

Long-term effects of repeated exposure to Subject Island constructions: Evidence for syntactic adaptation

Rui P. Chaves, Department of Linguistics, University at Buffalo, US, rchaves@buffalo.edu

Elaine J. Francis, Department of English, Purdue University, US, ejfranci@purdue.edu

Repeated exposure to Subject Island violations can lead to increased acceptability ratings and faster reading times (Chaves & Dery, 2014, 2019; Clausen, 2011; Francom, 2009; Hiramatsu, 2000; Lu et al. 2021; Lu et al., 2022). However, it remains unclear what the nature of this effect and the driving mechanism is. The present paper describes a longitudinal investigation to test whether the effect of repeated exposure to Subject Island constructions is short-lived or whether it can spread over three weeks' time, as measured by offline measures (Likert acceptability ratings) and online measures (self-paced reading). Using more observations and more sensitive methodologies, our work builds and improves on the only previous longitudinal study on such islands, Snyder (2022). We uncover evidence suggestive of gradual and strategic (by-construction and by-region) adaptation to Subject Island violations, indicated by faster response times, as well as higher acceptability ratings following repeated exposure, most consistent with a syntactic adaptation effect over and above task adaptation.



1. Introduction

Ross (1967) identified several syntactic environments which resist extraction, known ever since as *syntactic islands*. In this paper we focus on a particular kind of island, dubbed Subject Island, which blocks extraction from subject phrases. We illustrate the constraint with (1). Whereas fronting the *wh*-phrase from the object phrase in (1a) is acceptable, the extraction from within the subject phrase in (1b) is not (Chomsky, 1977, p. 106).

- (1) a. Who did you hear [stories about _]?
(Chomsky, 1973, (86b))
- b. *Who did [stories about _] terrify John?
(Chomsky, 1973, (92b))

Following the early proposals by Chomsky (1973, 1977), numerous analyses have been proposed to subsume the Subject Island under more general syntactic principles. For example, one class of approaches attempts to unify constraints on extraction from subject phrases (Subject Islands) with constraints on extraction from adjunct phrases (Adjunct Islands) (Bošković, 2016; Huang, 1982), while another class of approaches (known as *freezing*) posits that obligatory fronting (A-movement) makes subject phrases opaque to any further extraction of their sub-components (Culicover & Wexler, 1977; Gallego & Uriagareka, 2007). Although different from each other, syntactic accounts share the idea that extraction from English subject phrases is impossible because of some fundamental constraint on movement. Thus, traditionally, the Subject Island has been classified as a ‘strong island’ (Szabolcsi, 2006). However, in recent years, it has become clear that the empirical facts are too nuanced to be fully captured by a categorical constraint. For example, Chomsky (2008, p. 147) notes relatively acceptable cases, as in (2a). He suggests that perhaps the relevant constraint targets the external argument (deep subject), which is not the surface subject in the case of passive and unaccusative predicates.¹ In addition, he notes that (2b) is also relatively acceptable, suggesting that even the subject of a transitive clause may license extraction in some cases when it is assigned a non-agentive theta role. However, he does not elaborate on this observation. Since this time, various authors have noted relatively acceptable instances of Subject Islands in interrogative clauses with various types of predicates (2c–f).

¹ Polinsky et al. (2013) tested Chomsky’s (2008) conjecture experimentally, using English Subject Island sentences with transitive, unergative, and unaccusative predicates. In an acceptability judgment task, they found a penalty for extraction from the subject phrase, but no advantage for unaccusative predicates. They did, however, find a reading time advantage for unaccusative predicates at the region following the main verb in a separate self-paced reading task. They propose a constraint interaction analysis similar to that of Haegeman et al. (2014), in which different constraints target the external argument and the surface subject in English. They further propose that the constraints operate differently in Russian, where a clear advantage for unaccusative predicates is shown. See also Omaki et al. (2020) for experimental evidence about the lack of Subject Island effects in Japanese, first noted in Ross (1967).

- (2)
- a. Of which car was [the driver _] awarded a prize?
(Chomsky, 2008, p. 147)
 - b. Of which books did [the authors _] receive the prize?
(Chomsky, 2008, p. 160, n. 39)
 - c. What did [the attempt to find _] end in failure?
(Hofmeister & Sag, 2010, p. 370)
 - d. Which problem will [a solution to _] never be found?
(Chaves & Dery, 2014)
 - e. Which car did [some pictures of _] cause a scandal?
(Jiménez-Fernández, 2009, p. 111)
 - f. Which President would [the impeachment of _] not shock most people?
(Chaves & Putnam, 2021, p. 80)

Following from these and other observations of gradient acceptability, a growing body of research suggests that subject phrases do not uniformly behave as strong islands, and, as a result, there is no categorical ban on extraction from subject phrases. Some authors account for such variation in terms of two or more interacting syntactic constraints (Haegeman et al., 2014; McInnerney & Sugimoto, 2022; Polinsky et al., 2013). For example, Haegeman et al. (2014) propose a set of independent constraints which apply cumulatively, such that the least acceptable instances of Subject Islands involve the simultaneous violation of multiple constraints. They suggest that some of the relevant constraints may be rooted in pragmatic or processing factors, while others are purely syntactic (2014, p. 135). Importantly, they assume that each of the relevant constraints is of the ‘soft’ variety in English (i.e. each has a low weighting), and that constraint weightings vary both across constraints and across languages (Haegeman et al., 2014, p. 133). This analysis follows the framework of Linear Optimality Theory, which has been used to account for other gradient phenomena, such as auxiliary selection (Keller & Sorace, 2003). Although Haegeman et al. (2014) reject the idea of a single categorical constraint, their proposal adopts many insights from mainstream Minimalist approaches to island constraints.

Taking a somewhat different approach to gradience, other authors propose that extraction from a subject phrase is fully grammatical in English and not subject to any syntactic constraints of either the soft or hard variety. In this view, unacceptability often arises due to syntax-external factors, such as clashes in discourse information structure (Abeillé et al., 2020; Chaves, 2013; Cuneo & Goldberg, 2023; Liu et al., 2022) or processing complexity (Culicover & Winkler, 2022; Kluender, 2004). In our view, the most compelling evidence for a syntactic-external approach comes not from interrogatives, as in (2), but rather from relative clauses. First, attested examples suggest that at least some cases of extraction from relative clause subjects are licit. Culicover and Winkler (2022) find many such examples in their corpus sample, including those in (3a–b).

- (3) a. There are some things which [fighting against _] is not worth the effort.
 b. I'm looking for someone who I click with. You know, the type of person who [spending time with _] is effortless.
 (Culicover & Winkler, 2022)

Similarly, Abeillé et al. (2020) demonstrated experimentally that Subject Island violations were more acceptable relative to control sentences in relative clauses vs. interrogative clauses. Specifically, they showed that extracting a PP (preposition phrase) from a relative clause subject, as in (4a), was more acceptable than extracting a PP from a (non-island) relative clause object, as in (4b). This was the opposite of the pattern they found for similar PP extractions from the subjects and objects of interrogative clauses, as in (4c–d).

- (4) a. The dealer sold a sports car, of which the color _ delighted the baseball player because of its surprising luminance.
 b. The dealer sold a sports car, of which the baseball player loved the color _ because of its surprising luminance.
 c. Of which sportscar did the color _ delight the baseball player because of its surprising luminance?
 d. Of which sports car did the baseball player love the color _ because of its surprising luminance?

Abeillé et al. (2020) argue that the relevant syntactic structures are the same for both clause types, but the discourse functions differ in a way that permits PP extraction from a subject more readily in the case of relative clauses.² These findings are difficult to explain in a purely syntactic account, as there is no structural reason for why fronting should behave differently in interrogative and relative clauses.

An additional manner in which Subject Islands manifest variability and gradience is that they have been shown to ameliorate with repeated exposure over the course of an experiment session (Chaves & Dery, 2014, 2019; Clausen, 2011; Francom, 2009; Hiramatsu, 2000; Lu et al., 2021; Lu et al., 2022).³ Such improvements due to repeated exposure exemplify a general phenomenon

² Specifically, Abeillé et al. (2020) argue that the extracted phrase is a discourse focus in *wh*-questions, but not in relative clauses. This status as a focused constituent conflicts with the ordinary discourse status of a subject phrase as the discourse topic.

³ Two studies find only a near-significant effect (Snyder, 2000, 2022), and two others found no effect at all (Crawford, 2012; Sprouse, 2009). The reasons for these null effects remain unclear. Possible reasons might include: (1) the number of exposures (e.g. Hiramatsu (2000) found no effect with 4 exposures but found it with 7); (2) the complexity of the items (e.g. the items in Sprouse (2009) contained an extra clausal embedding); (3) the statistical methodology. With respect to the latter, Lu et al. (2024) conducted a meta-analysis that included data from several published studies. Using a common statistical method across all studies, they found significant amelioration in Subject Islands for the experiments in Sprouse (2009), even though the original study reported null effects. To maximize the chances of success, the current study was modeled on a previous study that found robust effects (Chaves & Dery, 2019); see Chaves and Putnam (2021, pp. 211–215) for a replication.

usually referred to as *syntactic satiation* (Snyder, 1994; Stromswold, 1986). For example, Chaves and Dery (2019) found that for sentence pairs like those in (5a–b), which were matched for plausibility and propositional content, the acceptability of Subject Island items like (5b) started out significantly lower than that of object extractions like (5a), but improved to reach the same level of acceptability by the end of the experiment session. Crucially, the filler items that were ungrammatical due to factors unrelated to islands did not exhibit this steep satiation profile.

- (5) a. Which committee does [the report of the experts] supposedly contradict the recommendations of _?
 b. *Which committee does [the report of _] supposedly contradict the recommendations of the experts?

In line with recent syntax-external accounts (Abeillé et al., 2020; Chaves, 2013; Culicover & Winkler, 2022; Liu et al., 2022), and following from the specific proposals of Chaves (2013), Chaves and Dery (2019) propose that Subject Island structures are syntactically well-formed but pragmatically anomalous, and therefore very rare and unexpected. On their account, satiation occurs when a possible but unexpected structure comes to be more expected. We note that their results are also compatible with accounts of gradience which assume that Subject Islands, as in (5b), violate one or more syntactic constraints of the ‘soft’ variety, as in the cumulative constraint violation approach of Haegeman et al. (2014). We will return to this issue when we discuss our own similar results for the current study.

Regardless of which view of Subject Islands one takes, the precise mechanisms accounting for gradient judgments and satiation effects remain unclear. One possibility is that satiation is so-called *syntactic adaptation* – a construction-specific form of implicit learning, consistent with both error-based learning models (Chang et al., 2006, 2012) and Bayesian belief-updating models (Fine & Jaeger, 2013; Lu et al., 2021).⁴ This type of learning involves a re-weighting of expectations about the likelihood of encountering a particular construction that is possible but relatively infrequent in usage. Another possibility is that it is caused by *task adaptation*, i.e. a decrease in reading times caused by mere familiarity with the experimental paradigm (Heathcote et al., 2000). Indeed, Harrington Stack et al. (2018), Prasad and Linzen (2019, 2021), and Dempsey et al. (2020) argue that the reduction in reading time due to syntactic adaptation is confounded with this more general adaptive phenomenon. Task adaptation does not directly depend on the

⁴ We will use the term *syntactic adaptation*, even though we are open to the possibility that other linguistic factors are also being attended to in such adaptation effects, including morphology, semantics, pragmatics, and intonation – properties which may be conventionally associated with a particular syntactic form in a constructionist view of grammar (e.g. Goldberg, 2006). In addition, there is evidence that speakers adapt to probabilistic information when processing a variety of linguistic input, including upcoming words (Altmann & Kamide, 1999; Arai & Keller, 2013; DeLong et al., 2005; Kutas & Hillyard, 1984), lexical categories (Gibson, 2006; Levy & Keller, 2013; Tabor et al., 1997), syntactic structures (Fine & Jaeger, 2013; Lau et al., 2006; Levy et al., 2012), semantics (Federmeier & Kutas, 1999; Kamide et al., 2003), and pragmatics (Mak et al., 2008; Ni, 1996; Roland et al., 2012).

syntactic structure of the sentence, but may instead depend on how slowly the sentence is read when encountered early in the experiment. Speed-up over time due to task adaptation should be greater for sentences that are (for any number of reasons) read more slowly at the beginning (Prasad & Linzen, 2021). Analogously, satiation in acceptability may depend merely on the level of acceptability at which a sentence is judged at the beginning of the experiment. For example, Brown et al. (2021) found that satiation effects were present across various grammatical and ungrammatical constructions, all of which started with an intermediate level of acceptability.⁵ As we will discuss below, it is unlikely that all satiation effects are reducible to task adaptation.

If satiation is a form of implicit learning, then it should be absent or weaker in familiar and expected structures, and it can potentially persist in time. That is, it may resemble long-term syntactic adaptation, as shown for structures known to be fully grammatical but infrequent in language use. For example, Kroczeck and Gunter (2017) demonstrate that syntactic adaptation to the relatively infrequent OSV word order in German is speaker-specific, cumulative, and long-lasting. In their study, participants listened to a series of German sentences with either SOV (canonical) or OSV (marked) word order. The sentences were produced by two different speakers, each of whom used the two orders with a different frequency distribution. Remarkably, their participants showed evidence of speaker-specific expectations even nine months after the initial experiment sessions.⁶

Thus far, Snyder (2022) is the only study that has investigated the possibility of long-term satiation effects in syntactic islands. Using *yes-no* acceptability judgment tasks on a variety of island types, Snyder (2022) tested the same participants at two separate times spaced four weeks apart. The only island type showing any satiation effects within the first session and across sessions was *Whether* Islands (e.g. *What does Henry wonder whether George discovered?*), which are among the weakest type of island in English. Snyder's (2022) results are, therefore, consistent with a syntactic adaptation account of satiation for *Whether* Islands. In contrast, Subject Island items showed no satiation effects at all in that study. However, as we shall see, there are several

⁵ An anonymous reviewer points out that the findings of Brown et al. (2021) showing amelioration in items rated intermediate in acceptability are not directly parallel to the findings of the reading time studies, which showed stronger task adaptation in items read more slowly at the beginning. Why, then, was stronger amelioration not shown in items of low acceptability? We think the reason for the discrepancy is likely to be that the studies of reading time used materials with fully grammatical sentences containing temporary ambiguities – i.e. items which would be unlikely to receive low ratings in a judgment task. Ungrammatical or otherwise anomalous sentences that receive low ratings in a judgment task exhibit less task adaptation, both in judgments and in reading time. For example, Left-Branch violations and That-trace violations never ameliorate (Francom, 2009; Goodall, 2011; Hiramatsu, 2000; Snyder, 2000; Snyder, 2022; Sprouse, 2009). Additional data are needed to decide this issue.

⁶ In a novel paradigm, probe trials were included in which disambiguating information was blanked out of the acoustic signal. In answering a comprehension question about the subject or the object, participants had to simply guess whether the sentence was SOV or OSV, based on prior expectations. This allowed Kroczeck and Gunter (2017) to measure syntactic adaptation to speaker-specific information within and across three experiment sessions.

design issues which may have led to a null effect. In this work, we revisit this question and employ a longitudinal design to investigate the long-term effects of exposure to Subject Island structures like (5b), as measured by offline tasks (Likert scale sentence acceptability) and by online tasks (self-paced reading). Taking into account the lack of satiation effects for Subject Island items in Snyder (2022), we use sentence materials similar to those in Chaves and Dery (2019), a previous study which showed clear facilitation effects for both acceptability ratings and reading times. Our results show that comprehenders gradually adapt to Subject Island constructions, and that the effect is detectable up to three weeks after the initial exposure. Crucially, the satiation effects are most pronounced in the Subject Island items, and do not occur in the ungrammatical control sentences, which involve subcategorization errors. We argue that these findings provide evidence in favor of an implicit learning account of satiation, over and above task adaptation.

