SQUIB

Investigating the relationship between individual differences and island sensitivity

Catherine Pham[1], Lauren Covey[2], Alison Gabriele[3], Saad Aldosari[4] and Robert Fiorentino[5]

[1] Department of Psychology, The Pennsylvania State University, University Park, PA, US

[2] Department of Linguistics, Montclair State University, Montclair, NJ, US

[3] Second Language Acquisition Laboratory, University of Kansas, Lawrence, KS, US

[4] College of Languages and Translation, Al Imam Mohammad Ibn Saud Islamic University, Riyadh, SA

[5] Neurolinguistics and Language Processing Laboratory, Department of Linguistics, University of Kansas, Lawrence, KS, US

Corresponding author: Lauren Covey (coveyl@montclair.edu)

It is well-attested that native speakers tend to give low acceptability ratings to sentences that involve movement from within islands, but the source of island effects remains controversial. The grammatical account posits that island effects result from syntactic constraints on *wh*-movement, whereas the resource-limitation view posits that low ratings emerge due to processing-related constraints on the parser, such that islands themselves present processing bottlenecks. We address this debate by investigating the relationship between island sensitivity and individual differences in cognitive abilities, as it has been argued that the two views make distinct predictions regarding whether a relationship should hold. Building directly on Sprouse et al. (2012a), we tested 102 native English speakers on four island types (*whether*, complex NP, subject, and adjunct islands) using an acceptability judgment task to quantify island sensitivity and three cognitive tasks to capture individual differences in working memory (via reading and counting span tasks) and attentional control (via a number Stroop task). Our results reveal strong island sensitivity effects across all island types. However, we did not find evidence that individual differences in working memory and attentional control modulated island sensitivity, which runs counter to the resource-limitation view (Kluender & Kutas 1993). These results are in line with grammatical accounts of island effects rather than due to processing difficulties.

Keywords: syntax; island constraints; acceptability judgments; individual differences; working memory; attentional control

1 Introduction

The current study addresses a theoretical debate regarding the source of syntactic island effects. Languages such as English allow for *wh*-movement, in which the *wh*-phrase (e.g. *who*) originates in one position and moves to a different position in the sentence, as in (1) (e.g. Chomsky 1981; 1986).

(1)   Who did you see __ yesterday?

The word *who* in (1) originates in the object position of the verb *see* and is then moved to the front of the sentence, leaving a gap at its original site. *Wh*-movement is argued to be subject to syntactic constraints, such that *wh*-phrases cannot be extracted out of certain syntactic structures called islands (Ross 1967). Example (2a) contains an embedded question, one type of island domain, and (2b) illustrates that extracting a *wh*-item out of the embedded question renders the sentence ungrammatical.

(2)   a.   You wonder [whether Kim read the book].
        b.   *What do you wonder [whether Kim read __]?

Native speakers have been shown to give low acceptability judgment ratings to sentences containing island violations (Sprouse 2007; Sprouse et al. 2011; Sprouse et al. 2012a), but there is little consensus regarding the source of these island effects (for further discussion see Sprouse & Villata in press). Proponents of the grammatical view postulate that syntactic constraints prevent extraction from islands (e.g. Phillips 2006; Wagers & Phillips 2009; Sprouse et al. 2012a; b; Yoshida et al. 2014; Sprouse et al. 2016; Kush et al. 2018), although related proposals have argued that island effects may be accounted for by semantic and pragmatic factors (e.g. Erteschik-Shir 1973; Kuno & Takami 1993; Szabolcsi & Zwarts 1993; Goldberg 2007; Truswell 2007; Abrusán 2014; Kush et al. 2019). In contrast to the grammatical view, proponents of the resource-limitation view argue that island effects arise when the processing costs associated with a sentence is too high, exceeding the individual's processing resources (e.g. Kluender & Kutas 1993; Kluender 1998; 2004; Hofmeister & Sag 2010; Hofmeister et al. 2012a; b; 2013). For example, Kluender and Kutas (1993) argued that island effects result from an "overload" of the limited capacity of processing resources available to the parser.

   This study builds directly on Sprouse and colleagues' (2012a) work addressing the debate between grammatical and resource-limitation accounts and examines the source of island effects by investigating the relationship between island sensitivity and individual differences in processing abilities, as Sprouse et al. argue that the two views make distinct predictions regarding whether a relationship should hold. The findings of the current study are poised to inform our understanding of the nature of island effects, as well as the extent to which individual differences affect language processing.

2 Grammatical vs resource-limitation view

One explanation of the grammatical view of islands is that island effects emerge due to violations of syntactic constraints that prohibit *wh*-extraction out of island structures (e.g. Ross 1967; Chomsky 1973; 1986; Huang 1982). These island constraints are assumed to be an innate part of a native speaker's mental grammar and cannot be reduced to processing-based explanations.

   In contrast to grammatical approaches, resource-limitation accounts (Kluender & Kutas 1993; Kluender 2004; Hofmeister & Sag 2010; Hofmeister et al. 2012a; b; 2013) claim that island effects arise not due to the violation of syntactic constraints, but instead due to processing difficulties. Under the resource-limitation theory first proposed by Kluender and Kutas (1993), it is assumed that the cost associated with processing a long-distance *wh*-dependency, which involves maintenance of the *wh*-filler while searching for the gap and retrieval of the *wh*-filler at the gap site, and the cost associated with processing an island structure both need to be active simultaneously for island effects to emerge. It also claims that the resources which are available for sentence processing are limited, and unacceptability emerges when the total processing cost necessary to parse a sentence exceeds the limited resources available. In short, islands are rejected because they are too difficult for the majority of native speakers to process.

   To investigate whether island effects arise due to processing difficulties, Hofmeister and Sag (2010) examined how linguistic properties of the *wh*-phrase affect native English speakers' processing of sentences containing island violations. Stimuli tested complex *wh*-fillers (e.g. *which employee*), which are argued to facilitate the processing of sentences containing island violations, given that more semantically and syntactically complex *wh*-fillers have stronger mental representations compared to bare *wh*-fillers (e.g. *who*) and are thus expected to be easier to retrieve from working memory at the gap position.

Therefore, sentences containing complex *wh*-fillers were expected to elicit faster reading times after the verb compared to sentences containing bare *wh*-fillers, which was confirmed by the results of a self-paced reading task. Hofmeister and Sag (2010) argued that their results provided evidence in support of the resource-limitation view of islands because non-structural factors (i.e. the complexity of the *wh*-filler) affected the processing and acceptability of ungrammatical sentences containing island violations. Studies on Danish (Christensen et al. 2013a; b; Christensen & Nyvad 2014) have also argued in support of a processing-based explanation of islands. In related work, Keshev and Meltzer-Asscher (2019) showed that decreased acceptability of ungrammatical sentences containing *wh*-islands in Hebrew is at least partially induced by processing costs.

