

## **Island effects and resumption in spoken Jordanian Arabic: an auditory acceptability judgment study**

Rania Al-Aqarbeh  
Mutah University  
Department of English Language and Literature  
Mu'tah, Karak, Jordan  
[rania11@mutah.edu.jo](mailto:rania11@mutah.edu.jo)

Jon Sprouse  
University of Connecticut  
Department of Linguistics  
Storrs, Connecticut, USA  
[jon.sprouse@uconn.edu](mailto:jon.sprouse@uconn.edu)

### **Abstract**

One central claim in the literature on resumptive pronouns is that resumptive pronouns ameliorate island effects. For languages that do not allow resumption as a grammatical option (intrusive resumption), a number of recent acceptability judgment studies have cast doubt on this claim by showing small or no amelioration effects. For languages that do allow resumption as a grammatical option (grammaticalized resumption), there have been relatively few judgment studies; but the few that exist have also shown small or no amelioration effects. In this article, we further investigate this issue for spoken Jordanian Arabic, a relatively understudied grammaticalized resumption language. We test four island types (adjunct, complex NP, *wh*, and *whether*) and two dependency types (bare *wh*-words and relative clauses) in an auditory judgment study. Our results revealed seven island effects with gaps. With resumption, one shows clear amelioration; one shows a novel form of amelioration involving a subliminal island effect; and five show no amelioration. This pattern of results raises the possibility that the claim that grammaticalized resumption ameliorates island effects is less universal, and therefore more complex, than previously thought. Though not the focus of our study, our results also suggest a number of new avenues of research on topics related to resumption and movement, such as the uniformity of *A'*-dependencies with respect to island effects, the role of movement in resumption, the cross-linguistic variation in island effects, and the existence of subliminal island effects.

**Keywords:** Jordanian Arabic, island effects, resumption, experimental syntax, acceptability judgments

## 1. Introduction

One central claim in the literature on resumptive pronouns is that they ameliorate island effects (Ross 1967; Kroch 1981; Chao and Sells 1983; Sells 1984; Engdahl 1985; et seq.; see Choueiri 2017 and McCloskey 2017 for reviews). Crucially, there are two components to this claim. The first is that, for languages that do not allow resumption as a grammatical option (*intrusive resumption*; Chao and Sells 1983; Sells 1984), resumption within an island will reduce the island effect size substantially, though perhaps not to zero. The second is that, for languages that allow resumption as a fully grammatical option (*grammaticalized resumption*), resumption within an island will completely eliminate the island effect. The claim about *intrusive resumption* has generated quite a bit of interest in the formal experimental literature. Table 1 below summarizes seven recent experimental studies on intrusive resumption: Alexopoulou and Keller 2007; Omaki and Nakao 2010; Heestand et al. 2011; Keffala 2013; Beltrama and Xiang 2016; Ackerman et al. 2018; Morgan and Wagers 2018. To facilitate comparison, the effect of resumption in each study has been transformed to a proportion of the size of the scale in the study. The results that emerge from these studies suggest that the intrusive resumption effect is relatively small, ranging from .01 to .08 of the scale for discrete scale tasks. And, in fact, the resumption effect is detected in the wrong direction (resumption worse than gaps) in about one third of the tests reviewed here (7 out of 19). It is only with a forced-choice task in which participants must explicitly choose between two sentences, as in Ackerman et al. 2018, that a sizeable resumption effect is detected. The consensus emerging from this literature is that intrusive resumption does reduce island effects, but the reduction is relatively small (though resumption may yield larger differences in production, as in Ferreira and Swets 2005, or in forced choice, as in Ackerman et al. 2018).

*Table 1: Seven prominent experimental studies on intrusive resumption. The islands are adjunct islands (adj), complex NP islands (np), subject islands (sub), relative clause islands (rc), wh-islands (wh), and polar questions headed by either whether or if (yn). Wh-simple means bare wh-words like who or what. Wh-complex means a wh-phrase like which language. The task column lists the task with the size of the scale in parentheses. Ratings were either taken from the text or estimated based on published plots. Ratings were linearly transformed to a [0,1] scale. ME stands for magnitude estimation. In the AK07 study, the ME responses were divided by the modulus and then log-transformed, so there are no bounds on the scale. We chose -.6 and .6 as a conservative bound (potentially leading to inflated effect sizes) for transformation purposes. Gray cells were not tested in the studies.*

study	language	dependency	adj	np	sub	rc	wh	yn	task (scale)
AK07	English	wh-simple				.03		-.17	acceptability (ME)
AFY18	English	wh-complex	.22			.44		.50	forced-choice (binary)
ON10	English	wh-complex				-.05			acceptability (1-7)
BX16	English	rel. clause				.08			comprehensibility (1-7)
BX16	English	rel. clause				.00			acceptability (1-7)
HXP11	English	rel. clause	-.05	-.03		-.02			acceptability (1-7)
K13	English	rel. clause				.07	.06		acceptability (1-11)
MW18	English	rel. clause	-.01	.07	.04			.01	acceptability (1-7)
AK07	German	wh-simple				.08		.02	acceptability (ME)
AK07	Greek	wh-simple				.00		-.06	acceptability (ME)
BX16	Italian	rel. clause				.07			comprehensibility (1-7)

The second component of the resumption claim – that *grammaticalized resumption* within an island will completely eliminate the island effect – has, to date, been investigated in relatively few formal experimental studies. Table 2 below reports our best attempt to collate formal experimental studies; we found only three: Farby et al. 2010; Keshev and Meltzer-Asscher 2017; Tucker et al. 2019. These studies explore only two languages, Hebrew and Modern Standard Arabic, and comparatively few island types. Nonetheless, the results of these three studies are tantalizingly similar to the results of the studies on intrusive resumption. Five out of eight tests show either an effect smaller than .05 of the scale or an effect in the wrong direction. In the remaining three tests the effect is more detectable, ranging from .12 to .20 of the scale, but as the authors themselves note in the papers, these small increases are not enough to move the sentences into the upper half of the range of acceptability as one might expect given the claim that grammaticalized resumption fully eliminates island effects. This raises the question of whether the claim about grammaticalized resumption should be revised as well.

*Table 2: Three experimental studies on grammaticalized resumption. (For details see the caption of Table 1.)*

study	language	dependency	adj	np	sub	rc	wh	yn	task (scale)
FDWBS10	Hebrew	rel. clause				.12			acceptability (1-5)
KMA17	Hebrew	rel. clause		.20					acceptability (1-7)
TISA19	MSA	wh-simple	-.05	.01			.03		acceptability (1-7)
TISA19	MSA	wh-complex	.18	.00			.04		acceptability (1-7)

Given the prominence of the claim that resumption eliminates island effects, and given the importance of grammaticalized resumption languages to that claim, we believe the previous results discussed above suggest a need to systematically test a wider range of island types and dependency types in grammaticalized resumption languages. To that end, we report here a series of auditory acceptability judgment experiments on spoken Jordanian Arabic – a relatively understudied grammaticalized resumption language. We test four island types (whether, wh-, complex NP, and adjunct islands) and two dependency types (bare wh-words and relative clauses) using the factorial definition of island effects (see section 2.1 for a complete description) to quantify the size of the island effects for both gaps and resumptive pronouns (and therefore the difference between them). Though the primary theoretical goal of this study is to test the effect of resumption on island effects, the design of the study also provides information that is relevant for several other aspects of syntactic theory, such as the uniformity of A'-dependencies with respect to island effects, the role of movement in resumption, the scope of cross-linguistic variation in island effects, and the existence of subliminal island effects. As such, we will discuss each of these issues in turn in the discussion in section 4.

## 2. The design of the experiments

### 2.1 Logic and design

For this study we used the factorial definition of island effects developed in Sprouse 2007, Sprouse et al. 2011, and Sprouse et al. 2012, and augmented it to include resumption (described in detail below). We chose the factorial definition because we believe it matches the logic that

has historically been used by syntacticians to define island effects, albeit translated into the factorial terminology that is typical of formal experiments. We note this choice at the outset because the factorial design has not been used in the literature on intrusive resumption for the obvious reason that, in intrusive resumption languages, it is impossible to create the non-island conditions that the factorial design calls for with resumptive pronouns, because resumption is not licensed in non-island structures in these languages. As such, the typical experimental design in the intrusive resumption literature is to use a pairwise comparison of an island condition with a gap to an island condition with a resumptive pronoun. This limitation does not exist for grammaticalized resumption languages (see also Tucker et al. 2019), therefore we employ the full factorial definition here to better match the standard logic of syntactic studies of island effects. One added benefit of using the factorial definition when it is possible to do so is that these results can be integrated into the growing cross-linguistic experimental literature on island effects that uses this design (a.o., Almeida 2014; Christensen et al. 2013; Kim and Goodall 2016; Ko et al. 2019; Kush et al. 2018; 2019; Lu et al. 2020; Omaki et al. 2019; Stepanov et al. 2018).

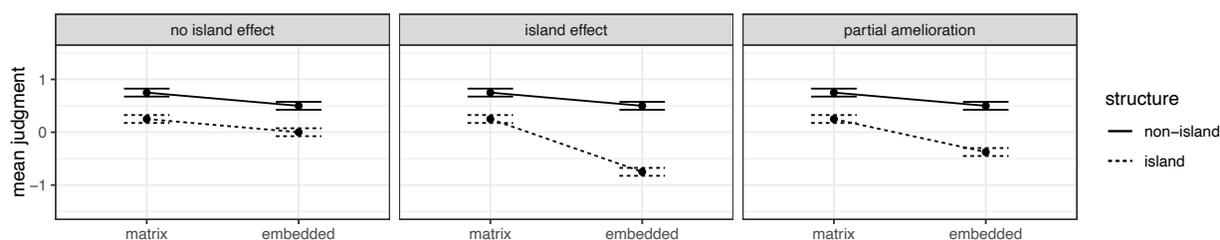
The standard version of the factorial design has two factors: DEPENDENCY manipulates the length of the dependency based on the location of the gap (matrix/embedded), and STRUCTURE manipulates the structure of the embedded clause (non-island/island). To test resumption, we added a third factor manipulating the tail of the dependency (gap/resumptive pronoun). In principle, this should yield eight conditions in a 2x2x2 design (DEPENDENCY x STRUCTURE x TAIL). However, it is not possible to have a resumptive pronoun in the matrix subject position in Jordanian Arabic. The result is therefore six conditions in a 2x2+2 design.

