A Meta-analysis of Syntactic Satiation in Extraction from Islands

Jiayi Lu Michael C. Frank Judith Degen

June 3, 2023

Abstract

Sentence acceptability judgments are often affected by a pervasive phenomenon called *satiation*: native speakers give increasingly higher ratings to initially degraded sentences after repeated exposure. Various studies have investigated the satiation effect experimentally, the vast majority of which focused on different types of island-violating sentences in English (sentences with illicit long-distance syntactic movements). However, mixed findings are reported regarding which types of island violations are affected by satiation and which ones are not. This paper presents a meta-analysis of past experimental studies on the satiation of island effects in English, with the aim to provide accurate estimates of the rate of satiation for each type of island, test whether different island effects show different rates of satiation, explore potential factors that contributed to the heterogeneity in past results, and spot possible publication bias. The meta-analysis shows reliable satiation for adjunct islands, the Complex NP Constraint (CNPC), subject islands, the *that*-trace effect, the *want-for* construction, and *whether*-islands undergo satiation, albeit at different rates. No evidence for satiation is found for the Left Branch Condition (LBC). Whether context sentences were presented in the original acceptability judgment experiments predicts the differences in the rates of satiation reported across studies. Potential publication bias is found among studies testing the CNPC and *whether*-islands. These meta-analytic results can be used to inform debates regarding the nature of island effects, and serve as a proof of concept that meta-analysis can be a valuable tool for linguistic research.

1 Introduction

Linguists have long relied on acceptability judgments by native speakers, collected either introspectively or experimentally to inform syntactic theory (Schutze, 1996). Recent studies found that acceptability judgments are affected by a pervasive phenomenon called *syntactic satiation*, or simply *satiation*: participants in acceptability rating experiments rate degraded sentences as increasingly acceptable as they see more instances of such sentences (Braze, 2002; Hiramatsu, 2001; Snyder, 2000, i.a.). The satiation effect has recently drawn much attention, and there is an abundance of experimental studies testing for the satiation of various unacceptable sentence types in English (Braze, 2002; Chaves and Dery, 2014, 2019; Crawford, 2012; Do and Kaiser, 2017; Francom, 2009; Goodall, 2011; Hiramatsu, 2001; Lu et al., 2021, 2022; Snyder, 2000, 2022a; Sprouse, 2009) and other languages (Abugharsa, 2016; Brown et al., 2021; Goodall, 2011; Myers, 2012; Sommer, 2022). However, the satiation literature is replete with mixed empirical findings and non-replications regarding which sentence constructions are affected by satiation (see Snyder, 2022b, for a review). Quantitative meta-analysis, a procedure commonly used to obtain more precise effect estimates by synthesizing past findings (Borenstein et al., 2009), has the potential to bring clarity to the empirical landscape on satiation.

In the current study, we present a meta-analysis of past findings in the satiation literature, with the aim to assess which sentence types reliably satiate. Specifically, we limit our focus to past studies examining the satiation of *island effects*: the degradation in acceptability of sentences that include long-distance syntactic movement operations that are illicit according to standard syntactic theories (Ross, 1967). There are two reasons to restrict the domain of study to island effects. First, the vast majority of past experimental studies on satiation has exclusively tested sentences with island violations, making island effects the only syntactic domain where a meta-analysis has sufficient statistical power and thus the potential to be informative. Second, findings from the literature on satiation have been extensively used to adjudicate between different theories in the domain of island effects. Therefore, the results of a meta-analysis on the satiation of island effects could help inform theoretical claims in the island literature.

In the remainder of the introduction, we introduce island effects and the satiation effect, respectively. We then report a meta-analysis we conducted on 25 island satiation experiments in Section 2. Finally, we discuss the implications of these results in Section 3, focusing especially on their potential for adjudicating between grammatical and processing accounts of island effects.

1.1 Island Effects

There is a long-standing generalization that certain structural domains restrict syntactic movement operations, a phenomenon termed "island effects" (Ross, 1967). Attempting to extract from islands results in degraded sentence acceptability, as in the examples in (1), all involving illicit wh-movement.

(1) Island-violating sentences (Snyder, 2000, p.576)

a. The Left Branch Condition
*How many$_i$ did John buy t_i books?

b. Adjunct island
*Who$_i$ did John talk with Mary after seeing t_i?

c. The Complex NP Constraint (CNPC)
*Who$_i$ does Mary believe the claim that John likes t_i?

d. Subject island
*What$_i$ does John know that a bottle of t_i fell on the floor?

e. The *that*-trace effect
*Who$_i$ does Mary think that t_i likes John?

f. The *want-for* construction
*Who$_i$ does John want for Mary to meet t_i?

g. *Whether*-island
*Who$_i$ does John wonder whether Mary likes t_i?

The nature of these island effects has been a long-standing source of debate in the linguistic literature. The degraded acceptability of island-violating sentences like (1a-g) has been variably attributed to constraints in *grammar* (Bresnan, 1976; Chomsky, 1964, 1973, 1977, 1986; Huang, 1982; Rizzi, 1990; Ross, 1967; Sag, 1976) or *processing* (Culicover et al., 2022; Hofmeister et al., 2013; Hofmeister and Sag, 2010; Keshev and Meltzer-Asscher, 2019; Kluender, 1991; Kluender and Kutas, 1993). Grammar-based approaches to island effects claim that sentences in (1) are ungrammatical because they violate certain syntactic constraints (e.g., the Subjacency Condition, the Phase Impenetrability Condition, etc.). Processing-based approaches to island effects claim that the sentences in (1) are grammatical, but are unacceptable due to the high processing burdens they incur (analogous to the difficulty in processing center-embedding constructions).

In addition to the debate over whether island effects are best explained as the result of grammatical or processing factors, there is a lack of consensus regarding whether certain islands form natural classes. For example, some have grouped subject and adjunct islands together as a natural class in opposition to the other island types, and have attributed the two distinct classes of islands to two different constraints in the grammar (Chomsky, 1986; Huang, 1982; Nunes and Uriagereka, 2000). Others, however, reject this grouping by either proposing that subject and adjunct island effects involve different grammatical constraints (Haegeman et al., 2014; Hiramatsu, 2001; Stepanov, 2007), or arguing for a unifying (syntactic, information structural, or processing-based) account for a larger set of island effects, including but not limited to subject and adjunct island effects (Abeillé et al., 2020; Bošković, 2016; Erteschik-Shir, 1973; Goldberg, 2013). In sum, the island literature lacks a consensus on the source and nature of island effects.

1.2 Satiation

The effect of repeated exposure on the perceived acceptability of island-violating sentences has been brought to bear on the debate over grammatical vs. processing accounts of island effects. A linking assumption that has been (implicitly or explicitly) adopted by many is that degraded acceptability due to grammatical violations should not be affect by repeated exposure. In contrast, if the source of degraded acceptability is processing difficulty, exposing participants to similar sentences of the same type should increase familiarity with this sentence type and ease the associated processing burden. In turn, acceptability should increase with repeated exposure (i.e., show the satiation effect).[1] If we accept this linking

---

[1] For discussions of this assumption and alternatives, see Snyder (2000), Hiramatsu (2001), Braze (2002), Hofmeister and Sag (2010), and Goodall (2011).

hypothesis, whether or not the acceptability of island-violating sentences increases with exposure can been used to diagnose whether certain island effects are grammatical or the result of processing constraints.

Acceptability increase after exposure, or the *satiation effect* (Stromswold, 1986), was first demonstrated experimentally for island-violating sentences in Snyder (2000).[2] The observation of satiation effects in island-violating sentences has subsequently been interpreted as evidence for the non-grammatical nature of islands, including the Complex-NP Constraint (Hofmeister et al., 2013; Hofmeister and Sag, 2010), the superiority effect (Hofmeister et al., 2011), and subject islands (Chaves, 2022; Chaves and Dery, 2014, 2019).

Other studies assume a different linking hypothesis for the satiation effect, whereby certain grammatical constraints may also be "unlearned" or weakened throughout repeated exposure to sentences violating those constraints, and differences in satiation profiles reflect differences in the types of grammatical constraints involved (Braze, 2002; Goodall, 2011; Hiramatsu, 2001; Snyder, 2000). Assuming this hypothesis, satiation results cannot be used to inform the grammar vs. processing debate regarding the nature of islands. Instead, satiation can be used to probe for natural classes formed by different islands. If two island types show different patterns of satiation (e.g., one satiates while the other does not), they are assumed to have different underlying sources of unacceptability and should not be grouped together as a natural class. This linking hypothesis underlies the argument against Huang (1982)'s proposal that subject and adjunct islands form a natural class (Hiramatsu, 2001; Stepanov, 2007).

In sum, while assuming slightly different linking hypotheses, the satiation effect has been used as evidence in multiple debates surrounding the nature of island effects. In the current study, our goal is not to further complicate the picture by taking sides in any of these debates. We also remain agnostic about which linking hypothesis for satiation is appropriate. Instead, we aim to clarify the empirical landscape on satiation so that satiation can be better leveraged as a source of evidence.

As mentioned earlier, the satiation literature abounds with mixed findings. Statistically significant satiation of island effects have been observed with some experimental procedures and items, but these effects have been inconsistent (see Snyder, 2022b, for a comprehensive overview). For example, the adjunct island is found to satiate in (Chaves and Putnam, 2020), but not in (Crawford, 2012; Francom, 2009; Hiramatsu, 2001; Snyder, 2000; Sprouse, 2009, *inter alia*). For a comprehensive list of past satiation studies and whether they observed satiation in each island type, the reader is referred to Tables 4 and 5 from Snyder (2022a). The current study reports a quantitative meta-analysis of past satiation studies, with the aim to summarize and aggregate past findings in the service of assessing which island types reliably satiate.