To further investigate the source of island effects and to tease apart grammatical and resource-limitation approaches, Sprouse et al. (2012a) examined the relationship between judgments of island violations and working memory in native English speakers. Sprouse et al. (2012a) argued that resource-limitation approaches should predict a relationship between an individual's processing resources and the size of island effects: under the resource-limitation view, those with greater working memory are expected to have more processing resources available and should find sentences containing island violations easier to process and more acceptable. In contrast, the grammatical view should not expect this relationship, as island violations are not permitted by the grammar and should thus be unacceptable, regardless of processing resources.

Sprouse et al. (2012a) tested working-memory using a serial-recall task and a *n*-back task (Kirchner 1958; Kane & Engle 2002; Jaeggi et al. 2008). To measure island sensitivity, participants completed a task in which they rated the acceptability of English sentences, utilizing a 7-point scale (Experiment 1) or magnitude estimation (Experiment 2). Four island types were tested: *whether*, complex NP, subject, and adjunct. The stimuli were created using a 2 × 2 factorial design, manipulating the presence/absence of an island structure and the *wh*-dependency length. An example set of stimuli for an adjunct island is depicted in (3).

(3)   Non-island/Matrix
      a.   Who __ suspects that the boss left her keys in the car?

       Non-island/Embedded
      b.   What do you suspect that the boss left __ in the car?

       Island/Matrix
      c.   Who __ worries [if the boss leaves her keys in the car]?

       Island/Embedded
      d. *What do you worry [if the boss leaves __ in the car]?

Sprouse et al. (2012a) calculated a differences-in-differences (DD) score for each island type. To calculate DD scores, the first difference score (D1) is calculated by subtracting the mean acceptability rating of the island/embedded condition (3d) from the non-island/embedded condition (3b). This D1 score quantifies the effect of an island structure in sentences with a long-distance dependency. The second difference score (D2) is calculated by subtracting the mean acceptability rating of the island/matrix condition (3c) from the non-island/matrix condition (3a). Finally, the DD score is calculated by subtracting D2 from D1. This score quantifies the strength of island effects on a long-distance dependency compared to a short-distance dependency. High DD scores indicate strong sensitivity to island effects and less acceptance of ungrammatical island violations. Low DD scores indicate weak sensitivity to island effects and greater acceptance of ungrammatical island violations.

As outlined by Sprouse et al., the resource-limitation approach would predict that individuals with better performance on the working memory tasks should show lower DD scores (less rejection of island violations). In contrast, grammatical approaches expect no such relationship. Sprouse et al. argued that their results revealed virtually no relationship between working memory and DD scores. In many cases, the relationships were not statistically significant, and in the cases in which there were significant effects, the amount of variance explained was nevertheless very small (R^2 value between 0.00–0.06). Thus, Sprouse et al. concluded that the perceived unacceptability of island violations cannot be reduced to processing difficulties and are likely due to the existence of grammatical constraints.

However, Hofmeister et al. (2012a; b) raised a number of criticisms regarding Sprouse et al. (2012a); the present study addresses several of these criticisms. One criticism was that the stimuli may have been too complex to process, given that the sentences included words like *who* in isolation without context. This could have masked the relationship between working memory and DD scores, such that even individuals with increased working memory resources may experience a processing breakdown given the extreme complexity. To alleviate some of the processing burden and increase the likelihood of finding a relationship between working memory and DD scores, the current study includes a background sentence prior to the target *wh*-question in order to establish context. We additionally utilize complex *wh*-fillers (e.g. *which worker*) instead of bare *wh*-fillers (e.g. *who*) since complex *wh*-fillers have been argued to facilitate the processing of *wh*-dependencies (Hofmeister & Sag 2010; Goodall 2015). Another criticism was that the serial-recall and *n*-back tasks used by Sprouse et al. (2012a) were not sufficient measures of working-memory capacity because these tasks are simple span tasks which do not include both storage and processing components. As Hofmeister and colleagues point out, the validity of Sprouse et al.'s (2012a) argument that there is no relationship between working memory and island sensitivity depends on the validity of the choice of working memory measure. To address this concern directly, we utilize a complex memory span measure used by Hofmeister and colleagues themselves (Hofmeister et al. 2014), which has been shown to predict language comprehension skills broadly (e.g. Daneman & Carpenter 1980; King & Just 1991; Just & Carpenter 1992).

3 The current study

We further investigate the role of individual differences on the processing of islands, building directly on Sprouse et al. (2012a) to account for the concerns of Hofmeister et al. (2012a; b).

3.1 Methodology

3.1.1 Participants

102 native English speakers from the University of Kansas (32 males) were tested. Participants ranged in age from 18–34 ($M = 20.7$).

3.1.2 Materials

The current study utilized sentences[1] from Aldosari (2015), who made two important modifications to Sprouse et al.'s sentences in order to address concerns raised by Hofmeister et al. (2012a). A declarative background sentence preceded each test sentence to provide a context for the *wh*-question. Secondly, a complex *wh*-filler (e.g. *which worker*) was used in place of a bare *wh*-filler (e.g. *who*) because it has been argued that the use of complex *wh*-fillers facilitates the processing of *wh*-dependencies (Hofmeister & Sag 2010; Goodall 2015). Following Sprouse et al. (2012a), four island types were tested: *whether,* complex

---

[1] A small number of sentences (5% of the total) in Aldosari's stimuli set were edited for clarity.

NP, subject, and adjunct islands. We created four conditions for each island type using a 2 × 2 factorial design manipulating presence of an island structure and the *wh*-dependency length. An example of the four conditions for the adjunct island type is shown in (4).

(4) Background sentence
The helpful worker thinks that the boss left her keys in the car.

Non-island/Matrix
a. Which worker __ thinks that the boss left her keys in the car?

Non-island/Embedded
b. Which keys does the worker think that the boss left __ in the car?

Island/Matrix
c. Which worker __ worries [if the boss leaves her keys in the car]?

Island/Embedded
d. *Which keys does the worker worry [if the boss leaves __ in the car]?

16 sets of sentences were created for each of the 4 conditions. In total, there were 64 sets distributed among 4 lists using a Latin-Square design; no filler sentences were included.[2] The sentences were divided into four blocks with the experimental sentences randomized in each block.

3.1.3 Tasks

During the acceptability judgment task, participants were first presented with the declarative background sentence and were instructed to press the space bar to advance to the next screen after reading it. The subsequent screen presented only the test sentence, and participants were asked to rate each target sentence using a 7-point scale ranging from *totally unnatural* to *perfectly natural*. There was no time limit for this task, and participants were provided with an *I do not know* option.