(1) A 2x2+2 factorial design for *whether*-islands with a *wh*-dependency.

- |    |   |          |            |     |
|----|---|----------|------------|-----|
| a. | mīn __ gāl          innu is-sīnama    ʕaradʕ-at          filim hindī?     | matrix   | non-island | gap |
|    | who __ said.3sgm that the-cinema presented-3sfg film Hindi                |          |            |     |
|    | ‘Who __ said that the cinema presented a Hindi film?’                     |          |            |     |
| b. | wāyʕ qaal          rāmī innu is-sīnama    ʕaradʕ-at __?                   | embedded | non-island | gap |
|    | what said.3sgm Rami that the-cinema presented-3sgf __                     |          |            |     |
|    | ‘What did Rami say that the cinema presented __?’                         |          |            |     |
| c. | mīn __ saʕal          iða    is-sīnama    ʕaradʕ-at          filim hindī? | matrix   | island     | gap |
|    | who __ asked.3sgm whether the-cinema presented-3sgf film Hindi            |          |            |     |
|    | ‘Who __ asked whether the cinema presented a Hindi film?’                 |          |            |     |
| d. | wāyʕ saʕal          rāmī iða    is-sīnama    ʕaradʕ-at __?                | embedded | island     | gap |
|    | what asked.3sgm Rami whether the-cinema presented-3sgf __                 |          |            |     |
|    | ‘What did Rami ask whether the cinema presented __?’                      |          |            |     |
- 
- |    |   |          |            |         |
|----|---|----------|------------|---------|
| e. | wāyʕ gāl          rāmī innu is-sīnama    ʕaradʕ-at- <i>uh</i> ?     | embedded | non-island | resump. |
|    | what said.3sgm Rami that the-cinema presented-3sgf- <i>it</i>       |          |            |         |
|    | ‘What did Rami say that the cinema presented <i>it</i> ?’           |          |            |         |
| f. | wāyʕ saʕal          rāmī iða    is-sīnama    ʕaradʕ-at- <i>uh</i> ? | embedded | island     | resump. |
|    | what asked.3sgm Rami whether the-cinema presented-3sgf- <i>it</i>   |          |            |         |
|    | ‘What did Rami ask whether the cinema presented <i>it</i> ?’        |          |            |         |

The value of the factorial definition is that it isolates the island effect in the interaction between DEPENDENCY and STRUCTURE (while subtracting out the main effects of those factors). If there is no island effect, we expect to see no interaction as illustrated in the left panel of Figure 1. If there is an island effect, we expect to see a superadditive interaction as illustrated in the center panel. Crucially, we can look for these interactions for both gaps and resumption using this design – to determine if an island effect that is present with gaps is eliminated with resumption. We can also look for either total amelioration, which will result in no interaction, or partial amelioration, which will result in a smaller interaction, as illustrated in the right panel of Figure 1.

Figure 1: Possible outcomes for the factorial design.



For space reasons, here we list only the island-violating sentences for each of the eight combinations of islands and dependency types, with glosses but not translations. The full list of materials is available on the authors' websites.

(2) Examples of the island-violating conditions for the four island types and two dependency types tested.

- a. wāyʃ saʔal rāmī iða is-sīnama ʃarad'-at \_\_? whether  
 what asked.3sgm Rami whether the-cinema presented-3sgf \_\_  
 'What did Rami ask whether the cinema presented \_\_?'
- b. wāyʃ saʔal bahā? lāyʃ amānī ʃaṭab-at \_\_? wh  
 what asked.3sgm Bahaa why Amani crashed-3sgf \_\_  
 'What did Bahaa ask why Amani crashed \_\_?'
- c. wāyʃ naʃar sāmī il-iʃaʃah innu intiʃār iʃtar-at \_\_? np  
 what spread.3sgm Sami the-rumor that Intisar bought-3sgf \_\_  
 'What did Sami spread the rumor that Intisar bought \_\_?'
- d. wāyʃ ziʃil nāʃir laʔinnu il-maktabih ʃabaʃ-at \_\_? adjunct  
 what got-angry.3sgm Naser because the-press printed-3sgf \_\_  
 'What did Naser get angry because the press printed \_\_?'
- 
- e. baʃrif il-mudīr illi is-skirtāyrah saʔal-at iða il-idarah ixtār-at \_\_. whether  
 know.1sg the-manager who the-secretary asked-3sgf whether the-board chose-3sgf \_\_  
 'I know the manager who the secretary asked whether the board chose \_\_.'

- f. ʃif-it il-asatiðeh illi ʃāhab-ak saʔal lāyʃ iṭ-ṭullāb waddaʃ-u \_\_. wh  
saw.1sg the-teachers who friend-your asked.3sgm why the-students said-goodbye.3plm \_\_  
'I saw the teachers who your friend asked why the students said goodbye to \_\_.'
- g. baʃrif is-sikirtāyrah illi il-katibih simiʃ-it il-ifāʃah innu il-mudīr idzawwaz \_\_. np  
know.1sg the-secretary who the-clerk heard.3sgm the-rumor that the-principal married.3sgm. \_\_  
'I know the secretary who the clerk heard the rumor that the principal married \_\_.'
- h. baʃrif il-mudīr illi il-binit imbaṣaṭ-at laʔinnu jār-na ʃazam \_\_. adjunct  
know.1sg the-manager who the-girl felt-happy-3sgf because neighbor-our invited.3sgm \_\_  
I know the manager who the girl felt happy because our neighbor invited \_\_.'

## 2.2 Choice of modality

The language context of Jordan is diglossic: Jordanian Arabic is the dominant native language, and is used for daily spoken conversation, and daily personal written communication like text messages and emails; Modern Standard Arabic is also learned by many citizens as a second language, as it is used to some degree for formal educational instruction, and for formal written products like books and magazines. Our goal is to study the native speaker judgments that our participants have for Jordanian Arabic, and not for Modern Standard Arabic (henceforth MSA). Since MSA is primarily a written language, we decided to employ an auditory version of a typical acceptability judgment task to minimize the possibility that the participants would apply MSA grammatical rules to their judgments, and ideally to maximize the possibility that they would engage their native speaker judgments of spoken Jordanian Arabic. Additionally, the first author spoke to the participants in Jordanian Arabic during the laboratory visit, and the instructions for the experiment directed the participants to imagine conversing with a friend in Jordanian Arabic, and to judge whether a native speaker of Jordanian Arabic could produce these sentences.

## 2.3 Materials and survey construction

Each participant completed a survey that consisted of 42 items: 6 practice items at the beginning of the survey (but not marked as such), followed by 12 experimental items and 24 filler items pseudorandomized to avoid related experimental items appearing in succession. The 24 experimental items consisted of 1 token of each of the 6 conditions for each of the two island/dependency combinations in their experiment. We chose one judgment per condition per participant because we wanted to maintain a filler-to-target ratio of 2:1 while still maintaining a reasonable experiment length. We compensated for the increased risk of noise with one judgment per condition by testing sample sizes (40+) that are likely to yield high statistical power for medium and larger effect sizes (Sprouse and Almeida 2017). We created 6 lexically matched sets of items per island/dependency combination. The items were then distributed among experimental lists using a Latin square procedure so that participants saw a unique lexical item in each condition. The 24 filler items consisted of 8 sentence types expected to receive low ratings, 8 expected to receive ratings in the middle of the scale, and 8 expected to receive high ratings. The first author recorded all items. We then used Praat (Boersma 2001) to normalize the volume to 70 dB and to gently ramp up the volume during the first 50ms of each sentence to avoid jarring onsets.

## 2.4 Participants and presentation

Participants were students at Mutah University in Jordan. They were all self-reported native speakers of Jordanian Arabic. They received course credit for their participation. For the four experiments, we recruited 40, 42, 43, and 40 participants, respectively. (For experiments 2, 3, and 4 this means that we have 42, 43, and 40 ratings per condition; a coding error in experiment 1 lead to minor deviations around 40 for the number of ratings per condition.) Participants completed the experiment during a visit to the first author's laboratory. The auditory experiments were implemented using PennController for IBEX (Drummond 2019; Zehr and Schwarz 2018). Each sentence received its own presentation screen with a 1 *غير مقبول إطلاقاً* (totally unacceptable) to 7 *مقبول تماماً* (perfectly acceptable) scale. The sentence played automatically upon advancement to the screen. Participants could replay the sentence by clicking on an icon. Participants indicated their rating by clicking on the appropriate number or by typing that number on the keyboard. Because this is a first exploration of Jordanian Arabic islands, we did not exclude any participants from analysis.

## 3. The results of the experiments

In this section we describe the results of the experiments, with a particular focus on the presence or absence of island effects, and the presence or absence of an amelioration effect of resumption. We will discuss the theoretical relevance of these results for theories of resumption, movement, and island effects in section 4. We z-score transformed the results of each participant prior to analysis to reduce the impact of common forms of scale bias. We first discuss the results for WH-dependencies, looking at both the overall pattern of island effect (section 3.1) and the finer-grained details of gaps versus resumption (section 3.2); we then do the same for RC-dependencies (sections 3.3. and 3.4).

