Island effects and amelioration by resumption in Jordanian Arabic: an auditory acceptability judgment study

Rania Al-Aqarbeh
Mutah University
Department of English Language and Literature
Mutah, Karak, Jordan
rania11@mutah.edu.jo

Jon Sprouse
New York University Abu Dhabi
Program in Psychology
Abu Dhabi, UAE
jon.sprouse@nyu.edu

Abstract

Our broad goal in this study is to bring evidence from Jordanian Arabic, a primarily spoken grammatical resumption language, into the (formal experimental) empirical base of both theories of island effects and theories of island amelioration by resumption. To that end, we report four auditory judgment studies exploring two dependency types (wh-questions and relative clauses), four island types (adjunct, complex NP, wh, and whether), and both gaps and resumption, yielding a total of 16 distinct quantified effects. Our experiments identified two sources of variation that raise challenges for existing theories: variation across dependency types in the sets of islands present with gaps and variation across island types in the presence of amelioration by resumption. We discuss the challenges these results raise for four major classes of theories of island effects (phase-based, intervention-based, information-structure-based, and processing-complexity-based), and point to paths forward for each. We also discuss the consequences of the variation in amelioration for theories of the source of resumption, concluding that both base generation and movement must be available options to learners of JA. We also observe some evidence of individual variation in the availability of resumption across dependency types that could be explored in future studies of spoken varieties of grammatical resumption languages.

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

1. Introduction

In this paper we report the results of four auditory judgment experiments, testing a total of 165 native speakers, that are designed to explore the pattern of island effects in Jordanian Arabic (JA) across two dependency types (wh-dependencies and relative clause dependencies), four island types (adjunct, complex NP, wh, and whether islands), and two tail types (gaps and resumptive pronouns). The result is a total of 16 experimentally quantified island effects. Our broad goal is to bring evidence from a relatively understudied, primarily spoken variety of Arabic into the experimental syntax literature and into the (formal experimental) empirical base of theories of island effects. Within that broad goal, we have two narrow theoretical goals. The first is to explore to what extent the pattern of island effects (with gaps) that we observe in JA can be explained by the four dominant classes of theories of island effects in the literature: phase-based theories (e.g., Chomsky 2000, 2001, Rackowski & Richards 2005, Müller 2010), intervention-based theories (e.g., Rizzi 1990, 2004, Szabolcsi & Zwarts 1993, 1997), information-structure-based theories (e.g., Erteschik-Shir 1973, Goldberg 2006, Ambridge & Goldberg 2008, Abeillé et al. 2020), and processing-complexity-based theories (e.g., Deane 1991, Kluender & Kutas 1993, Hofmeister & Sag 2010, O'Grady 2010). Anticipating the discussion slightly, we observe a novel type of variation in the set of island effects between wh-dependencies and rc-dependencies that raises challenges for all four classes of theories of island effects, and discuss potential paths forward for each class. The second goal is to explore to what extent resumptive pronouns, which are a grammatical option in JA, alter the pattern of island effects, and what this might reveal about theories of island amelioration by resumption. Again, we observe a novel pattern of amelioration, suggesting that the mechanism underlying resumption may vary across island types – base generation for those that show amelioration, movement for those that do not (see, e.g., Salzmann 2017 and Choueiri 2017 for broader reviews; and Malkawi & Guilliot 2007 for evidence from reconstruction). Finally, though it was not one of our theoretical goals, our experiments suggest that there could be previously unnoticed patterns of variability with respect to the preference for gaps and resumption across wh-dependencies and rc-dependencies in JA. Taken as a whole, we believe that these experiments show that there is much to be learned about the cross-linguistic variation in island effects, amelioration by resumption, and possibly even grammatical resumption through experimental syntax studies of spoken varieties of Arabic.

We organize the paper as follows. In section 2, we present the theoretical and empirical background motivating the study, including background on JA, reviews of the four major theories of island effects, a summary of recent formal experimental work on island effects, and a review of two major questions in the amelioration by resumption literature. In section 3, we discuss the specific design of the experiments. In section 4, we describe the results of the experiments through three analyses: an analysis of trends in the full sample of participants, an analysis based on filtered subsets of participants according to their preferences for gaps and resumption in each dependency, and an exploratory analysis of individual variation in the preference for gaps or resumption in each dependency. In section 5, we discuss the consequences of the observed patterns for theories of island effects, the consequences for theories of amelioration by resumption, and the relevance of the four patterns of preferences for gaps and resumption for future studies. Section 6 concludes.