Because Hofmeister et al. (2012a; b) argued that the serial-recall and *n*-back tasks used by Sprouse et al. (2012a) were not true measures of working memory capacity, the counting span task (Case et al. 1982) and the reading span task (Daneman & Carpenter 1980) were used because these tasks contain both a memory component and a processing component. Crucially, the reading span task has been used to investigate the relationship between working-memory capacity and the processing of *wh*-dependencies in work by Hofmeister and colleagues (2014) as well as in other research (Johnson et al. 2016). In the counting span task, following Conway et al. (2005), participants were presented with a screen depicting a random arrangement of target objects (dark blue circles) and distractor objects (dark blue squares and light green circles). They were asked to count the number of target objects aloud and remember the number. After 2 to 6 screens, participants were prompted to input the total number of target shapes counted from the last set of arrays in order. For the reading span task, participants read sentences aloud, provided a semantic judgment about the sentence, said the letter presented on the screen out loud, and were asked to remember the letters. After 2 to 5 sentences, participants were prompted to input the letters from the last set of sentences (Conway et al. 2005). Accuracy on the working memory tasks was measured as the percentage of numbers/letters participants correctly recalled in order.

---

[2] The present study is part of a larger research project that uses event-related potentials (ERPs) to investigate the online processing of *wh*-dependencies. The acceptability judgment task and the three cognitive tasks were administered in a single testing session that included an EEG experiment. In order to make the length of the session manageable for participants, the acceptability judgment task did not include the filler items used in Aldosari (2015); we recognize this as a limitation of the research design.

Additionally, we included a measure of attentional control given that it has been shown to capture individual variability in language processing (e.g. Hutchison 2007; Boudewyn et al. 2012; 2013; Zirnstein et al. 2018). Relevant to the current study, Johnson (2015) found that individuals with increased attentional control resources were more likely to engage in gap prediction during the processing of long-distance *wh*-dependencies. Following Johnson (2015), we included the number Stroop task in order to examine whether individuals with increased attentional control resources find island violation sentences (which contain a long-distance dependency) easier to process, resulting in increased acceptance of sentences with island violations. The number Stroop task measures participants' ability to attend to the target task despite interfering visual information. Following Bush et al. (2006), participants counted the total number of words presented on the screen, which ranged from 1 to 4, and pushed the corresponding button on a button box. For congruent trials, the words on the screen were monosyllabic animal words (e.g. *cat cat*), and participants were instructed to press the corresponding button (e.g. 2). For incongruent trials, the words on the screen were monosyllabic number words (e.g. *one one one*), and participants were instructed to press the button corresponding to the number of words on the screen (e.g. 3), inhibiting the meaning of the words. Reaction times and accuracy were recorded.

During the experiment, the three cognitive tasks were administered before the acceptability judgment task in counterbalanced order. The presentation software Paradigm (Tagliaferri 2005) was utilized to administer all tasks.

3.2 Predictions

The current study tests the predictions of the grammatical view and resource-limitation view regarding the source of island effects. The resource-limitation view hypothesizes that ungrammaticality of island violations is the result of an overload in processing resources (Kluender & Kutas 1993; Kluender 1998; 2004). Under this view, it is possible that a negative correlation between individual differences in cognitive abilities and acceptability of island violations will emerge, such that as performance on the cognitive tasks increases, DD scores, which index island sensitivity, should decrease. In other words, individuals with increased working memory and/or attentional control resources may show lower DD scores. The grammatical view of islands hypothesizes no such correlation between individual differences in working memory and attentional control and the acceptability of island violations.

3.3 Results

3.3.1 Acceptability judgment task analyses

The mean acceptability ratings for each condition are provided in Table 1, with higher scores reflecting greater acceptability.[3]

Table 1: Means and standard deviations of raw acceptability ratings for each condition.

	whether	complex NP	subject	adjunct
Non-Island/Matrix	6.47 (0.75)	6.46 (0.82)	6.09 (0.97)	6.02 (0.97)
Non-Island/Embedded	5.87 (1.12)	5.04 (1.36)	6.11 (0.93)	5.59 (1.15)
Island/Matrix	6.15 (0.95)	6.21 (0.96)	5.94 (1.06)	5.86 (1.20)
Island/Embedded	4.22 (1.45)	2.47 (1.28)	2.04 (1.02)	2.33 (1.30)

---

[3] A reviewer raised a concern about possible floor effects in the ungrammatical island/embedded conditions. An analysis of the individual responses to the four island/embedded conditions showed that participants used the entirety of the 7-point scale, with responses ranging from 1–7 across the four island/embedded conditions. Furthermore, across the four conditions, between 79–96% of participants had a mean rating above the lowest rating of '1'. The standard deviation values in Table 1 reflect this variability in mean ratings.

Prior to statistical analysis, each participant's acceptability judgment ratings were z-score transformed. We used linear mixed effects models to investigate whether participants were sensitive to island effects in the acceptability judgment task. Data were analyzed using R's lme4 and lmerTest packages (Bates et al. 2015; Kuznetsova et al. 2017; R Core Team 2019). For each island type, a full model was constructed which included the fixed effects Island Structure (non-island, island), Dependency Length (matrix, embedded), and the interaction term Island Structure × Dependency Length, as well as random intercepts for item and participant, as well as by-item and by-participant random slopes for each factor and the interaction term. The full model was simplified stepwise and likelihood ratio tests determined whether the inclusion of these random and fixed effects improved model fit.[4]

The best-fitting model for each island type revealed a significant main effect of *wh*-dependency length (*whether*: est = 0.92, *SE* = 0.07, *t* = 12.52, *p* < .001; complex NP: est = 1.83, *SE* = 0.10, *t* = 17.88, *p* < .001; subject: est = 1.93, *SE* = 0.01, *t* = 19.35, *p* < .001; adjunct: est = 1.72, *SE* = 0.09, *t* = 18.86, *p* < .001). A main effect of island structure for each island type was also significant (*whether*: est = 0.77, *SE* = 0.07, *t* = 10.77, *p* < .001; complex NP: est = 1.28, *SE* = 0.11, *t* = 11.70, *p* < .001; subject: est = 2.03, *SE* = 0.09, *t* = 21.49, *p* < .001; adjunct: est = 1.61, *SE* = 0.08, *t* = 19.11, *p* < .001). These main effects reflect the fact that sentences with longer *wh*-dependencies were rated lower than those with shorter (matrix) *wh*-dependencies, and sentences with islands were rated lower than non-island sentences.

Crucially, a significant interaction between *wh*-dependency length and island structure was also found for each island type (*whether*: est = –0.61, *SE* = 0.09, *t* = –6.968, *p* < .001; complex NP: est = –1.14, *SE* = 0.14, *t* = –7.87, *p* < .001; subject: est = –1.96, *SE* = 0.13, *t* = 15.20, *p* < .001; adjunct: est = –1.51, *SE* = 0.11, *t* = –13.96, *p* < .001). This interaction resulted from low acceptability ratings of the ungrammatical island violation condition compared to the other three grammatical conditions for each island type. In other words, the effect of the island structure was greater in sentences with a long *wh*-dependency length than in sentences with a short *wh*-dependency length, indicating that native English speakers were sensitive to island effects in all four island types. Interaction plots for each island type are shown in Figure 1.

In sum, superadditive effects were observed across all island types, such that the combination of a long *wh*-dependency and an island structure yielded lower acceptability than the sum of individually processing a long *wh*-dependency and individually processing an island structure. Under the grammatical view, this superadditivity would be taken to reflect the violation of an island constraint, while under the resource-limitation view, it would reflect a processing overload due to the simultaneous burdens of processing a long-distance dependency and an island structure.