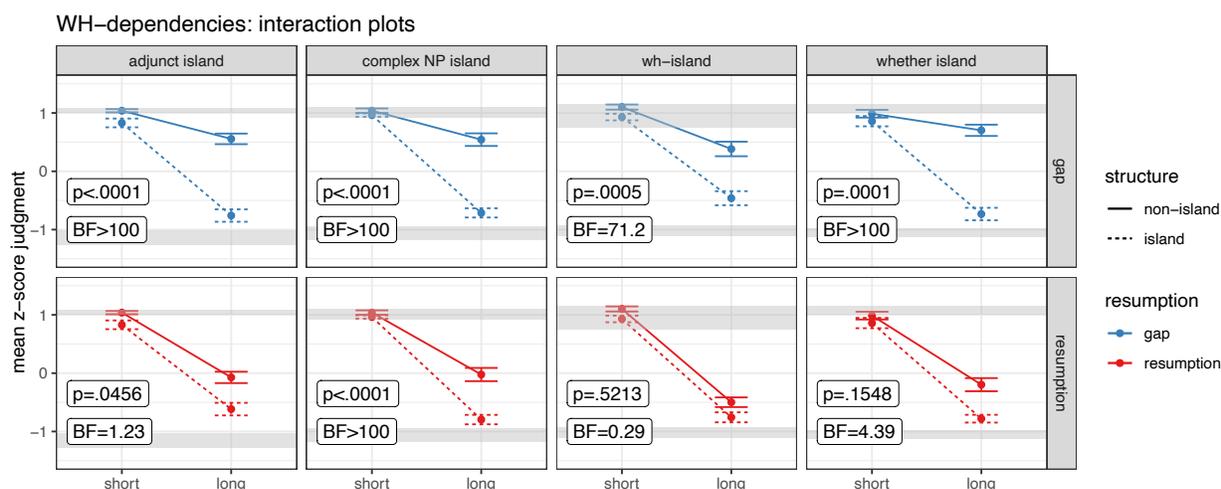
For statistical analysis, we ran two sets of tests. In the first set, we constructed linear mixed effects models with DEPENDENCY and STRUCTURE as fixed effects and participant and item as random effects (intercepts only) for each island, dependency type, and tail type using the lme4 package in R (Bates et al. 2015). We calculated  $p$ -values using the lmerTest package, which uses the Satterthwaite approximation for degrees of freedom (Kuznetsova et al. 2017). The full set of results are in the appendix. We have added the interaction term  $p$ -value to each cell of the plot. In the second set, we derived Bayes factors for the interaction term for the fixed effects in the linear models using the BayesFactor package (Morey and Rouder 2018). The Bayes factors reported here are of the  $BF_{10}$  type: they report the ratio of the likelihood of the data under the experimental hypothesis ( $H_1$ ) to the likelihood of the data under the null hypothesis ( $H_0$ ). (For example, a  $BF_{10}$  of 3 indicates that the data is 3x more likely under a theory in which the interaction is present than one in which the interaction is absent.) One advantage of including Bayes factors in addition to null hypothesis tests is that Bayes factors can be used to evaluate the null hypothesis directly (i.e., a  $BF_{10}$  of .33 would indicate that the data is 3x more likely under the null hypothesis than the experimental hypothesis). We will therefore look for three patterns in the statistical results: a  $p$ -value less than .05 and a  $BF_{10}$  greater than 3 (following Jeffries 1961) to show for the presence of an interaction ( $H_1$ ); a  $p$ -value greater than .05 and a  $BF_{10}$  less than 0.33 (following Jeffries 1961) to show evidence that there is no interaction ( $H_0$ ); and a  $p$ -value greater than .05 and a  $BF_{10}$  between 0.33 and 3 to indicate a lack of evidence for either

hypothesis. For readers interested in other statistical analyses, the full results are available for download and re-analysis on the authors' websites.

### 3.1 Island effects for WH-dependencies in Jordanian Arabic

Figure 2 reports the means and standard errors for WH-dependencies in Jordanian Arabic, arranged in an interaction plot as described in Figure 1 in section 2.1. The top row reports the results for gap conditions, and the bottom row reports the results for resumption conditions. In the gap conditions (top row), we see large superadditive interactions for all four island types (adjunct, complex NP, wh, and whether islands) that match the (monotonic) superadditivity pattern that we take to be the hallmark of an island effect, with the island-violating (island/long) condition in the lower half of the z-score scale, and the other three conditions in the upper half of the scale. These interactions are confirmed by both null hypothesis tests with  $p$ -values that are substantially lower than the conventional threshold of .05, and by Bayes factors that are substantially greater than the conventional threshold of 3.

*Figure 2: Interaction plots for island effects with WH-dependencies. Points are condition means. Error bars represent estimated standard error. For space reasons,  $p$ -values are rounded to a floor of .0001 and Bayes factors are rounded to a ceiling of 100. The horizontal gray bars represent the range of mean ratings for the eight unacceptable and eight acceptable filler types.*



In the resumption conditions (bottom row), the results are more complicated. Adjunct islands show what appears to be a small superadditive interaction. This interaction yields a  $p$ -value that is less than the conventional threshold of .05, but the Bayes factor of 1.23 suggests that the data is nearly equally likely under either hypothesis (and below the conventional threshold of 3). Therefore it is difficult to draw a firm conclusion about whether there is an island effect present or not. Complex NP islands show a superadditive interaction pattern that is confirmed by both a  $p$ -value that is substantially less than the conventional threshold of .05 and a Bayes factor that is substantially greater than the conventional threshold of 3. This suggests there is a complex NP island effect with resumption. Wh-islands show no interaction pattern as the lines are nearly parallel. The lack of interaction is confirmed by a  $p$ -value that is substantially above the conventional threshold of .05, and perhaps more importantly, a Bayes factor that is below the

conventional threshold of 0.33 in support of the null hypothesis (suggesting that the data is more than 3 times more likely under the null hypothesis than the under the experimental hypothesis). This suggests there is no wh-island effect with resumption. Finally, whether islands show what appears to be a small superadditive interaction; however, our two statistical tests yield conflicting results: the  $p$ -value is substantially greater than the conventional threshold of .05, suggesting no evidence of an effect, while the Bayes factor is greater than the conventional threshold of 3, suggesting that the data is more than 3 times more likely under the experimental hypothesis that there is an interaction. Therefore, it is difficult to draw a firm conclusion about the presence or absence of whether islands with resumption.

Though all four islands show island effects with gaps, and three out of the four islands show no island effects or inconclusive results with resumption, we would argue that there is no evidence of the classic amelioration effect of resumption in the WH-dependency results. This is because the classic amelioration effect is defined as the elimination of island effects due to the island-violating condition (island/long) increasing in acceptability with resumption (ideally to the upper half of the acceptability scale), whereas the results of our experiments suggest that the island-violating condition does not increase much at all with resumption in WH-dependencies. Instead, the differences between the gap conditions and the resumption conditions seem to be driven almost entirely by the non-island/long conditions (which are declarative CPs): the non-island/long conditions are rated in the top half of the acceptability scale with gaps, but decrease to the midpoint of the scale (near 0) with resumption. Because the non-island/long conditions do not involve islands, any variability we see between gaps and resumption is most likely due to participants' preferences for either gaps or resumption in WH-dependencies. The pattern that we see in Figure 2 suggests that gaps are generally preferred with WH-dependencies, and resumption may be dispreferred. We investigate this in more detail in the next subsection.

### 3.2 The preference for gaps versus resumption in WH-dependencies

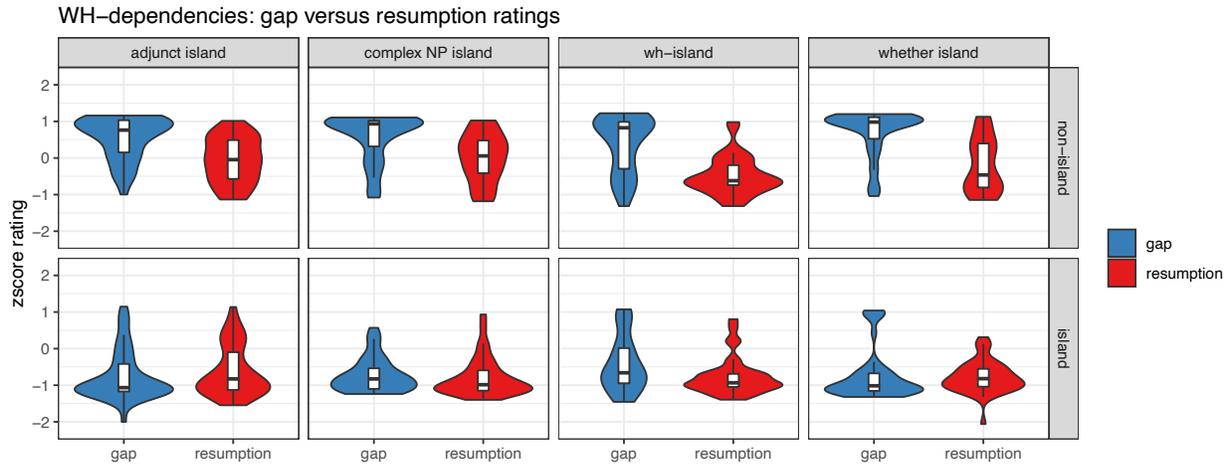
To see the pattern of preference for gaps versus resumption more clearly, we can plot the distributions of gap and resumption conditions side-by-side. Figure 3 uses violin plots to show the (mirrored) distribution of judgments for each condition with a boxplot overlaid to show the median rating (horizontal line) and interquartile range (white box). The columns report each island type. The top row reports the non-island/long conditions (declarative CPs). Here we see that the gap versions (in blue) have most of their probability mass in the upper portion of the scale (near 1), whereas the resumption version (in red) have a much flatter distribution spanning both halves of the scale for adjunct, complex NP, and whether islands, resulting in a mean rating near the midpoint of the scale. Wh-islands have substantial probability mass in the lower half of the scale, resulting in a lower mean rating.<sup>1</sup> The bottom row reports the island/long conditions,

---

<sup>1</sup> The non-island/long conditions for all four islands are structurally identical – they are declarative CPs. Therefore the differences between the columns represent only lexical differences in the specific sentences tested, and individual differences in the participants (because these are separate experiments with separate participant samples). In short, the non-island/long results are non-identical replications of each other. Therefore we are reluctant to conclude anything about resumption based on this difference in the wh-island experiment. It may simply reflect differences among the specific participants in that experiment.