The structure of the paper is as follows. In Section 2, we provide a brief survey of past satiation research, and, more specifically, on Subject Island satiation, and current accounts thereof. In Section 3, we describe a longitudinal study which expands on Snyder (2022) in various ways, including more items, longer exposures, more participants, and more sensitive methodology. In Section 4, we provide general discussion and future directions.

2. Background: Satiation in Subject Islands

As noted above, various studies have found that Subject Island violations in interrogative clauses show effects of so-called syntactic satiation, whereby acceptability ratings improve significantly with mere repetition over the course of an experiment (Chaves & Dery, 2014, 2019; Clausen, 2011; Francom, 2009; Hiramatsu, 2000; Lu et al. 2021; Lu et al., 2022). For example, Experiment 1 of Chaves and Dery (2019) found that Subject Island violations like those in (6a) gradually become more acceptable throughout the experiment session, to the point of becoming as highly rated as near-synonymous grammatical controls like (6b). Importantly, the acceptability of the ungrammatical filler sentences in the same experiment did not improve with repeated exposure, leading the authors to conclude that Subject Islands, though infrequently used, instantiate a possible structure of English. This result was replicated by Chaves and Putnam (2021, p. 213), and arose in the experiments reported below, using similar items.

- (6) a. Which stock does [the value of _] often parallel the price of the dollar?
(Subject Island violation)
- b. Which stock does the value of the dollar often parallel [the price of _]?
(Object sub-extraction control)

All of these results are from offline judgment experiments, but what about online evidence? It is well established that comprehenders apply an “active filler” strategy. That is, they postulate gaps as soon as possible during online processing of long-distance dependencies (Crain & Fodor,

1985; Frazier, 1987; Stowe, 1986; Stowe et al., 1991). This includes gaps within NP objects (e.g. *Who did you hear stories about _?*), as Tollan and Heller (2016) show. However, two classic experiments by Stowe (1986) and Experiment 2 of Pickering et al. (1994) found no evidence of comprehenders postulating gaps inside subject phrases. For example, using a self-paced reading task, Stowe (1986, p. 241) found no significant slowdown in reading time at the regions in bold for sentences like (7a) relative to (7b), suggesting that no gap was postulated upon encountering the preposition *about*.

- (7) a. The teacher asked what [the silly story about **Greg's** older brother] was supposed to mean.
 b. The teacher asked if [the silly story about **Greg's** older brother] was supposed to mean.

But more recent studies find that subject-internal gaps are sometimes postulated. For example, Phillips (2006) conducted a self-paced reading task with a plausibility mismatch manipulation (Traxler & Pickering, 1996), whereby comprehenders slowed down in items like (8) when reading *expand* when the filler is semantically incompatible (*which high school students*), relative to when reading *expand* when the filler phrase is semantically compatible (*which schools*).

- (8) The school superintendent learned {which schools, which high school students} the proposal to **expand** drastically and innovatively upon the current curriculum would overburden _ during the following semester.

Phillips (2006) argues that these findings refute previous claims that the parser is incapable of positing gaps within a Subject Island (Pritchett, 1991).⁷ Building upon this idea, Chaves and Dery (2019) hypothesized that repeated exposures to Subject Island sentences like (9a–c), which actually contain a subject-internal gap, should lead comprehenders to adapt and eventually begin to posit gaps within the subject phrase. In their Experiment 2, they tested this hypothesis using a self-paced reading experiment in which there were two groups of participants. Those in the Subject group read 15 Subject Island violation sentences and 30 distractor sentences in their first block, while participants in the Object group read 45 distractors in their first block. The second block was identical for both groups, and contained 10 Subject Island violation sentences, and 20 distractors. The examples in (9) illustrate the Subject Island items that participants read. The symbol | indicates the regions shown during the experiment.

⁷ Phillips (2006) argues that it was the possibility for a grammatical continuation with a second gap that prompted participants to anticipate a gap within the subject phrase. This conclusion was supported with evidence that participants did not posit a gap within a subject phrase containing a tensed clause – a configuration assumed to be ungrammatical with or without a second gap. However, Phillips' (2006) conclusion is disputed by Chaves and Dery (2019), who showed that readers can, in fact, anticipate a gap in such phrases, using the same paradigm, although with different stimuli items which better controlled for semantic felicity within the critical regions (Experiment 4).

- (9) a. Which animal₁ | does₂ | the song of₃ | reportedly₄ | mimic₅ | the Gray Catbird's sounds?₆ |
- b. Which athlete₁ | does₂ | the manager of₃ | clearly₄ | resemble₅ | Tiger Woods' agent ?₆ |
- c. Which company₁ | do₂ | the employees of₃ | allegedly₄ | reject₅ | salary increases?₆ |

Of all six regions, only Region 5 exhibited a significant effect, suggesting that participants in the Subject group (i.e. who saw fifteen Subject Island violation sentences in Block 1) read the region of interest of the Subject Island violations in Block 2 faster than participants in the Object group (i.e. who saw zero Subject Island violation sentences in Block 1). Region 5 was the region of interest, because it is the first region that unambiguously requires the reader to interpret the question phrase (e.g. *which animal*) as being associated with a gap inside the subject phrase (e.g. *the song of*). Until this point, it is still possible to anticipate a noun acting as a complement of the preposition *of* following the adverb. In short, these findings show that participants who had been previously exposed to Subject Island sentences began to anticipate a gap within the subject phrase.

Chaves and Dery (2019) argue that facilitation effects for Subject Islands, as they reported for both acceptability judgments (Experiment 1) and self-paced reading (Experiment 2), help support a syntax-external explanation of Subject Island effects and provide evidence for adaptation to subject-internal gaps: "...by increasing the frequency of such constructions, comprehenders can adapt and revise their expectations about the plausibility of subject internal gaps, all else being equal" (2019, p. 491). Following from the observation that judgments of ungrammatical control sentences did not satiate while judgments of Subject Islands did, they surmise that participants had gradually adapted to a possible but less familiar syntactic structure in the case of Subject Islands. We note that their findings are also compatible with theories that assume 'soft constraints' within the grammar, such as the cumulative constraint violation approach of Haegeman et al. (2014). This approach differs from Chaves and Dery's (2019) syntax-external approach in maintaining a partially syntactic explanation for the observed penalty in acceptability and reading time that Subject Island sentences incur early in the experiment session.

Syntactic adaptation of the type reported by Chaves and Dery (2019) may be a form of implicit learning in which comprehenders come to expect a particular structure that was previously unexpected, based on contingencies observed in the linguistic input (Chang et al., 2006, 2012; Fine et al., 2010, 2013; Fine & Jaeger, 2013; Sikos et al., 2016). With reference to experiments showing that English speakers adapted to unusual garden-path sentences following repeated exposures, Fine et al. (2013) formalize this type of learning using Bayesian belief-updating models. The basic idea of Bayesian belief-updating is that comprehenders manage the variability in the input by maintaining a representation of the probability distribution of syntactic structures over observations of utterances by particular speakers in particular contexts. This

representation guides their expectations about structural properties of the input and helps make processing more efficient. With each new observation, comprehenders update this representation to better reflect the totality of the input they have observed. Importantly, utterances with an unexpected syntactic structure (e.g. garden-path sentences) require a greater adjustment to this representation than utterances with an expected structure.⁸

Fine et al. (2013) identified two additional properties of syntactic adaptation that follow from their formal assumptions: (i) comprehenders should be sensitive to the distributional patterns associated with individual speakers; and (ii) with sufficient exposure, effects of adaptation may endure over days or weeks. The authors point to previous studies showing experimental evidence for both predictions. For example, regarding the first prediction, Kamide (2012) used a visual world experiment to test participants' resolutions of structurally ambiguous noun phrases (with either high or low attachment relative clause modifiers) in English. During the training phase, participants were exposed to three different speakers, each of whom was associated with a particular attachment option (always high, always low, or equally mixed). During the test phase, results showed that listeners were able to anticipate a particular resolution (as measured by eye movements) based on speaker identity.

Regarding the second prediction, Wells et al. (2009) conducted a longitudinal study in which participants completed four experiment sessions spaced out over a period of 3–4 weeks. Results of self-paced reading tasks given in the first and final sessions showed that reading times for object relative clauses (a difficult structure involving non-canonical word order) improved more for those participants who had been exposed to sentences containing many examples of relative clauses during the intervening sessions as compared to participants who had been exposed to other complex sentence structures. Similarly, Luka and Choi (2012) found that English speakers gave higher acceptability ratings to grammatical sentences with unusual structures if sentences with the same unusual structures had been previously encountered in a read-aloud task seven days earlier. More recently, many additional studies have provided similar evidence for syntactic adaptation across a range of dependent measures and sentence types; see Kaan and Chun (2018) for a review. Notably, a study by Kroczeck and Gunter (2017), introduced briefly above, showed evidence for syntactic adaptation to OSV word order in German which was both long-term (over a nine-month period) and speaker-specific. See also Yano (2024) for evidence of rapid adaptation to morphosyntactic and aspectual violations using event-related potentials as an index of sentence processing costs. This method avoids confounding adaptation effects with response familiarization involved in sentence plausibility judgment and self-paced reading tasks.

⁸ For an example of a belief-updating model of syntactic adaptation that includes the relevant mathematical formulas, we refer the reader to Kleinschmidt et al. (2012, pp. 600–602). Error-driven learning models of implicit learning, as proposed to account for structural priming (Chang et al., 2006, 2012), make similar empirical predictions (i.e. greater priming effects for less expected structures).

As mentioned above, syntactic adaptation may be confounded with task adaptation due to increased familiarity with the experimental paradigm (Dempsey et al., 2020; Prasad & Linzen, 2021). Since the two phenomena share some common properties, their independent effects may be tricky to tease apart. In the studies cited above, long-term effects were argued to support syntactic adaptation. However, there is still the possibility that long-term effects could be observed also for task adaptation. To our knowledge, this possibility has not been studied. Here, we remain agnostic about whether syntactic adaptation needs to be long-lived and about whether task adaptation is necessarily short-lived. Rather, our concern is whether satiation effects in Subject Islands are long-lived or not, and whether they are compatible with task adaptation. Below we show that they can be long-lived as well as construction-specific (i.e. they can specifically target sentences with Subject Islands, rather than ungrammatical distractors), suggesting that syntactic adaptation to subject-embedded gaps drives the effect, not a more general effect of task adaptation.

Regarding satiation in Subject Islands, existing evidence for the predictions from belief-updating models is limited to just a few studies. Regarding the first prediction for speaker-specific effects, Lu et al. (2021) reported acceptability rating experiments showing that satiation of Subject Islands and *Whether* Islands is modulated by speaker identity. Using a design much like that of Kamide (2012) with an initial training phase followed by a test phase, they found that participants were sensitive to both the abstract structural information (as shown by overall improvements) and its association with a particular speaker (as shown by an additional boost in acceptability when speaker identity was matched between training and test phases). Although theirs is currently the only study of speaker identity and satiation in island constructions, Lu et al. (2021) present encouraging evidence showing that satiation of Subject Islands and *Whether* Islands likely involves implicit learning and is consistent with the predictions of belief-updating models.

With respect to the longevity of satiation for Subject Islands, the evidence is again very limited. Snyder (2022) is the only island longitudinal study ever done, to our knowledge. Participants were asked to rate the acceptability of two sets of sentences with various types of island violations, four weeks apart, using binary (yes-no) judgments. Results showed that ratings of Subject Island sentences in the first part of the second session showed no evidence of improvement, as compared to the first part of the first session four weeks earlier. In contrast, ratings of *Whether* Islands in the same experiment (**Who did you ask whether he invited?**) showed significant improvement across sessions. Snyder's (2022) study thus appears to support a syntactic adaptation account of satiation effects for *Whether* Islands (consistent with the findings of Lu et al., 2021), but not for Subject Islands. Importantly, however, Subject Islands also showed no short-term satiation effects over the course of each session, in contrast to Chaves and Dery (2019), Lu et al. (2021), and several other studies. In fact, Snyder's (2022) study is one of a few

which have obtained null effects of within-session satiation for Subject Islands (Crawford, 2012; Goodall, 2011; Snyder, 2000, 2022; Sprouse, 2009). Null effects can be the result of a number of experimental and design factors.

Besides speaker-specific effects and long-lasting effects of satiation, another relevant area of research bearing on the source of satiation effects for Subject Islands is structural priming. Like satiation, structural priming can be characterized as a type of facilitation due to repetition of a syntactic structure. Structural priming across intervening trials has been argued to involve implicit learning (Bock & Griffin, 2000). Therefore, evidence for structural priming of Subject Island structures across intervening items, if found, would be consistent with a syntactic adaptation (implicit learning) account of satiation. Whereas satiation studies have measured cumulative improvements in acceptability over time, structural priming studies have measured facilitation within prime-target pairs. In addition, structural priming studies, unlike satiation studies, have rarely investigated island constructions. In a study that aimed to bridge these two areas of research, Do and Kaiser (2017) investigated priming across one vs. five intervening trials for Subject Islands and Complex NP Islands. They assumed that priming across longer distances, if shown, would provide solid evidence in favor of implicit learning. We will focus here on their findings for Subject Islands. In a Likert scale acceptability rating task, their analysis of prime-target pairs showed no significant improvement in ratings of Subject Island sentences following a Subject Island prime. This was true even when only one sentence intervened between the prime and the target. However, in a second experiment that used self-paced reading, they found faster reading times for Subject Island sentences following a Subject Island prime. When one sentence intervened between the prime and the target, faster reading times were shown across multiple regions. When five sentences intervened, faster reading times were shown in just one region. Despite the latter finding, the authors interpret these results as tentative support for a lingering activation account (which predicts priming effects, but only over shorter distances) and against an implicit learning account. Overall, we take these results to be inconclusive with respect to the source of facilitation, as reported in several previous studies of satiation in Subject Islands.

In the current study, we revisit the question of whether English Subject Islands show any lasting effects of satiation. We build upon the longitudinal design of Snyder (2022), with a few adjustments to increase the chances of satiation effects. Specifically, our stimuli and design are directly adapted from Chaves and Dery (2019), a study which exhibited robust satiation effects but lacked a longitudinal component. It remains unclear which specific factors led to Chaves and Dery's (2019) robust effect and Snyder's (2022) null results. Compared to Snyder (2022), we included more exposures to Subject Islands (18 per session across 3 sub-experiments instead of 5), more participants (80 instead of 22), and a more sensitive methodology (Likert-scale

acceptability judgment and self-paced reading tasks instead of a yes-no judgment task). Regardless of the possible reasons for Snyder’s null results, the results from our first session showed robust satiation effects, as in Chaves and Dery (2019). With respect to the longitudinal comparison, our results suggest that Subject Islands do, in fact, exhibit moderate long-term syntactic satiation effects. Although the ameliorative effect begins to wane once participants are no longer exposed to such sentences, the effect is still detectable three weeks after the initial exposure. Crucially, ungrammatical controls do not exhibit such a strong satiation profile, which speaks against task adaptation being the sole driver of Subject Island satiation, and suggests that practice with the experimental task is not all that matters; speakers can learn to adapt to the presence of subject-embedded gaps.

As we will elaborate below, our results are most consistent with a learning-based model of syntactic adaptation (Fine et al., 2013; Lu et al., 2021), and with conceptions of linguistic competence which include probabilistic information (Culicover et al., 2022; Fanselow et al., 2006; Francis, 2021; Lau et al., 2017; Manning, 2003; Villata & Tabor, 2022; Villata et al., 2019; Wasow, 2002). In the following section, we describe our methods and results.