2. Background and motivation

In this section we provide a brief review of the theoretical and empirical background that motivates our study. This section also provides a starting point for the discussion of the results in section 5.

2.1 Jordanian Arabic

Jordanian Arabic is the most commonly spoken native language in Jordan. It is primarily a spoken language, with its written form mostly confined to informal contexts like text messages or emails. As is common in Arabic speaking countries, speakers of JA tend to use Modern Standard Arabic (MSA) for formal written contexts like school assignments or business communications. We decided to investigate JA because, to our knowledge, there have not yet been any systematic formal experimental studies of island effects and resumption in spoken varieties Arabic (but see Tucker et al. 2019 for a formal experimental study of island effects and resumption in MSA). We also know of no theoretical studies that directly investigate island effects in JA. The one previous study that discusses island effects in JA is Malkawi & Guilliot's (2007) investigation of reconstruction with resumption in JA. They report examples that contain adjunct island structures (using *because*-clause) and wh-island structures (using embedded *why*-questions), with left-dislocation/topicalization dependencies and resumptive pronouns. The sentences that they report are marked as fully acceptable; however, there is a clear implication in the text that these sentences would be unacceptable if the tails of the dependencies were gaps instead of resumptive pronouns. Malkawi & Guilliot 2007 can, therefore, be interpreted as indirectly proposing that there are both adjunct and wh-islands in JA, and that both are fully ameliorated by resumption. Our study is thus intended to systematically corroborate and extend these claims, and provide a first full integration of a spoken variety of Arabic into the formal experimental literature on island effects.

One empirical issue that arises when investigating languages that allow resumption as a grammatical option is that resumption is not necessarily equally possible in all dependency types (see Choueiri 2017 and Salzmann 2017 for reviews). Though there is an implication in Malkawi & Guilliot 2007 that resumption may not be compatible with wh-dependencies in JA (because they chose to use left-dislocation/topicalization dependencies for their examples), we nonetheless decided to treat this as an empirical question by testing both tail types for both dependency types. We include analyses to account for (section 4.2), and quantify (section 4.4), the preference for gaps and resumption across both dependency types in JA.

2.2 Four classes of theories of island effects

The first goal of this project is to determine the pattern of island effects in JA, and then explore the consequences of that pattern for the four dominant classes of theories: phase-based, intervention-based, information-structure-based, and processing-complexity-based theories. The four classes differ substantially in the source of island effects, which in turn impacts to what extent each theory can explain variation across dependency types and across island types. Here we briefly review the four classes of theories and the predictions that they make regarding patterns of variation.

2.2.1 Phase-based theories

Phase-based theories (Chomsky 2000, 2001, and elaborated by many others) postulate special syntactic domains, called phases, that limit the application of syntactic operations. The basic insight is that a syntactic operation can only target two items if they are within the same phase or if one is within a phase and the other is within the "edge" of the next more deeply embedded phase (where "edge" is typically defined as the specifier or the head of the phase). It is easy to see how phases can give rise to something like island effects – there simply needs to be a relationship (possibly identity) between island structures and phases, along with a reason why the phase edge is not available to the moving element. What makes phase-based theories more than just a terminological variant of the descriptive statement that island effects exist is that phases are thought to be both general and grounded. They are general in that they constrain all syntactic operations, not just movement. They are grounded in the sense that their existence (and their impenetrability to syntactic operations) is thought to derive from constraints on (syntactic) computational efficiency that limit syntactic operations to relatively local domains (see Citko 2014 for a review of phase theory in general, and Boeckx 2013 and Müller 2021 for a review of phase-based theories of island effects). In this way, phase-based theories represent a reinterpretation of a number of insights gained from earlier theories of constraints on syntactic operations, including Subjacency (Chomsky 1973), the Condition on Extraction Domains (Huang 1982), the Empty Category Principle (e.g., Lasnik and Saito 1984), Barriers (Chomsky 1986), and multiple spellout (Uriagereka 1999) based on the motivating assumptions of the Minimalist Program (Chomsky 1995).