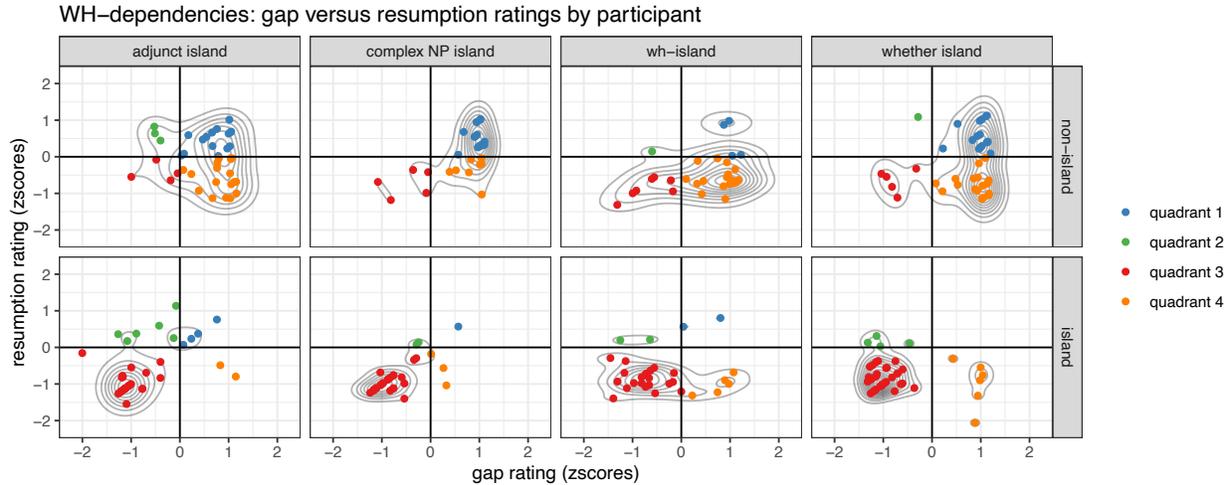
that is, the island-violating conditions. Here we clearly see the invariance between the gap and resumption versions reported in the previous section: both show substantial probability mass in the lower half of the acceptability scale (near -1) for all four islands, resulting in low mean ratings.

Figure 3: Violin plots of the distributions of judgments for the non-island/long and island/long conditions for WH-dependencies. Boxplots indicate medians and the interquartile range.



We can also view the preference for gaps versus resumption within each participant by plotting each participant's judgments of gap conditions against the judgments of resumption conditions in a scatterplot. We do this in Figure 4, with gap ratings on the x-axis and resumption ratings on the y-axis, for both the non-island/long conditions (top row) and the island/long conditions (bottom row), and for all four islands (columns). We divide each plot into four quadrants: a point in the top right quadrant (quadrant 1) represents a participant who rated the gap version of the condition in the upper (positive) half of the z-score scale, and also rated the resumption version of the condition in the upper half of the scale; a point in the top left quadrant (quadrant 2) represents a participant who rated the gap condition in the lower half of the scale, and the resumption condition in the upper half (as in a classic resumption effect); the bottom left (quadrant 3) represents a low judgment for both conditions; and the bottom right (quadrant 4) represents a high judgment for the gap condition and a low judgment for resumption. We have added two features to make the plot a bit easier to read: unique colors for the points in each of the four quadrants, and two-dimensional (joint) probability density estimates to draw attention to the density of the points in each location. Similar to a topographic map, concentric circles that are closer together represent higher density.

Figure 4: Scatterplots of gap versus resumption versions of the non-island/long and island/long conditions for WH-dependencies. The lines represent two-dimensional (joint) probability density estimates. The points are colored according to the quadrants defined by the midpoint (0) of the z-score scales.



We turn first to the non-island/long conditions (top row), as this shows us each participant's preference for gaps versus resumption without the interfering complication of islands: participants in the top-right (blue) allow both; participants in the top-left (green) prefer resumption; participants in the bottom-left disprefer both (red); and participants in the bottom-right (orange) prefer gaps. As mentioned above, all four island types use the same non-island structure – a declarative CP. This means that the four columns represent four (non-identical) replications of extraction from declarative CP. With that in mind, we can see that the first and fourth columns show a relatively even split between participants who allow both gaps and resumption (quadrant 1), and participants who prefer gaps (quadrant 4); while the second column shows primarily participants who allow both (quadrant 1), and the third column shows participants who primarily prefer gaps (quadrant 4). There are a few participants in the other two quadrants, but given the centers of gravity in quadrants 1 and 4, these are likely outliers (as we did no outlier removal prior to analysis). From this we conclude that gaps are universally allowed in WH-dependencies in Jordanian Arabic, and that participants appear to split into two groups when it comes to resumption – those that allow it, and those that do not.

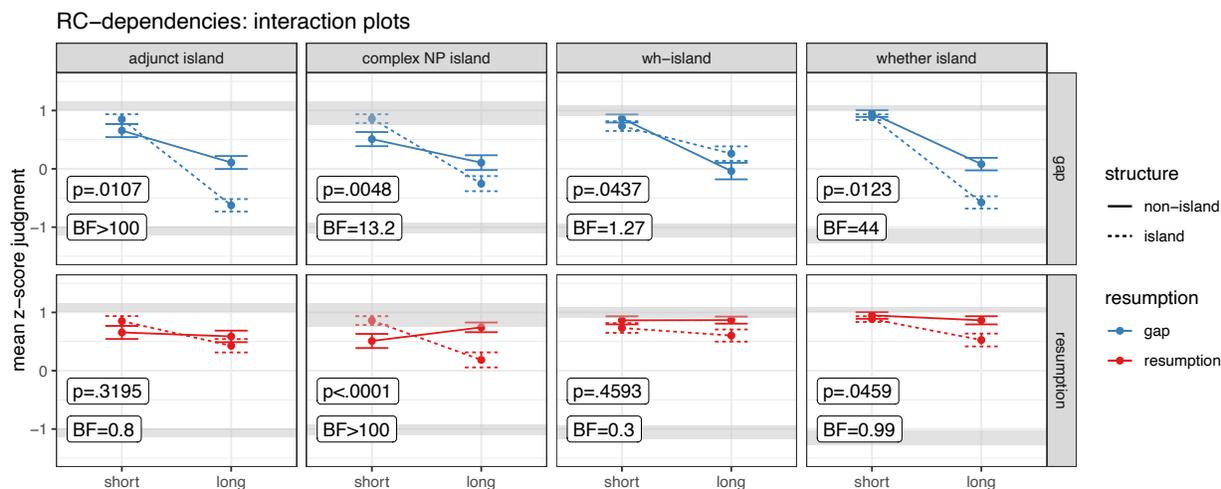
Turning next to the island/long conditions, that is, the island-violating conditions (bottom row), we can use the relationship between gap and resumption judgments to classify each participant according to the classic amelioration effect. We would expect a participant who shows the classic amelioration effect to rate the gap version low and the resumption version high, and therefore to be in the top-left quadrant (quadrant 2, in green). We would expect a participant who shows no amelioration effect to rate both the gap version and the resumption version low, and therefore to be in the bottom-left quadrant (quadrant 3, in red). The other two quadrants would contain participants who did not rate the gap version low, which suggests either that they do not have island effects or that these trials were outliers. What we see in Figure 3 is that all four island types show large concentrations of participants in the bottom-left quadrant (quadrant 3, in red), suggesting that the vast majority of participants rated the island-violating sentences low with both gaps and resumption, indicating no amelioration effect. Though there are a few

participants in the amelioration quadrant (top-left, quadrant 2, in green), the number in that quadrant is roughly equal to the number in the two quadrants that are likely populated by outliers, which leads us to conclude that the participants in quadrant 2 are also likely outliers.

### 3.3 Island effects for RC-dependencies in Jordanian Arabic

We can perform the same set of analyses for RC-dependencies. Figure 5 reports the means and standard errors for RC-dependencies in Jordanian Arabic.

*Figure 5: Interaction plots for island effects with RC-dependencies in Jordanian Arabic. Points are condition means. Error bars represent estimated standard error. For space reasons,  $p$ -values are rounded to a floor of .0001 and Bayes factors are rounded to a ceiling of 100. The horizontal gray bars represent the range of mean ratings for the eight unacceptable and eight acceptable filler types.*



Turning first to the gap conditions (top row), we see that there are superadditive interactions indicative of island effects for adjunct islands and whether islands. These interactions are confirmed by  $p$ -values that are less than the conventional threshold of .05 and Bayes factors that are substantially greater than the conventional threshold of 3. There is also an interaction for complex NP islands, confirmed by a  $p$ -value that is less than the conventional threshold and a Bayes factor that is greater than the conventional threshold, but the shape of the interaction is not the canonical form. The canonical form of an island effect is a monotonic interaction driven primarily by an exceptionally low rating in the island-violating condition. The interaction that we see for complex NP islands in the gap conditions is non-monotonic, and is driven by the combination of a higher-than-expected island/short condition and a moderately low island-violating condition (only -0.25 compared to -0.6 for adjunct and whether islands). It is thus difficult to draw a firm conclusion about the presence or absence of a complex NP island effect. Finally, wh-islands show an interaction in a direction that is opposite than the direction predicted – the island/long condition is rated higher than the non-island/long condition. This is not an island effect. It is also not clear if this interaction is robust, as the two statistical tests yield contradictory results: the  $p$ -value is below the conventional threshold of .05, but the Bayes factor is very close to 1, suggesting that the data is equally likely under both hypotheses. We thus

conclude that there is strong evidence of adjunct and whether islands, inconclusive evidence for complex NP islands, and no evidence for wh-islands.