---

[2] The term "syntactic satiation" was first used in Stromswold (1986), and was defined as the decrease in certainty in participants' judgments for sentences after repeated exposure. This is different from the working definition for satiation that we adopt: following the various studies since Snyder (2000), we define syntactic satiation as the increase in acceptability rating throughout repeated exposure.

2 A Meta-Analysis of Island Effect Satiation

Meta-analysis is a way to systematically synthesize evidence from multiple studies to estimate effect size more precisely than is possible in an individual study, and discover inconsistencies between studies (Borenstein et al., 2009). This practice is particularly important in a field like linguistics, where many experimental studies are plagued by low statistical power, leading to mixed results and non-replications (Vasishth and Gelman, 2021; Vasishth et al., 2018).

The main goal of a meta-analysis is to give an effect size estimate informed by the effect size estimates from multiple studies. In particular, the effect size estimate is taken to be the average of all effect sizes weighted by how reliable or informative they are based on their variance. Different statistical models can be used to achieve this goal, which makes different assumptions about the homogeneity of the effect: fixed-effect or random-effect meta-analytic models.[3]

The fixed-effect model provides an estimate of the effect size $\hat{\mu}$ as an average of each study's point estimate of the effect $\hat{\theta}_i$ weighted by the inverse of the variance of the data, as shown in equation (1). This weighted average approach is intuitively justified: larger and more informative studies with less variance should be given more weight in the model compared to smaller studies with greater variance.

$$\hat{\mu} = \frac{\sum_i \frac{1}{\hat{\sigma}_i^2}\hat{\theta}_i}{\sum_i \frac{1}{\hat{\sigma}_i^2}} \tag{1}$$

In many cases, the estimated effect may vary across the studies included in a meta-analysis due to differences in experimental methods or the sampling process. The fixed-effect meta-analytic model, which assumes a single population effect size for all studies, does not take such heterogeneity into account. In contrast, a random-effect meta-analytic model assumes the population effects of all studies come from a normal distribution with mean μ and standard deviation τ, as shown in (2).

$$\theta_i = \mu + \epsilon_i, \text{ where } \epsilon_i \sim N(0, \tau^2) \tag{2}$$

The model then provides estimates for τ in addition to the effect estimate μ. As shown in equation 3, the random-effect meta-analytic model estimates the effect size as an average of each study's point estimate, weighted by the inverse of each study's variance adjusted by the estimated between-study variance $\hat{\tau}^2$.

$$\hat{\mu} = \frac{\sum_i \frac{1}{\hat{\tau}^2+\hat{\sigma}_i^2}\hat{\theta}_i}{\sum_i \frac{1}{\hat{\tau}^2+\hat{\sigma}_i^2}} \tag{3}$$

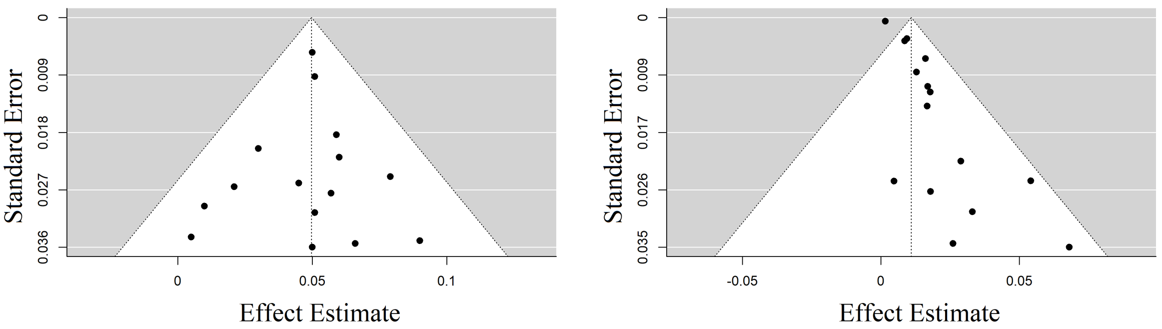
Since there is variation in methods and designs across satiation studies, we use the random-effect model instead of the fixed-effect model for our meta-analysis.

---

[3] These are not to be confused with fixed-effect and mixed-effects regression models.

The random-effect meta-analytic model does not provide a structured analysis of the factors that contribute to the cross-study heterogeneity. To investigate heterogeneity, one can include different moderators (study-level factors that may affect effect size) to form a mixed-effect meta-analytic model as in equation (4), where x_{ij} represents the j^{th} moderator for the i^{th} study. From the parameter estimates β_1 to β_j, we can infer whether the moderators influence the effect size.

$$\theta_i = \mu + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_j x_{ij} + \epsilon_i, \text{ where } \epsilon_i \sim N(0, \tau^2) \tag{4}$$



(a) A hypothetical funnel plot without publication bias.

(b) A hypothetical funnel plot with publication bias.

Figure 1: Hypothetical funnel plots showing standard error against effect estimate for each individual study. The dashed vertical line indicates the meta-analytic effect size estimate.

Finally, meta-analyses can also be used to detect publication bias. One simple way to do so is by creating a "funnel plot": a scatter plot of effect size against standard error (Light and Pillemer, 1984).[4] In absence of heterogeneity or publication bias, studies with smaller standard errors are expected to have effect sizes closer to the meta-analytic estimate, and those with larger standard errors to spread out further. Thus, the scatter plot should show a funnel shape (hence the name "funnel plot"), as shown in the hypothetical plot in Figure 1a. The white funnel-shaped area in the plot represents the 95% confidence interval for the observed effects calculated based on the meta-analytic estimate and the standard error and serves as a visual aid for what the funnel-shaped distribution should look like in absence of any publication bias. In contrast, if there is publication bias, studies with positive effect estimates are more likely to be reported. As a result, the funnel plot should show an asymmetric distribution of the shape of a right-skewed triangle rather than a funnel. An example is shown in Figure 1b. The existence of a publication bias can be statistically confirmed using Egger's regression test on funnel plot asymmetry (Egger et al., 1997), which detects a correlation between the effect size and standard error. A significant correlation suggests that the funnel plot is asymmetric and that there is potentially a publication bias.

---

[4] There are also more sophisticated models for testing publication bias available. See Hedges and Vevea (2005) for examples of non-graphical methods of detecting publication bias.

In the study reported below, we conducted a meta-analysis of satiation in seven different island types. We report analyses of satiation effect estimates, heterogeneity, and publication bias for each island type.

2.1 Method

2.1.1 Dataset Selection

The study selection process is summarized in Figure 2, which depicts a PRISMA flow chart (Moher et al., 2009). Our goal was to include as many studies as possible on island satiation effects in English. To this end, we first collected all results returned on the first 20 pages of Google Scholar with the search keywords "syntactic satiation" (200 entries). Excluding 3 duplicate entries, 3 non-English entries, and 20 entries without links to full text or abstract, we then screened the abstracts of the remaining 174 entries and narrowed the selection down to 24 entries reporting experimental studies on satiation. After accessing the full text of these 24 entries, we excluded 5 entries that did not study the satiation of island effects in English, and 3 entries whose experimental results were also reported in other publications by the same authors.

Since not all studies used the same method of data processing and statistical analysis, we reached out to all authors for the raw data files, so that we could extract effect size estimates from the data in a systematic and comparable manner. Our final analysis included all selected papers whose raw data files were kindly made available by the authors, in addition to those that directly reported the measurements we planned to use in the meta-analysis (unit satiation per repetition), which we introduce in detail in the next section. Five papers were excluded from the meta-analysis because the relevant effect sizes were not reported and could not be computed from the reported statistics, and the data files were not made available.

For the purpose of standardizing effect estimates across different studies, we only summarized experiments that employed acceptability judgment experiments with closed rating scales.[5] Although this meta-analysis was not pre-registered, all inclusion criteria are determined before computing meta-analytic estimates. The following studies were included based on our selection criteria: Snyder (2000), Experiments 1 and 2 of Francom (2009), the three replication experiments reported in section 2 of Sprouse (2009) (labeled Exp.1a, b, c respectively in the discussion below), and the two experiments reported in section 4 of the same paper (labeled Exp.2a and b below), experiments 1 and 2 of Francom (2009), Hofmeister and Sag (2010), Crawford (2012), experiments 1 and 2 of Chaves and Dery (2014), experiment 1 of Do and Kaiser (2017) (the two sub-experiments "lag 1" and "lag 5" in the original paper are labeled as Exp.1a and Exp.1b respectively in the discussion below for simplicity), Chaves and Dery (2019), the experiments reported in chapters 6.2.2

---

[5] We did not include open-scale acceptability judgment experiments (e.g., magnitude estimation experiments) because there is no well-justified way of converting such scales to a closed interval such that the results can be aggregated with the other studies for meta-analytic purposes. Five experiments from (Sprouse, 2009) were excluded for this reason. Note that the majority of studies on syntactic satiation utilized some form of the closed rating scale (e.g. 5/7-point Likert scale, 0-1 continuous slider scale).