3. Experiments

The current study employs a longitudinal design with the structure seen in **Figure 1**. Here, we present a brief overview of the study design before describing the methods in greater detail. The first session consisted of Experiments 1a, 2a, and 3a, conducted back-to-back, and the follow-up session consisted of the analogue Experiments 1b, 2b, and 3b, similarly conducted back-to-back. The same participants completed both sessions, three weeks apart.

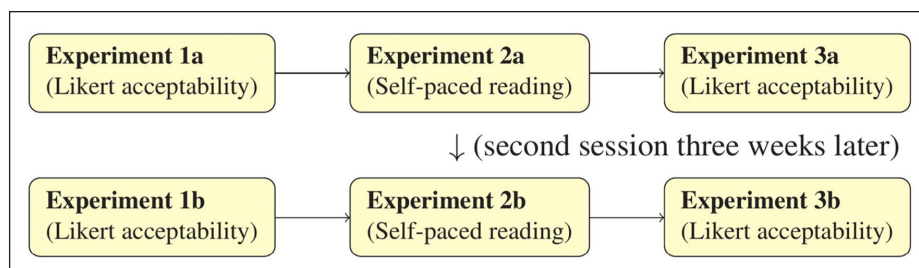


Figure 1: Longitudinal study consisting of two series of three experiments.

In Experiment 1a, participants were asked to rate 10 interrogative English sentences, using a Likert scale of 1–7, in the two conditions shown in (10), counterbalanced across two lists, pseudo-randomized, and interspersed with 10 interrogative distractor sentences. Each participant saw a different order of items.

- (10) a. Which artist does the son of frequently collaborate with the daughter of the Governor?
(Subject Island condition)
- b. Which artist does the daughter of the Governor frequently collaborate with the son of?
(Object condition)

Each participant, therefore, saw 5 Subject Island sentences, 5 Object (non-island) sentences (the controls), and 10 distractors. Our prediction is that the ratings for the Subject Island condition items start out as being significantly lower than the ratings for the Object condition, but gradually converge towards the latter.

In Experiment 2a, participants were asked to perform a self-paced reading experiment with a semantic plausibility manipulation, using items like those in (11). Crucially, in the Subject Island condition, the *wh*-phrase is a plausible filler for a gap in the subject phrase (i.e. *the manager of which factory*), whereas in the Object condition, the *wh*-phrase is an implausible filler for a gap in the subject phrase (i.e. *#the manager of which letter*).

- (11) a. Which **factory** ₁ | did ₂ | the **manager of** ₃ | likely ₄ | agree to ₅ | the demands ₆ | of ₇ | the union? ₈ |
(Subject Island condition)
- b. Which **letter** ₁ | did ₂ | the **manager of** ₃ | likely ₄ | the best ₅ | player ₆ | in the roster ₇ | make public? ₈ |
(Object condition)

There were 16 critical items, evenly distributed across the two conditions, pseudo-randomized, and counterbalanced across two lists. All items were interspersed with two grammatical distractor sentences. Each participant saw a different order of items.

Our prediction for Experiment 2a is that, initially, comprehenders should be relatively unlikely to posit a subject-internal gap in region 3 of either condition. However, we acknowledge that they might be influenced by exposure to five Subject Island items in the preceding acceptability task. Thus, while gap-creation within the subject phrase should be somewhat facilitated even at the beginning of the experiment, we would expect more gap postulation and a larger plausibility effect toward the end of Experiment 2a in the first session. This evidence is predicted to manifest in regions 3 and 5, as elaborated on below. In addition, if adaptation is long-lived, we would expect to see evidence of subject gap postulation more in Experiment 2b, three weeks later, compared to Experiment 2a.

In the first session (Experiment 2a), participants should initially exhibit difficulty in the disambiguating regions 5 and/or 6 in the Subject Island condition, because those continuations are unexpected unless a gap was postulated after the preposition in region 3. Following repeated exposures, however, we expect comprehenders to slowly revise their expectations about subject-embedded gaps and begin to postulate a gap at region 3. This should, then, cause the response

times at the disambiguating regions of Subject Island sentences (but not the Object sentences) to become faster. We therefore predict that there should be an interaction between condition and presentation order, such that reading times at regions 5 and/or 6 should speed up more in the Subject Island condition. Conversely, we expect that response times at region 3 of the Object sentences should become slower than before, as subject-internal gaps begin to be postulated, due to the implausibility of the *wh*-phrase as a filler for the subject-internal gap (*#the manager of which letter*). That is, we expect a similar interaction between condition and presentation order, such that reading times at region 3 should slow down more in the Object condition. We remain neutral as to the possibility for speed-up due to task adaptation in both conditions and acknowledge that task adaptation could potentially neutralize the expected slow-down for Object items in region 3. However, even in this case, an interaction between condition and presentation order should still be present.

After completing the self-paced reading task, participants were directed to Experiment 3a, which was a sentence acceptability rating experiment identical to Experiment 1a in terms of conditions, size, and structure, but composed of different sentences entirely (different main verbs, different *wh*-filler phrases, and different subject phrases). Our prediction for Experiment 3a is that the ratings for the Subject Island items start out as being significantly higher than those in Experiment 1a, and more easily converge towards the ratings of the Object controls.

Three weeks later, the same participants were asked to complete Experiments 1b, 2b, and 3b. In all regards, these experiments were identical to Experiments 1a, 2a, and 3a in terms of conditions, size, and structure, but crucially composed of different sentences entirely (different main verbs, different *wh*-filler phrases, and different subject phrases).

The main predictions are as follows: (i) the acceptability ratings for Subject Island items in Experiment 1b will be higher than in Experiment 1a, while ratings for the Object items will not differ between the two experiments; and (ii) reading times for Subject Island items in the disambiguating regions (5–6) will be faster in Experiment 2b than in Experiment 2a, compared to the speed-up for Object conditions (an interaction between condition and experiment). Finally, a slowdown in region 3 is predicted over time in the Object condition. As for Experiment 3b (the last part of the second session), we predict Subject Islands are rated higher than in Experiment 1b. In what follows, we describe the participants, materials, and procedures for these six experiments in more detail.

3.1 Methods

3.1.1 Participants

After obtaining IRB approval, we recruited 80 self-reported native speakers of American English via Prolific (<https://www.prolific.co/>) to participate in two online experiment sessions. Participants were compensated at a \$15 hourly rate. Mean total duration for each of the two

batches of experiments was 19 minutes (batch one being Experiments 1a, 2a, and 3a, and batch two being the follow-up Experiments 1b, 2b, and 3b). All participants were registered in Prolific as U.S. residents and participated from U.S. IP addresses.

There were 8 participants whose accuracy scores on comprehension questions (see below) were lower than the 75% threshold. Data collected from these participants were excluded from the analysis. The remaining 72 participants had accuracy levels of at least 75% in comprehension questions, with a mean accuracy level of 93%. Of these 72 participants who completed Experiments 1a, 2a, and 3a, a total of 52 participants returned 3 weeks later and completed the follow-up Experiments 1b, 2b, and 3b. Finally, a different set of 73 participants were recruited in the same way as described above, and were asked to complete Experiments 1b, 2b, and 3b only, without participating in the first session. The data from these participants serve as additional controls. 16 of these 73 participants were excluded from the analysis for having comprehension accuracy below 75%.

3.1.2 Design and materials

There were two sentence acceptability experiments in the first session (Experiment 1a and Experiment 3a), and two other sentence acceptability experiments in the second session, three weeks later (Experiment 1b and Experiment 3b). Each of these experiments consisted of five item blocks, each consisting of four types of interrogative sentences, as illustrated in (12).

- (12)
- a. Which celebrity does the wife of reportedly quarrel with the fiancée of the mayor?
(Subject Island condition)
 - b. Which politician does the son of Clint Eastwood sometimes socialize with the ex-wife of?
(Object condition)
 - c. Which rule did the Code of Conduct apparently contradict in the latest revision?
(Grammatical distractor)
 - d. Which boat does the report unexpectedly reveal that the soldiers were thinking?
(Ungrammatical distractor)

The Subject Island and Object items were counterbalanced, such that no two participants saw both versions of the same item. For example, (13) was the counterpart of the item block in (12).

- (13)
- a. Which celebrity does the fiancée of the mayor reportedly quarrel with the wife of?
(Object condition)
 - b. Which politician does the son of sometimes socialize with Clint Eastwood's ex-wife?
(Subject Island condition)
 - c. Which agent did the Russians repeatedly accuse of spying for the Americans?
(Grammatical distractor)
 - d. Which policy does the Department of Homeland Security mention revise next month?
(Ungrammatical distractor)

The ungrammatical distractors contained various subcategorization errors (i.e. an inappropriate complement or an extra complement not selected by any verb or preposition). In (12d), the verb *think* selects a PP headed by *about*, but, here, the preposition *about* is missing. In (13d), the verb *mention* selects a *to*-infinitive (*mention a policy to revise next month*) but the infinitival *to* is missing. In addition to counterbalancing the items across two lists, the order of the items within each item set was randomized, as was the order of item sets, so that no two participants saw the same critical items in the same order. Moreover, the items appearing in Experiment 1a and Experiment 3a were randomized across the two blocks, so that half the participants saw a given item in Experiment 1a and the other half, in Experiment 3a.

The items used were an expanded version of those used in Experiment 1 from Chaves and Dery (2019), with the difference being that the NP with the gap was held constant (i.e. both in (13a) and in (14a), the gap is in the phrase *the wife of* _). Thus, participants saw 5 Subject Island items and 5 Object controls in Experiment 1a, and another brand new 5 Subject Island items and Object controls in Experiment 3a. Three weeks later, the same participants saw 5 new Subject Island items and 5 new Object controls in Experiment 1b, and 5 new Subject Island items and Object controls in Experiment 3b.

Between Experiments 1a and 3a, as well as between Experiments 1b and 3b, participants were tasked with a self-paced reading experiment with a semantic plausibility manipulation. For each of these self-paced reading experiments (i.e. Experiments 2a and 2b), there were a total of 16 critical items, 8 Subject Islands and 8 Object controls, counterbalanced across two lists. A critical item is shown in (14). In the Subject Island condition, the *wh*-phrase was a plausible filler for a gap in the subject phrase, whereas in the Object condition it was not.

- (14) a. Which **disease**₁ | did₂ | the **cure for**₃ | virtually₄ | change₅ | modern₆ | medicine₇ | overnight?₈ |
(Subject Island condition)
- b. Which **lab**₁ | did₂ | the **cure for**₃ | virtually₄ | all₅ | allergic₆ | diseases₇ | come from?₈ |
(Object condition)

Each critical item was accompanied by two interrogative distractors, for a total of 48 experimental items per session. Within each group of items, one of the distractors was presented alone (as in (15)), and the other was presented together with a comprehension question (as in (16)). After each trial of the type in (16), participants received feedback about the accuracy of their comprehension question responses. Distractors in both types of trials consisted of fully grammatical interrogative sentences which varied in length, *wh*-phrase type, and verb tense.

- (15) a. Which artifact₁ | does₂ | the Museum₃ | of Fine Arts₃ | wish₄ | to purchase?₅ |
b. Where₁ | did₂ | the audience₃ | see₄ | the ace₅ | of spades₆ | reappear?₇ |
c. Which surfer₁ | does₂ | a headstand₃ | while₄ | chasing₅ | dolphins₆ | in the water?₇ |

- (16) a. Which animal ₁ | do ₂ | the villagers ₃ | allegedly ₄ | eat ₅ | as their ₆ | staple food? ₇ |
The diet of the local inhabitants consists uniquely of grains and herbs.
True/False?
- b. Which jury ₁ | members ₂ | did ₃ | the defendant ₄ | voice ₅ | various ₆ | complaints ₇ |
about?" ₈ |
The jury members complained about identifying the voice of the defendant.
True/False?
- c. Which consultant ₁ | did ₂ | the spokesperson ₃ | of the company ₄ | sue ₅ | for defamation? ₅ |
The company's consultant sued the spokesperson for defamation.
True/False?

As in all other experiments, items were pseudo-randomized, so that each participant saw a unique order of items. As already discussed, Session 1 (Experiments 1a, 2a, and 3a) and Session 2 (Experiments 1b, 2b, and 3b) were spaced three weeks apart.

3.1.3 Procedure

Presentation of the experimental stimuli was done on a web-based interface using PCIBex (Zehr & Schwarz, 2018), such that participants completed the experiments remotely and did not directly interact with the researchers. At the beginning of each experiment session, participants first read an IRB-approved consent form and checked a box to indicate their consent to participate. For Experiments 1a, 3a, 1b, and 3b, participants were instructed to judge how natural each sentence was, by giving it a rating from 1 (*very unnatural*) to 7 (*very natural*). Participants completed four practice trials at the beginning of each experiment to (re-)familiarize themselves with the task. Each experiment immediately followed the practice trials.

For Experiments 2a and 2b, participants were asked to read sentences on a self-paced moving window display (Just et al., 1982). For half of the distractor trials (one third of all trials), they were additionally required to answer a comprehension question. Each trial began by presenting a sequence of dashes representing the non-space characters in the sentences. Pressing the spacebar caused the dashes corresponding to the first region to be replaced by words. Subsequent presses revealed subsequent regions, while the previous region reverted to dashes. Reading times between each pair of button presses were recorded. The comprehension question that was presented after half of the distractor items had only two possible answers, *True* or *False*, which appeared on the screen in random order. The responses to the comprehension questions were recorded, and participants were immediately informed of any incorrect answers. Participants completed four practice trials at the beginning of each experiment to familiarize themselves with the task. The experiment immediately followed the practice trials.

3.2 Results

3.2.1 Sentence acceptability

The mean response for the Subject Island condition in Experiment 1a was 2.88 ($SD = 1.58$), and for the Object condition, 4.18 ($SD = 1.69$). The mean response for the grammatical distractors was 4.99 ($SD = 1.86$), and 2.75 ($SD = 1.63$) for the ungrammatical distractors. For Experiment 3a, the mean response for the Subject Island condition rose to 3.87 ($SD = 1.75$), and for the Object condition, 4.49 ($SD = 1.68$). Grammatical distractors were rated 6.51 ($SD = 0.92$), and ungrammatical distractors, 3.15 ($SD = 1.82$).

To account for the possibility of different participants using the Likert scale differently, we transformed the ratings into z-scores, by participant, before conducting any testing. The results for Experiments 1a and 3a in Session 1 are illustrated in **Figure 2**. Calculating z-scores by participant, by Experiment, and/or by Session instead of just by participant yielded no qualitative differences in the results.

