysis, increased acceptance rates of –Gap target items increased the rate of false hits, which lowered d′-scores. Thus, on my analysis, it was predicted that filler items should show less reduction in sensitivity compared to the target items with the inclusion of the memory recall task, which increased task difficulty overall, and the length manipulation, which increased task difficulty for the target items only.

6.2.2 Filler sensitivity scores

To test this, I fit a linear mixed effects model to the d′-scores for the filler items with the same structure. Memory Load had no effect on the d′-scores (β = –0.01, SE = 0.22, t(216) = –0.06, p = 0.95), nor did Length (β = 0.15, SE = 0.22, t(222) = 0.68, p = 0.51). However, there was a marginal interaction effect between Memory Load and Length (β = –0.54, SE = 0.31, t(222) = –1.75, p = 0.09). Pairwise comparisons revealed that + Memory Load, Short had reduced d′-scores compared to –Memory Load, Short (β = 0.79, SE = 0.31, t(223) = 2.60, p = 0.01), but there was no difference between + Memory Load and –Memory Load within the level Long *(p > 0.10)*. The mean d′-scores by experiment and by item type (Filler vs. Target) is shown in Figure 8.

6.3 Discussion

Overall, this meta-analysis suggested that the cross-experiment manipulations of memory load and dependency length affected the acceptability of –Gap sentences. Moreover, it confirmed the general pattern for preferring resumptive dependencies over anaphoric dependencies, when no syntactically licensed resolution site is available. Secondly, I argued that comparing the sensitivity scores between target items and filler items across the four experiments revealed a more extreme profile on the target items. Although performance may have been noisier with more complex experimental paradigms, overall, the effect was magnified for the –Gap sentences in the target items. This suggests that

---

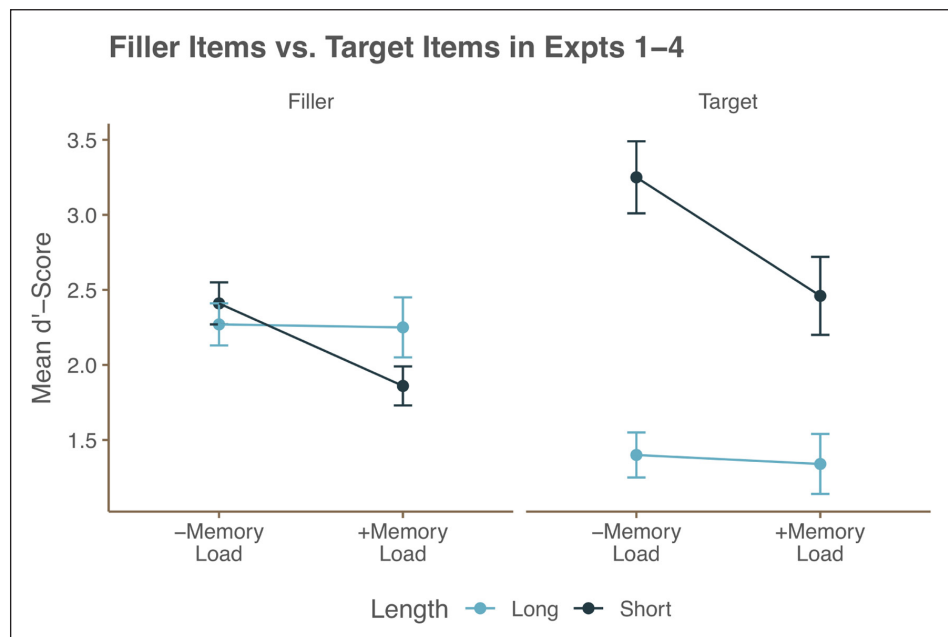[6] In `lme4` syntax: D′ ~ Memory Load * Length + (1 |Participant).

Figure 8: Comparisons of d′-scores across Experiments 1–4. Error bars correspond to one standard error from the mean. The results from Experiments 1 and 2 were coded as -Memory Load, and Experiments 3 and 4 were coded as +Memory Load. The results from Experiments 1 and 3 were coded as Short, and the results from Experiments 2 and 4 were coded as Long.

the cross-experiment manipulation of processing difficulty specifically impacted comprehenders' ability to detect an unresolved filler-gap dependency.

Finally, it is worth pointing out that there was no interaction between Pronoun and the critical factors that manipulated processing difficulty, Memory Load and Length. My account would likely be more strongly supported with such an interaction. On my account, it is expected that the facilitatory effect of co-reference between a pronoun and a filler phrase should be increased when the comprehender is placed under processing difficulty. However, this meta-analysis did not demonstrate this. This may partially be due in part to the lack of resumptive effect in Experiments 1 and 3, and the small magnitude of the effect overall. Thus, I leave investigating this more systematically for future research.

7 General discussion

An important and robust finding in sentence processing is that comprehenders actively construct filler-gap dependencies (Fodor 1978; Crain & Fodor 1985; Stowe 1986; Traxler & Pickering 1996; Kaan et al. 2000; Aoshima et al. 2004; Phillips et al. 2005; Omaki et al. 2015; Chacón et al. 2016), and that these processes appear to be suppressed in syntactic island configurations (Stowe 1986; Traxler & Pickering 1996; Phillips 2006; Wagers & Phillips 2009; Yoshida et al. 2014; Chacón 2015). These findings have been argued to reflect rapid deployment of fine-grained grammatical constraints. However, the real-time processing of resumption, as described by Hofmeister & Norcliffe (2013), appears to challenge this generalization. Given that resumption in English is ungrammatical (Kroch 1981; Chao & Sells 1983; Heestand et al. 2011; Asudeh 2012), the real-time construction of resumptive dependencies implies that comprehenders pursue syntactically unlicensed interpretations. To reconcile these findings, I argued that the acceptability of resumption depends on a reference relation constructed in contexts where typical grammatically-constrained active gap formation processes fail. Upon encountering the filler NP, the comprehender commits to memory a representation of an upcoming gapped structure. However, this representation may degrade due to processing difficulty. In these contexts,

a coreference relation between the filler NP and pronoun may facilitate recovery of the interpretation. In the four studies presented here, it was shown that increasing the length of filler dependencies and externally straining working memory through the addition of a word list recall task favored resumption. Moreover, I showed that these factors independently decreased sensitivity to the ungrammaticality of unresolved filler dependencies, which I argued partially subserves the perception of acceptability of resumptive dependencies. Thus, these data are consistent with this account.