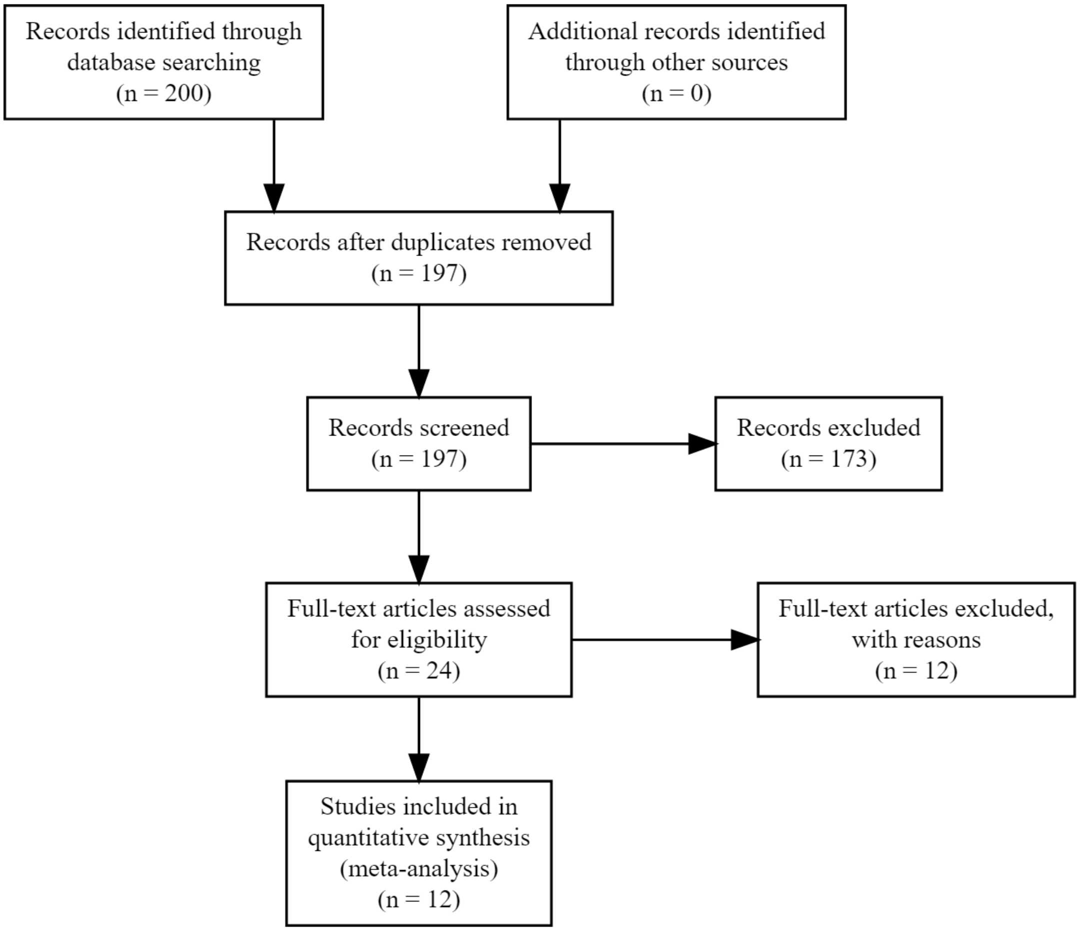


Figure 2: PRISMA flow chart summarizing the study selection process.

Study	Sample size	Reported satiating island(s)	Reported non-satiating island(s)
Snyder (2000)	22	CNPC, SI*, WI	AI, LBC, TT, WF
Francom (2009) Exp.1	205	SI, WF, WI	AI, CNPC, LBC, TT
Francom (2009) Exp.2	17	SI	CNPC, AI, TT, LBC
Sprouse (2009) Exp.1a	21		AI, CNPC, LBC, SI, TT, WF, WI
Sprouse (2009) Exp.1b	21		AI, CNPC, LBC, SI, TT, WF, WI
Sprouse (2009) Exp.1c	22		AI, CNPC, LBC, SI, WI
Sprouse (2009) Exp.2a	25		AI, CNPC, CSC, LBC, RC, SI, WI
Sprouse (2009) Exp.2b	25		AI, CNPC, RC, WI
Hofmeister and Sag (2010)	22	CNPC	
Crawford (2012)	22	WI	AI, SI
Chaves and Dery (2014) Exp.1	60	SI	
Chaves and Dery (2014) Exp.2	51	SI	
Do and Kaiser (2017) Exp.1a	44	CNPC	SI
Do and Kaiser (2017) Exp.1b	40		CNPC, SI
Chaves and Dery (2019)	48	SI	
Chaves and Putnam (2020) Exp.1	74	SI	
Chaves and Putnam (2020) Exp.2a	40	AI	
Chaves and Putnam (2020) Exp.2b	40	AI	
Chaves and Putnam (2020) Exp.3	106	AI	
Lu, Lassiter, and Degen (2021) Exp.1	106	CNPC, SI, WI	
Lu, Lassiter, and Degen (2021) Exp.2	102	CNPC, SI, WI	
Lu, Wright, and Degen (2022) Exp.1	294	SI, WI	
Lu, Wright, and Degen (2022) Exp.2	311	SI, WI	
Snyder (2022) Exp.1	20	WI	CNPC, LBC, SI
Snyder (2022) Exp.3	151	CNPC, WI	AI, LBC, SI, TT, WF

*: Only marginal significance reported.

Table 1: Summary of all studies included in the meta-analysis. For some studies, experiment numbering is added for ease of presentation (see section 2.1.1 for details). Each acronym refers to a type of island effect tested. AI: adjunct island; CNPC: complex NP constraint; CSC: coordinate structure constraint; LBC: left branch condition; RC: relative clause island; SI: subject island; TT: *that*-trace effect; WF: *want-for* construction; WI: *whether*-island.

(labeled Exp.1 below), 6.2.4 (the two sub-experiments labeled Exp.2a and 2b below), and 6.2.5 of Chaves and Putnam (2020) (labeled Exp.3 below), experiments 1 and 2 of Lu et al. (2021), experiments 1 and 2 of Lu et al. (2022), and experiments 1 and 3 of Snyder (2022a).[6] See Table 1 for a summary of the studies included and their reported findings. Please note that the studies in Table 1 have vastly different sample sizes, and used different data processing and hypothesis testing methods to reach their conclusions on which island effects satiate. Therefore, any "vote-counting" procedure by comparing the numbers of studies finding significant satiation effects for each island type would not be informative.

---

[6] As noted by Snyder (2022a), the CNPC condition in Francom (2009)'s experiment 1 includes several sentence tokens that in fact do not contain CNPC violation. Therefore, the CNPC satiation results in that study might be confounded. However, We still include all the data from Francom (2009) for systematicity. Note that removing the data from Francom (2009) does not lead to any qualitative change in the meta-analytic results.

2.1.2 Meta-Analytic Methods

We first grouped the selected studies by the island types they investigated. When a paper contained multiple experiments, each experiment was treated as a different study. When a study tested multiple variants of the same island effect type (e.g., the subject island effect induced by extraction from DP subjects vs. CP subjects), the different variants were treated as the same island type for the purpose of our analysis. A total of seven island types (those shown in (1)), each studied in at least three experiments, were included in the meta-analysis.

We defined satiation as a positive main effect of the number of repetitions of an island-violating sentence type on the acceptability of that sentence type.[7] Most meta-analyses are conducted across standardized unitless effect size measures (e.g.: Cohen's d) to ensure comparability of effects across studies. However, we are interested in a particular effect size that has interpretable units: change in acceptability (0-1) per sentence repetition. If we computed standardized effects as is typical, we would run the risk of combining effects from studies with different definitions for satiation, using different satiation manipulations, and measuring acceptability with different scales. In contrast, our acceptability per repetition effect is much more interpretable in its magnitude across studies than a standardized effect. Therefore, we depart from standard meta-analyses, and compute effect sizes and measures of variation directly from raw data files of the original studies, either available in the public domain or provided by the authors upon request.

To compute the quantity needed for meta-analysis, we first linearly transformed the acceptability ratings for each study to a value between 0 and 1 through min-max scaling, with 0 representing the "completely unacceptable" endpoint of the scale, and 1 representing the "completely acceptable" endpoint. For studies that directly reported the repetition number effects on acceptability ratings, we directly used the reported estimates, standard errors, and sample sizes in the meta-analysis. For the rest of the studies, we fit linear mixed-effect regression models predicting the adjusted acceptability ratings of each island-violating sentence type with a fixed effect of repetition number. Each model also included random by-participant and by-item intercepts and slopes for the fixed effect when the participant and item information was provided in the data files. In cases of non-convergence, random effects with the least variance were removed until convergence. We recorded the repetition number effect estimates and standard errors for meta-analysis. The process was repeated for each type of island effect.

Using the *metafor* package (Viechtbauer, 2010) in R, we fit a random-effect meta-analytic model for studies testing each island effect type. Then, Cochran's Q test (Cochran, 1950) was used to detect any cross-study heterogeneity. In case of significance, we fit a mixed-effect meta-analytic model to examine different moderators as possible sources of heterogeneity. Snyder (2022a) speculated that differences in scale types, total numbers of repetitions, and the use of context sentences [8] could contribute to different findings in

---

[7] We chose the number of repetitions instead of the overall experimental trial number because different studies included different numbers of fillers, and not all raw data files provided by the authors contained filler information.

[8] Context sentences used in satiation experiments are usually just the declarative form of the interrogative

satiation experiments. Therefore, we included these three factors as moderators.

Finally, a funnel plot (a plot of standard errors against point estimates) was created for each island type. The Egger's regression test (Egger et al., 1997) on funnel plot asymmetry was conducted to detect publication bias.

2.2 Results

Below, we report the meta-analysis results for satiation studies examining the seven different island effects listed above: the Left Branch Condition (LBC), adjunct islands, the complex NP constraint (CNPC), subject islands, the *that*-trace construction, the *want-for* construction, and *whether*-islands.