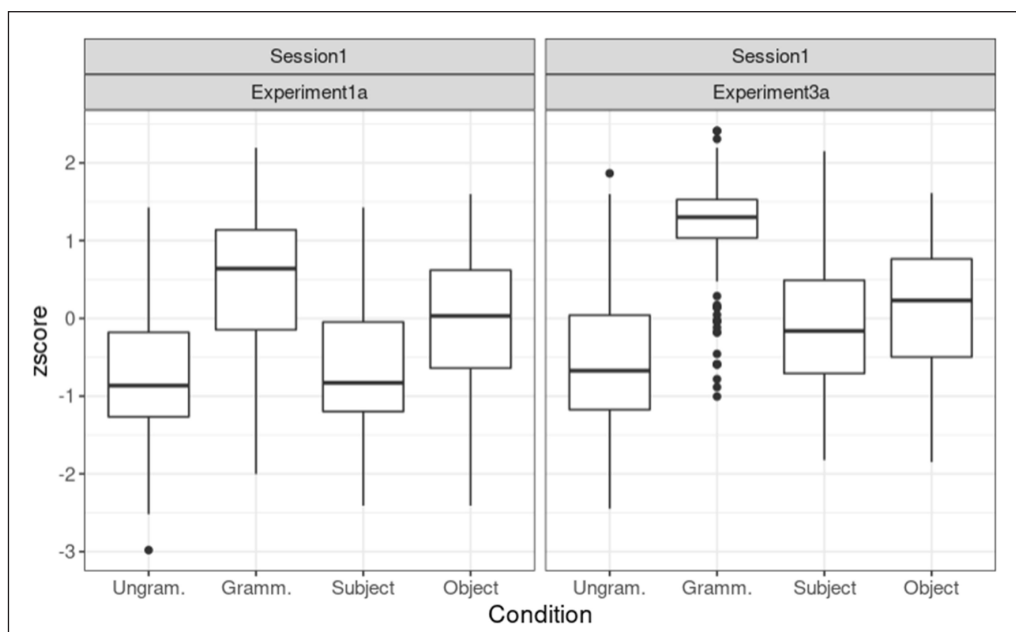


Figure 2: Z-scored acceptability ratings in Experiment 1a (left) and in Experiment 3a (right).

Linear Mixed Effects (LMER) models, with z-score ratings as the dependent variable and Experiment as the predictor (allowing the intercept to be adjusted by Sentence, and Participant), were fit separately for each of the four conditions: Subject Island, Object control, ungrammatical distractors, and grammatical distractors. Results indicated that Subject Island items were rated lower in Experiment 1a than in Experiment 3a ($\beta = 0.51$, $SD = 0.11$, $t = 4.68$, $p < 0.0001$). No such effect occurred for the Object items ($\beta = 0.18$, $SD = 0.14$, $t = 1.03$, $p = 0.2$), suggesting

no major change in acceptability for the Object Condition items across the two experiments. Similarly, no change of acceptability was registered for ungrammatical distractor items ($\beta = 0.21$, $SD = 0.19$, $t = 1.15$, $p = 0.264$).⁹ Grammatical distractor items were significantly more acceptable in Experiment 3a than in Experiment 1a ($\beta = 0.79$, $SD = 0.2$, $t = 3.97$, $p = 0.0008$).

To compare the Subject Island and Object control conditions, LMER models with z-score ratings as the dependent variable and Condition as the predictor (allowing the intercept to be adjusted by Sentence and Participant) were fit separately for Experiment 1a and Experiment 3a. Results indicated that Subject Island items were rated lower than Object control items both in Experiment 1a ($\beta = 0.62$, $SD = 0.12$, $t = 5.08$, $p < 0.0001$) and in Experiment 3a ($\beta = 0.29$, $SD = 0.13$, $t = 2.22$, $p = 0.03$), although the effect size was halved in the latter. These within-experiment comparisons can also be seen in **Figure 2**.

The effect of presentation order on acceptability ratings throughout the first experiment session (Experiments 1a and 3a) is shown in **Figure 3**. The dips at the end of Experiment 3a might be due to fatigue or lack of motivation, as the experiment wrapped up.

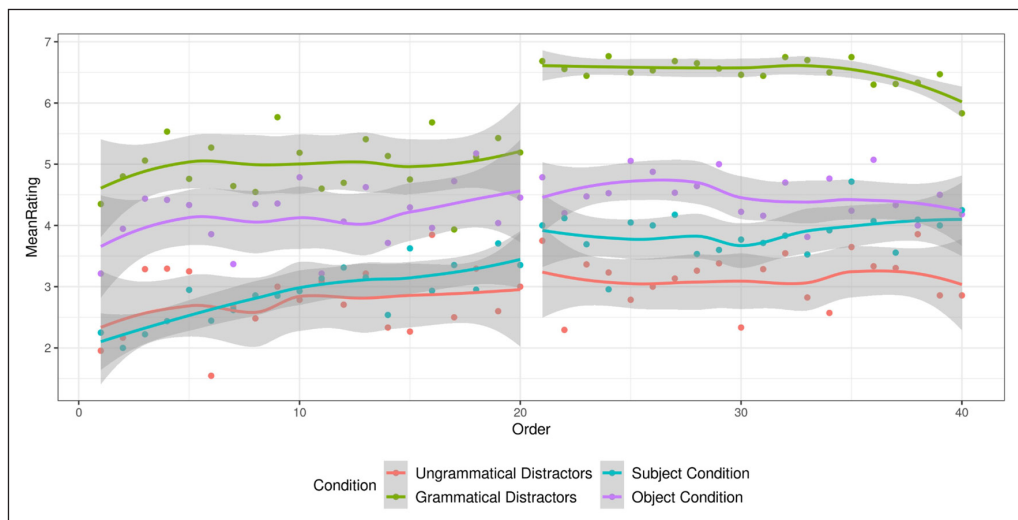


Figure 3: Mean acceptability ratings in Experiments 1a and 3a in Session 1 as a function of presentation order.

⁹ As can be seen in the supplemental materials, the subcategorization violations in our ungrammatical distractors varied in type. An anonymous reviewer suggests that maybe if the distractors had been more homogeneous, then they too might have ameliorated. While this remains a possibility, we suspect none of our distractor types would ameliorate with repeated exposure to the same error type, given how structurally unsound they are. And even if they did ameliorate, it would be extremely unlikely that they would come to be rated as high as the grammatical controls, given previous studies of satiation of ungrammatical structures. Finally, some ungrammatical structures never ameliorate at all (Francom, 2009; Goodall, 2011; Hiramatsu, 2000; Snyder, 2000; Snyder, 2021; Sprouse, 2009). See also Chaves and Putnam (2021, pp. 230–233) for satiation studies with sentence agreement attraction errors and with nominative object case assignment errors. On the other hand, the lack of amelioration found in our (heterogeneous) ungrammatical items suggests that task adaptation played little role in the ratings of these items.

Separate LMER models with z-score ratings as the dependent variable and Presentation Order as the predictor (allowing the intercept to be adjusted by Sentence and Participant) were fit for each of the conditions in Experiment 1a. The results, as shown in **Table 1**, suggest that Subject Island items and Object items became more acceptable as the experiment progressed, but the distractors did not change.

Table 1: Effect of presentation order on each of the item types for Experiment 1a.

| Item Type | β | SD | t | p |
|--------------------------|--------------|--------------|--------------|-------------------|
| Subject Island | 0.035 | 0.005 | 6.245 | <0.0001 |
| Object | 0.014 | 0.006 | 2.255 | 0.024 |
| Grammatical distractor | 0.0006 | 0.0005 | 1.116 | 0.26 |
| Ungrammatical distractor | 0.01 | 0.006 | 1.54 | 0.12 |

Crucially, an LMER model like those above was fit to probe the interaction between Presentation Order and Condition. This analysis showed that the increase of acceptability for Object items as a function of presentation order was less than that of Subject Island items ($\beta = -0.02$, $SD = 0.009$, $t = -2.248$, $p = 0.02$).

Similar models were then fit for Experiment 3a, shown in **Table 2**. At this stage, none of the conditions exhibited any statistically significant change. The only exception was the grammatical distractors, which exhibited a negative effect of presentation order. However, it is clear from **Figure 3** that this phenomenon is limited to the last quarter of Experiment 3a.

Table 2: Effect of presentation order on each of the item types for Experiment 3a.

| Item Type | β | SD | t | p |
|--------------------------|---------------|--------------|---------------|--------------|
| Subject Island | 0.006 | 0.005 | 1.144 | 0.250 |
| Object | -0.008 | 0.005 | -1.414 | 0.158 |
| Grammatical distractor | -0.009 | 0.003 | -2.308 | 0.021 |
| Ungrammatical distractor | -0.009 | 0.005 | -1.548 | 0.123 |

Moving on to the second session three weeks later, the mean response for the Subject Island condition in Experiment 1b was 3.22 ($SD = 1.75$), and for the Object condition, 4.13 ($SD = 1.84$). The mean response for the grammatical distractors was 5.61 ($SD = 1.48$), and 2.24 ($SD = 1.49$) for the ungrammatical distractors. For Experiment 3b, the mean response for the Subject

Island condition rose to 3.87 ($SD = 1.75$), and for the Object condition, 4.49 ($SD = 1.68$). Grammatical distractors were rated 6.51 ($SD = 0.92$), and ungrammatical distractors, 3.15 ($SD = 1.82$). The z-scored ratings, computed by participant, are depicted in **Figure 4**.

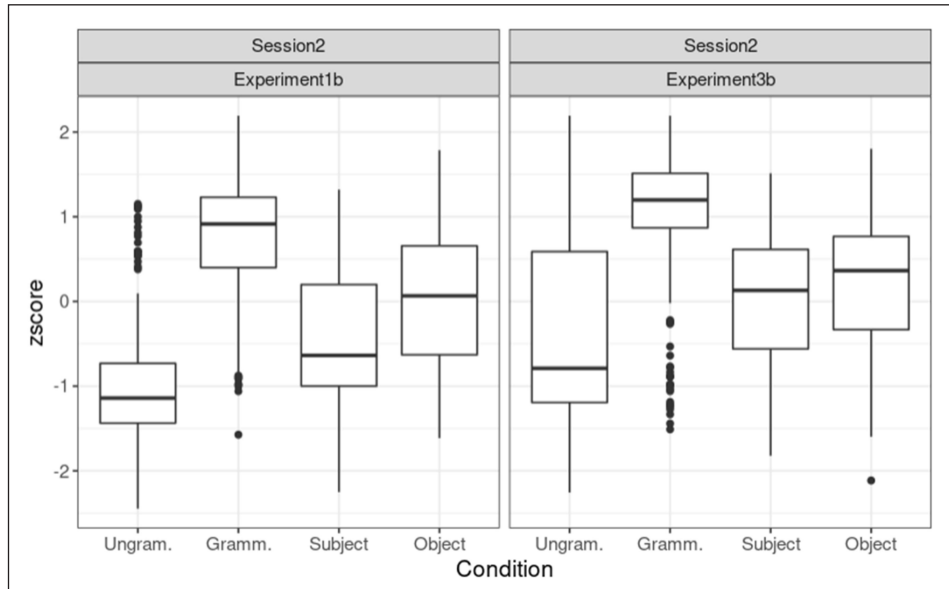


Figure 4: Z-score acceptability ratings in Experiment 1a (left) and in Experiment 3a (right).

A LMER model comparing Experiments 1b and 3b was fit separately for each condition, as described above for Experiments 1a and 3a. Results show that Subject Island items were rated higher in Experiment 3b than the Subject Island items in Experiment 1b ($\beta = 0.48$, $SD = 0.14$, $t = 3.328$, $p = 0.003$). As before, no such effect occurred for the Object items ($\beta = 0.13$, $SD = 0.17$, $t = 0.742$, $p = 0.46$), suggesting no major change in acceptability for the Object items across the two experiments in Session 2. Ungrammatical distractors were rated higher in Experiment 3b compared to Experiment 1b, with a difference approaching significance ($\beta = 0.58$, $SD = 0.28$, $t = 2.06$, $p = 0.053$), but grammatical distractors showed no significant differences between the two experiments ($\beta = 0.28$, $SD = 0.19$, $t = 1.43$, $p = 0.16$).

LMER models comparing the two critical conditions (Subject Island and Object control) were fit separately for Experiment 1b and Experiment 3b. These results indicated that Subject Island items were rated lower than Object items in Experiment 1b ($\beta = 0.52$, $SD = 0.18$, $t = 2.81$, $p = 0.01$) but, most importantly, not in Experiment 3b ($\beta = 0.14$, $SD = 0.15$, $t = 0.94$, $p = 0.35$). In other words, the Subject Island effect vanished in Experiment 3b.

We now turn to the effect of presentation order in Session 2, as shown in **Figure 5**.

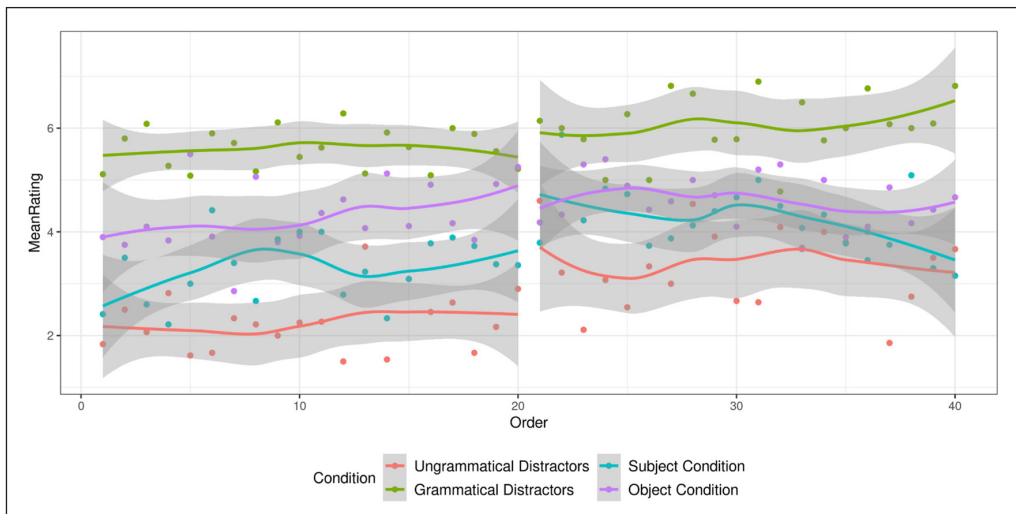


Figure 5: Mean acceptability ratings in Experiments 1b and 3b in Session 2 as a function of presentation order.

As before, separate LMER models were fit for each of the four conditions in Experiment 1b, to examine the effect of presentation order on the ratings of the items in the experiment. The results, as displayed in **Table 3**, suggest that only Subject Island items became slightly more acceptable as the experiment progressed.

Table 3: Effect of presentation order on each of the item types for Experiment 1b.

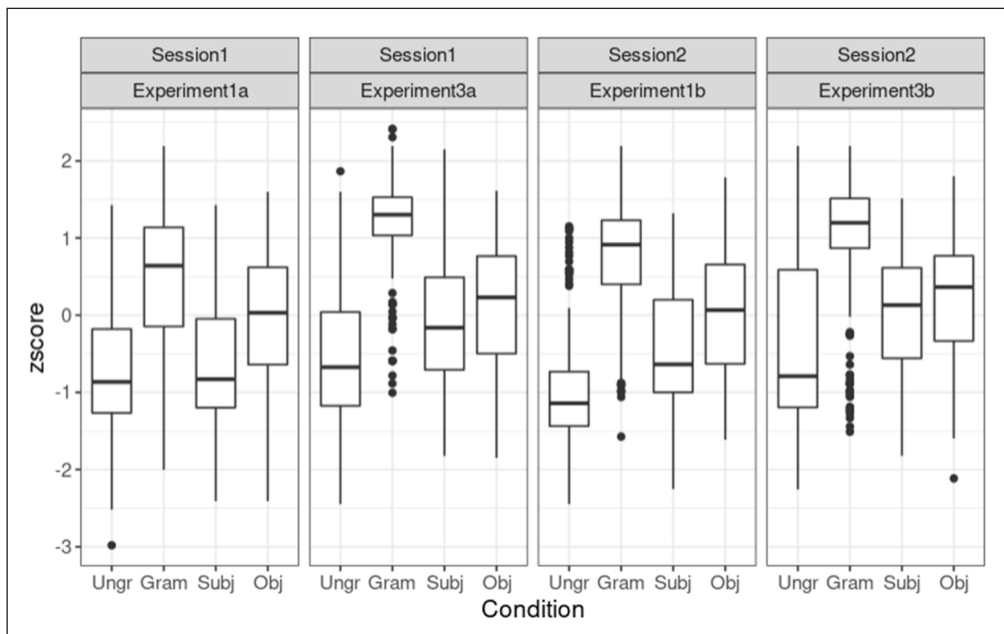
| Item Type | β | SD | t | p |
|--------------------------|--------------|--------------|-------------|--------------|
| Subject Island | 0.016 | 0.007 | 2.09 | 0.038 |
| Object | 0.01 | 0.007 | 1.366 | 0.174 |
| Grammatical distractor | 0.0005 | 0.0007 | 0.702 | 0.48 |
| Ungrammatical distractor | 0.009 | 0.008 | 1.135 | 0.258 |

As in the case of Experiment 3a, the very end of Experiment 3b exhibits some quick changes, perhaps due to fatigue, as the experiment wrapped up. Interestingly, there is a differentiation trend whereby acceptable items (Object condition and grammatical distractors) experience an upwards trend, while less acceptable items (Subject Island and ungrammatical distractors) experience a downwards trend. However, the results for LMER models fit to each of the four conditions suggest that there were no significant changes due to presentation order within Experiment 3b (**Table 4**).