Because phases are grounded in general principles of computational efficiency, the set of phases should be universal. That means that variation in island effects cannot be driven by variation in the set of phases, but rather by the availability of a(n intermediate) landing site for movement at the edge of the phase in question. For example, Rackowski & Richards (2005) propose tying the availability of extraction through the edge of the phase to agreement, such that phases that agree with a higher phase head allow extraction, while those that do not agree with a higher phase head do not. Taking a different tack, Müller 2010 proposes tying the availability of extraction to derivational timing. The edge features that license movement to the edge of a phase can only be added to a phase head while that head has other active syntactic features. In practice, this means that phases close to extraction when the final specifier of the phase is merged into the phase, thus eliminating the phase head's last active syntactic feature. In this way, last-merged specifiers all become islands to movement. Finally, taking a structural approach, Nyvad et al. 2017 propose trying the availability of extraction to the availability of multiple specifiers for certain phases. Phases with multiple specifiers can allow extraction even when one specifier is occupied by another lexical item. Though these approaches can explain variation across island types, to our knowledge, phase-based theories cannot easily accommodate variation in island effects across dependency types (wh-questions, relative clauses, etc). This is because the existence of a phase and the availability of its edge as a landing site are both typically determined independently of the dependency type. Capturing variation across dependencies would either require postulating different mechanisms underlying the dependencies, such as movement versus base-generation, or require the availability of edge features to somehow be determined by the dependency type.

2.2.2 Intervention-based theories

Intervention-based theories posit that certain island effects arise when a special lexical item *intervenes* between the head of the dependency and the tail of the dependency. The classic example would be a wh-island violation such as **What did you wonder who invented __?*, in which *who* intervenes between landing and launching sites of *what*, thus blocking the dependency. This description, of course, raises three theoretical questions: (i) Which items can be possible *interveners*?, (ii) How do we define the *intervenes* relationship?, and (iii) Why does the intervener block the dependency? There are two approaches to intervention in the literature currently: Relativized Minimality, which explores syntactic answers to these questions (Rizzi 1990, 2004; see Rizzi 2013 for a review), and Scope Theory, which explores semantic answers to these questions (Szabolcsi & Zwarts 1993, 1997; see Szabolcsi & Lohndal 2017). We focus exclusively on Relativized Minimality in this study because Scope Theory as currently configured is intended to explain the lack of island effects for complex argument wh-phrases (like *which inventor*) compared to the presence of island effects for adjunct wh-phrases (like *how quickly*). It does this by postulating that non-individual-denoting wh-phrases (like *how quickly*) are incompatible with the Boolean operations required by embedded questions, while individual-denoting wh-phrases (like *which inventor*) are compatible. It is not clear if Scope Theory is intended to extend to simple wh-words like those in this study (the JA equivalent of *what* and *who*), as these are individual-denoting but still show island effects.

Relativized Minimality (RM) defines interveners as items that carry the same syntactic feature(s) as the moved item. In the example above, both *what* and *who* might carry a +wh feature, therefore *who* is a potential intervener for the movement of *what*. RM defines intervention using c-command: an item intervenes in a dependency if the head of the dependency c-commands the item and the item c-commands the tail of the dependency. In the example above, *what* c-commands *who*, and *who* c-commands the launching site of *what* (the underscore), so *who* is an intervener. The underlying insight of RM is that intervention blocks the dependency because the dependency violates a preference for "minimal" dependencies – the dependency that could potentially hold between the intervener and the left peripheral position is shorter than the existing dependency, so the existing dependency is not minimal. RM is thus part of the much broader trend in generative syntax to explore metrics of computational efficiency in grammar.

RM captures variation in the pattern of island effects, both across languages and across dependencies within a single language, by defining classes of equivalent features, and allowing the classes of equivalent features to vary (e.g., Friedman et al. 2009, Belleti et al. 2012, Abels 2012). For example, embedded wh-questions could intervene in rc-dependencies, leading to wh-island effects, if the feature driving the rc-dependency is in the same equivalence class as the +wh feature in the embedded question. Similarly, rc-dependencies would be insensitive to wh-islands if the feature driving the rc-dependency is not in the same equivalence class as the +wh feature in the embedded question. Variation in island effects is thus driven by (and evidence for) variation in equivalence classes of features. However, one limitation of RM and other intervention-based theories is that they tend to only cover the set of island effects called *weak* or *selective* islands, which includes wh-islands, whether islands, and negative islands. These islands are called *weak* or *selective* because it has been claimed that they are only islands to certain dependencies, such as bare wh-words like *who*, but not to other dependencies, such as complex wh-phrases like *which inventor* (see Szabolcsi & Lohndal 2017 for a review). Despite their limited application, we continue to include intervention-based theories in our investigation

because we test both wh-islands and whether islands in this study, and crucially observe a theoretically relevant pattern of variation between them.