3.3.2 Individual differences analyses

To investigate the source of the island effects, we next examined whether sensitivity (quantified by DD scores) was modulated by individual differences in cognitive abilities (quantified by performance on the cognitive tasks). Recall that a relationship is expected under the resource-limitation view, such that better cognitive abilities should lead to lower island sensitivity (i.e. greater acceptability of island violations), but no such relationship is expected by the grammatical view.

---

[4] The best-fitting LME model for each island type is provided below. Note that in R syntax * denotes fully crossed effects (i.e. both main effects and the interaction term).

z-score rating ~ IslandStructure*DependencyLength + (1 + IslandStructure*DependencyLength| Participant) + (1|Item).
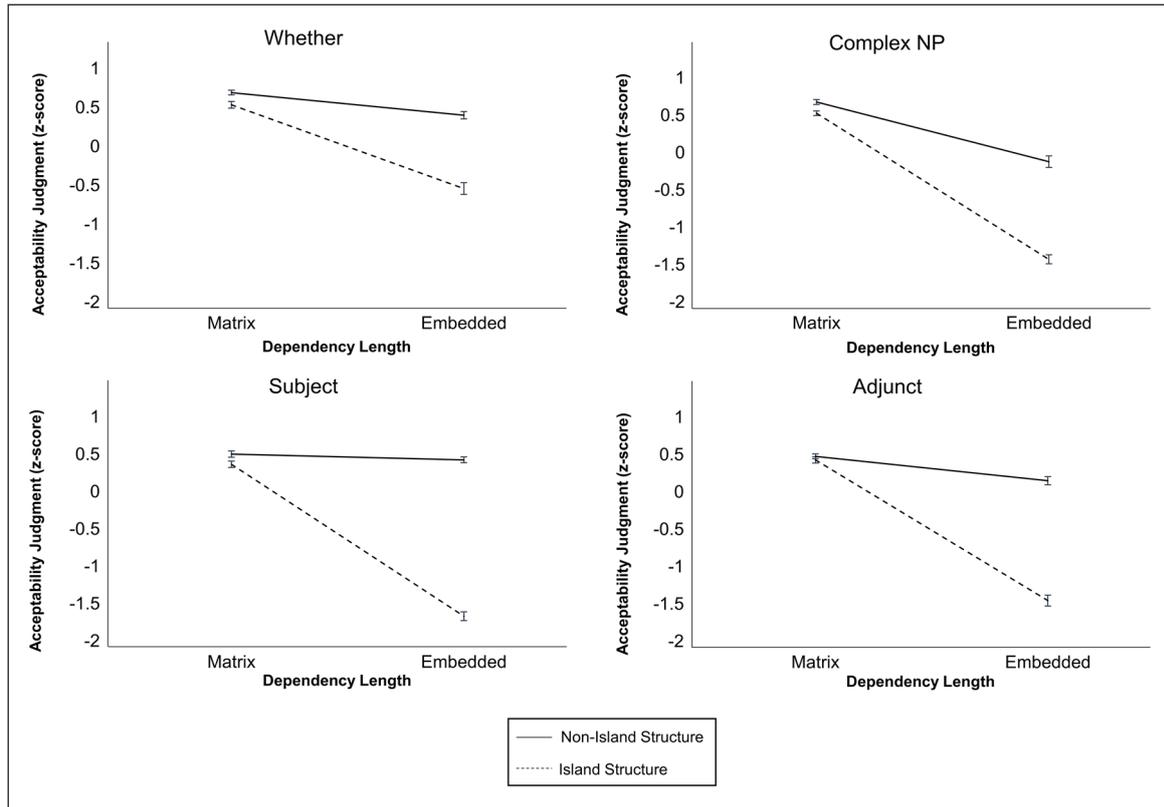
Figure 1: Interaction plots for each island type.

Following Sprouse et al. (2012a), two sets of linear regressions were conducted for each of the four island types. The first linear regression was run with the complete set of DD scores for each island type. The second linear regression was run with DD scores $\geq$ 0 for each island type. DD scores below zero are indicative of subadditive effects, indicating that the effect the island structure in a sentence with a long *wh*-dependency was less than the effect of the island structure in a sentence with a short *wh*-dependency. As Sprouse et al. (2012a) note, because neither theory predicts subadditive effects, inclusion of these negative DD scores might potentially mask the ability to observe a relationship between the individual difference measures and DD scores. Therefore, negative DD scores were excluded from the second analysis. Sprouse et al. also note that negative DD scores could represent individuals who do not experience sensitivity to typical superadditive effects, in which case the inclusion of these negative DD scores might increase the likelihood of finding a negative correlation between individual difference measures and DD scores. Both analyses are reported here following Sprouse et al. This second analysis resulted in the exclusion of two participants for *whether* islands, twelve participants for complex NP islands, three participants for subject islands, and five participants for adjunct islands.

In addition to the linear regressions, Bayes factors (BF) were used to assess the strength of evidence with respect to hypothesis testing (Dienes 2014). A BF < .33 is considered substantial evidence for the null hypothesis over the alternative hypothesis, which would be in line with the grammatical account of islands, which predicts no relationship between DD scores and working memory/attentional control scores. A BF > 3 would be considered substantial evidence for the alternative hypothesis over the null hypothesis, which would be expected under a resource-limitation view. BF between .33 and 3 indicate that the data do not provide substantial evidence to distinguish the null and alternative hypotheses. For

each island type, a Bayes factor analysis was conducted using the JZS prior with the R package BayesFactor (Morey & Rouder 2018).

3.3.2.1 Working memory

On the counting span task, participants' scores ranged from 22.22 to 94.44 ($M = 61.37$; $SD = 13.56$). For the reading span task, participants' scores ranged from 26.25 to 95 ($M = 63.15$; $SD = 14.29$). Because scores on the two tasks were significantly correlated ($r = .36, p < .001$), participants' scores on each task were z-score transformed and added to create a composite working memory variable. DD scores are plotted as a function of working memory scores in Figure 2.

Results from the linear regressions are reported in Table 2. For each island type, the line of best fit, goodness of fit, and significance of the slope are provided. In both sets of linear regressions, none of the best-fit slopes were significantly different from zero across all island types. Additionally, the R^2 value, which measures how much of the variance in DD scores can be explained by the working memory scores, was very low for each island type.