An anonymous reviewer suggests another interpretation of these results. The target items all contained a (potential) resumptive pronoun in the subject genitive position (e.g., *her friend*). In McKee & McDaniel (2001), subject genitive resumptive pronouns were heavily dispreferred, especially compared to resumptive pronouns in other contexts. In their study, subject genitive resumptive pronouns were accepted less than 20% of the time, similar to the results in Experiment 1. McKee & McDaniel (2001) attributed this to the availability of salient, grammatical alternative with pied-piped *whose* phrases, e.g., *This butler said that this is the babysitter **whose friend** really liked kids*. If resumption in subject genitive position is dispreferred, then the results found in this paper do not necessarily generalize to all cases of resumption. Similarly, this may explain the lack of weak crossover effects observed in Experiments 1, 2, or 4. That is, if participants did not interpret the filler as binding the pronoun, then there were no weak cross-over violations also binds the gap. Relatedly, the reviewer suggested that preference for resumption was not observed for the in Experiments 1 and 3 because shorter resumptive dependencies are less likely than longer resumptive dependencies.

Importantly, evidence for the weak cross-over effect was observed in Experiment 3, and in the larger meta-analysis. Similarly, the interaction effect between $\pm$ Gap and Pronoun in Experiment 2, Experiment 4, and the meta-analysis suggested that participants were capable of analyzing the subject genitive pronoun as resumptive. Moreover, I assumed that participants typically attempted to find an antecedent for a pronoun in the same sentence, without any other context to support alternative interpretations (cf. Badecker & Straub 2002). Put differently, coreference between filler NP and the pronoun would be strongly favored within the context of the experiment, especially in the Subject conditions. This may mitigate the bias against construing subject genitive pronouns as resumptive. Similarly, although it is possible that participants were biased against understanding the pronoun as resumptive in the shorter dependencies, it must be explained why resumption is dispreferred and rare for shorter dependencies in English. I submit that theories that seek to derive the distribution of acceptable resumption from processing difficulty interfering with filler-gap dependency processing is a reasonable step forward to explaining this pattern.

Regardless, if active dependency formation processes depend on prediction, then it is still possible that comprehenders selectively generate predicted structures containing resumption. For instance, upon detecting a filler NP, participants may not generate a prediction for a structure containing a resumptive pronoun in the subject genitive position at a short distance from the filler NP, but they may generate predictions for resumption at greater distances. Importantly, this may explain the lack of increased acceptance for resumption in Experiment 3. In Experiment 3, participants were placed under memory strain with the word list recall task. However, they did not demonstrate improved tolerance for resumption, as predicted. Future research may benefit from exploration of resumptive pronouns in contexts in which they are assigned higher ratings than in subject genitive positions. For instance, the object genitive position (McKee & McDaniel 2001; Zukowski & Larsen 2004) and ECP contexts (McDaniel & Cowart 1999; Omaki & Nakao 2010) show a stronger preference for resumption. In these configurations, participants may accept

resumption even without increased dependency length or without external strain on memory resources, supporting this alternative analysis. Similarly, providing context in the experiment may affect the preference for resumption, which may provide a foothold for more systematically isolating the effect of filler NP accessibility (Erteschik-Shir 1992; Ariel 1999), which is conflated with length in these results.

7.1 What's stored in memory, and how?

The account I provide here makes some assumptions that are not shared by other work on the processing of filler dependencies. In this paper, I argued that filler-gap dependency processing depends on active maintenance of a predicted structure. Typically, this process is characterized as active maintenance of a filler NP while searching for a resolution site (Wanner & Maratsos 1978; Crain & Fodor 1985; Frazier 1987; Wagers & Phillips 2014). On this view, active gap formation is motivated by a need to discharge the dependency, so that a thematic role may be assigned to the filler NP (Aoshima et al. 2004), or to alleviate the burden of maintaining it in memory. On this generally accepted view, the eagerness to discharge the dependency results in the filled-gap effects, as discussed in Section 1, and other evidence of early interpretation, such as the plausibility mismatch effect (Traxler & Pickering 1996). Similarly, EEG results show a sustained left anterior negativity, which has been interpreted as reflecting active maintenance of the filler NP in working memory (King & Kutas 1995; Fiebach et al. 2001; 2002; Phillips et al. 2005). Finally, as discussed in section 1, previous accounts of resumption specifically attribute the availability of resumption to low accessibility of the filler NP (Ariel 1999).

There are two reasons why it's important that the predicted resolution site is maintained in working memory for my account. First, I argued that acceptability of resumption in part depends on a diminished sensitivity to the grammatical requirement that a filler binds a gap, as reflected in the variable acceptance rates for –Gap target items across Experiment 1–4. On my view, detecting the ungrammaticality of an unresolved filler dependency relies on comparing the predicted structure in memory against the bottom-up input. On the traditional view, it presumably relies on a failed prospective search for a gap. Moreover, in order for resumption to facilitate acceptability, the filler NP must be a sufficiently accessible antecedent for the pronoun (Erteschik-Shir 1992; Ariel 1999). If resumption is only licensed when the filler NP has sufficiently degraded in memory, then it is also less likely that the filler NP will be retrieved as an antecedent.