2.2.3 Information-structure-based theories

Information-structure-based theories of island effects begin with Erteschik-Shir's (1973) seminal dissertation. Erteschik-Shir proposed that the clauses or phrases that allow extraction are what she called semantically *dominant* – where dominance is defined as not presupposed to be true, and not referentially related to a previously uttered clause in the context. Conversely, the clauses or phrases that are islands are either presupposed to be true or are referentially related to a previous clause in the utterance. Erteschik-Shir 1973 presents a systematic investigation of a number of different clause types in both English and Danish (which show markedly different island patterns) to demonstrate that there is a strong correlation between the possibility of extraction and the results of several diagnostics for dominance. The functional insight of this approach is that the dominant clause contains the information under discussion, and therefore should be amenable to the semantic and pragmatic functions of various dependencies, such as question formation through wh-depedencies or modification through rc-dependencies. Non-dominant clauses are presupposed or pre-uttered, so it would be odd to apply these semantic or pragmatic functions to them. More recently, Erteschik-Shir's dominance approach has been recast in terms of focus and backgroundedness by Goldberg 2007, Ambridge & Goldberg 2008, and Ambridge et al. 2014, with island effects explained as a pragmatic clash that occurs when a focus operation (like wh-movement) targets an item that is within a backgrounded clause.

Within information-structure-based theories, variation in the pattern of island effects across languages reduces to a question of the variation in dominance (or backgroundedness). Erteschik-Shir argues that dominance is driven by multiple factors, including the meaning (and possibly complexity) of the verb that selects the clause, the placement and meaning of certain adverbs, and the location of intonational stress (which signals focus). Erteschik-Shir explores the variation between English and Danish along these lines (though she also notes complexities in the variation that may require additional mechanisms). Variation across dependencies reduces to a question of whether the dependency creates a pragmatic clash with the dominance (or backgroundedness) of the clause. Abeillé et al. 2020 leverage this approach to explain a difference in extractability between wh-dependencies and rc-dependencies for subject islands in French and English. One possible consequence of this approach is that we might expect that variation across dependencies within a single language will be all or nothing – either a dependency creates a clash, and therefore shows all of the islands that exist in that language, or it does not create a clash, and therefore shows no islands in the language. We might not expect to observe two dependencies yielding two distinct patterns of island effects within the same language. This is something we already observe in some languages (see section 2.3), so this is one aspect of information-structure-based theories that may require modification.

2.2.4 Processing-complexity-based theories

Processing-complexity-based theories posit that the unacceptability that we call an island effect is not caused by a grammatical violation (indeed, the sentences in question are posited to be fully grammatical), but rather by the dynamics of simultaneously processing a long-distance dependency and the structures that we call islands. There have been a number of specific

proposals along these lines (e.g., Wannter & Maratsos 1978, Deane 1991, Hawkins 1999, Hofmeister & Sag 2010, O'Grady 2010, a.o), but perhaps the proposal with the most well-worked out mechanisms and broadest coverage of island effects is the working-memory-capacity theory developed in Kluender 1991, Kluender & Kutas 1993, Kluender 1998, Kluender 2004, and Kluender & Gieselman 2013. Kluender's working memory capacity theory proposes that both the processing of long-distance dependencies and the processing of the syntactic structures that we call islands draw on the same pool of working memory resources. When deployed simultaneously, these two sets of processes require more resources than are available in the pool of working memory, creating a processing failure that speakers perceive as unacceptability.

One parsimonious property of processing-complexity-based theories is that variation in island effects, both across languages and across dependencies within a single language, can be explained through variation in the processing dynamics of the island structures or dependencies in question (such as requiring fewer or greater working memory resources). The empirical challenge for processing-complexity-based theories of variation is to independently identify those processing dependents (outside of island effects), such that the patterns of variation that we observe with island effects can be independently predicted from those components. To our knowledge, there is relatively little work on processing-complexity-based theories of cross-linguistic variation in island effects (but see Christensen et al. 2013 for one prominent example of a processing-complexity-based theory of island effects in Danish).