Bayes factors for each island type are also provided in Table 2 and provide adequate evidence in line with the null hypothesis for most of the linear regressions, with Bayes factors below or around .33 for most island types. One exception to this is for the *whether* island in the overall linear regression (BF = 1.018), which did not show conclusive evidence for either the null or alternative hypotheses.
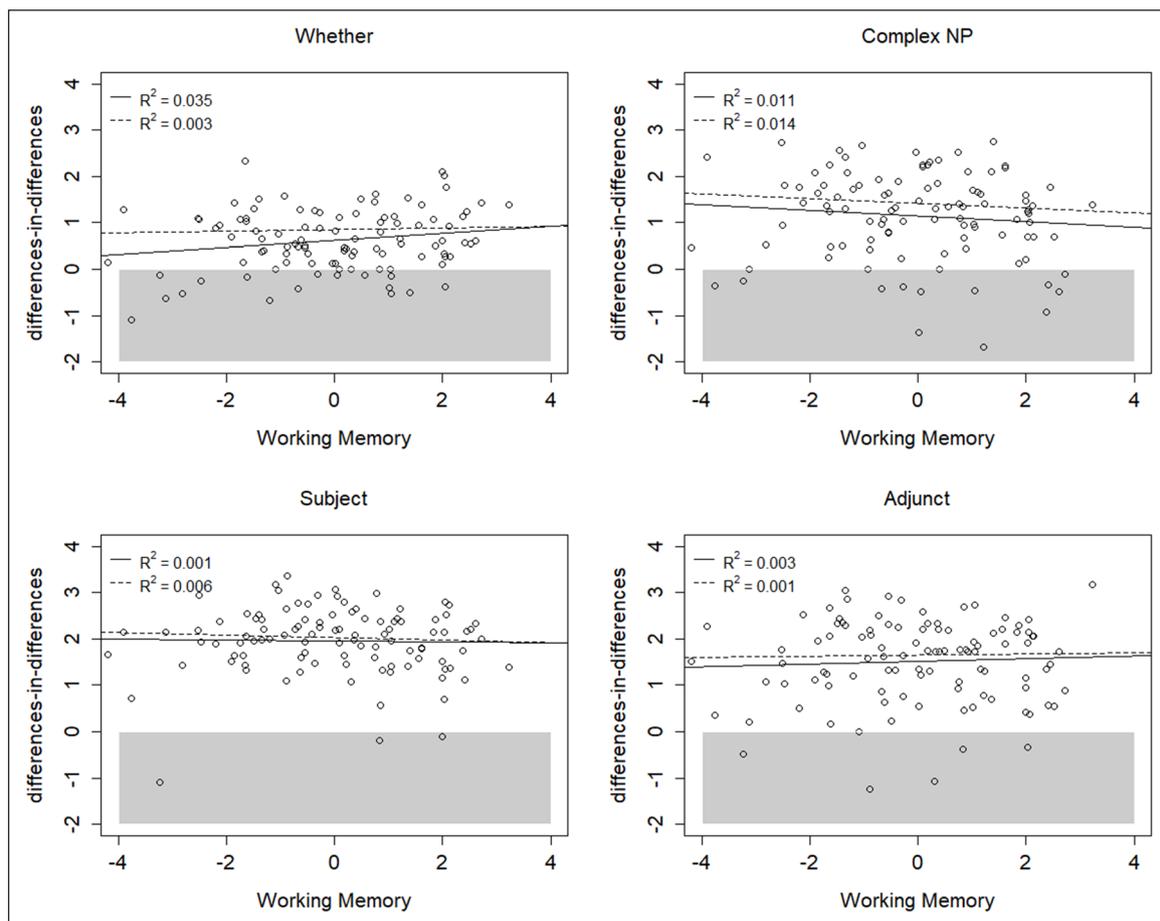
Figure 2: DD scores plotted as a function of composite working memory scores (reading span, counting span). The solid line represents the line of best fit for all DD scores. The dashed line represents the line of best fit when DD scores below zero are removed from analysis (shaded gray). R^2 for each trend line is reported in the legend.

Table 2: Linear regressions modeling DD scores as a function of working memory scores.

	ISLAND	LINE OF BEST FIT		GOODNESS OF FIT	SIGNIFICANCE TEST		BAYES FACTOR
		Intercept	Slope	R^2	t-statistic	p-value	
All DDs	*whether*	0.615	0.075	0.035	1.894	0.061	1.018
	complex NP	1.144	−0.060	0.011	−1.056	0.294	0.343
	subject	1.960	−0.011	0.001	−0.247	0.805	0.215
	adjunct	1.516	0.029	0.003	0.528	0.598	0.237
DDs ≥ 0	*whether*	0.848	0.019	0.003	0.528	0.599	0.261
	complex NP	1.424	−0.050	0.014	−1.089	0.279	0.376
	subject	2.034	−0.027	0.006	−0.748	0.457	0.271
	adjunct	1.648	0.012	0.001	0.261	0.795	0.221

Together, these results indicate that there is no robust relationship between DD scores and working memory, contrary to expectations of the resource-limitation view.

3.3.2.2 Attentional control

Participants' reaction times and accuracy were recorded in order to calculate the Stroop interference effect. In this calculation, larger positive scores are indicative of higher attentional control. Participants' Stroop reaction time interference effect scores ranged from –239.97 ms to 110.70 ms ($M = $ –86.81 ms; $SD = $ 60.80 ms). Stroop accuracy interference effect scores ranged from –21.25 to 5.00 ($M = $ –3.55; $SD = $ 4.12). These two variables were marginally correlated ($r = $ .18, $p = $ .07), and given the fact that they measure the same task, a composite variable was created for the linear regression analysis.[5] DD scores are plotted as a function of attentional control composite scores in Figure 3.

Results from the linear regressions are reported in Table 3. In both sets of linear regressions, none of the best-fit slopes were significantly different from zero across all island types. The R^2 value for each model was very low, and thus the goodness-of-fit results and significance tests of the slopes (p-values) indicate a lack of significant relationship between DD scores and attentional control. Bayes factors for each island type similarly provided evidence in line with the null hypothesis for most of the linear regressions, with Bayes factors below or around .33 for most island types, with the exception of the second *whether* island model which indicated that there was not substantial evidence to support the null hypothesis or reject the null hypothesis in favor of the alternative (BF = 0.415). Together, these analyses provide no evidence of a robust relationship between attentional control and island sensitivity.

3.4 Discussion

Our findings do not suggest a strong relationship between individual differences in working memory, as assessed via counting and reading span tasks, and attentional control, assessed via number Stroop task, and island sensitivity, which is contrary to what is predicted by the resource-limitation view (Kluender & Kutas 1993; Kluender 1998; 2004). When negative DD scores were included and excluded from the linear regressions, the results overall showed no relationship between these cognitive measures and DD scores for the four island types ($p > $ .05). No more than 3.5% of the variance in DD scores was accounted for by any of the cognitive measures. Furthermore, Bayes factors for these linear regression analyses