To clarify the difference, consider the contrast in (13). On my proposal, (13a) is recognized as acceptable when the comprehender is able to detect that the predicted structure matches the input, and (13b) is recognized as unacceptable when the comprehender detects that the expected gap is not available. If this prediction has sufficiently degraded, then (13a) may be perceived as a selectional violation.[7] In (13b), if the expectation for a gap has sufficiently degraded, then the sentence will be perceived as more acceptable. Then, coreference between the pronoun and the filler NP may further increase acceptance, by means of recovering a coherent interpretation. On theories that posit that the filler NP is a fragile representation maintained in memory, the explanation for identifying the grammaticality of (13a) and the ungrammaticality of (13b) is largely the same. By contrast, in (13b), if the filler NP has been forgotten, then the pronoun would be most likely interpreted as a pronoun lacking an antecedent. On the assumption that antecedent retrieval relies on a cue-based retrieval strategy that is sensitive to the accessibility of target representations in memory (cf. Lewis & Vasishth 2005), a degraded filler NP should be

---

[7] Interestingly, the meta-analysis in Section 6 demonstrated that longer dependencies overall lowered ratings for +Gap conditions, which may be interpreted as consistent with this proposal. However, it may also reflect an overall trend towards noisier responses.

difficult to retrieve as the antecedent of *it*, leaving *it* without an antecedent. Thus, in order for resumption to be perceived as acceptable while producing a coherent interpretation, it is important to distinguish the status of the filler NP in memory from the ability to detect whether a filler dependency has grammatically resolved.

(13)   a.   This is **the case** that … Dale solved __.
         b.   This is **the case** that … Dale solved it.

The existing evidence on active gap formation processes is largely equivocal on what information is stored in working memory. For instance, filled gap effects are sometimes characterized as supporting an account in which the comprehender maintains a representation of the filler NP in memory while prospectively searching for a gap. However, filled gap effects could follow from incrementally checking the predicted structure maintained in memory against the input sentence. Similarly, sustained anterior negativity results may reflect the active maintenance of the structured prediction, and not the filler NP as such.

One reason to suspect that the comprehender actively maintains features of the filler NP is the profile of "D-linked" *wh*-phrases. D-linked *wh*-phrases presume a set of alternatives available in the discourse (Pesetsky 1987). Additionally, they are argued to be more robustly represented in memory, which in turn mitigates the impact that distance and syntactic islands have on dependencies headed by these structures (Goodall 2015; see also Frazier & Clifton Jr. 2002; Diaconescu & Goodluck 2004; Hofmeister & Sag 2010). Compellingly, Alexopoulou & Keller (2007; 2013) demonstrated that D-linking and resumption interact, such that D-linking may increase the acceptability of resumption in syntactic islands. Similarly, it has been demonstrated that feature overlap between the the filler NP and subsequent linguistic material results in interference effects (Rizzi 2013; Atkinson et al. 2015; Villata et al. 2016), suggesting that these features are maintained in working memory. Thus, I take these findings to indicate that some working memory resources are devoted to maintenance of the filler NP. However, it is still possible that the comprehender also maintains an active prediction of the resolution site with some representation of the filler NP.

Another tension between my proposal and previous literature is the interpretation of length effects. On my proposal, length adversely affects maintenance of predicted structure in memory, and this predicted structure is required for determining whether a filler dependency has a syntactically licensed resolution. This predicts that length may diminish filled-gap effects. Wagers & Phillips (2014) and Chow & Zhou (2018) examined filler-gap dependencies that cross variable distances. In both sets of studies, they found evidence of active dependency formation at all dependency lengths. Wagers & Phillips (2014) found that filled-gap effects were observed over variable distances, but detection of implausible filler-verb relations was delayed until after the reactivating the filler NP for longer dependencies. By contrast, Chow & Zhou (2018) found evidence of sensitivity to implausible filler-verb relations at all distances.

On my account, it is unlikely that the representation of the gap completely degrades. For instance, if the syntactic prediction had completely degraded, then the –Gap target items should have been rated near ceiling. Conversely, the +Gap target items should have been accepted at lower rates. This is because all +Gap items contained verbs or prepositions that required objects. Thus, if the prediction for a gap had completely disappeared, then these items should have been perceived as selectional violations, e.g., an obligatorily transitive verb missing a required argument NP. However, even in Experiment 4, there was still a preference for +Gap over –Gap, implying that comprehenders still distinguished between grammatical and ungrammatical resolutions. This could reflect residual

representation of the predicted structure in memory, at least on some trials. Even so, my account predicts that active dependency formation should diminish over time, if not vanish. However, this does not seem to match the findings by Wagers & Phillips (2014) and Chow & Zhou (2018). The plausibility mismatch effect in Chow & Zhou (2018) was similar in magnitude across filler dependency lengths, implying no substantial degradation of the semantic information of the filler NP. It's crucial to point out that this tension extends beyond my proposal. If comprehenders are capable of executing active dependency formation across long distances, then the findings by Alexopoulou & Keller (2007) and Hofmeister & Norcliffe (2013) similarly need explanation. If grammatically-sensitive active dependency formation processes extend across long distances without losing sensitivity, then it is unclear why these processes entertain ungrammatical dependencies for long dependencies.