Table 4: Effect of presentation order on each of the item types for Experiment 3b.

| Item Type | β | SD | t | p |
|--------------------------|---------|--------|--------|-------|
| Subject Island | -0.009 | 0.0076 | -1.240 | 0.217 |
| Object | -0.009 | 0.0075 | -1.248 | 0.213 |
| Grammatical distractor | 0.0002 | 0.0006 | 0.339 | 0.69 |
| Ungrammatical distractor | 0.0034 | 0.008 | -0.424 | 0.672 |

For completeness, we provide in **Figure 6** a plot displaying all conditions across all four acceptability judgment experiments. As reported above, Subject Island violations exhibited robust increases of acceptability between Experiments 1a and 3a, and moderate increases between Experiments 1b and 3b. The Object condition exhibited no such changes. Similarly, no such effect was observed for the ungrammatical distractors.

**Figure 6:** Z-score ratings in all experiments and sessions for Subject Island and Object conditions.

3.2.1.1 Longitudinal comparison

The last, and perhaps most critical, comparison we performed was to fit an LMER model pitting the Subject Island ratings from Experiment 1a against those from Experiment 1b, three weeks later. Indeed, we found a significant difference ($\beta = 0.17$, $SD = 0.06$, $t = 2.814$, $p = 0.005$),

suggesting that the acceptability of Subject Island violations three weeks later was higher than that of the initial exposure. This means that the acceptability increase caused by repeated exposure is still present three weeks later for sentences containing Subject Island violations. No such difference was found for Object conditions ($\beta = 0.08$, $SD = 0.06$, $t = 1.33$, $p = 0.182$). The opposite effect was found for ungrammatical distractors ($\beta = -0.24$, $SD = 0.06$, $t = -3.832$, $p = 0.0001$), suggesting a decrease of acceptability in Experiment 1b relative to Experiment 1a. Finally, a positive effect was found for the grammatical distractors ($\beta = 0.31$, $SD = 0.06$, $t = 5.097$, $p < 0.0001$).

As already mentioned, the three experiments in Session 2 were conducted a second time, with a new group of participants that had not participated in Session 1. These experiments, which we refer to as Experiments 1b', 2b', and 3b', serve as controls, allowing us to see the effect of having completed Session 1 on the ratings of the exact same stimulus sentences from Session 2 (recall that Experiments 1a, 1b, 3a, and 3b have the same conditions but different sentences). The plot in **Figure 7** illustrates the results. We once again see the positive effect of exposure on the Subject Island items.

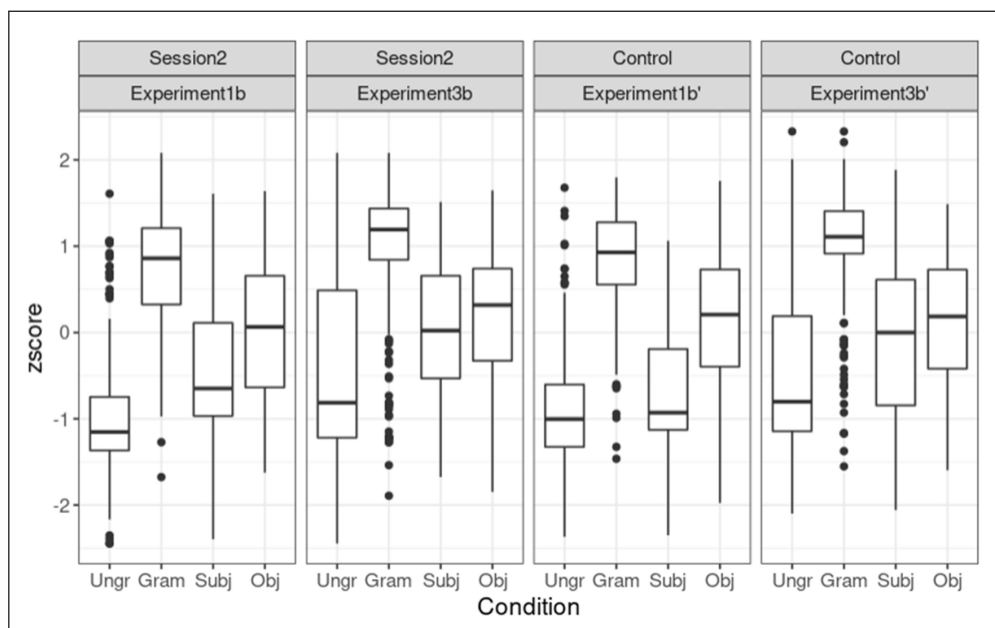


Figure 7: Z-score ratings for Session 2 and for the Control Session experiments.

More importantly, an LMER model reveals that Subject Island items were rated lower in Experiment 1b' (Control) than in the Experiment 1b counterpart, suggesting that having completed Session 1 led to a boost in acceptability ($\beta = -0.19$, $SD = 0.07$, $t = -2.472$, $p = 0.01$). In contrast, an LMER model for the Object condition items showed no difference between

Experiment 1b' (Control) and the Experiment 1b counterpart ($\beta = 0.057$, $SD = 0.07$, $t = 0.754$, $p = 0.45$). No statistical difference was found between the Subject Island item ratings across Experiment 1a and Experiment 1b' (Control), with $\beta = 0.051$, $SD = 0.07$, $t = 0.69$, $p = 0.48$, nor between the Object item ratings across Experiment 1a and Experiment 1b' (Control) with $\beta = -0.101$, $SD = 0.07$, $t = -1.31$, $p = 0.19$. This is as predicted.

3.2.1.2 Discussion

The results are consistent with our predictions. The ratings for the Subject Island items in Experiment 1a gradually converged towards the Object condition, to the point of making the island effect vanish, just as it did in Experiment 1 of Chaves and Dery (2019). Crucially, the acceptability ratings for Subject Island items in Experiment 1b, three weeks later, were higher than those of Subject Island items in Experiment 1a. Moreover, the latter were stable and exhibited no further acceptability increases. No other condition exhibited this acceptability rating profile, including the ungrammatical distractors, suggesting that the amelioration effect is not due to task adaptation, but rather to syntactic adaptation to a specific structure involving a subject-embedded gap. We note that the ungrammatical distractors, which involved various types of subcategorization errors, started at the same level of acceptability as the Subject Island items at the beginning of the first session, but did not significantly improve with repetition (**Figure 3**).

3.2.2 Reading time analysis

We now turn to the analysis of the self-paced reading studies, Experiments 2a and 2b. Recall that our critical items involved two conditions: one in which there was a *wh*-phrase that was a plausible filler for a gap located inside the subject phrase (Subject Island condition), and another in which the *wh*-phrase was not a plausible filler for a (globally illicit) gap inside the subject phrase (Object control condition). In the latter condition, the real gap was located outside an island environment. In response to an anonymous reviewer's comment, we conducted a norming experiment with 80 participants to elicit plausibility ratings for each of the subject phrases alone (e.g., *the cure for the lab*). Results showed that, contrary to our expectations, three of the original Object control items were statistically indistinguishable from their Subject Island counterparts. Accordingly, responses to these three items were excluded from the analysis that follows, although the removal of these data had no qualitative impact on the results.¹⁰

¹⁰ In the norming experiment, 80 self-reported native speakers of English rated a series of simple English phrases with respect to naturalness on a scale of 1 to 5. In addition to the subject phrases from the Subject Island and Object Control conditions of the main experiment, the stimuli included two times as many fillers representing a variety of plausible and implausible noun phrases, verb phrases, and adjective phrases. Each participant saw one of two lists. Items on each list were counterbalanced and pseudo-randomized.

All observations with reading times less than 100 ms or greater than 2000 ms were excluded, leading to the loss of 3% of observations. We then computed residual reading times by subtracting from the actual reading time for a region the reading time predicted by a regression equation relating region length to reading time (Trueswell et al., 1994). This regression equation was computed separately for each participant, using all regions in the experimental and distractor items.

We begin with an analysis of presentation order effects in the RTs for each region in Experiment 2a of Session 1. We use Bayesian Mixed Effects Linear Regression (BMER) models, fit with the BRMS package (Bürkner, 2017), as they are more robust than frequentist linear models with respect to convergence issues, and provide credible intervals rather than point estimates. Credible intervals are more intuitive to interpret than their frequentist counterparts, as they provide a range of values where the true outcome is expected to fall with a certain probability.

Since there are no prior longitudinal studies of online processing of Subject Islands, and existing theories do not make quantitative predictions, we resort to default (flat) priors, to keep the model very conservative. All our models were checked for convergence after fitting them with four chains and 6,000 iterations, half of which were the warm-up phase, by verifying that the R-hat values were close to one, and visually inspecting the chains.

Probing for an interaction between Condition (Subject Island, Object control) and Order in predicting RTs, with sentence and participant as random slopes, we found evidence for an effect in critical region 3, and in the spill-over regions 6 and 7. The results are given in **Table 5**, including estimate errors, credible intervals, evidence ratio, and posterior probabilities. Although the credible intervals include zero, these have moderate to strong Evidence Ratios close or above 10 (see Lee & Wagenmakers, 2014), as they are located at the marginal upper end of the interval and result in a +90% posterior probability. This is a non-negligible value which should not be summarily dismissed by attempts to dichotomize significance. Stronger results could be obtained with more data and more informative priors.¹¹

The residual reading times are shown in **Figure 8**, as a function of presentation order. We re-introduce example (14) as (17) to better illustrate the regions referenced in **Figure 8**.

- (17) a. Which **disease**₁| did₂| the **cure for**₃| virtually ₄| change ₅| modern ₆| medicine ₇|
overnight?₈|
(Subject Island condition)

¹¹ For example, Boland et al. (2023) find evidence of syntactic adaptation to the “needs + participle” regional construction in self-paced reading relative to a control, estimating with a significant $\beta = -11.99$ interaction coefficient (Experiment 1, exposure block). Even if we halve that estimate and assume a prior of $N(-5,4)$ for our interaction, rather than a flat prior, in order to reflect the expectation of some adaptation to Subject Island environments as exposure increases, our interaction CI shifts to $[-9.98,-0.55]$, with $\beta = -5.22$, $ER = 29$, and $P(\beta < 0) = 0.97$.

- b. Which **lab**₁ | did₂ | the **cure for**₃ | virtually₄ | all₅ | allergic₆ | diseases₇ | come from?₈ |
(Object condition)

Table 5: Bayesian Mixed Effects Linear Regression models for Condition * Presentation Order interactions in Session 1 (Experiment 2a).

| Region | β | Err | CI | ER | $P(\beta < 0)$ |
|--------|---------|------|----------------|-------|----------------|
| 3 | -5.57 | 4.19 | [-12.3, 1.2] | 9.9 | 0.91 |
| 4 | -0.95 | 4.81 | [-8.93, 6.91] | 1.35 | 0.58 |
| 5 | 1.86 | 3.67 | [-4.13, 7.9] | 0.44 | 0.31 |
| 6 | -4.46 | 3.29 | [-9.89, 1.01] | 10.16 | 0.91 |
| 7 | -4.51 | 2.91 | [-9.27, 0.29] | 15.57 | 0.94 |
| 8 | -3.73 | 4.38 | [-10.86, 3.43] | 4.13 | 0.81 |

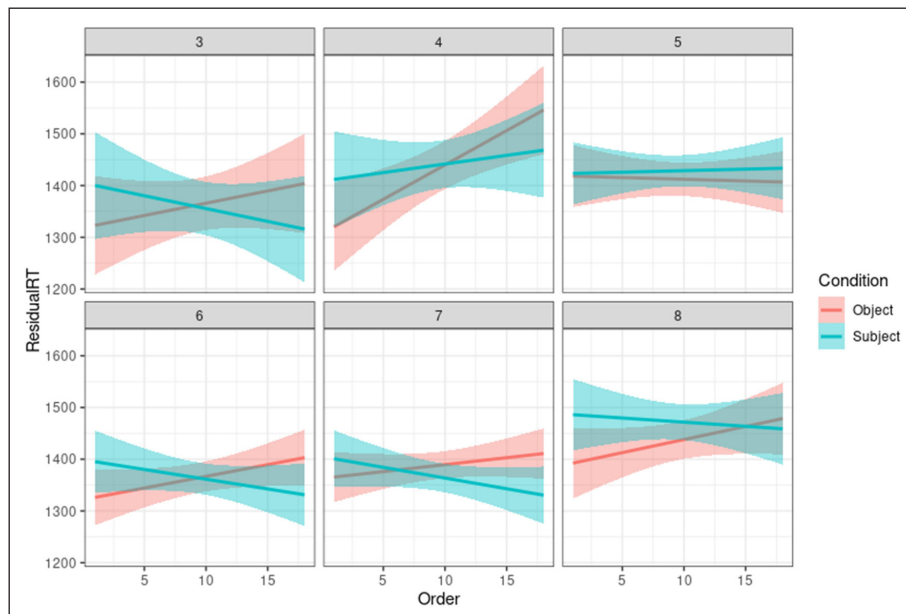


Figure 8: Residual RTs for regions 3 through 8 in Session 1 (Experiment 2a).

We interpret the interactions in region 3 (the plausibility cue) and in regions 6–7 (the spillover for the disambiguating region) as adaptation to a parse where a subject-internal gap is postulated. These results are in line with our predictions that RTs for Subject Island items should speed up over time in the disambiguating regions, while RTs for Object items should begin to slow down over time at the plausibility cue. More specifically, such findings are consistent with our prediction of a plausibility effect in region 3, caused by the filler phrase being semantically incongruent with a subject gap, as participants begin to postulate a subject-internal gap in the

Object items. Similarly, the interaction observed in the spillover regions 6 and 7 is consistent with participants adapting to the presence of a subject-internal gap in the Subject Island items. Failing to postulate a subject-embedded gap in the Subject Island condition causes the rest of the sentence (starting in region 5) to be incoherent.

As expected, Experiment 2b exhibited a very different profile, as seen in **Table 6**. BMER models with the same structure as those used for Experiment 2a found a mere trend for an interaction between Condition and Presentation Order effect in regions 4 through 6, in the same direction as those of Experiment 2a. The slopes are shown in **Figure 9**.