2.3 Previous formal experimental studies of island effects

Table 1 below summarizes the results of 15 relatively recent formal experimental investigations of island effects (with gaps in the tails of the dependencies), spanning 11 languages, that either directly or indirectly employ the same factorial design that we use in our study. Though we intend this list to be fairly comprehensive, it may not be exhaustive, as the literature is growing quickly. We present this table as an empirical counterpart to section 2.2 to help motivate the scope of our four experiments.

Table 1: 15 formal experimental studies of island effects (with gaps in the tails of the dependencies) using a factorial design. A plus (+) indicates an island effect, a dash (–) indicates no island effect, and a gray box indicates that the island type was not tested in the study. Parentheses around the plus indicates a subliminal island effect – an effect with the island-violating condition at or above the midpoint of the scale.

study	language	dependency	adj	np	sub	rc	wh	whether
Almeida 2014	Br. Portuguese	wh-simple						(+)
Almeida 2014	Br. Portuguese	topicalization						–
Lu et al. 2019	Chinese	wh-arg-in-situ				+		
Lu et al. 2019	Chinese	wh-adj-in-situ				+		
Christensen et al. 2013	Danish	wh-simple					+	
Poulsen 2008	Danish	topicalization	+					
Sprouse et al. 2016	English	wh-simple	+	+	+			+
Sprouse et al. 2016	English	wh-complex	+	+	+			+
Sprouse et al. 2016	English	rel. clause	–	+	+			+
Sprouse et al. 2011	English	wh-arg-in-situ	–	–	–			–
Sprouse et al. 2016	Italian	wh-simple	+	+	+			+

Sprouse et al. 2016	Italian	rel. clause	+	+	–			+
Omaki et al. 2019	Japanese	np scrambling			–			
Sprouse et al. 2011	Japanese	wh-arg-in-situ	–	–	–			–
Kim & Goodall 2016	Korean	wh-arg-in-situ	–					+
Kim & Goodall 2016	Korean	wh-arg-in-situ	–					+
Ko et al. 2019	Korean	np scrambling	–			–	–	
Tucker et al. 2019	MS Arabic	wh-complex	+	+				+
Kush et al. 2018	Norwegian	wh-simple	+	+	+	+		+
Kush et al. 2018	Norwegian	complex wh	+	+	+	+		+
Kush et al. 2019	Norwegian	topicalization	+	+	+	+		–
Stepanov et al. 2018	Slovenian	wh-simple			+		–	
Pañeda et al. 2020	Spanish	wh-simple	+	–	+			+
Pañeda & Kush 2021	Spanish	wh-complex	+			+		(+)
Pañeda & Kush 2021	Spanish	wh-complex	+			+	+	

There are three trends in Table 1 that help to guide our study. First, to our knowledge, there have been no formal experimental studies of a spoken variety of Arabic yet (only Modern Standard Arabic). It is our hope that JA is the first of many to be added to this literature. Second, we can see in this table that there is variation across dependencies within individual languages, including Brazilian Portuguese, English, Italian, Norwegian, and Spanish. This underscores that variation across dependencies is an actively growing component of the empirical base for the theory of island effects. Here we investigate both wh-depedencies (wh-simple in the table) and rc-depedencies. Third, we see that some of the most informative studies test four or more island types. We selected two strong islands (adjunct and complex NP) and two weak islands (wh-islands and whether) for our study. All four have shown variation in at least one previous study. We did not select subject islands because extraction from the subject position in many varieties of Arabic can trigger a cleft-like structure when the extracted item is non-animate, which could introduce additional differences between subject islands and other island types. We did not select rc-islands because they have not shown variation in previous studies.

2.4 Theories of island amelioration by resumption

The second goal of our study is to explore to what extent resumptive pronouns alter the pattern of island effects in JA, and what this might reveal about theories of island amelioration by resumption (Ross 1967; Kroch 1981; Chao & Sells 1983; Sells 1984; Engdahl 1985; et seq.). We see this as a secondary goal because a full investigation of theories of resumption, which are constructed to explain a wide variety of phenomena beyond island effects (including reconstruction, strong crossover, and various reflexes of successive cyclic movement), is far beyond the scope of our study of island effects (see Salzmann 2017 for a comprehensive review of the complex empirical landscape of resumption). For this study, we will focus on two questions about amelioration by resumption. The first is empirical: what is the pattern of amelioration in JA?. The second is theoretical: what does the pattern suggest about the grammatical mechanism(s) underlying resumption in JA? (For broader discussions of questions about amelioration by resumption see Rouveret 2011, McCloskey 2017, and Salzmann 2017)