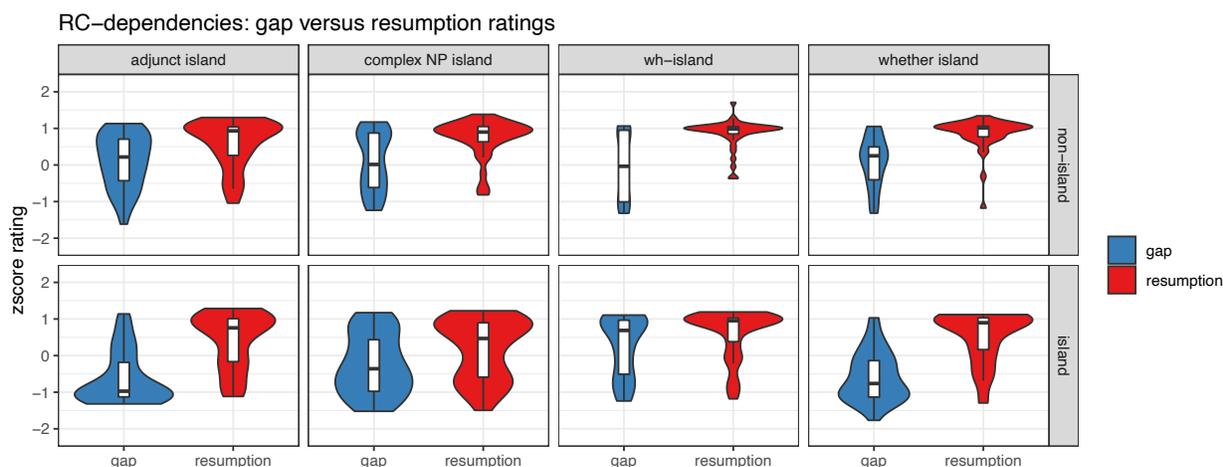
Turning next to the resumption conditions (bottom row), we find a relatively complex pattern. First, adjunct islands show no interaction pattern visually, and no evidence of an interaction is detected by the null hypothesis test (the  $p$ -value is greater than the conventional threshold). However, the Bayes factor is 0.8, which suggests that the data is roughly equally likely under both hypotheses. Thus we conclude that it is probable that there is no island effect with adjunct islands, but that we likely need to run a replication study to confirm this. For complex NP islands, we once again find an interaction that is confirmed by both of our statistical tests, but with the non-canonical (i.e., non-monotonic) form. In this case, the island-violating sentence is rated 0.19, which is within the upper half of the z-score scale. Similar to the complex NP island results with gaps, we are left without a firm conclusion about the presence or absence of the island effect. That said, if one were to decide to interpret this non-monotonic interaction as the presence of an island effect, it would fall within the definition of a subliminal island (Almeida 2014; Kush et al. 2018; Keshev and Meltzer-Asscher 2019), because the island-violating condition is in the upper half of the acceptability scale. We discuss subliminal islands in more detail in section 4.5. For wh-islands, despite a small amount of separation in the long conditions, we find no evidence of an interaction through the null hypothesis test, and find a Bayes factor of 0.3, which suggests that the data is more than 3 times more likely under the null hypothesis than the experimental hypothesis. We thus conclude that there is no island effect with wh-islands. Finally, for whether islands, we see a small visual pattern of an interaction, but the two statistical tests give contradictory results: the  $p$ -value is below the conventional significance level of .05, but the Bayes factor is 0.99, which suggests that the data is equally likely under both hypotheses. Furthermore, the island-violating condition has a mean of 0.52, which is substantially within the upper half of the acceptability scale. Thus we conclude either that there is no island effect with whether islands, or that there is a very small subliminal island; however, we likely need to run a replication study to confirm this subliminal island.

The final question we can ask about the island effects is whether we see evidence of the amelioration effect of resumption. For adjunct and whether islands, the answer is a qualified yes. We see clear evidence of island effects in the gap conditions, and we see trends toward no island effect in the resumption conditions, with the elimination of the effect primarily driven by an increase in the island/long (island-violating) condition. The two qualifications are (i) that our two statistical tests for each island are contradictory, suggesting that it might be prudent to replicate these findings in the future, and (ii) that the amelioration for whether islands may not be total, but rather it may yield a subliminal island effect in the resumption conditions. For complex NP islands, the answer is complicated. On the one hand, there is a clear (albeit non-monotonic) interaction for both the gap and resumption conditions. If we interpret this as an island effect, it would suggest no amelioration effect of resumption. On the other hand, the island-violating condition is rated near the middle of the acceptability scale for both gap conditions and resumption conditions, which may suggest that this is a subliminal island effect rather than a classic island effect. It is not clear whether we expect subliminal island effects to be ameliorated by resumption as they have not yet been widely explored in the literature, so it may not be fair to ask whether there is a resumption effect for complex NP islands. Finally, for wh-islands, we see no island effect with either gap conditions or resumption conditions.

### 3.4 The preference for gaps versus resumption in RC-dependencies

Beyond the potential amelioration effects, the mean ratings in Figure 5 also suggest that there is a general preference for resumption over gaps in RC-dependencies. The non-island/long conditions (declarative CPs) appear to be rated higher with resumption than gaps. We can see this more directly by again plotting the distributions of gap conditions and resumption conditions side-by-side. Figure 6 does this once again using violin plots overlaid with boxplots for both the non-island/long (declarative CP) conditions and island/long (island-violating) conditions.

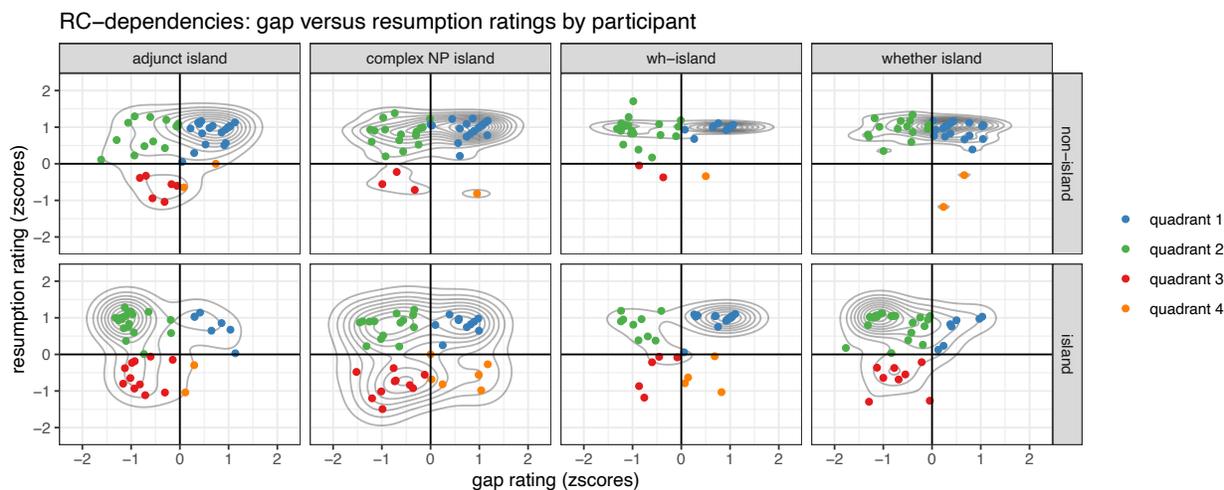
Figure 6: Violin plots of the distributions of judgments for the non-island/long and island/long conditions for RC-dependencies. Boxplots indicate medians and the interquartile range.



For the non-island/long (declarative CP) conditions (top row), we see that gap conditions lead to broad distributions that span both halves of the acceptability scale, leading to a median rating near the midpoint of the scale, while resumption conditions have most of their probability mass substantially within the upper half of the acceptability scale (near 1). This is indicative of a preference for resumption; and, interestingly, is the inverse of the pattern we saw with WH-dependencies. For the island/long (island-violating) conditions (bottom row), we see patterns that match the discussion in the previous subsection about island effects: the gap conditions for adjunct and whether islands show substantial probability mass in the lower half of the scale, while the resumption conditions show substantial probability mass in the upper half of the scale; the gap condition for complex NP islands is broadly distributed spanning both halves of the scale, perhaps with a small bias (a mode) toward the lower half of the scale, while the resumption condition shows two concentrations (two modes), a smaller one in the lower half of the scale and a larger one in the upper half of the scale; finally, both conditions for wh-islands show a concentration (a mode) in the upper half of the scale.

We can also view the preference for gaps versus resumption within each participant by plotting each participant's judgments of gap conditions against the judgments of resumption conditions in a scatterplot. We do this for RC-dependencies in Figure 7, just as we did for WH-dependencies in Figure 4.

Figure 7: Scatterplots of gap versus resumption versions of the non-island/long and island/long conditions for rc-dependencies. The lines represent two-dimensional (joint) probability density estimates. The points are colored according to the quadrants defined by the midpoint (0) of the z-score scales.



Turning first to the non-island/long condition (top row), which we note again are all the same structure (declarative CP) and therefore the columns each represent a between-participants replication, we see that the largest concentration of participants allow both gaps and resumption (quadrant 1), with a secondary concentration of participants who prefer resumption over gaps (quadrant 2). This presents a potentially interesting contrast with WH-dependencies, which showed participants either allowing both gaps and resumption or preferring gaps over resumption. We will discuss how these patterns fit within the typology of preferences for gap-versus-resumption in Arabic in section 4.4.

Turning next to the island/long conditions (bottom row), we see different patterns based on the island structure. For adjunct and whether islands, we see a large mass showing a resumption effect (quadrant 2), a smaller mass showing no resumption effect (quadrant 3), and a few participants accepting both, which indicates no island effect at all (quadrant 1). For wh-islands, we see a large concentration of participants who accept both gaps and resumption (indicating no island effect; quadrant 1). This confirms the overall conclusion that there is no island effect with wh-islands. There are also a small number of participants in each of the other three quadrants, suggesting either that those participants are outliers, or that there are smaller populations of speakers showing the other patterns. For complex NP islands, we see three roughly equal groups: those that show no amelioration (quadrant 3), those that show amelioration (quadrant 2), and those that show no island effect at all (quadrant 1). Much like the results for grand means in Figure 5, this further suggests that the status of complex NP islands may be complicated in Jordanian Arabic, perhaps involving three dialects.