2.2.1 Satiation effect estimates

Figure 3 summarizes the satiation effect estimates and 95% confidence intervals of the estimates obtained from the random-effect meta-analytic models for each island type. The effect sizes represent the estimated increase in acceptability on a 0-1 scale per repetition. A positive effect estimate with a 95% confidence interval not overlapping with 0 is taken as evidence for satiation. Based on the random-effect meta-analysis, we found significant evidence for the satiation of adjunct islands, the CNPC, subject islands, the *that*-trace construction, the *want-for* construction, and *whether*-islands. There was no evidence for the satiation of the LBC. Figures 4-10 are forest plots summarizing all selected studies that tested each of the seven island types.

In Figure 3, there appear to be varying rates of satiation among the six island effect types that do satiate. This observation is confirmed by an analysis whereby we pooled the results of the satiating island types (all but LBC) and fit a mixed-effect meta-analytic model predicting the rate of satiation with island type as a Helmert-coded predictor (ordered by satiation effect size estimates from small to large, as shown in Figure 3). The analysis revealed significance for CNPC ($z = 2.28$, $p < 0.05$) and *whether*-islands (z=2.06, $p < 0.05$), suggesting that these two island types each shows a significantly higher satiation rate compared to the means of the previous levels. We can thus conclude that among the satiating island types, there are varying rates of satiation.

2.2.2 Heterogeneity

The random-effect meta-analytic models revealed significant cross-study heterogeneity in three of the seven island types tested: subject islands, the CNPC, and *whether*-islands. Results from the Cochran's Q test of heterogeneity for each island type are shown in Table 3. For the three islands showing significant heterogeneity, we fit mixed-effect meta-analytic models with three moderators to explore the sources of heterogeneity: context (whether or not a context sentence was provided in the experiment), the total number of repetitions, and scale type (binary vs. multi-point). Categorical moderators were sum coded. The

---

test sentences, presented along with the test sentences to participants. We follow Snyder (2022a) in calling these sentences "context sentences".

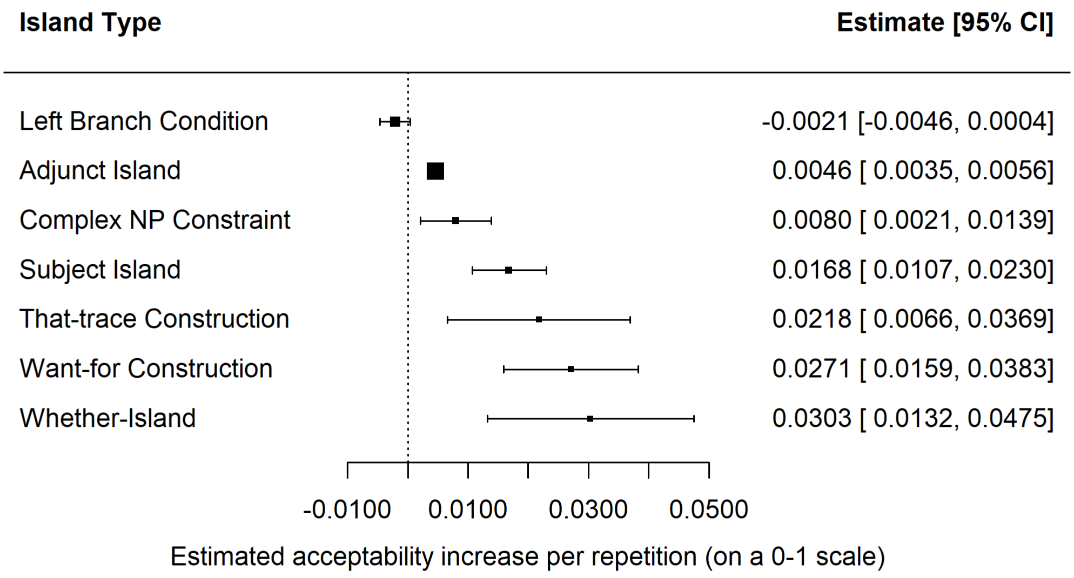


Figure 3: Forest plot summarizing estimates of satiation rate for all seven island types. Error bars represent 95% CIs. Area of squares represents the weight given to each study based on its standard error and the estimated cross-study variance.

moderator analyses results are shown in Table 4. For subject islands and *whether*-islands, inclusion of a context sentence in the task resulted in greater satiation effects. None of the other moderator-island pairs reached significance. For all three island types, there was significant residual heterogeneity even when the three moderators were included in the meta-analytic model, suggesting that additional moderators might modulate the rate of satiation.

2.2.3 publication bias

To assess whether these results are affected by publication bias, Figures 11a-11g contain funnel plots for the selected studies of each island type. The funnel plots for CNPC and *whether*-island are visibly asymmetric[9]. The asymmetry is further confirmed by Egger's regression test results, as shown in Table 2. This suggests that there is possible publication bias among the CNPC and *whether*-island studies, so we should take the positive effect estimates for these two island types with a grain of salt.

---

[9] The funnel plot for subject island also appears asymmetric, but Egger's test does not show significance. The high number of studies that fall outside the funnel area in Figure 11a could result from high cross-study heterogeneity.

Island type	Egger's results	
	t	*p*
Subject Island	1.36	< 0.190
Complex NP Constraint	2.08	< 0.059
Whether-Island	1.91	< 0.081
Adjunct Island	-0.66	< 0.521
Left Branch Condition	0.61	< 0.564
That-trace Effect	0.74	< 0.495
Want-for Construction	0.39	< 0.720

Table 2: Egger's regression test results for each island type. Marginally significant effects (suggesting possible publication bias) are shaded in light gray.

Island type	Heterogeneity measures		
	Cochran's Q	*p*	I^2
Subject Island	346.4	<0.001	94.06%
Complex NP Constraint	31.19	<0.006	69.07%
Whether-Island	326.91	<0.001	98.38%
Adjunct Island	8.65	<0.733	0.00%
Left Branch Condition	2.21	<0.978	0.00%
That-trace Construction	7.31	<0.293	9.26%
Want-for Construction	3.28	<0.657	0.00%

Table 3: Heterogeneity measures from random-effect meta-analytic models. Statistically significant effects are shaded in gray.

Island Type	Moderators in the mixed-effect models						Residual heterogeneity	
	Context inclusion		Scale type		Repetition number			
	t	*p*	*t*	*p*	*t*	*p*	Cochran's Q	*p*
Subject Island	2.37	<0.031	-1.17	<0.259	1.02	<0.321	202.49	<0.001
Complext NP Constraint	1.12	<0.287	-1.43	<0.180	-0.08	<0.936	22.31	<0.023
Whether Island	2.99	<0.014	-1.55	<0.154	-0.34	<0.738	60.52	<0.001

Table 4: Mixed-effect meta-analytic model results for island types showing significant heterogeneity in the random-effect analyses. Statistically significant effects are shaded in gray.

3 General Discussion

3.1 Summary

In this study, we conducted a meta-analysis of past experimental studies on the satiation of island-violating sentences in English. The results of this study provide answers to the following three questions. First, which types of islands reliably satiate? Second, which island types display heterogeneity in satiation, and which factors contribute to this heterogeneity? Third, is there evidence for a publication bias in the satiation literature?

To answer the first question, random effects meta-analytic models revealed significant acceptability increases for repetition in sentences violating constraints on adjunct islands, complex noun phrase islands, subject islands, the *that*-trace effect, the *want-for* construction, and *whether*-islands. In contrast, the models did not reveal evidence of acceptability increases for sentences violating the Left Branch Condition (see Figure 3).

To answer the second question, significant cross-study heterogeneity was detected among the CNPC studies, the subject island studies, and the *whether*-island studies. Following speculation by Snyder (2022a), we tested the contribution of three moderators (presence of context sentences in the experiment, total number of repetitions, and the scale type) to the heterogeneity. For CNPC studies, there was no evidence for any of the moderators modulating satiation; for subject island studies and *whether*-island studies, the presence of a context sentence increased the rate of satiation, while the other two moderators did not reach significance. For all three groups of studies, residual heterogeneity persisted even when the three moderators were included in the mixed-effect model, suggesting that additional moderators contribute to cross-study heterogeneity.

One such moderator may be the number of fillers. The increase in acceptability of the critical items due to exposure might gradually decay when participants see many unrelated filler items. Thus, the rate of satiation might be smaller when more filler items are included in the experiment. However, since not all selected studies reported filler information in the original papers or in the data files shared by the authors, we could not include the number of fillers as a moderator. Another factor that could affect satiation is the inclusion of multiple island types in the same experiment. It has been demonstrated that exposure to one island-violating sentence type leads to a change in the acceptability of another island-violating sentence type (Lu et al., 2022). Therefore, it is possible that when multiple island sentence types are tested together in a single experiment, they might influence each other's rate of satiation. Cross-study heterogeneity could then arise as a result of different studies testing different sets of island sentence types.

To answer the third question, there is possible publication bias favoring the studies reporting significant satiation for sentences violating the CNPC and the *whether*-island. However, the evidence is not strong and is based on marginally significant results from Egger's regression test of funnel plot asymmetry. We did not find a funnel plot asymmetry in any other group of studies, suggesting that there is no evidence for a publication bias in those groups of studies. This suggests that the evidence for the satiation of adjunct islands, subject islands, the *that*-trace construction, and the *want-for* construction is reliable.

3.2 Implications

The results of this meta-analysis are valuable in at least four different ways: in the debate over grammatical vs. processing accounts of islands, in the debate between different linking hypotheses for satiation, in the debate over whether subject and adjunct islands form a natural class in the taxonomy of islands, and in revealing the varying rates of satiation of different islands as a future direction of research.