Table 6: Interaction between Condition and Presentation Order in Session 2 (Experiment 2b).

| Region | β | Err | CI | ER | $P(\beta < 0)$ |
|--------|---------|------|----------------|------|----------------|
| 3 | 4.68 | 6.33 | [-5.64, 15.11] | 0.3 | 0.23 |
| 4 | -6.2 | 6.38 | [-16.69, 4.3] | 5.02 | 0.83 |
| 5 | -4.42 | 5.03 | [-12.5, 3.91] | 4.22 | 0.81 |
| 6 | -2.8 | 4.34 | [-9.93, 4.43] | 2.86 | 0.74 |
| 7 | 2.28 | 3.95 | [-4.25, 8.76] | 0.4 | 0.29 |
| 8 | 0.3 | 6.12 | [-9.6, 10.43] | 0.92 | 0.48 |

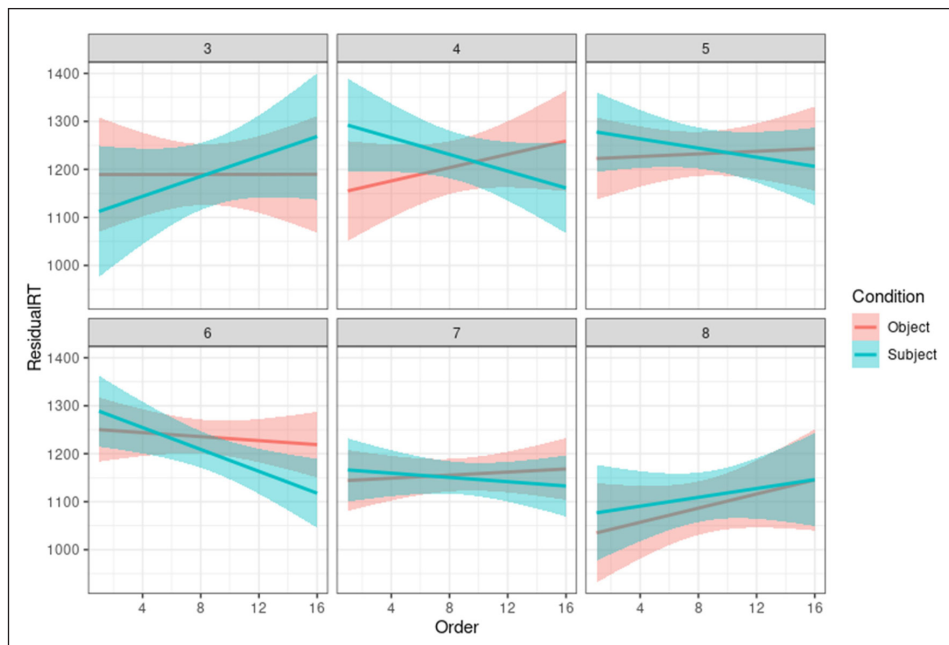


Figure 9: Residual Reading Times for regions 3 through 8 in Session 2 (Experiment 2b).

These results suggest that most adaptation occurred during the first session, which is consistent with a response to the tension between two opposing forces: (i) the fact that the Subject Island items require the postulation of a subject-internal gap for such items to be felicitously parsed, and (ii) the plausibility penalty caused by attempting to fill a gap inside the subject in the Object condition. Participants may have adjusted their expectations based on Session 1 and become more cautious about deciding whether a gap should be postulated at region 3, causing the effect of plausibility to be delayed until regions 4 and 5.

3.2.2.1 Longitudinal comparison

We now turn to longitudinal comparisons between Experiment 2a and Experiment 2b. If adaptation is short-lived, we would expect the RTs to return to baseline after three weeks. A BMER model was fit with Condition and Experiment as interacting predictors of RTs, and list, sentence, and participant as random slopes. Results found some evidence for an effect in spillover region 6, as seen in **Table 7** and in **Figure 10**, whereby Subject Island items were overall read faster in Experiment 2b (second session), relative to Object items, than they were in Experiment 2a (first session). Again, the evidence is nuanced, but stronger results are discussed below.

Table 7: Longitudinal comparison of Subject Island/Object conditions across Experiments 2a and 2b.

| Region | β | Err | CI | ER | $P(\beta < 0)$ |
|--------|---------------|--------------|-----------------------|------------|----------------|
| 3 | 13.21 | 38.29 | [-43.94, 81.01] | 0.47 | 0.32 |
| 4 | 19.27 | 43.47 | [-51.59, 91.03] | 0.49 | 0.33 |
| 5 | -16.13 | 32.51 | [-69.3, 36.67] | 2.56 | 0.69 |
| 6 | -35.75 | 27.03 | [-80.43, 8.33] | 9.8 | 0.91 |
| 7 | 15.82 | 26.77 | [-28.33, 59.12] | 0.38 | 0.28 |

To examine the effects of presentation order across the two sessions, and gain a sense of the reading time dynamics during the sessions, BMER models were fit separately for each condition, with Experiment and Order as an interaction (and with list, sentence, and participant as random slopes). For the Subject Island condition, strong effects were found in region 5 ($\beta = -5.61$, Err = 4.15, CI = [-12.36, 1.3], ER = 10.51, $P(\beta < 0) = 0.91$) and region 6 ($\beta = -8.1$, ER = 3.95, CI = [-14.63, -1.57], ER = 47.58, $P(\beta < 0) = 0.98$), suggesting that participants sped up more over time in Experiment 2b compared to Experiment 2a, despite the use of flat priors. Crucially, no such effect was detected for region 5 of the Object condition ($\beta = 0.56$, Err = 5.19, CI = [-7.96, 9.08], ER = 0.84, $P(\beta < 0) = 0.46$). Although there was a significant effect in region 6 ($\beta = -8.61$, ER = 4.08, CI = [-15.4, -1.92], ER = 58.11, $P(\beta < 0) = 0.98$), this was because the Object condition was processed increasingly slower in the first session. Object Items stopped increasing and stabilized in the second session. See **Figure 11** for an illustration.

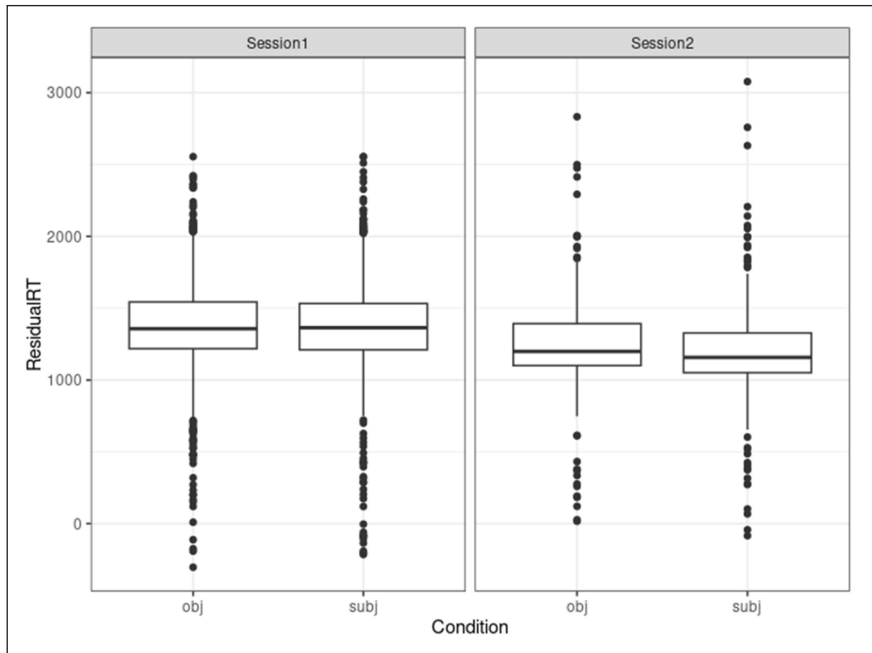


Figure 10: Residualized reading times in region 6 across both sessions and conditions.

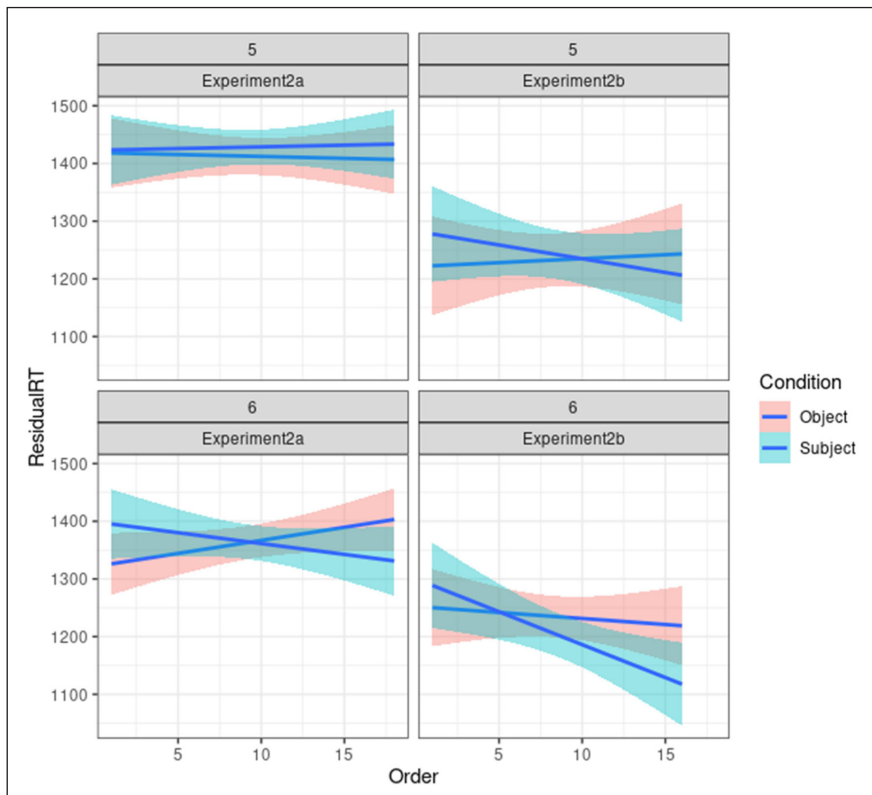


Figure 11: Side-by-side comparison of regions 5 and 6 across sessions.

No other regions yielded significant effects, including regions 3 and 7, suggesting that only the spillover regions were of strategic value for processing the items in both experiments, but more so for Subject Island items, since the latter enabled faster reading times. The overall dip in processing time observed in Experiment 2b is consistent with task adaptation, but the reading time interaction between the Subject Island and Object conditions suggests syntactic adaptation.

We next compared the reading times of Experiment 2a with those of Control Experiment 2b' (Table 8), probing for an interaction between Condition and Experiment, with Control as the baseline. Recall that for both experiments, these were the first sessions the participants were exposed to. The only differences between these experiments are the participants and the lexical items in the stimuli, which are different. As expected, there were no interactions. The evidence ratio was far below 10 for all cases.

Table 8: Interaction between Condition and Experiment (Experiment 2a and Experiment 2b').

| Region | β | Err | CI | ER | $P(\beta > 0)$ |
|--------|---------|-------|------------------|------|----------------|
| 3 | -50.2 | 43.23 | [-121.81, 21.33] | 0.12 | 0.13 |
| 4 | -27.64 | 41.05 | [-95.33, 39.71] | 0.33 | 0.25 |
| 5 | 17.7 | 31.79 | [-34.49, 69.73] | 2.57 | 0.72 |
| 6 | -5.94 | 28.16 | [-51.47, 40.44] | 0.72 | 0.42 |
| 7 | 4.51 | 27.8 | [-41.6, 50.1] | 1.31 | 0.57 |

Comparing the reading times of Experiment 2b with those of Control Experiment 2b' (Table 9), there was evidence suggestive of an interaction between the Condition and Experiment, as expected. The interaction in region 6 indicates that Subject Island items were read faster than Object items in Experiment 2b, relative to the Control Experiment 2b'.

Table 9: Interaction between Condition and Experiment (Experiment 2b and Experiment 2b').

| Region | β | Err | CI | ER | $P(\beta > 0)$ |
|--------|---------------|--------------|-----------------------|--------------|----------------|
| 3 | 47.44 | 45.96 | [-28.61, 122.55] | 5.74 | 0.85 |
| 4 | 14.47 | 48.71 | [-66.03, 93.98] | 1.65 | 0.62 |
| 5 | -0.58 | 36.6 | [-60.89, 59.29] | 0.98 | 0.5 |
| 6 | -44.42 | 29.34 | [-3.63, 93.14] | 14.75 | 0.94 |
| 7 | 20.15 | 36.65 | [-39.61, 80.31] | 0.41 | 0.29 |

The box-and-whiskers plot in Figure 12 illustrates these findings for region 6. Recall that Session 2 (Experiment 2b) and the Control (Experiment 2b') were identical, except for the

participants. If the participants in Session 2 had not adapted to the Subject Island sentences from the earlier session, then the results for Session 2 and the Control should have been the same. The results are, therefore, consistent with syntactic adaptation across sessions.

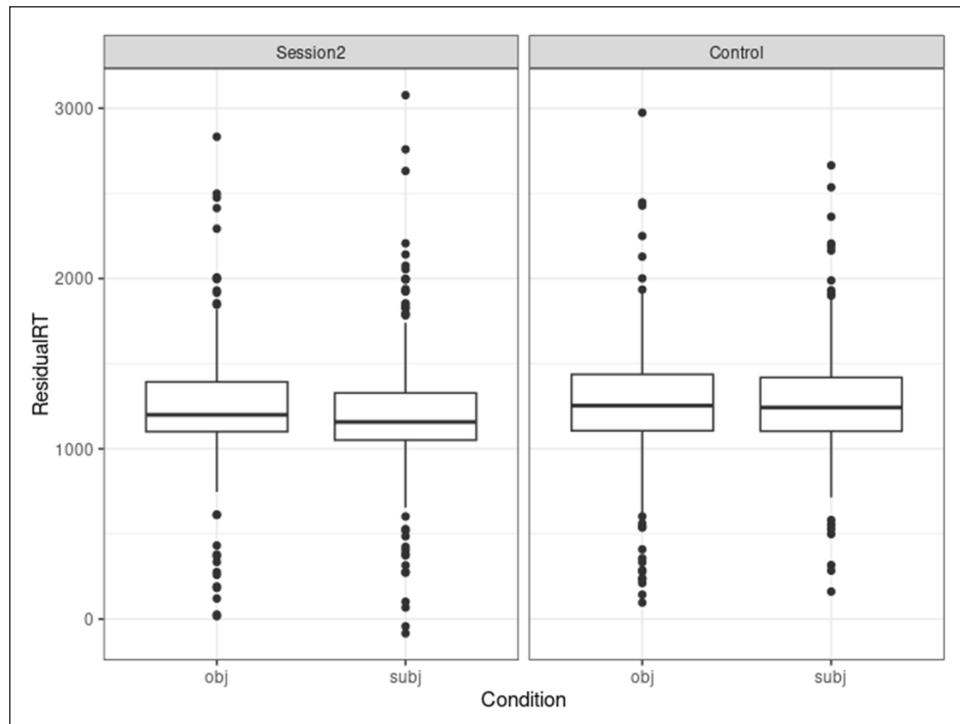


Figure 12: Overall residualized reading times across both sessions and conditions for region 6.

Finally, models probing for the effect of presentation order and experiment found a significant interaction between Experiment and Order (with list, sentence, and participant as random effects) for the Subject Island condition in region 6 ($\beta = -8.86$, $\text{Err} = 3.99$, $\text{CI} = [-15.31, -2.22]$, $\text{ER} = 65.67$, $P(\beta < 0) = 0.98$), suggesting that the participants in the second session (Experiment 2b) sped up more than those in the control session (Experiment 2b'), reading exactly the same items. No other regions were significant, and no such effect arose for Object items. **Figure 13** serves to illustrate the RTs across the two conditions and experiments. In sum, items are overall faster in Session 2b, likely in part due to task adaptation, but Subject Island items exhibit a trend for faster adaptation than in the Control experiment.