---

[5] We conducted additional linear regression analyses examining each Stroop measure separately (reaction time, accuracy). We observed the same results for each analysis as in the models utilizing the composite Stroop score, with very low R^2 values and non-significant p-values for the slopes for each island type.
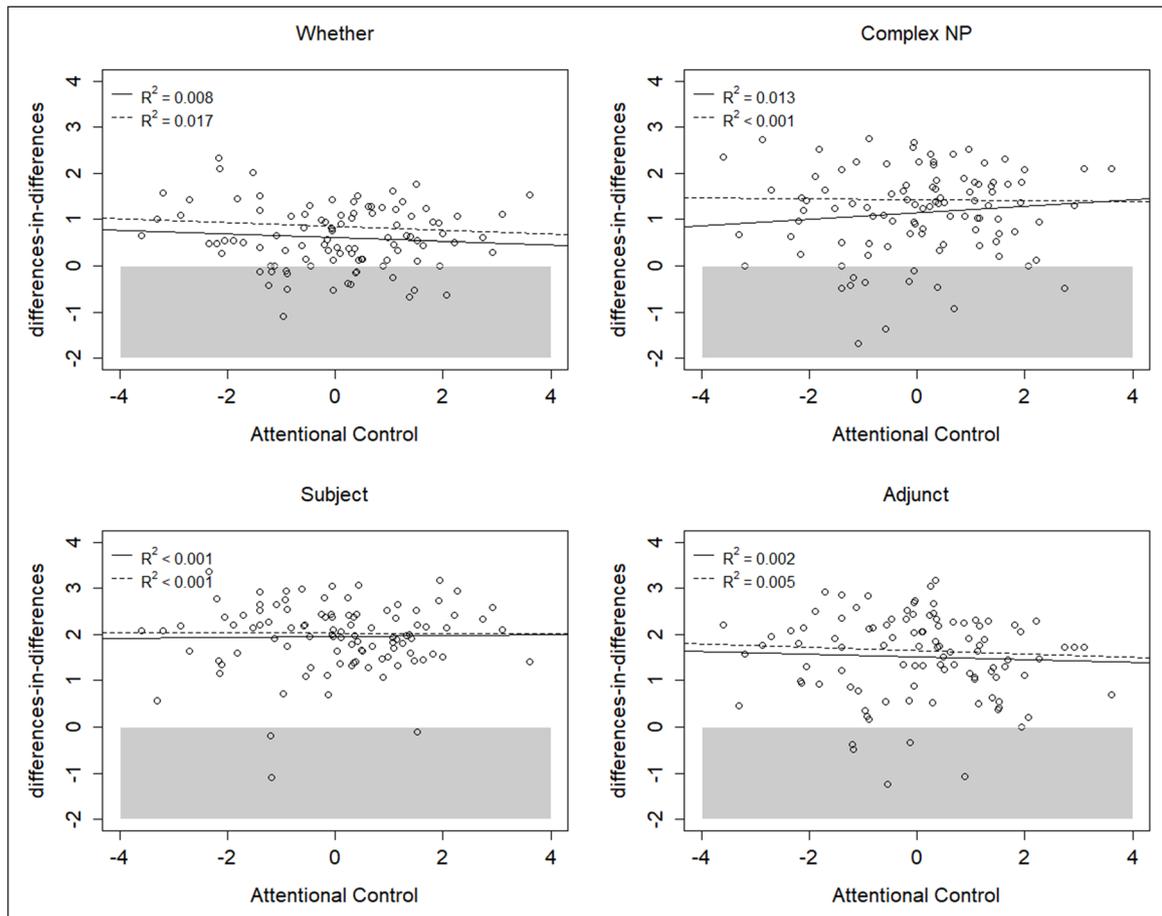
Figure 3: DD scores plotted as a function of composite attentional control scores (Stroop reaction time and accuracy). The solid line represents the line of best fit for all DD scores. The dashed line represents the line of best fit when DD scores below zero are removed from analysis (shaded gray).

Table 3: Linear regressions modeling DD scores as a function of attentional control scores.

	ISLAND	LINE OF BEST FIT		GOODNESS OF FIT	SIGNIFICANCE TEST		BAYES FACTOR
		Intercept	Slope	R^2	t-statistic	p-value	
All DDs	*whether*	0.615	−0.039	0.008	−0.919	0.360	0.304
	complex NP	1.144	0.069	0.013	1.130	0.261	0.369
	subject	1.960	0.010	<0.001	0.212	0.833	0.213
	adjunct	1.516	−0.028	0.002	−0.489	0.626	0.232
DDs ≥ 0	*whether*	0.851	−0.040	0.017	−1.161	0.249	0.415
	complex NP	1.425	−0.009	<0.001	−0.180	0.857	0.227
	subject	2.034	−0.003	<0.001	−0.090	0.929	0.212
	adjunct	1.648	−0.035	0.005	−0.712	0.478	0.269

largely supported the null hypothesis. In light of these results, we argue that island effects are not reducible to processing costs of working memory or attentional control. Thus, the current study provides further evidence in favor of the grammatical view of island effects in line with Sprouse et al. (2012a), as well as with recent work in Italian (Sprouse et al. 2016), Norwegian (Kush et al. 2018; 2019), and Akan (Goodluck et al. 2017) (see also Michel 2014; Yoshida et al. 2014; Aldosari 2015).

Although our results indicate significant island effects across all four island types tested, sentences in the *whether* island violation condition were rated higher on average (4.22) compared to the other island violation types tested (mean of 2.28 for complex NP, subject, and adjunct). Smaller, more variable island effects have been observed for *whether* islands, which are considered weak islands, or those which selectively allow extraction (Chomsky 1986; Cinque 1990; Szabolcsi 2006). Native English speakers have shown variation in their willingness to reject sentences containing extraction from *whether* islands (Johnson & Newport 1991; Marthardjono 1993; Aldosari 2015), a finding echoed in recent work on *whether* islands in Norwegian (Kush et al. 2018). In the individual difference analyses, R^2 and *p*-values for linear regressions indicated a lack of relationship between working memory scores and DD scores for *whether* islands; however, the Bayes factor for this regression revealed that there was insufficient evidence to support the null hypothesis. Thus, the results suggest, in line with previous studies, that island effects differ across constructions and that the null hypothesis, which supports the grammatical account of islands, is most strongly supported by the results of the complex NP, subject, and adjunct islands.

We believe the results provide strong support of Sprouse et al.'s proposal given that our methods addressed several of the concerns outlined by Hofmeister et al. (2012a). First, in the present study, a background declarative sentence preceded each target sentence to eliminate the pragmatic oddity of being presented a question in isolation. In addition, the stimuli used complex *wh*-fillers, which may facilitate the processing of *wh*-dependencies (Hofmeister & Sag 2010; Goodall 2015) as they have been argued to be more Discourse-linked (D-linked) than bare *wh*-fillers, meaning the noun phrase refers to some previously introduced entity. Although D-linking has been argued to ameliorate island effects, recent work by Sprouse et al. (2016) found that the amelioration through D-linking was unable to overcome superadditive island effects. This is in line with our results given that island effects emerged across all four island types.