Relatedly, an anonymous reviewer points out that the findings by Wagers & Phillips (2009) may be problematic for my account. As I discussed in Section 1, they found that active dependency formation persists into the second conjunct in ATB constructions. This may be surprising on my account, since conjoined VPs are arguably complex configurations. Wagers & Phillips (2009) may have detected continued active dependency formation in the second conjunct of ATB configurations for two possible reasons. First, it may be that conjoined VPs are not particularly complex, for some relevant metric of complexity. Second, it may be that after detecting the conjunct *and*, the comprehender generated a new prediction for upcoming structure, conditioned on how the first conjunct was produced. Previous work has argued for a processing-based parallelism preference on ATB configurations (Frazier et al. 2000; Frazier & Clifton 2001; Apel et al. 2007; Sturt et al. 2010; Knoeferle 2014), that is, a preference that a gap in the subsequent conjuncts is located in the same relative position as it was in the first conjunct. This parallelism preference may be attributed to a predictive mechanism (Frazier et al. 2000). Upon detecting the conjunct *and*, the comprehender may generate a new prediction before entering the second VP. If so, then this second prediction is likely to be very active. Importantly, Parker (2017) found that active dependency formation in ATB configurations obeys parallelism, i.e., comprehenders expect gaps in the the second conjunct to be located in the same position as in the first conjunct. If active dependency formation partially depends on structured prediction, then this may be expected.

Finally, I have remained agnostic about the nature of the degradation of representations in memory and the structure of short-term memory. Classically, theories of working memory postulate privileged memory buffers for maintaining information accessible to computations (Baddeley & Hitch 1974; Baddeley 2000). Limitations on working memory follow from limited working memory resources, and storing representations in these buffers over time is costly. For instance, Wanner & Maratsos (1978) proposed that processing of filler-gap dependencies requires maintaining the filler in a specialized buffer for later retrieval at the gap. However, more recent cue-based retrieval models reject privileged memory buffers (Anderson 1983; Lewis & Vasishth 2005; Jonides et al. 2005; McElree 2006). On these accounts, recently built representations are stored in memory alongside all others. If recently built structures are more accessible, this is due to residual activation from having been recently processed in the focus of attention. However, as memory representations go unused or unactivated, they may degrade over time (*decay*). Additionally, difficulty can arise form trying to access a representation that shares properties with other active representations in memory (*interference*; see, e.g., Van Dyke & Lewis 2003).

The results of this study are likely uninformative for arbitrating between these approaches to short term memory. For instance, the effects of length could either be cast as decay of the predicted structure in memory or as interference between the predicted structure, the

linguistic material processed after the filler NP, and the word list. Similarly, the effect of the word list could be cast as interference between the word list and the material stored in working memory, or as competition for limited working memory resources.

However, if the impact of the word list recall task is ultimately due to interference, then the form of the stimuli in the list may have differential effects on acceptability. For instance, asking participants to memorize sequences of non-linguistic stimuli, such as numbers or shapes, may be less likely to improve ratings for gapless filler dependencies. If so, this may suggest that linguistic interference underlies the improved acceptability of Memory Load in these studies. Similarly, one way to directly isolate decay as a factor would be to keep the linguistic form of the sentences the same, but modulate the stimulus onset synchrony (SOA), such that there is increased time between the filler NP and the gapless predicate. If increasing the length of time improves the ratings of –Gap predicates, then this demonstrates that decay of the representation over time uniquely contributes to the loss of the predicted structure in memory. Finally, it may be possible to directly modulate the effect of interference on the prediction maintained in memory while keeping sentence length the same by selecting different island types, given an independent measure of syntactic complexity. Alternatively, interference may be increased without affecting sentence length by manipulating the degree of feature overlap between the filler NP and the subsequent linguistic material, to induce similarity-based interference.

8 Conclusion

In this paper, I provided a mechanistic account of resumption in English. I argued that resumption in English is a complex phenomenon that relies on anaphoric processing and failure to execute typical filler-gap dependency processing. Shortly upon detecting the filler, comprehenders construct a prediction for gapped structure. However, as further material is processed over time, this representation becomes more difficult to access. This results in loss of the representation, which in turn means that comprehenders are less likely to notice if the prediction goes unsatisfied. Because no filler-gap dependency is computed, a reference dependency between the pronoun and filler NP allows the comprehender to relate the filler NP to the meaning of the sentence. This allows the comprehender to build a a coherent interpretation of the ungrammatical sentence. In four studies, I compared the resumptive pronouns with anaphoric pronouns, and resumptive dependencies resolving in islands with later filler-gap dependencies. Across experiments, I manipulating the length of the dependency, and increased strain on memory with the addition of the memory recall task. I showed that increased demand on memory resources decreased sensitivity to whether a filler dependency resolves. Moreover, I showed that coreference between the filler NP and the pronoun improves ratings in a subset of these cases. Additionally, I argued that this proposal is consistent with the generalization that comprehenders deploy grammatical constraints rapidly and effectively in typical sentence-processing, because resumptive dependencies are constructed only when typical, grammatically-constrained processing fails.

Abbreviations

F = (stereotypically) feminine, M = (stereotypically) masculine

Additional File

The additional file for this article can be found as follows:

- **Supplementary file 1.** 'Materials'. DOI: https://doi.org/10.5334/gjgl.839.s1

Ethics and Consent

Acknowledgements

Competing Interests

The author has no competing interests to declare.