First, the results may inform the debate over grammatical and processing accounts of islands. As discussed in Section 1, it is often implicitly assumed that when a degraded sentence type satiates, the source of the degradation should come from non-grammatical factors like high working memory burden (Hofmeister et al., 2013; Hofmeister and Sag, 2010) or low frequency (Chaves and Dery, 2014, 2019). Assuming that this linking hypothesis is accurate (and we shall return to the possibility that it is not), the results reported here can be used to inform theories of island effects: the reliable satiation effects for adjunct islands, CNPC, subject islands, the *that*-trace construction, the *want-for* construction, and *whether*-islands suggests they should be considered grammatical but degraded due to processing factors. In contrast, LBC does not satiate, and therefore should be considered ungrammatical. These grammaticality statuses pose a challenge to syntactic theories that predict the ungrammaticality of the satiating island types. This includes the Government and Binding (GB) theory, which attributes island effects to grammatical constraints like the Subjacency Condition and the Empty Category Principle (Chomsky, 1986; Huang, 1982), and syntactic theories under the minimalist program that attribute island effects to the cyclic nature of spell-out and linearization(Fox and Pesetsky, 2005; Nunes and Uriagereka, 2000). In contrast, our results are in general compatible with syntactic theories without non-local syntactic constraints. For example, in certain versions of the Head-driven Phrase Structure Grammar (HPSG) (Boas and Sag, 2012; Michaelis, 2013), only island effects that can be framed in terms of local constraints are predicted to exist without arbitrary stipulations of filtering constraints. All the island effects that reliably satiate according to this meta-analysis are predicted to be grammatical in this framework (see Chaves and Putnam (2020) for a detailed discussion of the predicted island effects under this framework). In contrast, the LBC, where we found no evidence for satiation, can be captured in HPSG by an independently motivated local constraint requiring that only elements in the ARG(UMENT)-ST(RUCTURE) feature list of a head can appear in the GAP feature list of the same head. Possessors and modifiers are not part of the ARG-ST list of an N head, and thus cannot appear in the GAP list of N (i.e., cannot be extracted from an NP, Chaves and Putnam, 2020; Sag, 2012).

Second, the discussion up to this point assumes the linking hypothesis that grammaticality decides the satiability of degraded sentences. However, this hypothesis is not unchallenged. Other factors have been suggested to affect the satiability of sentences, including whether the violated grammatical constraint is part of Universal Grammar (UG) (Braze, 2002; Hiramatsu, 2001; Snyder, 2000), the surface similarity with a grammatical alternative (Sprouse, 2007), and sentence interpretability (Francom, 2009). Instead of assuming a particular linking hypothesis under which to test theories of island effects, one can also use the meta-analysis results to inform the linking hypothesis itself. For example, the hypothesis that grammatical principles in UG determine satiability is rejected by our results. Under this hypothesis, LBC, which is the only non-satiating island type among the ones investigated, should be a principle in the UG. However, LBC is in fact subject to variation cross-linguistically. Left branch extraction (extraction of modifiers or possessors out of an NP) is permitted in many Slavic languages (Bošković, 2005). For example, Serbo-Croatian allows the equivalent of "Whose$_i$ does Petko like t$_i$ car?" as shown in (2). This shows that LBC cannot be a grammatical principle encoded in UG, thus rejecting the hypothesis that satiation diagnoses UG principles.

(2) Left Branch Extraction in Serbo-Croation (Bošković, 2005, pp.3)

Čija$_i$	xaresva	Petko	t$_i$	kola?
Whose$_i$	like	Petko	t$_i$	car

"Whose car does Petko like?

Third, our results also have implications for the debate regarding whether subject and adjunct islands are reducible to the same underlying constraint. In syntactic theories, among the various types of island effects, adjunct islands and subject islands are traditionally considered to form a natural class. For example, Huang (1982) attributes both adjunct and subject island effects to a single syntactic principle: the Condition on Extraction Domains (CED), which states that constituents that are not properly governed restrict extraction from within. Subject DPs and adjuncts are not properly governed, and therefore are both CED islands.[10] Chomsky (1986), which aims to provide an analysis for all island effects using the concept of "barriers", also grouped subject and adjunct islands together. Movements out of adjuncts and subjects need to cross two barriers, whereas movements out of other islands (e.g., *whether*-islands, complex NPs) cross only one barrier. Studies including Hiramatsu (2001) and Stepanov (2007) question Huang (1982)'s and Chomsky (1986)'s accounts on the grounds that subject island sentences satiate while adjunct island sentences do not. Assuming a linking hypothesis under which structures with a common source of unacceptability should show a similar satiation profile (Braze, 2002; Goodall, 2011; Snyder, 2000), we would expect subject and adjunct islands to either both satiate or both resist satiation, contrary to Hiramatsu (2001)'s observation. However, as evident in the current meta-analysis, there is reliable evidence that both subject and adjunct island sentences satiate. Although this result cannot distinguish between the CED or the barriers account, at least it provides sufficient grounds for rejecting Hiramatsu (2001)'s counterargument.

Fourth, the results point to a future direction of research. As mentioned in Section 2.2.1, there appear to be varying rates of satiation among the satiating island types. The differences in the rates of satiation might signal differences in the linguistic properties underlying these sentence types and could potentially become a useful diagnostic tool for experimental syntacticians. Differences in rates of satiation could reflect different sources of unacceptability (see Goodall, 2011, for a similar proposal). Further research is needed to determine which factors govern the rate of satiation.

---

[10] See Nunes and Uriagereka (2000) for a more modern rendering of the CED in the minimalist framework.

Finally, the current study demonstrates that meta-analysis, already widely employed in disciplines such as medicine and psychology, can be a valuable research tool for linguists. A common issue for quantitative studies in linguistics is low statistical power due to small sample sizes or poor research design, giving rise to mixed findings and non-replications (Prasad and Linzen, 2021; Sönning and Werner, 2021; Vasishth and Gelman, 2021; Vasishth et al., 2018). Moreover, academic journals typically discourage the publication of null experimental results, leading to widespread publication bias (Roettger, 2021; Vasishth et al., 2018). The current study shows that meta-analysis can address these issues by synthesizing results from individual studies, even if they are underpowered, without the need for new experiments with substantially larger sample sizes. Moreover, as demonstrated, meta-analysis can help identify publication bias. Overall, meta-analytic methods can improve the quality and rigor of quantitative research in linguistics and should be considered an essential component of the linguistic research toolkit.

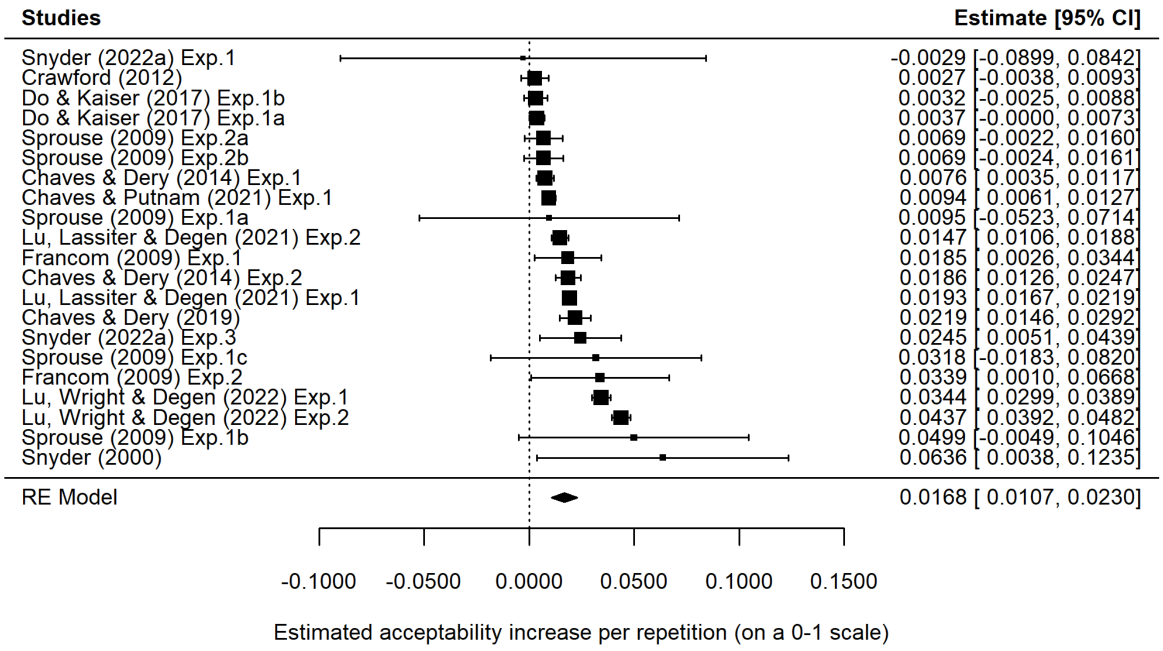


Figure 4: Forest plot of studies testing subject island satiation. Effect size captures acceptability increase per repetition. Error bars represent 95% CI, and the area of squares represents the weight given to each study based on its standard error and the estimated cross-study variance.