#### 4. General discussion

We conducted four auditory judgment experiments to test resumption in four island types and two dependency types in spoken Jordanian Arabic – to our knowledge the broadest formal experimental test of island effects and resumption in a grammaticalized resumption language to

date. For WH-dependencies, we found (i) that in non-island structures, participants either prefer gaps to resumption or allow both (i.e., participants never prefer resumption to gaps), (ii) that all four island types lead to large island effects with both gaps and resumption, and (iii) that there is no amelioration of island-violating sentences with resumption – contrary to the general claim in the literature for grammaticalized resumption languages. For RC-dependencies, we found (i) that in non-island structures, participants either prefer resumption to gaps or allow both (i.e., participants never prefer gaps to resumption), (ii) that only whether and adjunct islands show clear island effects with gaps, and (iii) that resumption leads to full amelioration for adjunct islands and to a potentially new form of amelioration in whether islands that results in a subliminal island effect. These island effect results are summarized in Table 3. In this section, we discuss the consequences of these results for a number of theoretical issues related to resumption, movement, and island effects.

*Table 3: Summary of the results of the experiments.*

		adjunct	complex np	wh	whether
WH-dependencies	gap	island	island	island	island
	resumption	(island)	island	none	(island)
RC-dependencies	gap	island	island*	none	island
	resumption	(none)	island*	none	(subliminal)

( ) corroborated by only one of the two statistical tests

\* non-monotonic interaction

#### 4.1 The existence of island amelioration by grammaticalized resumption

The primary empirical question motivating this study was whether we would find unequivocal island amelioration effects with resumption in Jordanian Arabic. Our results suggest that we do see at least one canonical amelioration effect for adjunct islands with RC-dependencies, as well as a new type of amelioration effect involving subliminal island effects for whether islands with RC-dependencies. This is reassuring, as it corroborates the classic view that resumption can fully ameliorate (at least some) island effects in grammaticalized resumption languages. However, it is important to note that the amelioration that we found was far from global, as we found an amelioration effect of resumption for at most two out of seven island effects that we observed with gaps. Combined with the small increases in acceptability found in previous studies of grammaticalized resumption (i.e., Farby et al. 2010; Keshev and Meltzer-Asscher 2017; and Tucker et al. 2019), this suggests that our current understanding of the amelioration effect of grammaticalized resumption is likely incomplete. We need additional systematic studies across both languages and island types, as the pattern of amelioration may turn out to be rather complex. This raises the possibility that the field has yet to uncover a theory that would capture exactly which islands and dependencies will show amelioration by grammaticalized resumption.

#### 4.2 The independence of WH-dependencies and RC-dependencies

Perhaps the most salient finding in our study is that WH-dependencies and RC-dependencies yield distinct patterns of amelioration. On the one hand, this is perhaps not particularly

surprising, as syntactic theories of the different varieties of Arabic have often cataloged distinct preferences for either gaps or resumption across these two dependency types, even within a single language (we discuss this in more detail in section 4.4). But, on the other hand, it is a bit surprising to see WH-dependencies and RC-dependencies yielding distinct island effect patterns because they are often analyzed as two instances of the same dependency type – what are often called A'-dependencies. Though it is rarely formulated explicitly, there is an assumption running through much of the literature on A'-dependencies that all types of A'-dependencies are subject to island constraints. One strong form of this would be that all A'-dependencies in a single language would be subject to the same set of island effects. We could call this the *unified A' hypothesis*. Our results provide evidence against the unified A' hypothesis in two ways. The first is that there are distinct patterns of island effects with gaps between the two dependencies: there are wh-island effects for WH-dependencies but not RC-dependencies. There have been a handful of previous studies that have demonstrated distinct sets of island effects for distinct A'-dependencies within Brazilian Portuguese (Almeida 2014), English and Italian (Sprouse et al. 2016), and Norwegian (Kush et al. 2018; 2019). Our results suggest that we can add Jordanian Arabic to this list.

The second piece of evidence against the unified A' hypothesis in our results is that there is a distinct pattern of amelioration by resumption between the two dependencies: there are amelioration effects for RC-dependencies (with adjunct and perhaps whether islands) but not for WH-dependencies. We note that there is no simple explanation of the fact that WH-dependencies do not show amelioration effects in Jordanian Arabic. One tempting analysis would be to say that WH-dependencies in Jordanian Arabic do not allow grammaticalized resumption at all; therefore, the lack of an amelioration effect is simply the lack of grammaticalized resumption. This would make resumption with WH-dependencies *intrusive resumption*, which has not typically shown amelioration effects in formal experiments (see Table 1). However, about half of our participants do allow resumption in grammatical contexts with WH-dependencies. This suggests that grammaticalized resumption is possible with WH-dependencies for some speakers of Jordanian Arabic. These speakers did not show amelioration effects (see Figure 4). This suggests that the restriction is truly that there is no amelioration effect for islands with WH-dependencies. This in turn suggests that both island effects and amelioration by resumption are properties that can vary across A'-dependencies within a single language. Investigating (and potentially rejecting) the unified A' hypothesis is beyond the scope of this study; but we do note that it raises a number of interesting questions not only for syntactic theory, but also for theories of acquisition (as the child must track evidence for constraints on these dependencies separately) and theories of sentence processing (as the parser may apply distinct sets of constraints in real-time based on the dependency).

### 4.3 Theories of resumption and movement

The island amelioration effect has been leveraged as a critical piece of evidence for the theory of resumptive pronouns – specifically for the question of whether the dependency linking a resumptive pronoun to the head of the dependency is created by movement or by some other variable binding mechanism (see McCloskey 2017 for a recent review). One classic approach in this literature is to postulate that gap dependencies are created by movement and that resumption dependencies involve base generation of both the head of the dependency and the resumptive pronoun in their surface positions, with a non-movement variable binding mechanism linking the

two (e.g., Perlmutter 1972; Morgan 1972; Bresnan & Grimshaw 1978; McCloskey 1979). If island effects only arise for movement dependencies, and not for these other variable binding dependencies, then this would explain why there are no island effects with resumption. Our results with WH-dependencies suggest that resumption can also lead to island effects. Within this classic approach, this would either suggest that the variable binding mechanism underlying resumption can also lead to island effects, or that resumption is generated by movement (and therefore leads to island effects). In either case, our results raise a challenge for this classic non-movement approach to resumption.

Our results are not the first to raise challenges for the classic non-movement analysis of resumption. The first set of challenges came from Vata, Gbadi (Koopman 1984), and Swedish (Engdahl 1985), which are all languages that show island effects with resumption. This led Koopman (1984) and Engdahl (1985) to propose that for some languages, resumptive pronouns do participate in movement dependencies. However, in those three languages, all resumptive pronouns yielded island effects. This allowed for the possibility that languages simply differ in whether resumptive pronouns are generated by movement or not, with Vata, Gbadi, and Swedish instantiating the movement mechanism (and showing island effects), and the classic grammaticalized resumption languages like Irish, Hebrew, and Arabic instantiating the non-movement mechanism (and not showing island effects). Aoun et al. 2001 argued that this picture must be complicated a bit further, as there are languages, like Lebanese Arabic, that show evidence of two types of resumptive pronoun dependencies – one type that is created by movement and one type that is base generated. Aoun and colleagues used reconstruction as a diagnostic for movement – they found reconstruction effects for resumptive pronouns that occur in non-island structures (indicating movement), and no reconstruction effects for resumptive pronouns that occur within islands (indicating no movement). Therefore, for Aoun et al., there are two types of resumption, and languages can either instantiate all of one type, as in Vata, Gbadi, and Swedish, or mix the two types, as in Lebanese Arabic. Under this view, our results would suggest that the preference for the non-movement type of resumption within island structures that is observed in Lebanese Arabic is not universal. Jordanian Arabic appears to use the movement type of resumption within all four types of island structures with WH-dependencies, and possibly also within complex NP islands with RC-dependencies.

The fact that Jordanian Arabic appears to employ the movement type of resumptive pronoun in island contexts raises the question of whether all resumption may involve movement, thus allowing us to eliminate the non-movement type of resumption from the theory of syntax. Our results could be explained by a single movement mechanism that simply tracks gaps and resumption separately for the purposes of variation in island effects. *Prima facie*, that does not seem more complex than postulating two types of resumption, with the type varying across dependencies and island types. The most obvious empirical challenge for an all-movement approach would be the observation by Aoun et al. that there is no reconstruction inside of islands with resumption in Lebanese Arabic. If reconstruction is a diagnostic of movement, then this would be evidence that there are non-movement resumptive pronouns in Lebanese Arabic. However, the idea that reconstruction is a diagnostic for movement has been challenged by a number of studies (see Rouveret 2011 and McCloskey 2017 for reviews). Among these are Guillot and Malkawi 2006 and Malkawi and Guillot 2007, which observe that Jordanian Arabic shows the opposite pattern than the one reported by Aoun et al. for Lebanese Arabic: non-island declarative CP structures show no reconstruction, and adjunct islands show reconstruction. Guillot and Malkawi, therefore, argue that we need to relax the link between movement and

reconstruction (a conclusion echoed by Rouveret 2011 and McCloskey 2017). It is tantalizing that this challenge comes from Jordanian Arabic. It may be the case that Lebanese Arabic, like Vata, Gbadi, and Swedish, simply instantiates a relatively uniform version of resumption, while Jordanian Arabic does not. To our minds, this suggests a need for simultaneous systematic investigations of reconstruction and resumption across varieties of Arabic and other grammaticalized resumption languages.