References

Ackerman, Lauren, Michael Frazier & Masaya Yoshida. 2018. Resumptive pronouns can ameliorate illicit island extractions. *Linguistic Inquiry* 49(4). 847–859. DOI: https://doi.org/10.1162/ling_a_00291

Alexopoulou, Theodora & Frank Keller. 2007. Locality, cyclicity, and resumption: At the Interface between the grammar and the human sentence processor. *Language* 83(1). 110–160. DOI: https://doi.org/10.1353/lan.2007.0001

Alexopoulou, Theodora & Frank Keller. 2013. What vs. who and which: Kind-denoting fillers and the complexity of *whether*-islands. In Jon Sprouse & Norbert Hornstein (eds.), *Experimental syntax and island effects*, 310–340. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139035309.005

Almeida, Diogo. 2014. Subliminal wh-islands in Brazilian Portuguese and the consequences for syntactic theory. *Revista da ABRALIN* 13. 55–93. DOI: https://doi.org/10.5380/rabl.v13i2.39611

Anderson, Jordan. 1983. *The architecture of cognition*. Cambridge: Harvard University Press.

Aoshima, Sachiko, Colin Phillips & Amy Weinberg. 2004. Processing fillergap dependencies in a head-final language. *Journal of Memory and Language* 51(1). 23–54. DOI: https://doi.org/10.1016/j.jml.2004.03.001

Aoun, Joseph, Lina Choueiri & Norbert Hornstein. 2001. Resumption, movement, and derivational economy. *Linguistic Inquiry* 32(3). 371–403. DOI: https://doi.org/10.1162/002438901750372504

Apel, Jens, Pia Knoeferle & Matthew W. Crocker. 2007. Processing parallel structure: Evidence from eye-tracking and a computational model. In Stella Vosniadou, Daniel Kayser & Athanassios Protopapas (eds.), *Proceedings of the European Cognitive Science Conference*, 125–131. Mahwah, NJ: Lawrence Erlbaum Associates.

Ariel, Mira. 1999. Cognitive universals and language conventions: The case of resumptive pronouns. *Studies in Language* 23(2). 217–269. DOI: https://doi.org/10.1075/sl.23.2.02ari

Arnold, Jennifer E. 2010. How speakers refer: The role of accessibility. *Linguistics and Language Compass* 4(4). 187–203. DOI: https://doi.org/10.1111/j.1749-818X.2010.00193.x

Arnold, Jennifer E., Janet G. Eisenbrand, Sarah Brown-Schmidt & John C. Trueswell. 2000. The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition* 76(1). B13–B26. DOI: https://doi.org/10.1016/S0010-0277(00)00073-1

Arnold, Jennifer E. & Shin-Yi C. Lao. 2015. Effects of psychological attention on pronoun comprehension. *Language, Cognition and Neuroscience* 30(7). 832–852. DOI: https://doi.org/10.1080/23273798.2015.1017511

Asudeh, Ash. 2012. *The logic of pronominal resumption*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199206421.001.0001

Atkinson, Emily, Aaron Apple, Kyle Rawlins & Akira Omaki. 2015. Similarity of wh-phrases and acceptability variation in wh-islands. *Frontiers in Psychology* 6(2048). DOI: https://doi.org/10.3389/fpsyg.2015.02048

Baddeley, Alan. 2000. The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences* 4(11). 417–423. DOI: https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, Alan D. & Graham Hitch. 1974. Working memory. *Psychology of Learning and Motivation* 8. 47–89. DOI: https://doi.org/10.1016/S0079-7421(08)60452-1

Badecker, William & Kathleen Straub. 2002. The processing role of structural constraints on interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28. 748–768. DOI: https://doi.org/10.1037//0278-7393.28.4.748

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. DOI: https://doi.org/10.1016/j.jml.2012.11.001

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Beltrama, Andrea & Ming Xiang. 2016. Unacceptable but comprehensible: The facilitation effect of resumptive pronouns. *Glossa: a journal of general linguistics* 1(1). 29. DOI: https://doi.org/10.5334/gjgl.24

Boeckx, Cedric. 2003. *Islands and chains: Resumption as stranding*. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/la.63

Chacón, Dustin Alfonso. 2015. *Comparative psychosyntax*. College Park, MD: University of Maryland dissertation. DOI: https://doi.org/10.13016/M29D2B.

Chacón, Dustin Alfonso, Mashrur Imtiaz, Shirsho Dasgupta, Sikder Monoare Murshed, Mina Dan & Colin Phillips. 2016. Locality and word order in active dependency formation in Bangla. *Frontiers in Psychology* 7(AUG). DOI: https://doi.org/10.3389/fpsyg.2016.01235

Chao, Wynn & Peter Sells. 1983. On the interpretation of resumptive pronouns. In Peter Sells & Charles Jones (eds.), *North eastern linguistics society 13*, 47–61. Amherst, MA: Graduate Linguistics Student Association.

Chomsky, Noam. 1981. *Lectures on government and binding: The Pisa lectures*. Dordrecht: Foris.

Chow, Wing-Yee & Yangzi Zhou. 2018. Eye-tracking evidence for active gapfilling regardless of dependency length. *Quarterly Journal of Experimental Psychology*. DOI: https://doi.org/10.1177/1747021818804988

Crain, Stephen & Janet Dean Fodor. 1985. How can grammars help parsers? In David R. Dowty, Lauri Karttunen & Arnold M. Zwicky (eds.), *Natural language parsing*, 94–128. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511597855.004

Creswell, Cassandre. 2002. Resumptive pronouns, wh-island violations, and sentence production. *Proceedings of the Sixth International Workshop on Tree Adjoining Grammar and Related Frameworks* 6. 101–109.

Diaconescu, Rodica Constanta & Helen Goodluck. 2004. The pronoun attraction effect for d(iscourse)-linked phrases: Evidence from speakers of a null subject language. *Journal of Psycholinguistic Research* 33(4). 303–319. DOI: https://doi.org/10.1023/B:JOPR.0000035103.19149.dc

Drummond, Alex. 2018. Ibexfarm. www.spellout.net/ibexfarm.