To ground our first question – what is the pattern of amelioration in JA? – it may help to review the broader typological classification of languages in the resumption literature. At the

highest level, languages are often divided into two types: *intrusive resumption languages* (Chao & Sells 1983 and Sells 1984), which do not allow resumption as a grammatical option in non-island structures, but may allow resumption within island structures as an exceptional device to ameliorate the island effect; and *grammatical resumption languages*, which allow resumption as a grammatical option in non-island structures, and may also allow resumption as a grammatical option within island structures to ameliorate the island violation. Because JA is a grammatical resumption language, we will focus exclusively on grammatical resumption in this project. That said, there is a large and growing literature using formal experiments to explore island amelioration in intrusive resumption languages that we would like to cite here for readers interested in exploring that aspect of the literature (see Dickey 1996, McDaniel & Cowart 1999, Frazier & Clifton 2002, Ferreira & Swets 2005, Alexopoulou & Keller 2007; Omaki & Nakao 2010; Heestand et al. 2011; Keffala 2013; Beltrama and Xiang 2016; Ackerman et al. 2018; Morgan and Wagers 2018, Chacón 2019). Grammatical resumption languages are often further divided into three subtypes: those that show amelioration with all island types (full set amelioration), those that amelioration only with weak islands (e.g., wh, whether; weak set amelioration), and those that show no amelioration (no amelioration; see McCloskey 2017 and Salzmann 2017 for recent reviews, and see Szalbolcsi & Lohdal 2017 for a review of the strong/weak distinction more generally). Within the theoretical literature, JA is reported to be a full set amelioration language based on the examples from Malkawi & Guilliot (2007) showing adjunct islands (a strong island) and wh-islands (a weak island) as fully acceptable with resumptive pronouns. We will investigate this claim directly by testing two strong islands (adjunct and complex NP) and two weak islands (wh and whether islands). Though this is the first formal experimental investigation of amelioration by resumption in a spoken variety of Arabic (to the best of our knowledge), readers interested in formal experimental work in Semitic languages should see Farby et al. 2010 and Keshev & Meltzer-Asscher 2017 for formal experiments on resumption in Hebrew, and Tucker et al. 2019 for formal experiments on resumption in MSA.

For our second question – what are the grammatical mechanisms underlying resumption in JA? – there are three classes of theories: base generation, movement, and mixed theories (see Salzmann 2017 and McCloskey 2017 for reviews; see also Asudeh 2012 for a non-transformational approach developed within Lexical-Functional Grammar that nonetheless divides resumption into two types). Under base generation theories, the head of the dependency and the resumptive pronoun are both generated in their surface positions (no movement), and linked through a semantic binding mechanism (Ross 1967, Morgan 1972, Perlmutter 1972, Givon 1973, Hayon 1973, Chomsky 1977, Bresnan & Grimshaw 1978, McCloskey 1979, 1990, Borer 1984, et seq.). Base generation theories provide a straightforward explanation for languages that do not show any island effects with resumption (full amelioration) because binding is generally assumed to be insensitive to islands, but require additional assumptions to explain languages that show island effects with resumption (weak set and no amelioration). Under movement theories, the link between the head of the dependency and the resumptive pronoun is established through syntactic movement, with the details varying by implementation: there are approaches that propose that the resumptive pronoun itself moves (possibly covertly, e.g., Demirdache 1991); there are approaches that propose that the head of the dependency moves and the copy or trace left behind is spelled out as a pronoun (e.g., Zaenen et al. 1981, Pesetsky 1998, Hornstein 2001, Bianchi 2004, Müller 2014, Sichel 2014, Hladnik 2015); and there are approaches that propose that the head and resumptive pronoun are merged together,