#### 4.4 Cross-linguistic variation

Though neither were a specific focus of our study, there are two dimensions of cross-linguistic variation that our results may be relevant for: the preference for gaps or resumption, and the inventory of island effects. Turning first to the preference for gaps or resumption, Table 4 recreates a table presented in Choueiri 2017 reporting the preferences for three varieties of Arabic: Lebanese (Choueiri 2017), Egyptian (Wahba 1984; Brustad 2000; Aoun et al. 2010; and Soltan 2011), and Moroccan (Nouhi 1996). We have added our results as the final row, based solely on the ratings of gap and resumption conditions in non-island structures (declarative CP).

*Table 4: The preference for gaps or resumption in four varieties of Arabic, adapted from Choueiri 2017 and including the results presented here.*

	WH-dependencies	RC-dependencies
Egyptian Arabic	resumption	resumption
Lebanese Arabic	both	resumption
Moroccan Arabic	gap	both
Jordanian Arabic	gap or both	resumption or both

Our results differ from those previously reported in that we observe two groups of participants for each dependency: for WH-dependencies, there is a group that prefers only gaps and a group that allows both; for RC-dependencies, there is a group that prefers only resumption and a group that allows both. One question this raises is whether there are similar group splits in other varieties of Arabic, and what the space of possible variation is. Another question is whether the preference for gaps or resumption correlates with the presence or absence island amelioration effects. Again, these questions suggest a need for systematic studies with large sample sizes across varieties of Arabic.

Turning next to variation in the inventory of island effects, we observe one potentially novel pattern: the absence of wh-island effects for RC-dependencies (with either gaps or resumption) and the presence of whether island effects for RC-dependencies (with gaps; and potential amelioration with resumption). To our knowledge, this pattern has not been previously reported in the island effects literature. The fundamental conclusion is that the property of embedded polar questions that causes them to show island effects cannot be the same as the property of embedded wh-questions that causes them to show island effects. We believe this is potentially consequential for all extent approaches to island effects, including syntactic approaches like Subjacency (Chomsky 1973 et seq.), Barriers (Chomsky 1986; et seq.), Relativized Minimality (Rizzi 1990; et seq.), and phases (Uriagereka 1999; Chomsky 2001; et seq.); semantic approaches (Szabolcsi and Zwarts 1993; Abrusán 2011), pragmatic approaches (Erteschik-Shir 1973; Goldberg 2006; Ambridge et al. 2014), and even working memory

approaches (Kluender and Kutas 1993; Hofmeister and Sag 2010). We leave it to future research to explore the range of modifications to each of these theories that might capture this pattern.

#### 4.5 Subliminal islands

One final notable finding in these results is that whether islands with RC-dependencies and resumption appear to show a subliminal island effect: a superadditive interaction that appears in the upper half of the scale of acceptability. To our knowledge there have only been two clear cases of subliminal islands previously reported in the literature: Almeida 2014 observed one for whether islands with WH-dependencies in Brazilian Portuguese; and Keshev and Meltzer-Asscher 2019 observed one for wh-islands with WH-dependencies in Hebrew. (Kush et al. 2018 observe a pattern that looks similar for whether islands with WH-dependencies in Norwegian, but it appears to ultimately be driven by averaging together two groups of responses: typical island effects and no island effects). The primary question in the subliminal islands literature is what is driving these effects. One possibility is that they are driven by a grammatical constraint. If that is the case, it means either that there is some extra factor in these sentences that is causing the size of the island effect to be reduced (some sort of ameliorating extra-grammatical effect), or that the grammar itself is gradient, and the constraints responsible for subliminal islands have very small effects (e.g., Keller 2000; Featherston 2005; Smolensky and Legendre 2006). Another possibility is that subliminal islands are not driven by the grammar at all, but instead are entirely the consequence of extra grammatical factors. For example, Keshev and Meltzer-Asscher 2019 argue that the subliminal island effect that they observed in Hebrew could potentially be explained by something like the Kluender and Kutas 1993 working memory approach to island effects. Since our study was not designed to investigate the source of subliminal islands, we can make no claims about the source of the subliminal island effect that we observed in Jordanian Arabic. For now we can only note that Jordanian Arabic may provide a third test case for future studies designed to explore the source of subliminal island effects.

#### 5. Conclusion

In this study, we sought to determine whether Jordanian Arabic shows the classic island amelioration effect of grammaticalized resumption. To that end, we conducted a series of auditory judgment experiments to explore four island types and two dependency types with both gaps and a resumption using the factorial definition of island effects. Our results did reveal one unequivocal case of island amelioration by resumption, as well as one novel case in which the amelioration leads to a subliminal island. However, there were five other island effects (with gaps) in this study that did not show amelioration by resumption. This raises the possibility that the theory of island amelioration with grammaticalized resumption is more complex than previously thought (which in turn suggests a need for more systematic data collection in grammaticalized resumption languages). Our results also suggest a number of new avenues of research on topics related to resumption and movement. We observed distinct patterns of both island effects and amelioration effects for WH-dependencies and RC-dependencies, raising new challenges for the unified A' hypothesis. We also observed island effects with resumptive pronouns, suggesting either that resumption within some islands in Jordanian Arabic is formed by movement, or that the mechanism underlying resumption is also constrained by islands. To our knowledge, this is the second time Jordanian Arabic has raised issues for the theory of

movement and resumption, with the other being reconstruction facts, suggesting a need for joint studies of reconstruction and island effects in varieties of Arabic. We also observed novel patterns in the preference for gaps or resumption in the two dependencies tested here (i.e., some participants preferring a single option, and some allowing both) that also suggest a need for additional systematic studies of varieties of Arabic to determine the range of variation in dependencies. Finally, our results also yielded two novel observations that appear relevant for the general theory of island effects. The first is a novel pattern in the inventory of island effects: the presence of whether islands and the absence of wh-islands. Capturing this pattern will likely require modifications to all existing theories of island effects, and new studies specifically designed to test the predictions of those modifications. The second observation is a new subliminal island effect, which we believe is the third to be reported in the literature. This suggests that Jordanian Arabic could be used as a new test case for exploring the source of subliminal island effects. Taken together, the results of our exploratory study point the way to several new avenues of research involving resumption, movement, island effects, and cross-linguistic variation among varieties of Arabic and beyond.

## References

- Abrusán, Márta. 2011. Presuppositional and negative islands: a semantic account. *Natural Language Semantics* 19: 257–321.
- Alexopoulou, Theodora, and Frank Keller. 2007. Locality, cyclicity, and resumption: at the interface between grammar and the human sentence processor. *Language* 83: 110–160.
- Ackerman, Lauren, Michael Frazier, and Masaya Yoshida. 2018. Resumptive pronouns can ameliorate illicit island extractions. *Linguistic Inquiry* 49: 847–859.
- Almeida, Diogo. 2014. Subliminal wh-islands in Brazilian Portuguese and the consequences for syntactic theory. *Revista da ABRALIN* 13: 55–93.
- Ambridge, Ben, Julian M. Pine, and Elena V. M. Lieven. 2014. Child language acquisition: Why universal grammar doesn't help. *Language* 90: e53–e90.
- Aoun, Joseph, Lina Choueiri, and Norbert Hornstein. 2001. Resumption, movement, and derivational economy. *Linguistic Inquiry* 32: 371–403.
- Aoun, Joseph, Elabbas Benmamoun, and Lina Choueiri. 2010. *The syntax of Arabic*. Cambridge: Cambridge University Press.
- Bates, Douglas, Martin Maechler, Ben Bolker, Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67: 1–48.
- Beltrama, Andrea, and Ming Xiang. 2016. Unacceptable but comprehensible: the facilitation effect of resumptive pronouns. *Glossa* 1: 1.

Boersma, Paul. 2001. Praat, a system for doing phonetics by computer. *Glott International* 5: 341-345.

Bresnan, Joan, and Jane Grimshaw. 1978. The syntax of free relatives in English. *Linguistic Inquiry* 9: 331-391.

Brustad, Kristen. 2000. *The syntax of spoken Arabic*. Washington, DC: Georgetown University Press.

Chao, Wynn, and Peter Sells. 1983. On the interpretation of resumptive pronouns. In *Proceedings of North Eastern Linguistic Society 13*, ed. P. Sells and C. Jones, 47–61. Amherst: Graduate Linguistics Student Association,

Choueiri, Lina. 2017. Resumption in varieties of Arabic. In the *Routledge Handbook of Arabic Linguistics*, eds. Elabbas Benmamoun and Reem Bassiouney, 131-154. Abingdon: Routledge.

Christensen, Ken Ramshøj, Johannes Kizach, and Anne Mette Nyvad. 2013. Escape from the island: Grammaticality and (reduced) acceptability of wh-island violations in Danish. *Journal of Psycholinguistic Research* 42: 51-70.

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

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

Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: a life in language*, ed. Michael Kenstowicz, 1–50. Cambridge: MIT Press.

Drummond, Alex. 2019. *Ibex Farm*. Available at: <http://spellout.net/ibexfarm>

Engdahl, Elisabet. 1985. Parasitic gaps, resumptive pronouns, and subject extractions. *Linguistics* 23: 3-44.

Erteschik-Shir, Nomi. 1973. On the nature of island constraints. Ph.D. thesis, MIT.

Farby, Shira, Gabi Danon, Joel Walters, and Michal Ben-Shachar. 2010. The acceptability of resumptive pronouns in Hebrew. In the *Proceedings of the Israel Association for Theoretical Linguistics (IATL)* 26.

Featherston, Sam. 2005. The Decathlon Model of empirical syntax. In *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*, eds. M. Reis and S. Kepser, 187-208. Berlin: Mouton de Gruyter.

Ferreira, Fernanda, and Benjamin Swets. 2005. The production and comprehension of resumptive pronouns in relative clause "island" contexts. In *Twenty-first century*

*psycholinguistics: Four cornerstones*, ed. Anne Cutler, 263-278. Mahwah, New Jersey: Laurence Erlbaum Associates.