3.2.2.2 Discussion

A different processing profile was found for Experiment 2b in the second session compared to Experiment 2a in the first session: in spillover region 6, the processing advantage for Subject Island violations became more pronounced during the second session in comparison to the first

session. In contrast, although Object items also showed faster reading times overall, there was no additional speed-up over time during the second session. Rather, reading times of Object items merely stabilized with exposure, rather than slowing down, as they did in the first session.

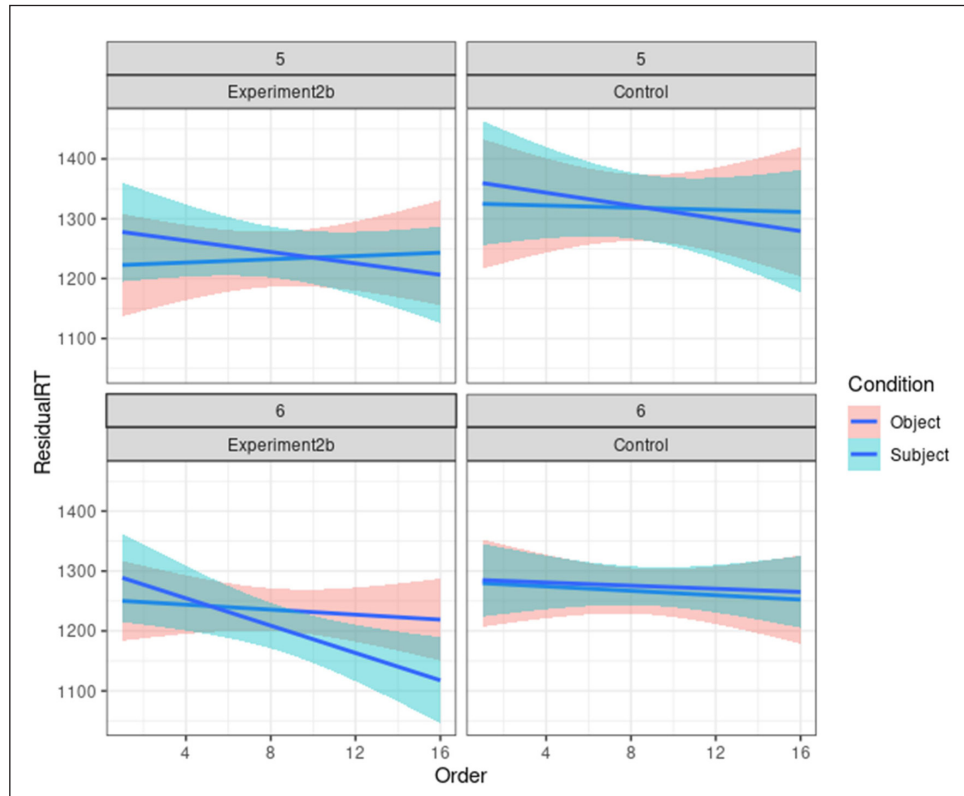


Figure 13: Item Residual RTs for regions 5 and 6 in Experiments 2b and 2b'(Control).

No speed-up was shown in region 6 for the Subject Island violations in the control group, which completed the second session experiments without having done the first session. Overall, the results suggest that adaptive behavior carried over between sessions, and across the span of the intervening three weeks.

4. General discussion

Previous research on syntactic satiation has shown that acceptability judgments of certain island-violating structures, including English Subject Islands, improve with repeated exposure over the course of an experiment session (Chaves & Dery, 2014, 2019; Clausen, 2011; Francom, 2009; Hiramatsu, 2000; Lu et al., 2021; Lu et al., 2022). Although the precise mechanisms underlying satiation in Subject Islands are debated, the extant satiation findings in Subject Islands align with the predictions of Bayesian belief-updating models, as proposed to account for adaptation effects

on reading times for English garden-path structures (Fine et al., 2010, 2013; Fine & Jaeger, 2013; Sikos et al., 2016). According to such models, language comprehenders adapt their expectations to align with the statistics of new linguistic input, to overcome processing disadvantages associated with mispredictions. Building on these previous studies, the current study tested a key prediction of belief-updating models: that adaptation may exhibit long-term effects. To accomplish this, we employed a longitudinal design with two sessions, spaced three weeks apart. Each session consisted of a self-paced reading task in between two sentence acceptability judgment tasks.

In the first session, the acceptability ratings for the Subject Island items gradually increased and converged towards the ratings for the Object condition, to the point of making the island effect vanish. Also in the first session, our findings for self-paced reading confirm that participants began to postulate subject-embedded gaps and adapted their expectations following repeated exposure to Subject Island violations. The gap site inside the subject (region 3) began exhibiting a plausibility effect, showing that participants posited a gap in this position. As comprehenders began to postulate subject-internal gaps, the processing of Subject Island violations was facilitated and the processing of Object controls was hampered, as the latter led to temporarily infelicitous parses. We also found gradual adaptation to subject-internal gaps manifesting as speed-up in the region where the island violation is perceived (regions 5 and 6). Since no subject-embedded gap is postulated in the Subject Island condition at the beginning of the experiment, the rest of the sentence becomes incoherent, leading to longer reading times in the spillover regions. But as participants adapted to such structures, the spillover effect was reduced in the Subject Island condition.

While the results from the first session are consistent with syntactic adaptation and in line with the previous findings of Chaves and Dery (2019), the most important evidence in support of an implicit learning account comes from the longitudinal comparisons. For the acceptability experiments, the ratings for Subject Island items were higher in the first part of the second session (Experiment 1b) compared to the first part of the first session (Experiment 1a) three weeks earlier, and they exhibited no further acceptability increases in the second part of the second session (Experiment 3b). No other condition exhibited this acceptability rating profile, including the ungrammatical distractors. Similarly, for the self-paced reading experiments, response times for Subject Island items in spillover region 6 got faster over time during each session, but sped up more in the second session compared with the first session. In contrast, response times for the Object items in the same region slowed down over time during the first session (consistent with a plausibility effect), and then showed no change over time during the second session. These patterns of results for both acceptability and self-paced reading are consistent with the presence of long-lived syntactic adaptation to Subject Islands. Since the effects were shown in Subject Island sentences but not ungrammatical fillers, they are unlikely to have resulted from task adaptation alone. This study thus supports an implicit learning account of satiation in Subject

Islands in which the typically low acceptability of Subject Islands is due to the probabilistic expectation that a gap is unlikely to occur within a subject phrase (Chaves & Dery, 2019).

What do these results imply for theoretical approaches to Subject Islands? We believe that they are equally compatible with syntax-external approaches and with syntactic approaches that assume gradient grammaticality. According to syntax-external approaches, extraction of a constituent from within a subject phrase is fully grammatical. Under this view, island effects result from various semantic, pragmatic, and processing factors which conspire to create a very low frequency of occurrence (Chaves, 2013; Chaves & Dery, 2019). It is, therefore, fully expected that an increase in input frequency, as induced by our experimental manipulations, should result in amelioration effects and possibly implicit learning. On the syntax-external view, the current results are in line with long-term syntactic adaptation, as shown, for example, with respect to fully grammatical but non-canonical OSV word order in German (Kroczek & Gunter, 2017).

Gradient grammaticality approaches differ from syntax-external approaches in that some (but not all) of the factors contributing to Subject Island effects are held to be syntactic in nature. These approaches, therefore, imply gradience within the competence grammar itself. As discussed in Section 1, Haegeman et al. (2014) follow Linear Optimality Theory (Keller & Sorace, 2003) in proposing a set of constraints that have language-specific weightings and which may apply cumulatively. Our stimuli from the Subject Island condition, as in (13a) above (*Which celebrity does the wife of reportedly quarrel with the fiancée of the mayor?*), violate some of their proposed constraints, but not others.¹² Importantly, repeated exposure to these types of sentences does not alter the number or nature of the constraints which are violated. However, because the constraints are formulated as ‘soft constraints’ which, by their nature, can be ameliorated by contextual factors, satiation effects and implicit learning can be easily accommodated.

Notably, Villata et al. (2019) and Villata and Tabor (2022) have more recently proposed a gradient grammaticality approach to island constraints which is similar in spirit, but differs in the details of implementation from that of Haegeman et al. (2014). In their self-organized sentence processing (SOSP) model, a dynamic and flexible structure-building system can create sub-optimal but interpretable parses for some island-violating structures. This ‘interpretable coercion’ is possible in cases where the island-violating structure shares certain features with a similar but fully grammatical structure. In the case of Subject Islands, Villata et al. (2019, p. 1183) claim that such coercion is not possible, due to the lack of any analogous grammatical structure in English. Thus, Subject Islands are predicted to be consistently unacceptable. Although this specific claim is difficult to reconcile with the current findings, and with other empirical facts

¹² Specifically, our stimuli violate the Freezing Condition, the Edge Condition, the Preposition Stranding Condition, the Inactivity Condition, and the Specificity Condition, while respecting the Argument Condition and the D-Linking Condition. We refer the reader to the original source for more details (Haegeman et al., 2014).

about such islands (e.g. exceptions like those listed in (2c–f) above in Section 1, as well as so-called parasitic circumventions; see Chaves and Putnam (2021, pp. 76–86) for an overview), the general idea of ‘interpretable coercion’ is not. For example, if we assume that extraction from a subject phrase is fully grammatical in relative clauses (Abeillé et al., 2020), then relative clauses could provide the analogous structure necessary for interpretable coercion of interrogative Subject Island sentences. Although the structure-building process is purely grammar-internal for Villata and colleagues, one can assume that ungrammatical structures which can undergo interpretable coercion should also be relatively amenable to grammar-external influences, such as frequency effects and repeated exposure effects. Thus, we see gradient grammaticality approaches such as Linear Optimality Theory and SOSP as compatible with our findings.¹³ Importantly, gradient grammaticality approaches readily incorporate syntactic constraints from current Minimalist or other formal generative theories.

In conclusion, the current results show that a low expectation for gaps within a subject phrase can be readjusted in response to changing patterns of input, potentially resulting in longer-term changes through implicit learning. While limited in scope, these findings contribute to a broader set of proposals that seek to reduce the explanatory burden of strict syntactic constraints and broaden the scope of linguistic competence to include gradient and probabilistic information (Culicover et al., 2022; Francis, 2021; Lau et al., 2017; Manning, 2003; Müller, et al. 2022; Villata & Tabor, 2022; Wasow, 2002).

¹³ Besides Linear Optimality Theory and SOSP, other gradient grammaticality approaches to extraction constraints include the Decathlon Model (Featherston, 2005) and Minimalist Gradient Harmonic Grammar (Müller et al., 2022).

Data accessibility statement

The authors have made available the data files and statistical analysis files at the following OSF repository: https://osf.io/m64ha/?view_only=e5c4197061ce47b2bb32bc0bf6062220

Ethics and consent

Prior to data collection, ethics approval was obtained from the Institutional Review Board (IRB) of Purdue University, the second author's institution (reference number: IRB-2022-254). In addition, an agreement was signed by an IRB designee of the University at Buffalo, the first author's institution, to cede IRB review and oversight to Purdue. Before beginning the experiment, online participants were required to read the IRB-approved consent screen and click to indicate their consent.

Acknowledgements

We gratefully acknowledge funding for this study from an Aspire Research Enhancement grant awarded by the College of Liberal Arts at Purdue University. We are grateful to Vanessa Sheu and Amy Hutchinson for their assistance with pilot testing the experiment. We appreciate the feedback we received from members of the ExLing lab at Purdue and from conference attendees at HSP 2023. Finally, we would like to thank three anonymous reviewers for *Glossa Psycholinguistics* for their constructive feedback, which helped us a lot in improving the readability and argumentation of the article.

Competing interests

The authors have no competing interests to declare.

Author contributions

The first author, Rui Chaves, developed the initial idea for the study and took the lead on experiment design, stimulus development, data interpretation, and manuscript writing (initial draft). He was solely responsible for implementing the online experiment, recruiting the participants on Prolific, and running the statistical analyses. The second author, Elaine Francis, assisted with experiment design, stimulus development, data interpretation, and manuscript writing (primarily the background and general discussion sections). The two authors contributed equally to securing funding and IRB approval for the study.

References

- Abeillé, A., Hemforth, B., Winckel, E., & Gibson, E. (2020). Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition*, *204*, 104293. DOI: <https://doi.org/10.1016/j.cognition.2020.104293>
- Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. DOI: [https://doi.org/10.1016/S0010-0277\(99\)00059-1](https://doi.org/10.1016/S0010-0277(99)00059-1)
- Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes*, *28*(4), 525–560. DOI: <https://doi.org/10.1080/01690965.2012.658072>
- Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, *129*(2), 177–192. DOI: <https://doi.org/10.1037/0096-3445.129.2.177>
- Boland, J. E., Atkinson, E., De Los Santos, G., & Queen, R. (2023). What do we learn when we adapt to reading regional constructions? *PLoS ONE*, *18*(4), e0282850. DOI: <https://doi.org/10.1371/journal.pone.0282850>
- Bošković, Ž. (2016). On the timing of labeling: Deducing Comp-trace effects, the subject condition, the adjunct condition, and tucking in from labeling. *The Linguistic Review*, *33*(1), 17–66. DOI: <https://doi.org/10.1515/tlr-2015-0013>
- Brown, J. M., Fanselow, G., Hall, R., & Kliegl, R. (2021). Middle ratings rise regardless of grammatical construction: Testing syntactic variability in a repeated exposure paradigm. *PLOS ONE*, *16*(5), e0251280. DOI: <https://doi.org/10.1371/journal.pone.0251280>
- Bürkner, P.-C. (2017). BRMS: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. DOI: <https://doi.org/10.18637/jss.v080.i01>
- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. DOI: <https://doi.org/10.1037/0033-295X.113.2.234>
- Chang, F., Janciauskas, M., & Fitz, H. (2012). Language adaptation and learning: Getting explicit about implicit learning. *Language and Linguistics Compass*, *6*(5), 259–278. DOI: <https://doi.org/10.1002/lnc3.337>
- Chaves, R. P. (2013). An expectation-based account of subject islands and parasitism. *Journal of Linguistics*, *49*(2), 285–327. DOI: <https://doi.org/10.1017/S0022226712000357>
- Chaves, R. P., & Dery, J. E. (2014). Which subject islands will the acceptability of improve with repeated exposure. In *Proceedings of the 31st West Coast Conference on Formal Linguistics* (pp. 96–106).
- Chaves, R. P., & Dery, J. E. (2019). Frequency effects in subject islands. *Journal of Linguistics*, *55*(3), 475–521. DOI: <https://doi.org/10.1017/S0022226718000294>
- Chaves, R. P., & Putnam, M. T. (2021). *Unbounded dependency constructions: Theoretical and experimental perspectives*. Oxford University Press. DOI: <https://doi.org/10.1093/oso/9780198784999.001.0001>