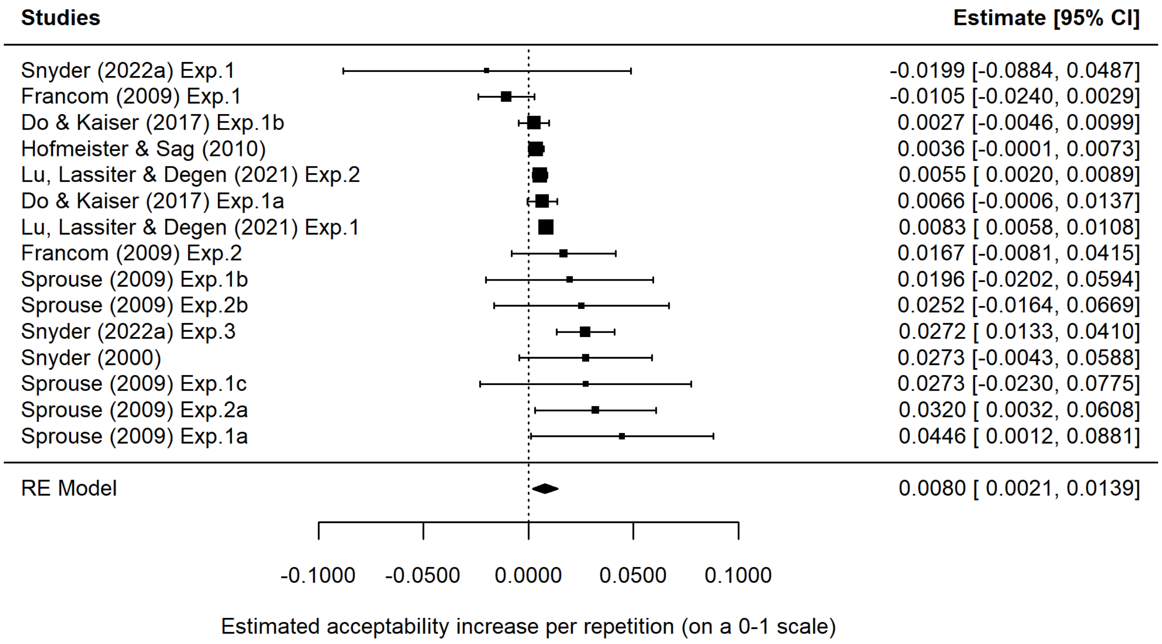


Figure 5: Forest plot of studies testing the satiation of CNPC. Effect size captures acceptability increase per repetition. Error bars represent 95% CI, and the area of squares represents the weight given to each study based on its standard error and the estimated cross-study variance.

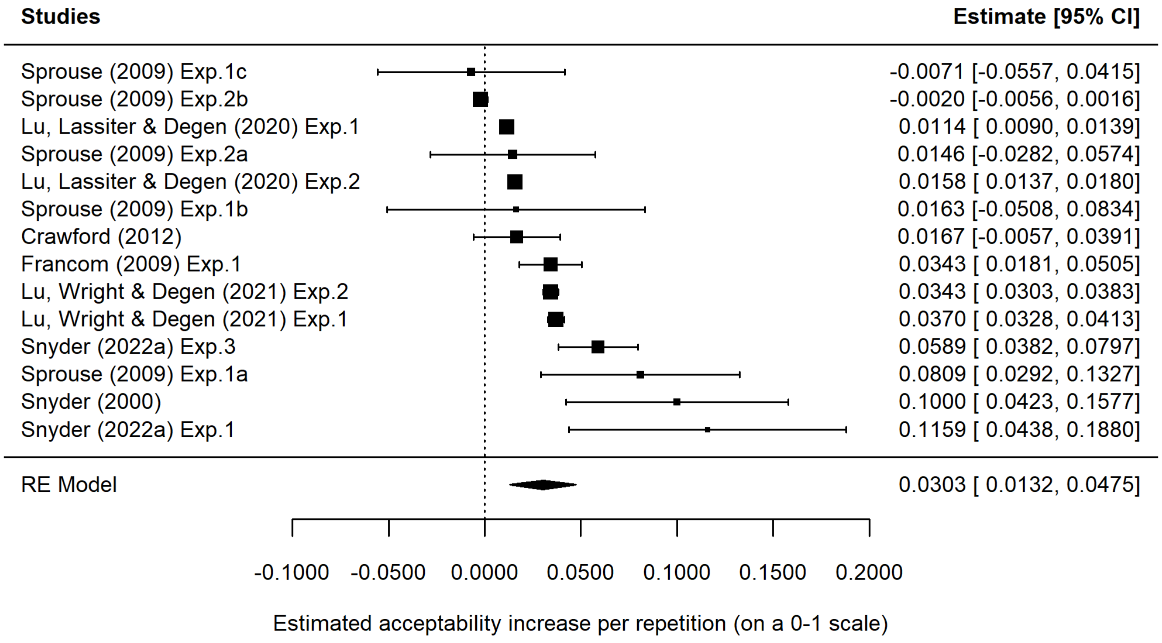


Figure 6: Forest plot of studies testing *whether*-island satiation. Effect size captures acceptability increase per repetition. Error bars represent 95% CI, and the area of squares represents the weight given to each study based on its standard error and the estimated cross-study variance.

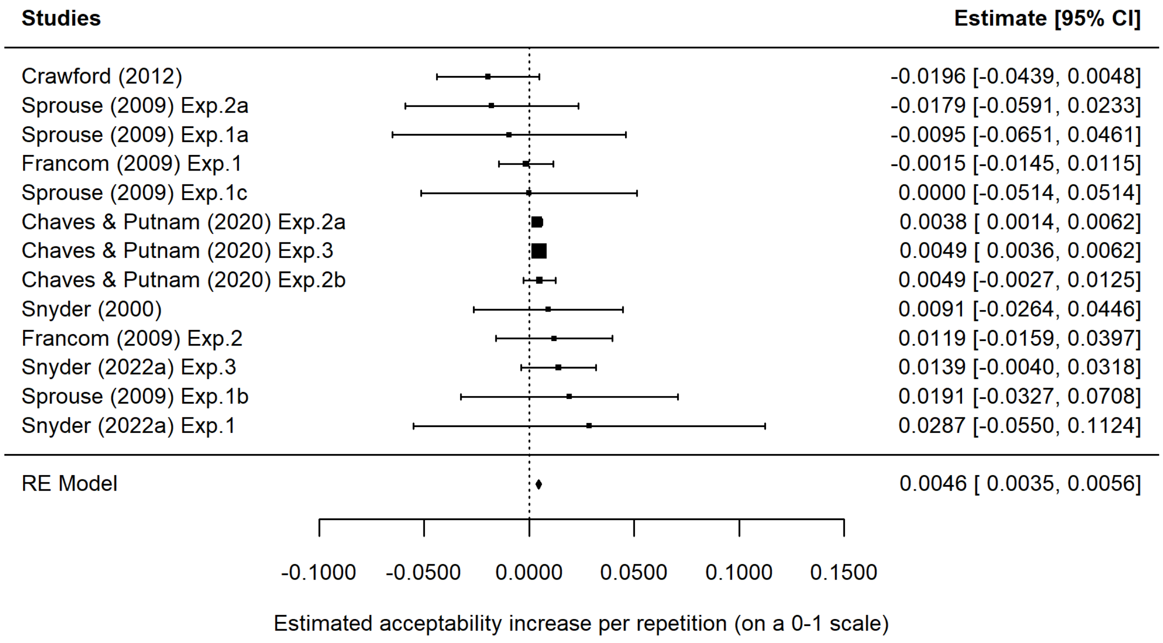


Figure 7: Forest plot of studies testing adjunct island satiation. Effect size captures acceptability increase per repetition. Error bars represent 95% CI, and the area of squares represents the weight given to each study based on its standard error and the estimated cross-study variance.

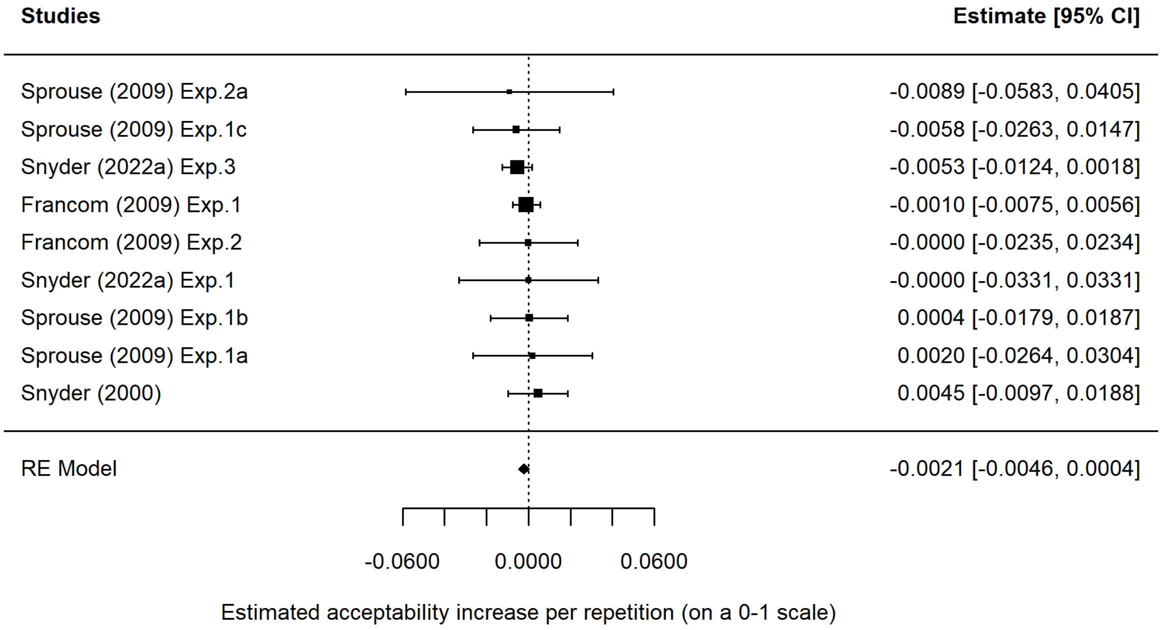


Figure 8: Forest plot of studies testing the satiation of LBC. Effect size captures acceptability increase per repetition. Error bars represent 95% CI, and the area of squares represents the weight given to each study based on its standard error and the estimated cross-study variance.