An additional criticism was that the serial-recall and *n*-back tasks used by Sprouse et al. (2012a) were not sufficient measures of working-memory capacity because they do not both include a processing component and a storage component. Hofmeister et al. (2012a; b) argued that this could account for Sprouse et al. (2012a)'s failure to find a robust relationship between working memory and island sensitivity. We therefore utilized the counting span and reading span tasks, both of which include a processing component and a storage component (Conway et al. 2005). We also included an additional cognitive task, the number Stroop task, to measure attentional control, which has been shown to capture individual variability in the processing of *wh*-dependencies (e.g. Johnson 2015). Despite the use of different working memory tasks and an additional attentional control measure, we still found no robust relationship between individual differences and island sensitivity, which suggests that the source of island effects is not due to processing difficulties.

Note that the reading and counting span tasks we utilized provide a measure of working memory capacity as described by capacity models of memory during language comprehension (Just & Carpenter 1992; Gibson 2000). As Pañeda et al. (2020) discuss in their recent work on island effects in Spanish, it may be the case that working-memory capacity size is an inadequate memory measure given the cue-based retrieval theory, which argues that the critical mechanism involved in language comprehension concerns how accurately a comprehender retrieves, rather than stores, information (e.g. Lewis et al. 2006; McElree et al. 2003; for a review see Parker et al. 2017). It is possible that we did not observe a relationship between working-memory capacity and island sensitivity due to the memory tasks we utilized; future research could employ working memory tasks which do not test serial order recall capacity (Gieselman et al. 2013).

4 Conclusion

We found no robust relationship between individual differences in working memory and attentional control and island sensitivity. The results of our study, which took several criticisms of Sprouse et al.'s (2012a) approach into account, provide further evidence in line with grammatical theories regarding the source of island effects. Although the extent to which different island phenomena may result from processing pressures remains an intriguing open question, the findings from the current study do not provide evidence supporting an attempt to recast grammatical island constraints for the four island types we tested as due to capacity-based limitations in cognitive resources such as working memory.

Competing Interests

The authors have no competing interests to declare.

References

Abrusán, Marta. 2014. *Weak island semantics*. New York, NY: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199639380.001.0001

Aldosari, Saad. 2015. *The role of individual differences in the acceptability of island violations in native and non-native speakers*. Lawrence, KS: University of Kansas dissertation.

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

Boudewyn, Megan A., Debra L. Long & Tamara Y. Swaab. 2012. Cognitive control influences the use of meaning relations during spoken sentence comprehension. *Neuropsychologia* 50(11). 2659–2668. DOI: https://doi.org/10.1016/j.neuropsychologia.2012.07.019

Boudewyn, Megan A., Debra L. Long & Tamara Y. Swaab. 2013. Effects of working memory span on processing of lexical associations and congruence in spoken discourse. *Frontiers in Psychology* 4. 60. DOI: https://doi.org/10.3389/fpsyg.2013.00060

Bush, George, Paul J. Whalen, Lisa M. Shin & Scott L. Rauch. 2006. The counting Stroop: a cognitive interference task. *Nature Protocols* 1. 230–233. DOI: https://doi.org/10.1038/nprot.2006.35

Case, Robbie, D., Midian Kurland & Jill Goldberg. 1982. Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology* 33(3). 386–404. DOI: https://doi.org/10.1016/0022-0965(82)90054-6

Christensen, Ken Ramshøj & Anne Mette Nyvad. 2014. On the nature of escapable relative islands. *Nordic Journal of Linguistics* 37(1). 29–45. DOI: https://doi.org/10.1017/S0332586514000055

Christensen, Ken Ramshøj, Johannes Kizach & Anne Mette Nyvad. 2013a. Escape from the island: Grammaticality and (reduced) acceptability of *wh*-island violations in Danish. *Journal of Psycholinguistic Research* 42. 51–70. DOI: https://doi.org/10.1007/s10936-012-9210-x

Christensen, Ken Ramshøj, Johannes Kizach & Anne Mette Nyvad. 2013b. The processing of syntactic islands – an fMRI study. *Journal of Neurolinguistics* 26(2). 239–251. DOI: https://doi.org/10.1016/j.jneuroling.2012.08.002

Chomsky, Noam. 1973. Conditions on transformations. In Stephen Anderson & Paul Kiparsky (eds.), *A Festschrift for Morris Halle*, 232–286. New York, NY: Holt, Rinehart and Winston.

Chomsky, Noam. 1981. Principles and parameters in syntactic theory. In Norbert Hornstein & David Lightfoot (eds.), *Explanation in Linguistics: The Logical Problem of Language Acquisition*, 32–75. London: Longman.

Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.

Cinque, Guglielmo. 1990. *Types of Ā-dependencies*. Cambridge, MA: MIT Press.

Conway, Andrew, Michael Kane, Michael Bunting, D. Hambrick, Oliver Wilhelm & Randall Engle. 2005. Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review* 12(5). 769–786. DOI: https://doi.org/10.3758/BF03196772

Daneman, Meredyth & Patricia A. Carpenter. 1980. Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 19(4). 450–466. DOI: https://doi.org/10.1016/S0022-5371(80)90312-6

Dienes, Zoltan. 2014. Using Bayes to get the most out of non-significant results. *Frontiers in Psychology* 5. 781. DOI: https://doi.org/10.3389/fpsyg.2014.00781

Erteschik-Shir, Nomi. 1973. *On the nature of island constraints*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In Alec Marantz, Yasushi Miyashita & Wayne O'Neil (eds.), *Image, Language, Brain: Papers from the First Mind Articulation Project Symposium*, 94–126. Cambridge, MA: MIT Press.

Gieselman, Simone, Robert Kluender & Ivano Caponigro. 2013. Isolating processing factors in negative island contexts. In Yelena Fainleib, Nicholas LaCara & Yangsook Park (eds.), *Proceedings of the 41st Annual Meeting of the North East Linguistic Society*. 233–246. Amherst, MA: Graduate Linguistic Student Association.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press. DOI: https://doi.org/10.1515/COGL.2009.005

Goodall, Grant. 2014. The D-linking effect on extraction from islands and non-islands. *Frontiers in Psychology* 5. 1493. DOI: https://doi.org/10.3389/fpsyg.2014.01493

Goodluck, Helen, Frank Tsiwah & Kofi Saah. 2017. Island constraints are not the result of sentence processing. *Proceedings of the Linguistic Society of America* 2. 1–5. DOI: https://doi.org/10.3765/plsa.v2i0.4068

Hofmeister, Philip, Laura Staum Casasanto & Ivan A. Sag. 2012a. How do individual cognitive differences relate to acceptability judgments? A reply to Sprouse, Wagers, and Phillips. *Language* 88(2). 390–400. DOI: https://doi.org/10.1353/lan.2012.0025

Hofmeister, Philip, Laura Staum Casasanto & Ivan A. Sag. 2012b. Misapplying working-memory tests: A reductio ad absurdum. *Language* 88(2). 408–409. DOI: https://doi.org/10.1353/lan.2012.0033

Hofmeister, Philip, Laura Staum Casasanto & Ivan A. Sag. 2013. Islands in the grammar? Standards of evidence. In Jon Sprouse & Norbert Hornstein (eds.), *Experimental Syntax and Island Effects*, 42–63. New York, NY: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139035309.004

Hofmeister, Philip, Laura Staum Casasanto & Ivan A. Sag. 2014. Processing effects in linguistic judgment data: (super-)additivity and reading span scores. *Language and Cognition* 6(1). 111–145. DOI: https://doi.org/10.1017/langcog.2013.7

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

Huang, Cheng-Teh James. 1982. *Logical relations in Chinese and the theory of grammar*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Hutchison, Keith A. 2007. Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33(4). 645–662. DOI: https://doi.org/10.1037/0278-7393.33.4.645

Jaeggi, Susanne M., Martin Buschkuehl, John Jonides & Walter J. Perrig. 2008. Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences* 105(19). 6829. DOI: https://doi.org/10.1073/pnas.0801268105

Johnson, Adrienne. 2015. *Individual differences in predictive processing: Evidence from subject filled- gap effects in native and non-native speakers of English*. Lawrence, KS: University of Kansas dissertation.