with the head leaving the pronoun behind analogous to preposition stranding (e.g., Aoun et al. 2001, de Vries 2002, Boeckx 2003, Belletti 2006, Boeckx & Hornstein 2008, Donati & Cecchetto 2011, Chidambaram 2013, Klein 2014, 2016). Movement theories provide a straightforward explanation for languages that show island effects with resumption (no amelioration) because movement is assumed to be island sensitive. But movement theories require additional assumptions to explain languages that show no island effects (full amelioration) or a reduced set of island effects (e.g., weak set amelioration). Mixed theories allow for resumption to be due to either movement or base generation plus binding (e.g., Demirdache 1991, Pesetsky 1998, Boeckx 2003, Müller 2014, and Klein 2016). Mixed theories do not require the postulation of an additional mechanism in the grammar, because it is assumed that the grammar requires both movement and binding independently; but mixed theories do increase the complexity of the acquisition problem, as learners must track resumption separately in each relevant syntactic context, and must encounter evidence that indicates which mechanism underlies each separately tracked context. As Salzmann 2017 lays out in his review, adjudicating among the three types of theories requires investigating a wide range of phenomena that go beyond the scope of our study, including strong crossover, parasitic gaps, and reflexes of successive cyclic movement. Therefore, we will not attempt to choose among the three types of theories for JA here, but rather discuss the consequences of the pattern of results that we observe for each of the three types of theories.

3. The experiments

The two goals of our study are (i) to determine the pattern of island effects in JA and its consequences for the four dominant classes of theories of island effects, and (ii) to determine the pattern of amelioration by resumption in JA and its consequences for theories of amelioration. The theoretical and empirical review in section 2 suggests that the most informative study would include multiple dependency types and multiple island types. Therefore, we decided to test both wh-dependencies and rc-dependencies, and four island types – two strong islands (adjunct and complex NP islands) and two weak islands (wh and whether islands). We also selected the specific adjunct and wh-islands to connect directly with the previous work in Malkawi & Guilliot 2007 (*because* clauses and *why* clauses, respectively).

For ease of exposition, here we provide a single paragraph, high-level summary of the design of the experiments, followed by a detailed description of each component of the experiments in the following subsections (leading to some repetitions). The two dependencies and four island types yield eight specific island effects. We created four experiments, with each experiment testing two island effects distinct in both island type and dependency. We recruited 165 self-reported native speakers of JA from Mutah University in the Karak region of Jordan, split roughly evenly among the four experiments. The participants completed the experiment during a visit to the first author's lab, and received course credit for their participation. Table 2 lists the island effects for each of the four experiments along with the number of participants recruited for that experiment.

Table 2: The distribution of dependency types and island types across the four experiments

	dependency	island	participants
experiment 1	wh-dependency rc-dependency	wh complex NP	40
experiment 2	wh-dependency rc-dependency	complex NP wh	42
experiment 3	wh-dependency rc-dependency	whether adjunct	43
experiment 4	wh-dependency rc-dependency	adjunct whether	40

The experiments were auditory judgment experiments using a 7-point rating scale (from 1 to 7), each 42 items long: 6 practice items at the beginning of the survey (but not marked as such), followed by 12 experimental items and 24 filler items (8 each of low, medium, and high acceptability) in a pseudorandomized order. We used a factorial design to quantify the island effects (Sprouse 2007, Sprouse et al. 2011, Sprouse et al, 2012) and the amelioration by resumption effects (Tucker et al. 2019). This design leads to six conditions per island type. We collected one judgment per condition per participant, yielding 40, 42, 43, and 40 judgments per condition depending on the experiment. These sample sizes have been shown to yield over 80% statistical power (a common target power level in experimental psychology) for medium effect sizes and 100% statistical power for large effect sizes (Sprouse & Almeida 2017, Marty et al. 2020). We use both linear mixed effects models and Bayes factors to analyze the results.

3.1 The factorial definition of island effects

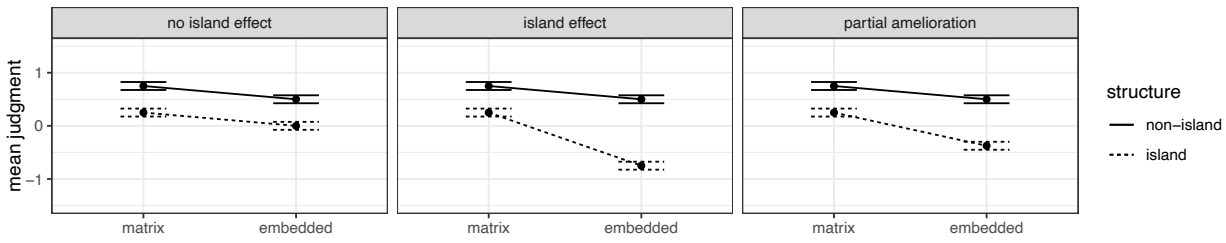
We chose the factorial definition because we believe it matches the logic that has historically been used by syntacticians to define island effects, albeit translated into the factorial terminology that is typical of formal experiments. The standard version of the factorial design for island effects 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 explore 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 JA. The result is therefore six conditions in a 2x2+2 design (see also Tucker et al. 2019). We illustrate the full design with whether islands and wh-dependencies in JA in (1):