Erteschik-Shir, Nomi. 1992. Resumptive pronouns in islands. In Helen Goodluck & Michael Rochemont (eds.), *Island constraints: Theory, acquisition, and processing*, 89–108. Dordrecht: Kluer. DOI: https://doi.org/10.1007/978-94-017-1980-3_4

Farby, Shira, Gabi Danon, Joel Walters & Michal Ben Shachar. 2010. The acceptability of resumptive pronouns in Hebrew. In Yehuda Falk (ed.), *Proceedings of Israel Association for Theoretical Linguistics 26*. http://linguistics.huji.ac.il/IATL/26/Farby_et_al.pdf.

Ferreira, Fernanda & Benjamin Swets. 2005. The production and comprehension of resumptive pronouns in relative clause "island" contexts. In Anne Cutler (ed.), *Twenty-first century psycholinguistics: Four cornerstones*, 263–278. Mahwah, NJ: Lawrence Erlbaum.

Fiebach, Christian J., Matthias Schlesewsky & Angela D. Friederici. 2001. Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI. *Journal of Psycholinguistic Research* 30(3). 321–338. DOI: https://doi.org/10.1023/A:1010447102554

Fiebach, Christian J., Matthias Schlesewsky & Angela D. Friederici. 2002. Separating syntactic memory costs and syntactic integration costs during parsing: The processing of german wh-questions. *Journal of Memory and Language* 47(2). 250–272. DOI: https://doi.org/10.1016/S0749-596X(02)00004-9

Fodor, Janet Dean. 1978. Parsing strategies and constraints on transformations. *Linquistic Inquiry* 9(3). 427–473.

Frazier, Lyn. 1985. Syntactic complexity. In David R. Dowty, Lauri Karttunen & Arnold M. Zwicky (eds.), *Natural language parsing*, 129–189. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511597855.004

Frazier, Lyn. 1987. Processing Dutch sentence structures. *Journal of Psycholinguistic Research* 22. 85–108. DOI: https://doi.org/10.1007/BF01067826

Frazier, Lyn, Alan Munn & Charles Clifton. 2000. Processing coordinate structures. *Journal of Psycholinguistic Research* 29(4). 343–370. DOI: https://doi.org/10.1023/A:1005156427600

Frazier, Lyn & Charles Clifton. 2001. Parsing coordinates and ellipsis: Copy alpha. *Syntax* 4(1). 1–22. DOI: https://doi.org/10.1111/1467-9612.00034

Frazier, Lyn & Charles Clifton Jr. 2002. Processing 'd-linked' phrases. *Journal of Psycholinguistic Research* 31. 633–660. DOI: https://doi.org/10.1023/A:1021269122049

Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In Alec P. Marantz, Yasushi Miyashita & Wayne O'Neil (eds.), *Image, language, brain: Papers from the first mind articulation project symposium*, 95–126. Cambridge, MA: MIT Press.

Gibson, Edward T. & James Thomas. 1999. Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes* 14(3). 225–248. DOI: https://doi.org/10.1080/016909699386293

Goodall, Grant. 2015. The d-linking effect on extraction from islands and non-islands. *Frontiers in Psychology* 5. 1493. https://www.frontiersin.org/article/10.3389/fpsyg.2014.01493. DOI: https://doi.org/10.3389/fpsyg.2014.01493

Gordon, Peter C., Randall Hendrick, Marcus Johnson & Yoonhyoung Lee. 2006. Similarity-based interference during language comprehension: Evidence from eye tracking

during reading. *Journal of Experimental Psychology: Learning Memory and Cognition* 32(6). 1304–1321. DOI: https://doi.org/10.1037/0278-7393.32.6.1304

Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein. 1995. Centering: Aframework for modeling the local coherence of discourse. *Computational Linguistics* 21(2). 203–225. DOI: https://doi.org/10.21236/ADA324949

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307. DOI: https://doi.org/10.2307/416535

Han, Chung-hye, Noureddine Elouazizi, Christina Galeano, Emrah Görgülü, Nancy Hedberg, Jennifer Hinnell, Meghan Jeffrey, Kyeong-Min Kim & Susannah Kirby. 2012. Processing strategies and resumptive pronouns in English. In Nathan Arnett & Ryan Bennett (eds.), *Proceedings of the 30th west coast conference on formal linguistics*. 153–161. Somerville, MA: Cascadilla Proceedings Project.

Hawkins, John A. 1999. Processing complexity and filler-gap dependencies across grammars. *Language* 75(2). 244–285. DOI https://doi.org/10.2307/417261

Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199252695.001.0001

Heestand, Dustin, Ming Xiang & Maria Polinsky. 2011. Resumption still does not rescue islands. *Linguistic Inquiry* 42(1). 138–152. DOI: https://doi.org/10.1162/LING_a_00032

Hofmeister, Philip & Elisabeth Norcliffe. 2013. Does resumption facilitate sentence comprehension? In Philip Hofmeister & Elisabeth Norcliffe (eds.), *The core and the periphery: Data-driven perspectives on syntax inspired by Ivan A. Sag*, 225–246. Stanford, CA: CSLI Publications.

Hofmeister, Philip & Ivan A. Sag. 2010. Cognitive constrainst and island effects. *Language* 86(2). 366–415. DOI: https://doi.org/10.1353/lan.0.0223

Jäger, Lena A., Felix Engelmann & Shravan Vasishth. 2017. Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language* 94. 316–339. DOI: https://doi.org/10.1016/j.jml.2017.01.004

Jonides, John, Steven C. Lacey & Derek Evan Nee. 2005. Processes of working memory in mind and brain. *Current Directions in Psychological Science* 14(1). 2–5. DOI: https://doi.org/10.1111/j.0963-7214.2005.00323.x

Kaan, Edith, Anthony Harris, Edward Gibson & Phillip Holcomb. 2000. The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes* 15(2). 159–201. DOI: https://doi.org/10.1080/016909600386084

Keenan, Edward L. & Bernard Comrie. 1977. Noun phrase accessibility hierarchy and Universal Grammar. *Linguistic Inquiry* 8(1). 63–99.