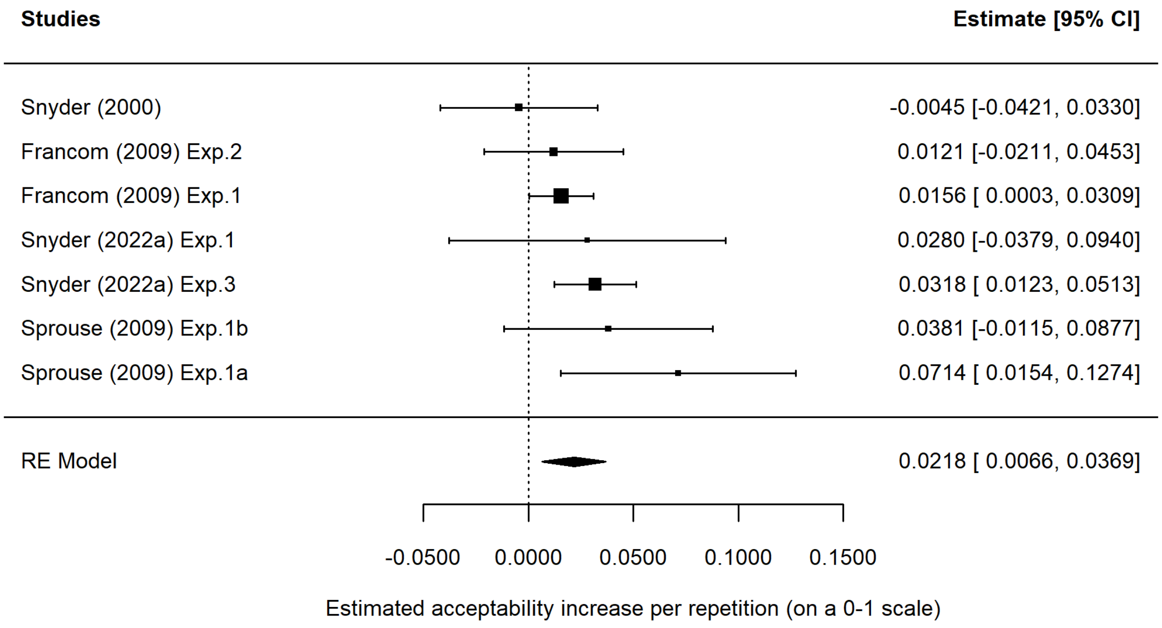


Figure 9: Forest plot of studies testing the satiation of the *that*-trace construction. Effect size captures acceptability increase per repetition. Error bars represent 95% CI, and the area of squares represents the weight given to each study based on its standard error and the estimated cross-study variance.

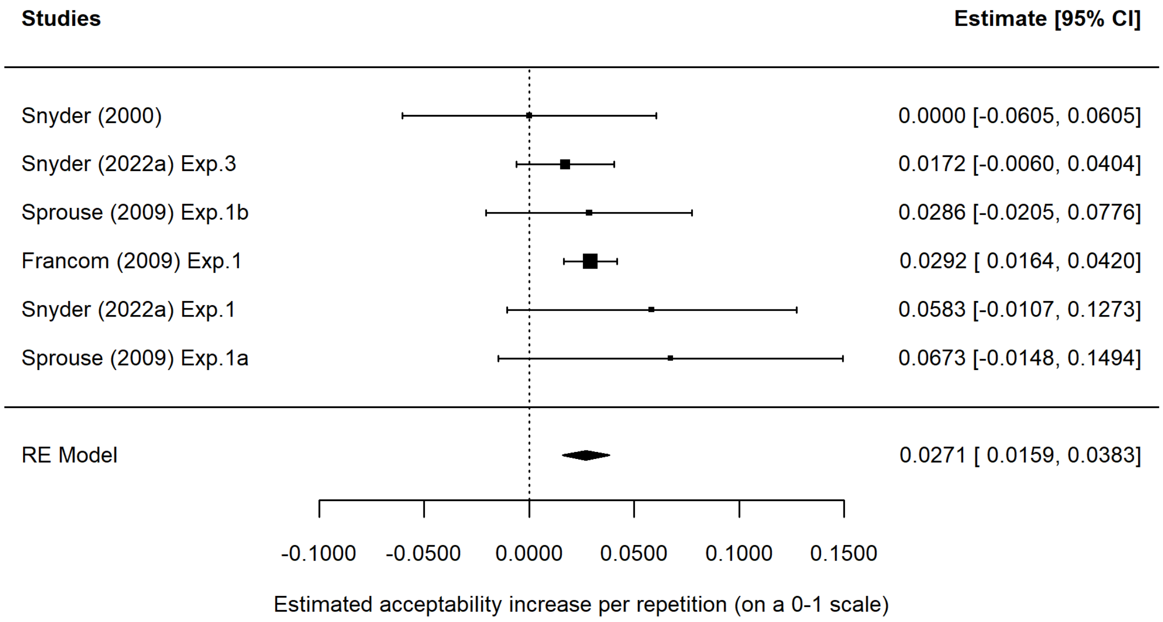
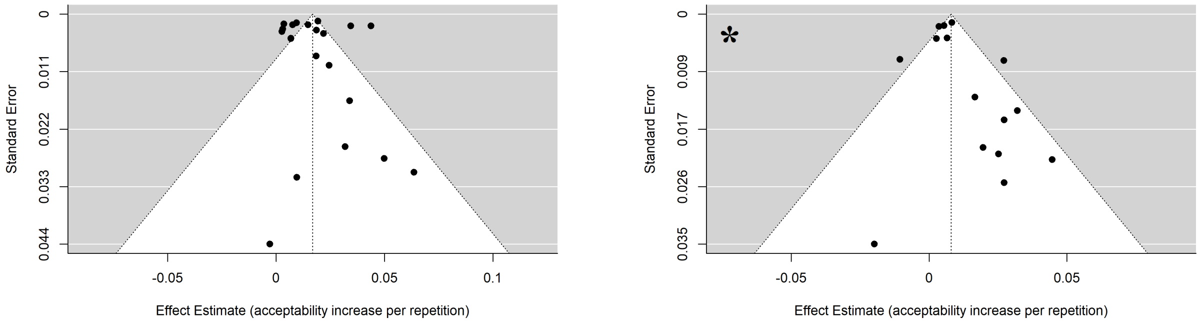
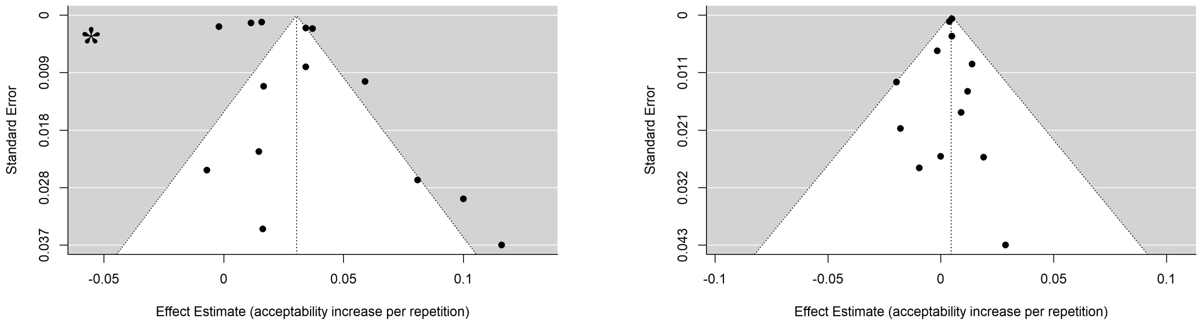


Figure 10: Forest plot of studies testing the satiation of the *want-for* construction. Effect size captures acceptability increase per repetition. Error bars represent 95% CI, and the area of squares represents the weight given to each study based on its standard error and the estimated cross-study variance.



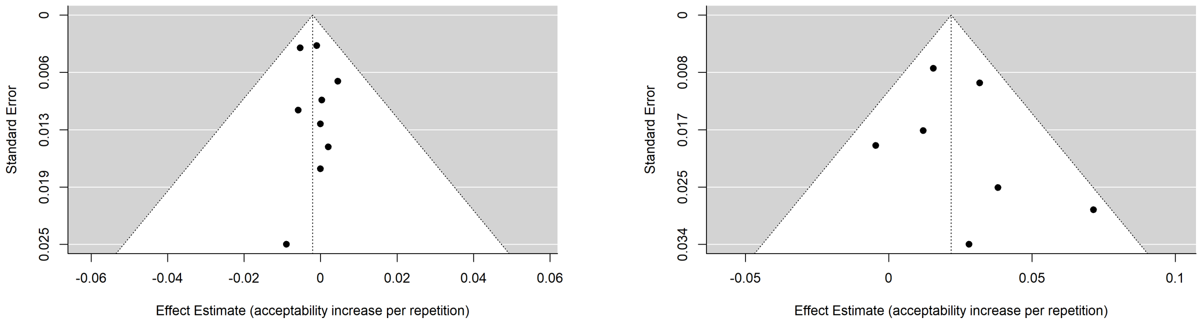
(a) Subject Island

(b) CNPC



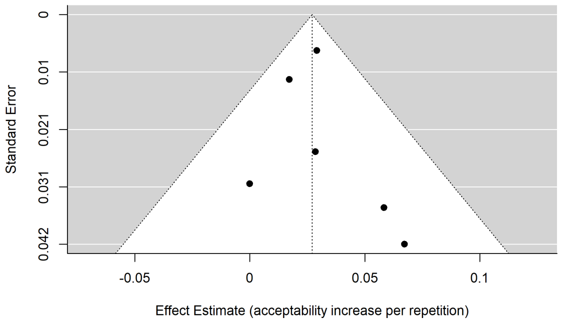
(c) *Whether*-Island

(d) Adjunct Island



(e) LBC

(f) *That*-trace Construction



(g) *Want-for* Construction

Figure 11: Funnel plots for studies testing the satiation of each island type. Stars represent marginal significance in Egger's test.