Johnson, Adrienne, Robert Fiorentino & Alison Gabriele. 2016. Syntactic constraints and individual differences in native and non-native processing of wh-movement. *Frontiers in Psychology* 7. 549. DOI: https://doi.org/10.3389/fpsyg.2016.00549

Johnson, Jacqueline S. & Elissa L. Newport. 1991. Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition* 39(3). 215–258. DOI: https://doi.org/10.1016/0010-0277(91)90054-8

Just, Marcel A. & Patricia A. Carpenter. 1992. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99(1). 122–149. DOI: https://doi.org/10.1037/0033-295X.99.1.122

Kane, Michael J. & Randall W. Engle. 2002. The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review* 9. 637–671. DOI: https://doi.org/10.3758/BF03196323

Keshev, Maayan & Aya Meltzer-Asscher. 2019. A processing-based account of subliminal wh-island effects. *Natural Language & Linguistic Theory* 37. 621–657. DOI: https://doi.org/10.1007/s11049-018-9416-1

King, Jonathan & Marcel Adam Just. 1991. Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language* 30. 580–602. DOI: https://doi.org/10.1016/0749-596X(91)90027-H

Kirchner, Wayne K. 1958. Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology* 55. 352–358. DOI: https://doi.org/10.1037/h0043688

Kluender, Robert. 1998. On the distinction between strong and weak islands: A processing perspective. In Peter Culicover & Louise McNally (eds.), *Syntax and Semantics Vol. 29: The Limits of Syntax*, 241–279. San Diego, CA: Academic Press. DOI: https://doi.org/10.1163/9789004373167_010

Kluender, Robert. 2004. Are subject islands subject to a processing account? In Angelo Rodríguez, Vineeta Chand, Ann Kelleher, & Benjamin Schmeiser (eds.), *Proceedings of West Coast Conference on Formal Linguistics 23*, 475–499. Somerville, MA: Cascadilla Press.

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

Kuno, Susumu & Ken-Ichi Takami. 1993. *Grammar and discourse principles: Functional syntax and GB theory*. Chicago, IL: University of Chicago Press.

Kush, Dave, Terje Lohndal & Jon Sprouse. 2018. Investigating variation in island effects: A case study of Norwegian wh-extraction. *Natural Language & Linguistic Theory* 36. 743–779. DOI: https://doi.org/10.1007/s11049-017-9390-z

Kush, Dave, Terje Lohndal & Jon Sprouse. 2019. On the island sensitivity of topicalization in Norwegian: An experimental investigation. *Language* 95. 393–420. DOI: https://doi.org/10.1353/lan.2019.0051

Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26. DOI: https://doi.org/10.18637/jss.v082.i13

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

Marthardjono, Gita. 1993. *Wh-movement in the acquisition of a second language: A crosslinguistics study of three languages with and without overt movement*. Ithaca, NY: Cornell University dissertation.

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

Michel, Daniel. 2014. *Individual cognitive measures and working memory accounts of syntactic island phenomena*. San Diego, CA: University of California, San Diego dissertation.

Morey, Richard D. & Jeffrey N. Rouder. 2018. BayesFactor: Computation of Bayes factors for common designs. *R package version 0.9.12-4.2*. https://CRAN.R-project.org/package=BayesFactor.

Parker, Dan, Michael Shvartsman & Julie Van Dyke. 2017. The cue-based based retrieval theory of sentence comprehension: New findings and new challenges. In Linda Escobar, Vicenç Torrens, & Teresa Parodi (eds.), *Language Processing and Disorders*, 121–144. Newcastle: Cambridge Scholars Publishing.

Pañeda, Claudia, Sol Lago, Elena Vares, João Veríssimo & Claudia Felser. 2020. Island effects in Spanish comprehension. *Glossa: A Journal of General Linguistics* 5(1). 21. DOI: https://doi.org/10.5334/gjgl.1058

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

R Core Team. 2019. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org.

Ross, John Robert. 1967. *Constraints on variables in syntax*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Sprouse, Jon. 2007. Rhetorical questions and wh-movement. *Linguistic Inquiry* 38(3). 572–580. DOI: https://doi.org/10.1162/ling.2007.38.3.572

Sprouse, Jon, Ivano Caponigro, Ciro Greco & Carlo Cecchetto. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language & Linguistic Theory* 34. 307–344. DOI: https://doi.org/10.1007/s11049-015-9286-8

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

Sprouse, Jon, Matt Wagers & Colin Phillips. 2012b. Working-memory capacity and island effects: A reminder of the issues and the facts. *Language* 88. 401–407. DOI: https://doi.org/10.1353/lan.2012.0029

Sprouse, Jon & Sandra Villata. in press. Island effects. In Grant Goodall (ed.), *The Cambridge Handbook of Experimental Syntax*. Cambridge University Press.

Sprouse, Jon, Shin Fukuda, Hajime Ono & Robert Kluender. 2011. Reverse island effects and the backward search for a licensor in multiple wh-questions. *Syntax* 14. 179–203. DOI: https://doi.org/10.1111/j.1467-9612.2011.00153.x

Szabolcsi, Anna. 2006. Strong vs. weak islands. In Martin Everaert & Henk van Riemsdijk (eds.), *The Wiley Blackwell Companion to Syntax*, 479–531. Malden, MA: Blackwell. DOI: https://doi.org/10.1002/9780470996591.ch64

Szabolcsi, Anna & Frans Zwarts. 1993. Weak islands and an algebraic semantics for scope taking. *Natural Language Semantics* 1. 235–284. DOI: https://doi.org/10.1007/BF00263545

Tagliaferri, Bruno. 2005. Paradigm. Perception Research Systems, Inc. www.perception-researchsystems.com.

Truswell, Robert. 2007. Extraction from adjuncts and the structure of events. *Lingua* 117. 1355–1377. DOI: https://doi.org/10.1016/j.lingua.2006.06.003

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

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

Zirnstein, Megan, Janet G. van Hell & Judith F. Kroll. 2018. Cognitive control ability mediates prediction costs in monolinguals and bilinguals. *Cognition* 176. 87–106. DOI: https://doi.org/10.1016/j.cognition.2018.03.001