Keffala, Bethany. 2013. Resumption and gaps in English relative clauses: Relative acceptability creates an illusion of 'saving'. In Chundra Catchart, I-Hsuan Chen, Greg Finley, Shinae Kang, Clare S. Sandy & Elise Stickles (eds.), *Proceedings of the 37th annual meeting of the Berkeley Linguistics Society*, 140–154. Berkeley, CA: Berkeley Linguistics Society. DOI: https://doi.org/10.3765/bls.v37i1.846

Kempen, Gerard & Theo Vosse. 1989. Incremental syntactic tree formation in human sentence processing: A cognitive architecture based on activation decay and simulated annealing. *Connection Science* 1(3). 273–290. DOI: https://doi.org/10.1080/09540098908915642

Kennison, Shelia M. & Jessie L. Trofe. 2003. Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research* 32(3). 355–378. DOI: https://doi.org/10.1023/A:1023599719948

Keshev, Maayan & Aya Meltzer-Asscher. 2017. Active dependency formation in islands: How grammatical resumption affects sentence processing. *Language* 93(3). 549–568. DOI: https://doi.org/10.1353/lan.2017.0036

King, Jonathan W. & Marta Kutas. 1995. Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience* 7(3). 376–395. DOI: https://doi.org/10.1162/jocn.1995.7.3.376

Kluender, Robert. 1998. On the distinction between strong and weak islands: A processing perspective. In P. Culicover & L. McNally (eds.), *Syntax and Semantics 29: The limits of syntax*, 241–279. San Diego, CA: Academic Press.

Kluender, Robert. 2004. Are subject islands subject to a processing account? In Vineeta Chand, Ann Kelleher, Angelo J. Rodriguez & Benjamin Schmeiser (eds.), *Proceedings of the 23rd West Coast Conference on Formal Linguistics*, 475–499.

Kluender, Robert & Marta Kutas. 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8(4). 573–633. DOI: https://doi.org/10.1080/01690969308407588

Knoeferle, Pia. 2014. Conjunction meaning can modulate parallelism facilitation: Eye-tracking evidence from German clausal coordination. *Journal of Memory and Language* 75. 140–158. DOI: https://doi.org/10.1016/j.jml.2014.05.002

Kroch, Anthony. 1981. On the role of resumptive pronouns in amnestying island constraint violations. In *Proceedings of the 17th annual meeting of the chicago linguistics society*, 125–135. Somerville, MA: Cascadilla Press.

Kush, Dave, Jeffrey Lidz & Colin Phillips. 2017. Looking forwards and backwards: The real-time processing of strong and weak crossover. *Glossa: a journal of general linguistics* 2(1). 70. DOI: https://doi.org/10.5334/gjgl.280

Lewis, Richard L. 1996. Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research* 25(1). 93–115. DOI: https://doi.org/10.1007/BF01708421

Lewis, Richard L. & Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29(3). 375–419. DOI: https://doi.org/10.1207/s15516709cog0000_25

Lewis, Richard L., Shravan Vasishth & Julie A. Van Dyke. 2006. Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences* 10(10). 447–454. DOI: https://doi.org/10.1016/j.tics.2006.08.007

Macmillan, Neil A. & C. Douglas Creelman. 2005. *Detection theory: A user's guide*. Mahwah, NJ: Lawrence Erlbaum.

McCloskey, James. 2002. Resumption, successive cyclicity, and the locality of operations. In David Epstein & T. Daniel Seely (eds.), *Derivation and explanation in the minimalist program*, 184–226. Malden, MA: Blackwell. DOI: https://doi.org/10.1002/9780470755662.ch9

McCloskey, James. 2006. Resumption. In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell companion to syntax*, 94–117. Malden, MA: Blackwell Publishing. DOI: https://doi.org/10.1002/9780470996591.ch55

McCloskey, James. 2017. New thoughts on old questions – resumption in irish. In Jason Ostrove, Ruth Kramer & Joseph Sabbagh (eds.), *Asking the right questions: Essays in honor of Sandra Chung*, 81–102. Santa Cruz, CA: UC Santa Cruz Previously Published Works.

McDaniel, Dana & Wayne Cowart. 1999. Experimental evidence for a Minimalist account of English resumptive pronouns. *Cognition* 70(2). 15–24. DOI: https://doi.org/10.1016/S0010-0277(99)00006-2

McElree, Brian. 2006. Accessing recent events. *Psychology of Learning and Motivation – Advances in Research and Theory* 46(06). 155–200. DOI: https://doi.org/10.1016/S0079-7421(06)46005-9

McElree, Brian, Stephani Foraker & Lisbeth Dyer. 2003. Memory structures that subserve sentence comprehension. *Journal of Memory and Language* 48. 67–91. DOI: https://doi.org/10.1016/S0749-596X(02)00515-6

McKee, Cecile & Dana McDaniel. 2001. Resumptive pronouns in English relative clauses. *Language Acquisition* 9(2). 113–156. DOI: https://doi.org/10.1207/S15327817LA0902_01

Morgan, Adam Milton & Matthew W. Wagers. 2018. English resumptive pronouns are more common where gaps are less acceptable. *Linguistic Inquiry* 49(4). 861–876. DOI: https://doi.org/10.1162/ling_a_00293

Omaki, Akira & Chizuru Nakao. 2010. Does English resumption really help to repair island violations? *Snippets* 21(4). 11–12. DOI: https://doi.org/10.1023/A:1023599719948

Omaki, Akira, Ellen Lau, Imogen Davidson White, Myles Louis Dakan, Aaron Apple & Colin Phillips. 2015. Hyper-active gap filling. *Frontiers in Psychology* 6(MAR). 1–18. DOI: https://doi.org/10.3389/fpsyg.2015.00384

Pablos, Leticia. 2008. Filled-gap effects in sentence processing: Different accounts compared. *International Journal of Basque Linguistics and Philology* 42(1). 309–332.