- Chomsky, N. (1973). Conditions on transformations. In S. R. Anderson & P. Kiparsky (Eds.), *A festschrift for Morris Halle* (pp. 232–286). Holt, Rinehart and Winston.
- Chomsky, N. (1977). *Essays on form and interpretation*. North-Holland.
- Chomsky, N. (2008). On phases. In R. Freidin, D. Michaels, C.P. Otero, & M. L. Zubizarreta (Eds.), *Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud* (pp. 133–165). MIT Press. DOI: <https://doi.org/10.7551/mitpress/9780262062787.003.0007>
- Clausen, D. R. (2011). Informativity and acceptability of complex Subject Islands [Poster presentation]. *24th Annual CUNY conference on human sentence processing*. Stanford, CA.
- Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? In D. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural language parsing: Psycholinguistic, computational, and theoretical perspectives* (pp. 94–128). Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511597855.004>
- Crawford, J. (2012). Using syntactic satiation to investigate subject islands. In *Proceedings of the 29th West Coast Conference on Formal Linguistics* (pp. 38–45). Cascadilla Proceedings Project.
- Culicover, P. W., Varaschin, G., & Winkler, S. (2022). The radical unacceptability hypothesis: Accounting for unacceptability without universal constraints. *Languages*, 7(2), 96. DOI: <https://doi.org/10.3390/languages7020096>
- Culicover, P. W., & Wexler, K. (1977). Some syntactic implications of a theory of language learnability. In P. W. Culicover, T. Wasow, & A. Akmajian (Eds.), *Formal syntax* (pp. 7–60). Academic Press.
- Culicover, P. W., & Winkler, S. (2022). Parasitic gaps aren't parasitic, or, the case of the uninvited guest. *The Linguistic Review*, 39(1), 1–35. DOI: <https://doi.org/10.1515/tlr-2021-2080>
- Cuneo, N., & Goldberg, A. E. (2023). The discourse functions of grammatical constructions explain an enduring syntactic puzzle. *Cognition*, 240, 105563. DOI: <https://doi.org/10.1016/j.cognition.2023.105563>
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117–1121. DOI: <https://doi.org/10.1038/nm1504>
- Dempsey, J., Liu, Q., & Christianson, K. (2020). Convergent probabilistic cues do not trigger syntactic adaptation: Evidence from self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(10), 1906–1921. DOI: <https://doi.org/10.1037/xlm0000881>
- Do, M. L., & Kaiser, E. (2017). The relationship between syntactic satiation and syntactic priming: A first look. *Frontiers in Psychology*, 8, 18–51. DOI: <https://doi.org/10.3389/fpsyg.2017.01851>
- Fanselow, G., Féry, C., Vogel, R., & Schlesewsky, M. (2006). *Gradience in grammar: Generative perspectives*. Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199274796.003.0001>
- Featherston, S. (2005). The decathlon model of empirical syntax. In S. Kepser & M. Reis (Eds.), *Linguistic evidence* (pp. 187–208). Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110197549.187>

- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, 41(4), 469–495. DOI: <https://doi.org/10.1006/jmla.1999.2660>
- Fine, A. B., & Jaeger, T. F. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science*, 37(3), 578–591. DOI: <https://doi.org/10.1111/cogs.12022>
- Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLOS One*, 8(10), e77661. DOI: <https://doi.org/10.1371/journal.pone.0077661>
- Fine, A., Qian, T., Jaeger, T. F., & Jacobs, R. (2010). Syntactic adaptation in language comprehension. In *Proceedings of the 2010 Workshop on Cognitive Modeling and Computational Linguistics* (pp. 18–26). Association for Computational Linguistics.
- Francis, E. J. (2021). *Gradient acceptability and linguistic theory*. Oxford University Press. DOI: <https://doi.org/10.1093/oso/9780192898944.001.0001>
- Francom, J. C. (2009). *Experimental syntax: Exploring the effect of repeated exposure to anomalous syntactic structure – Evidence from rating and reading tasks* [Unpublished doctoral dissertation]. The University of Arizona.
- Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory*, 5, 519–559. DOI: <https://doi.org/10.1007/BF00138988>
- Gallego, Á., & Uriagereka, J. (2007). Conditions on sub-extraction. In L. Eguren & O. Fernández-Soriano (Eds.), *Coreference, modality, and focus: Studies on the syntax–semantics interface* (pp. 45–70). John Benjamins. DOI: <https://doi.org/10.1075/la.111.04gal>
- Gibson, E. (2006). The interaction of top-down and bottom-up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language*, 54, 363–338. DOI: <https://doi.org/10.1016/j.jml.2005.12.005>
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199268511.001.0001>
- Goodall, G. (2011). Syntactic satiation and the inversion effect in English and Spanish wh-questions. *Syntax*, 14(1), 29–47. DOI: <https://doi.org/10.1111/j.1467-9612.2010.00148.x>
- Haegeman, L., Jiménez-Fernández, Á. L., & Radford, A. (2014). Deconstructing the subject condition in terms of cumulative constraint violation. *The Linguistic Review*, 31(1), 73–150. DOI: <https://doi.org/10.1515/tlr-2013-0022>
- Harrington Stack, C. M., James, A. N., & Watson, D. G. (2018). A failure to replicate rapid syntactic adaptation in comprehension. *Memory & Cognition*, 46, 864–877. DOI: <https://doi.org/10.3758/s13421-018-0808-6>
- Heathcote, A., Brown, S., & Mewhort, D. J. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, 7(2), 185–207. DOI: <https://doi.org/10.3758/BF03212979>
- Hiramatsu, K. (2000). *Assessing linguistic competence: Evidence from children's and adults' acceptability judgments* [Unpublished doctoral dissertation]. University of Connecticut.

- Hofmeister, P., & Sag, I. A. (2010). Cognitive constraints and island effects. *Language*, 86(2), 366–415. DOI: <https://doi.org/10.1353/lan.0.0223>
- Huang, C.-T. J. (1982). *Logical relations in Chinese and the theory of grammar* [Unpublished doctoral dissertation]. Massachusetts Institute of Technology.
- Jiménez-Fernández, Á. (2009) On the Composite Nature of Subject Islands: A Phase-Based Approach. *SKY Journal of Linguistics*, 22, 91–138.
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111(2), 228–238. DOI: <https://doi.org/10.1037/0096-3445.111.2.228>
- Kaan, E., & Chun, E. (2018). Syntactic adaptation. In K. D. Federmeier & D. G. Watson (Eds.), *The psychology of learning and motivation: Current topics in language* (pp. 85–116). Elsevier Academic Press. DOI: <https://doi.org/10.1016/bs.plm.2018.08.003>
- Kamide, Y., Scheepers, C., & Altmann, G. T. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, 32, 37–55. DOI: <https://doi.org/10.1023/A:1021933015362>
- Keller, F., & Sorace, A. (2003). Gradient auxiliary selection and impersonal passivization in German: An experimental investigation. *Journal of Linguistics*, 39(1), 57–108. DOI: <https://doi.org/10.1017/S0022226702001676>
- Kleinschmidt, D. F., Fine, A. B., & Jaeger, T. F. (2012). A belief-updating model of adaptation and cue combination in syntactic comprehension. In N. Miyake, D. Peebles, & R.P. Cooper (Eds.), *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 599–604). Austin, TX: Cognitive Science Society.
- Kluender, R. (2004). Are subject islands subject to a processing account? In B. Schmeiser, V. Chand, A. Kelleher, & A. Rodriguez (Eds.), *Proceedings of the 23rd West Coast Conference on Formal Linguistics* (pp. 101–125). Cascadia Press.
- Kroczek, L. O., & Gunter, T. C. (2017). Communicative predictions can overrule linguistic priors. *Scientific Reports*, 7(1), 17581. DOI: <https://doi.org/10.1038/s41598-017-17907-9>
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307(5947), 161–163. DOI: <https://doi.org/10.1038/307161a0>
- Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language*, 98(1), 74–88. DOI: <https://doi.org/10.1016/j.bandl.2006.02.003>
- Lau, J. H., Clark, A., & Lappin, S. (2017). Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge. *Cognitive Science*, 41(5), 1202–1241. DOI: <https://doi.org/10.1111/cogs.12414>
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9781139087759>
- Levy, R., Fedorenko, E., Breen, M., & Gibson, E. (2012). The processing of extraposed structures in English. *Cognition*, 122(1), 12–36. DOI: <https://doi.org/10.1016/j.cognition.2011.07.012>

- Levy, R., & Keller, F. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language*, 2(68), 199–222. DOI: <https://doi.org/10.1016/j.jml.2012.02.005>
- Liu, Y., Winckel, E., Abeillé, A., Hemforth, B., & Gibson, E. (2022). Structural, functional, and processing perspectives on linguistic island effects. *Annual Review of Linguistics*, 8, 495–525. DOI: <https://doi.org/10.1146/annurev-linguistics-011619-030319>
- Lu, J., Frank, M. C., & Degen, J. (2024). A meta-analysis of syntactic satiation in extraction from Islands. *Glossa Psycholinguistics*, 3(1), 1–33. DOI: <https://doi.org/10.5070/G60111425>
- Lu, J., Lassiter, D., & Degen, J. (2021). Syntactic satiation is driven by speaker-specific adaptation. In *Proceedings of the 43rd annual conference of the Cognitive Science Society* (pp. 1493–1499). eScholarship Publishing. DOI: <https://escholarship.org/uc/item/1dn4v3b4>
- Lu, J., Wright, N., & Degen, J. (2022). Satiation effects generalize across island types. In *Proceedings of the 44th annual conference of the Cognitive Science Society* (pp. 2724–2730).
- Luka, B. J., & Choi, H. (2012). Dynamic grammar in adults: Incidental learning of natural syntactic structures extends over 48h. *Journal of Memory and Language*, 66(2), 345–360. DOI: <https://doi.org/10.1016/j.jml.2011.11.001>
- Mak, W. M., Vonk, W., & Schriefers, H. (2008). Discourse structure and relative clause processing. *Memory & Cognition*, 36, 170–181. DOI: <https://doi.org/10.3758/MC.36.1.170>
- Manning, C. D. (2003). Probabilistic syntax. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 289–341). MIT Press. DOI: <https://doi.org/10.7551/mitpress/5582.003.0011>
- McInerney, A., & Sugimoto, Y. (2022). On dissociating adjunct island and subject island effects. *Proceedings of the Linguistic Society of America*, 7(1), 5207–5207. DOI: <https://doi.org/10.3765/plsa.v7i1.5207>
- Müller, G., Englisch, J., & Opitz, A. (2022). Extraction from NP, frequency, and minimalist gradient harmonic grammar. *Linguistics*, 60(5), 1619–1662. DOI: <https://doi.org/10.1515/ling-2020-0049>
- Ni, W., Crain, S., & Shankweiler, D. (1996) Sidestepping garden paths: Assessing the contributions of syntax, semantics and plausibility in resolving ambiguities. *Language & Cognitive Processes* 11(3), 283–334. DOI: <https://doi.org/10.1080/016909696387196>
- Omaki, A., Fukuda, S., Nakao, C., & Polinsky, M. (2020). Subextraction in Japanese and subject-object symmetry. *Natural Language & Linguistic Theory*, 38, 627–669. DOI: <https://doi.org/10.1007/s11049-019-09449-8>
- Phillips, C. (2006). The real-time status of island phenomena. *Language*, 82(4), 795–823. DOI: <https://doi.org/10.1353/lan.2006.0217>
- Pickering, M., Barton, S., & Shillcock, R. (1994). Unbounded dependencies, island constraints, and processing complexity. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 199–224). Lawrence Erlbaum Associates, Inc.
- Polinsky, M., Gallo, C. G., Graff, P., Kravtchenko, E., Milton Morgan, A., & Sturgeon, A. (2013). Subject islands are different. In J. Sprouse & N. Hornstein (Eds.), *Experimental syntax and island effects* (pp. 286–309). Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9781139035309.015>

- Prasad, G., & Linzen, T. (2019). Reassessing the evidence for syntactic adaptation from self-paced reading studies [Poster presentation]. *The 32nd annual CUNY conference on human sentence processing*. Boulder, CO.
- Prasad, G., & Linzen, T. (2021). Rapid syntactic adaptation in self-paced reading: Detectable, but only with many participants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(7), 1156–1172. DOI: <https://doi.org/10.1037/xlm0001046>
- Pritchett, B. L. (1991). Subjacency in a principle-based parser. In R. C. Berwick, S. P. Abney, & C. Tenny (Eds.), *Principle-based parsing: Computation and psycholinguistics* (pp. 301–345). Springer. DOI: https://doi.org/10.1007/978-94-011-3474-3_12
- Roland, D., Mauner, G., O'Meara, C., & Yun, H. (2012). Discourse expectations and relative clause processing. *Journal of Memory and Language*, 66(3), 479–508. DOI: <https://doi.org/10.1016/j.jml.2011.12.004>
- Ross, J. R. (1967). *Constraints on variables in syntax* [Unpublished doctoral dissertation]. Massachusetts Institute of Technology.
- Sikos, L., Martin, H., Fitzgerald, L., & Grodner, D. (2016). Memory-based limits on surprisal-based syntactic adaptation [Paper presentation]. *The 29th annual CUNY conference on human sentence processing*. Gainesville, FL.
- Snyder, W. (1994). A psycholinguistic investigation of weak crossover, islands, and syntactic satiation effects: Implications for distinguishing competence from performance [Poster presentation]. *The 7th annual CUNY conference on human sentence processing*. New York, NY.
- Snyder, W. (2000). An experimental investigation of syntactic satiation effects. *Linguistic Inquiry*, 31(3), 575–582. DOI: <https://doi.org/10.1162/002438900554479>
- Snyder, W. (2021). Satiation. In Grant Goodall (ed.), *The Cambridge Handbook of Experimental Syntax*, 154–180. Cambridge University Press. DOI: <https://doi.org/10.1017/9781108569620.007>
- Snyder, W. (2022). On the nature of syntactic satiation. *Languages*, 7(1), 38. DOI: <https://doi.org/10.3390/languages7010038>
- Sprouse, J. (2009). Revisiting satiation: Evidence for an equalization response strategy. *Linguistic Inquiry*, 40(2), 329–341. DOI: <https://doi.org/10.1162/ling.2009.40.2.329>
- Stowe, L. A. (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>
- Stowe, L. A., Tanenhaus, M. K., & Carlson, G. N. (1991). Filling gaps on-line: Use of lexical and semantic information in sentence processing. *Language and Speech*, 34(4), 319–340. <https://doi.org/10.1177/002383099103400402>
- Stromswold, K. (1986). *Syntactic satiation* [Unpublished manuscript]. Massachusetts Institute of Technology.
- Szabolcsi, A. (2006). Strong vs. weak islands. In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell companion to syntax* (pp. 479–531). Wiley Blackwell. DOI: <https://doi.org/10.1002/9780470996591.ch64>

- Tabor, W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, 12(2–3), 211–271. DOI: <https://doi.org/10.1080/016909697386853>
- Tollan, R. & Heller, D. (2016). Elvis Presley on an island: Wh dependency formation inside complex NPs. In C. Hammerly & B. Prickett (Eds.), *Proceedings of the 46th annual meeting of the North East Linguistic Society* (pp. 221–232). GLSA Publications.
- Traxler, M. J., & Pickering, M. J. (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>
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33(3), 285–318. DOI: <https://doi.org/10.1006/jmla.1994.1014>
- Villata, S., Sprouse, J., & Tabor, W. (2019). Modeling ungrammaticality: A self-organizing model of islands. In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st annual conference of the Cognitive Science Society* (pp. 1178–1184).
- Villata, S., & Tabor, W. (2022). A self-organized sentence processing theory of gradience: The case of islands. *Cognition*, 222, 104943. DOI: <https://doi.org/10.1016/j.cognition.2021.104943>
- Wasow, T. (2002). *Postverbal behavior*. CSLI Publications.
- Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, 58(2), 250–271. DOI: <https://doi.org/10.1016/j.cogpsych.2008.08.002>
- Yano, M. (2024). The adaptive nature of language comprehension. In M. Koizumi (Ed.), *Volume 2: Interaction between linguistic and nonlinguistic factors* (pp. 115–132). De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110778939-007>
- Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). DOI: <https://doi.org/10.17605/OSF.IO/MD832>