Goldberg, Adele. 2006. *Constructions at work*. London: Oxford University Press.

Guilliot, N. and Malkawi, N., 2006. When resumption determines reconstruction. In: D. Baumer, D. Montero, and M. Scanton, eds. *Proceedings of the 25th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Proceedings Project, 168–176.

Heestand, Dustin, Ming Xiang, and Maria Polinsky. 2011. Resumption still does not rescue islands. *Linguistic Inquiry* 42:138–152.

Hofmeister, Philip, and Ivan A. Sag. 2010. Cognitive constraints and island effects. *Language* 86: 366–415.

Jeffreys, Harold. 1961. *Theory of probability* (3rd ed.). Oxford: Oxford University Press.

Keffala, Bethany. 2013. Resumption and gaps in English relative clauses: Relative acceptability creates an illusion of ‘saving’. In *Proceedings of the 37th annual meeting of the Berkeley Linguistics Society*, eds. Chundra Catchart, I-Hsuan Chen, Greg Finley, Shinae Kang, Clare S. Sandy, and Elise Stickle, 140–154. Berkeley, CA: Berkeley Linguistics Society.

Keller, Frank. 2000. *Gradience in grammar: experimental and computational aspects of degrees of grammaticality*. Ph.D. thesis, University of Edinburgh.

Keshev, Maayan, and Aya Meltzer-Asscher. 2017. Active dependency formation in islands: How grammatical resumption affects sentence processing. *Language* 93: 549–68.

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

Kim, Boyoung, and Grant Grant. 2016. Islands and non-islands in native and heritage Korean. *Frontiers in psychology* 7: 134.

Kluender, Robert, and Marta Kutas. 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8: 573–633.

Ko, Heejeong, Han-Byul Chung, Kitaek Kim, and Jon Sprouse. 2019. An experimental study on scrambling out of islands: To the left and to the right. *Language & Information Society* 37.

Koopman, Hilda. 1984. *The syntax of verbs: From verb-movement rules in the Kru languages to Uni- versal Grammar*. Dordrecht: Foris Publications.

Kroch, Anthony. 1981 On the role of resumptive pronouns in amnestying island violations. In *CLS 17: Papers from the 17th Regional Meeting of the Chicago Linguistic Society*, 125-135. Chicago, Illinois: Chicago Linguistic Society.

Kush, Dave, Terje Lohndal, and Jon Sprouse. 2018. Investigating variation in island effects: A case study of Norwegian wh-extraction. *Natural Language and Linguistic Theory* 36: 743–79.

Kush, Dave, Terje Lohndal, and Jon Sprouse. 2019. On the Island Sensitivity of Topicalization in Norwegian: An experimental investigation. *Language* 95: 393-420.

Kuznetsova, Alexandra, Per B. Brockhoff, Rune H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82: 1–26.

Lu, Jiayi, Cynthia Thompson, and Masaya Yoshida. 2020. Chinese Wh-in-situ and Islands: A Formal Judgment Study. *Linguistic Inquiry* 51: 611-623.

Malkawi, Nouman, and Nicolas Guilliot, 2007. Reconstruction and islandhood in Jordanian Arabic. In *Perspectives on Arabic linguistics XX: Papers from the 20th Annual Symposium on Arabic linguistics*, ed. M. Mughazy, 87-104. Amsterdam: John Benjamins.

McCloskey, James. 1979. *Transformational syntax and model theoretic semantics: A case-study in Modern Irish*. Dordrecht: D. Reidel Publishing Company.

McCloskey, James. 2017. Resumption. In *The Blackwell companion to syntax*, eds. M. Everaert and H. van Riemsdijk, 84–117. Malden, MA: Blackwell.

Morey, Richard D., and Jeffrey N. Rouder. 2018. *BayesFactor: Computation of Bayes Factors for Common Designs*. <http://CRAN.R-project.org/package=BayesFactor>.

Morgan, Jerry L. 1972. Some aspects of relative clauses in English and Albanian. In *The Chicago Which Hunt: Papers from the Relative Clause Festival*, eds. Paul M. Peranteau, Judith N. Levi, and Gloria C. Phares, 63-72. Chicago, IL: Chicago Linguistic Society.

Morgan, Adam M. and Matthew W. Wagers. 2018. English resumptive pronouns are more common where gaps are less acceptable. *Linguistic Inquiry* 49: 861-876.

Nouhi, Youssef. 1996. Wh-constructions in Moroccan Arabic. MA Thesis, University of Ottawa.

Omaki, Akira, and Chizuru Nakao. 2010. Does English resumption really help to repair island violations? *Snippets* 21: 11–12.

Omaki, Akira, Shin Fukuda, Chizuro Nakao, and Maria Polinsky. 2019. Subextraction in Japanese and subject-object symmetry. *Natural Language and Linguistic Theory* 38: 627-669.

Perlmutter, David. 1972. Evidence for shadow pronouns in French relativization. In *The Chicago Which Hunt: Papers from the Relative Clause Festival*, eds. Paul M. Peranteau, Judith N. Levi, and Gloria C. Phares, 73-105. Chicago, IL: Chicago Linguistic Society.

Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge: MIT Press.

- Ross, J. R. 1967. Constraints on variables in syntax. Cambridge, MA: MIT dissertation.
- Rouveret, Alain. 2011. Some issues in the theory of resumption: A perspective on early and recent research. In *Resumptive pronouns at the interfaces*, vol. 5. *Language Faculty and Beyond: Internal and External Variation in Linguistics*, 1-62. Philadelphia: John Benjamins Publishing Company.
- Sells, Peter. 1984. Syntax and semantics of resumptive pronouns. Ph.D. thesis, University of Massachusetts, Amherst.
- Shlonsky, Ur. 1992. Resumptive pronouns as last resort. *Linguistic Inquiry* 23: 443–468.
- Smolensky, Paul, and Géraldine Legendre. 2006. *The harmonic mind: From neural computation to optimality-theoretic grammar (Cognitive architecture)*, Vol. 1. Cambridge: MIT press.
- Soltan, Usama. 2011. On strategies of question-formation and the grammatical status of the Q-particle huwwa in Egyptian Arabic Wh-questions. In *University of Pennsylvania Working Papers in Linguistics* 17: 24.
- Sprouse, Jon. 2007. A program for experimental syntax. Ph.D. thesis, University of Maryland.
- Sprouse, Jon, Shin Fukuda, Hajime Ono, and Robert Kluender. 2011. Reverse island effects and the backward search for a licenser in multiple wh-questions. *Syntax* 14: 179–203.
- Sprouse, Jon, Matthew Wagers, and Colin Phillips. 2012. A test of the relation between working memory capacity and syntactic island effects. *Language* 88: 82–123.
- Sprouse, Jon, Ivano Caponigro, Ciro Greco, and Carlo Cecchetto. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language and Linguistic Theory* 34: 307–44.
- Stepanov, Arthur, Manca Mušič, Penka Stateva. 2018. Two (non-) islands in Slovenian: A study in experimental syntax. *Linguistics* 56: 435-476.
- Szabolcsi, Anna, and Frans Zwarts. 1993. Weak islands and an algebraic semantics for scope taking. *Natural Language Semantics* 1: 235–284.
- Tucker, Matthew, Ali Idrissi, Jon Sprouse, and Diogo Almeida. 2019. Resumption ameliorates different islands differentially: Acceptability data from Modern Standard Arabic. *Perspectives on Arabic Linguistics XXX: Papers from the annual symposia on Arabic Linguistics*, eds. Amel Khalfaoui and Matthew A. Tucker, 159–93. Amsterdam: John Benjamins.
- Uriagereka, Juan. 1999. Multiple spell-out. In *Working minimalism*, eds. Samuel David Epstein and Norbert Hornstein, 251–282. Cambridge: MIT Press.

Wahba, Wafaa. 1984. *Wh*-constructions in Egyptian Arabic. Ph.D. Thesis, University of Illinois, Urbana-Champaign.

Zehr, Jérémy, and Florian Schwarz. 2018. PennController for Internet Based Experiments (PCIBEX). <https://doi.org/10.17605/OSF.IO/MD832>

## Appendix

Table A1: Results of 2x2 linear mixed effects models for DISTANCE x STRUCTURE for both dependency types and tail types (separately), using the lme4 and lmerTest packages, and treatment coding (with short and non-island as reference levels).

dependency	tail	model factor	adjunct		np		wh		whether	
			$\beta$	<i>p</i>	$\beta$	<i>p</i>	$\beta$	<i>p</i>	$\beta$	<i>p</i>
WH	gap	intercept	1.04	.001	1.04	.001	1.10	.001	1.01	.001
		distance	-0.48	.001	-0.50	.001	-0.72	.001	-0.30	.118
		structure	-0.21	.068	-0.79	.507	-0.17	.207	-0.18	.335
		dist x struct	-1.11	.001	-1.18	.001	-0.67	.001	-1.26	.001
	resumption	intercept	1.04	.001	1.04	.001	1.10	.001	1.01	.001
		distance	-1.11	.001	-1.06	.001	-1.60	.001	-1.17	.001
		structure	-0.21	.075	-0.08	.427	-0.17	.090	-0.18	.356
		dist x struct	-0.33	.046	-0.70	.001	-0.09	.521	-0.42	.155
RC	gap	intercept	0.68	.001	0.51	.001	0.87	.001	0.95	.001
		distance	-0.59	.015	-0.40	.020	-0.91	.001	-0.88	.001
		structure	0.17	.442	0.35	.039	-0.13	.376	-0.06	.699
		dist x struct	-0.88	.011	-0.72	.005	0.43	.044	-0.58	.012
	resumption	intercept	0.68	.001	0.51	.001	0.87	.001	0.95	.001
		distance	-0.06	.792	0.23	.075	0.01	.982	-0.08	.379
		structure	0.17	.478	0.35	.008	-0.13	.299	-0.06	.506
		dist x struct	-0.34	.320	-0.91	.001	-0.13	.459	-0.27	.046