Parker, Dan. 2017. Processing multiple gap dependencies: Forewarned is forearmed. *Journal of Memory and Language* 97. 175–186. DOI: https://doi.org/10.1016/j.jml.2017.08.003

Pesetsky, David. 1987. Wh-in-situ: Movement and unselective binding. In Eric Reuland & Alice ter Meulen (eds.), *The representation of (in)definiteness*, 98–129. Cambridge, MA: MIT Press.

Phillips, Colin. 2006. The real-time status of island phenomena. *Language* 82(4). 795–823. DOI: https://doi.org/10.1353/lan.2006.0217

Phillips, Colin. 2013. On the nature of Island constraints I: Language processing and reductionist accounts. In Jon Sprouse & Norbert Hornstein (eds.), *Experimental syntax and island effects*, 64–108. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139035309.005

Phillips, Colin, Nina Kazanina & Shani H. Abada. 2005. ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research* 22(3). 407–428. DOI: https://doi.org/10.1016/j.cogbrainres.2004.09.012

Postal, Paul. 1971. *Crossover phenomena*. New York: Holt, Rinehart, & Wilson.

Prince, Ellen. 1990. Syntax and discourse: A look at resumptive pronouns. In David J. Costa (ed.), *Proceedings of the sixteenth annual meeting of the Berkeley Linguistics Society*, 482–497. DOI: https://doi.org/10.3765/bls.v16i0.1719

R Development Core Team. 2008. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org. ISBN 3-900051-07-0.

Rizzi, Luigi. 2013. Locality. *Lingua* 130. 169–186. DOI: https://doi.org/10.1016/j.lingua.2012.12.002

Ross, John Robert. 1967. *Constraints on variables in syntax*. Cambridge, MA: MIT dissertation. DOI: https://doi.org/10.1023/A:1011298526791

Runner, Jeffrey T., Rachel S. Sussman & Michael K. Tanenhaus. 2006. Processing reflexives and pronouns in picture noun phrases. *Cognitive science* 30(2). 193–241. DOI: https://doi.org/10.1207/s15516709cog0000_58

Shlonsky, Ur. 1992. Resumptive pronouns as a last resort. *Linguistic Inquiry* 23(3). 443–468.

Sprouse, Jon, Matt Wagers & Colin Phillips. 2012. A test of the relation between working-memory capacity and syntactic island effects. *Language* 88(1). 82–123. DOI: https://doi.org/10.1353/lan.2012.0004

Stowe, Laurie. 1986. Parsing WH-Constructions: Evidence for on-line gap location. *Language and Cognitive Processes* 1(3). 227–245. DOI: https://doi.org/10.1080/01690968608407062

Sturt, Patrick, Frank Keller & Amit Dubey. 2010. Syntactic priming in comprehension: Parallelism effects with and without coordination. *Journal of Memory and Language* 62(4). 333–351. DOI: https://doi.org/10.1016/j.jml.2010.01.001

Traxler, Matthew J. & Martin J. Pickering. 1996. Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language* 35(3). 454–475. DOI: https://doi.org/10.1006/jmla.1996.0025

Van Dyke, Julie A. & Richard L. Lewis. 2003. Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language* 49(3). 285–316. DOI: https://doi.org/10.1016/S0749-596X(03)00081-0

Vasisth, Shravan, Katja Suckow, Richard L. Lewis & Sabine Kern. 2010. Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures. *Language and Cognitive Processes* 25(4). 533–567. DOI: https://doi.org/10.1080/01690960903310587

Villata, Sandra, Luigi Rizzi & Julie Franck. 2016. Intervention effects and relativized minimality: New experimental evidence from graded judgments. *Lingua* 179. 76–96. DOI: https://doi.org/10.1016/j.lingua.2016.03.004

Wagers, Matthew W. & Colin Phillips. 2009. Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics* 45(2). 395–433. DOI: https://doi.org/10.1017/S0022226709005726

Wagers, Matthew W. & Colin Phillips. 2014. Going the distance: Memory and control processes in active dependency construction. *Quarterly Journal of Experimental Psychology* 67(7). 1274–1304. DOI: https://doi.org/10.1080/17470218.2013.858363

Wagers, Matthew W., Ellen F. Lau & Colin Phillips. 2009. Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language* 61(2). 206–237. DOI: https://doi.org/10.1016/j.jml.2009.04.002

Wanner, Eric & Michael Maratsos. 1978. An ATN Approach to Comprehension. In Morris Halle, Joan Bresnan & George A. Miller (eds.), *Linguistic theory and psychological reality*, 119–161. Cambridge, MA: MIT Press.

Wasow, Thomas. 1972. *Anaphoric Relations in English*. MIT dissertation.

Yoshida, Masaya, Nina Kazanina, Leticia Pablos & Patrick Sturt. 2014. On the origin of islands. *Language, Cognition and Neuroscience* 29(7). 761–770. DOI: https://doi.org/10.1080/01690965.2013.788196

Zukowski, Andrea & Jaiva Larsen. 2004. The production of sentences that we fill their gaps. *Poster presented at 29th Annual CUNY Conference on Human Sentence Processing.*