4 Acknowledgments

We would like to thank Rui Chaves, Jean Crawford, Monica Do, Jerid Francom, Phillip Hofmeister, William Snyder, and Jon Sprouse for kindly making their data files available. Thanks to Rui Chaves, Dan Lassiter, William Snyder, Tom Wasow, and audience at LSA 2023 and CAMP 2023 for valuable comments and feedback.

References

Abeillé, A., Hemforth, B., Winckel, E., and Gibson, E. (2020). Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition*, 204:104293.

Abugharsa, A. F. (2016). *The Usefulness of Explicit Grammar Teaching: An lnvestigation of Syntactic Satiation Effects and Acceptability Judgements in Libyan EFL Contexts.* PhD thesis, Middlesex University.

Boas, H. C. and Sag, I. A. (2012). *Sign-based construction grammar.* CSLI Publications/-Center for the Study of Language and Information Stanford, CA.

Borenstein, M., Hedges, L. V., Higgins, J. P., and Rothstein, H. R. (2009). *Introduction to meta-analysis.* John Wiley & Sons.

Bošković, Ž. (2005). On the locality of left branch extraction and the structure of NP. *Studia Linguistica*, 59(1):1–45.

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.

Braze, F. D. (2002). *Grammaticality, acceptability and sentence processing: A psycholinguistic study.* PhD thesis, University of Connecticut.

Bresnan, J. W. (1976). On the form and functioning of transformations. *Linguistic Inquiry*, 7(1):3–40.

Brown, J. M., Fanselow, G., Hall, R., and Kliegl, R. (2021). Middle ratings rise regardless of grammatical construction: Testing syntactic variability in a repeated exposure paradigm. *PLOS One*, 16(5):e0251280.

Chaves, R. P. (2022). Sources of discreteness and gradience in island effects. *Languages*, 7(4):245.

Chaves, R. P. and 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*, pages 96–106.

Chaves, R. P. and Dery, J. E. (2019). Frequency effects in subject islands. *Journal of Linguistics*, 55(3):475–521.

Chaves, R. P. and Putnam, M. T. (2020). *Unbounded dependency constructions: Theoretical and experimental perspectives*, volume 10. Oxford University Press, USA.

Chomsky, N. (1964). *Current issues in linguistic theory.* De Gruyter Mouton.

Chomsky, N. (1973). Conditions on transformations. In Anderson, S. R. and Kiparsky, P., editors, *A Festchrift for Morris Halle*, pages 232–286.

Chomsky, N. (1977). On wh-movement. In Culicover, P. W., Wasow, T., and Akmajian, A., editors, *Formal Syntax*, pages 71–132.

Chomsky, N. (1986). *Barriers.* MIT press.

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, 37(3/4):256–266.

Crawford, J. (2012). Using syntactic satiation to investigate subject islands. In *Proceedings of the 29th West Coast Conference on Formal Linguistics*, pages 38–45. Cascadilla Proceedings Project Somerville, MA.

Culicover, P. W., Varaschin, G., and Winkler, S. (2022). The radical unacceptability hypothesis: Accounting for unacceptability without universal constraints. *Languages*, 7(2):96.

Do, M. L. and Kaiser, E. (2017). The relationship between syntactic satiation and syntactic priming: A first look. *Frontiers in Psychology*, 8:18–51.

Egger, M., Smith, G. D., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, 315(7109):629–634.

Erteschik-Shir, N. (1973). *On the nature of island constraints.* PhD thesis, Massachusetts Institute of Technology.

Fox, D. and Pesetsky, D. (2005). Cyclic linearization of syntactic structure.

Francom, J. C. (2009). *Experimental Syntax: Exploring the effect of repeated exposure to anomalous syntactic structure–evidence from rating and reading tasks.* PhD thesis, The University of Arizona.

Goldberg, A. E. (2013). Backgrounded constituents cannot be "extracted". *Experimental syntax and island effects*, 221.

Goodall, G. (2011). Syntactic satiation and the inversion effect in english and spanish wh-questions. *Syntax*, 14(1):29–47.

Haegeman, L., Jiménez-Fernández, Á. L., and Radford, A. (2014). Deconstructing the subject condition in terms of cumulative constraint violation. *The Linguistic Review*, 31(1):73–150.

Hedges, L. V. and Vevea, J. (2005). Selection method approaches. In Rothstein, H. R., Sutton, A. J., and Borenstein, M., editors, *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, pages 145–174. Wiley Online Library.

Hiramatsu, K. (2001). *Accessing linguistic competence: Evidence from children's and adults' acceptability judgments.* PhD thesis, University of Connecticut.

Hofmeister, P., Casasanto, L. S., and Sag, I. A. (2013). Islands in the grammar? Standards of evidence. In Sprouse, J. and Hornstein, N., editors, *Experimental syntax and island effects*. Cambridge University Press.

Hofmeister, P., Jaeger, T. F., Arnon, I., Sag, I. A., and Snider, N. (2011). The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgments. *Language and Cognitive Processes*, 28(1-2):48–87.

Hofmeister, P. and Sag, I. A. (2010). Cognitive constraints and island effects. *Language*, 86(2):366–415.

Huang, C.-T. J. (1982). *Logical relations in Chinese and the theory of grammar.* PhD thesis, Massachusetts Institute of Technology.

Keshev, M. and Meltzer-Asscher, A. (2019). A processing-based account of subliminal wh-island effects. *Natural Language & Linguistic Theory*, 37(2):621–657.

Kluender, R. (1991). *Cognitive constraints on variables in syntax.* PhD thesis, University of California, San Diego.

Kluender, R. and Kutas, M. (1993). Subjacency as a processing phenomenon. *Language and Cognitive Processes*, 8(4):573–633.

Light, R. J. and Pillemer, D. B. (1984). *Summing up: The science of reviewing research.* Harvard University Press.

Lu, J., Lassiter, D., and Degen, J. (2021). Syntactic satiation is driven by speaker-specific adaptation. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 43.

Lu, J., Wright, N., and Degen, J. (2022). Satiation effects generalize across island types. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 44.

Michaelis, L. A. (2013). Sign-based construction grammar. In Hoffman, T. and Trousdale, G., editors, *The Oxford Handbook of Construction Grammar*.

Moher, D., Liberati, A., Tetzlaff, J., and Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *Annals of Internal Medicine*, 151(4):264–269.

Myers, J. (2012). Testing adjunct and conjunct island constraints in chinese. *Language and Linguistics*, 13(3):437.

Nunes, J. and Uriagereka, J. (2000). Cyclicity and extraction domains. *Syntax*, 3(1):20–43.

Prasad, G. and 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.

Rizzi, L. (1990). *Relativized minimality.* The MIT Press.

Roettger, T. B. (2021). Preregistration in experimental linguistics: Applications, challenges, and limitations. *Linguistics*, 59(5):1227–1249.

Ross, J. R. (1967). *Constraints on variables in syntax.* PhD thesis, Massachusetts Institute of Technology.

Sag, I. A. (1976). *Deletion and logical form.* PhD thesis, Massachusetts Institute of Technology.

Sag, I. A. (2012). Sign-based construction grammar: An informal synopsis. In *Sign-based Construction Grammar*, pages 69–202.

Schutze, C. T. (1996). *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology.* University of Chicago Press.

Snyder, W. (2000). An experimental investigation of syntactic satiation effects. *Linguistic Inquiry*, 31(3):575–582.

Snyder, W. (2022a). On the nature of syntactic satiation. *Languages*, 7(1):38.

Snyder, W. (2022b). Satiation. In Goodall, G., editor, *The Cambridge Handbook of Experimental Syntax*, pages 154–180. Cambridge University Press.

Sommer, M. (2022). A further look: Is syntactic priming a plausible underlying mechanism for syntactic satiation? Master's thesis, Radboud University.

Sönning, L. and Werner, V. (2021). The replication crisis, scientific revolutions, and linguistics. *Linguistics*, 59(5):1179–1206.

Sprouse, J. (2007). Syntactic satiation: Toward an etiology of linguist's disease. In *30th Annual GLOW Colloquium. University of Tromso, Norway.*

Sprouse, J. (2009). Revisiting satiation: Evidence for an equalization response strategy. *Linguistic Inquiry*, 40(2):329–341.

Stepanov, A. (2007). The end of CED? Minimalism and extraction domains. *Syntax*, 10(1):80–126.

Stromswold, K. (1986). Syntactic satiation. unpublished ms.

Vasishth, S. and Gelman, A. (2021). How to embrace variation and accept uncertainty in linguistic and psycholinguistic data analysis. *Linguistics*, 59(5):1311–1342.

Vasishth, S., Mertzen, D., Jäger, L. A., and Gelman, A. (2018). The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language*, 103:151–175.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the *metafor* package. *Journal of Statistical Software*, 36(3):1–48.