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

		DEPEN.	STRUCT.	TAIL
a.	mi:n __ ga:l-Ø innu is-si:nama ʕaraðˁ-at filim hindi:? who __ said-3SGM that the-cinema presented-3SGF film Hindi 'Who __ said that the cinema presented a Hindi film?'	matrix	non-island	gap
b.	weiʃ ga:l-Ø ra:mi: innu is-si:nama ʕaraðˁ-at __? what said-3SGM Rami that the-cinema presented-3SGF __ 'What did Rami say that the cinema presented __?'	embedded	non-island	gap
c.	mi:n __ saʔal-Ø iða is-si:nama ʕaraðˁ-at filim hindi:? who __ asked-3SGM whether the-cinema presented-3SGF film Hindi 'Who __ asked whether the cinema presented a Hindi film?'	matrix	island	gap
d.	weiʃ saʕal-Ø ra:mi: iða is-si:nama ʕaraðˁ-at __? what asked-3SGM Rami whether the-cinema presented-3SGF __ 'What did Rami ask whether the cinema presented __?'	embedded	island	gap
e.	weiʃ ga:l-Ø ra:mi: innu is-si:nama ʕaraðˁ-at-***uh***? what said.3SGM Rami that the-cinema presented-3SGF-***it*** 'What did Rami say that the cinema presented ***it***?'	embedded	non-island	resump.
f.	weiʃ saʔal-Ø ra:mi: iða is-si:nama ʕaraðˁ-at-***uh***? what asked.3SGM Rami whether the-cinema presented-3SGF-***it*** 'What did Rami ask whether the cinema presented ***it***?'	embedded	island	resump.

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 as in the left panel, or partial amelioration, which will result in a smaller interaction, as illustrated in the right panel.

Figure 1: Possible outcomes for the factorial design.



For space reasons, in (2) we list the island-violating conditions for each of the eight combinations of islands and dependency types. The full list of materials is available in the supporting materials.

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

a. weiʃ saʔal-Ø ra:mi: iða is-si:nama ʕaraðˁ-at __? whether
what asked-3SGM Rami whether the-cinema presented-3SGF __
'What did Rami ask whether the cinema presented __?'

b. weiʃ saʔal-Ø baha:ʔ leiʃ ama:ni: ʃatˁab-at __? wh
what asked-3SGM Bahaa why Amani crashed-3SGF __
'What did Baha' ask why Amani crashed __?'

c. weiʃ naʃar-Ø sa:mi: il-iʃaʕah innu intisˁa:r iʃtar-at __? np
what spread-3SGM Sami the-rumor that Intisar bought-3SGF __
'What did Sami spread the rumor that Intisar bought __?'

d. weiʃ ziʕil-Ø na:sˁir laʔinnu il-maktabih tˁabaʕ-at __? adjunct
what got-angry-3SGM Naser because the-press printed-3SGF __
'What did Naser get angry because the press printed __?'

---

e. baʕrif il-mudi:r illi is-sikiriteirah saʔal-at iða il-idarah ixta:r-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-asatðeh illi sˁa:ħab-ak saʔal-Ø leiʃ itˁ-tˁulla:b waddaʕ-u __. wh
saw.1SG the-teachers who friend-your asked.3SGM why the-students said-goodbye.to-3PLM _
'I saw the teachers who your friend asked why the students said goodbye to __.'

g. baʕrif is-sikiriteirah illi il-katibih simiʕ-it il-iʃa:ʕah innu il-mudi:r idʒawwaz-Ø__ np
know.1SG the-secretary who the-clerk heard-3SGF the-rumor that the-principal married-3SGM __
'I know the secretary who the clerk heard the rumor that the principal married __.'

h. baʕrif il-mudi:r illi il-binit imbasatˁ-at laʔinnu dʒa:r-na ʕazam-Ø __. 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 __.'

There are two additional properties of our design to note. The first is that we attempted to select embedded verbs that are obligatorily transitive in JA in order to ensure that the gap conditions received the intended interpretation. The second is that the resumptive pronouns in our resumption conditions are clitics. Malkawi & Guilliot 2007 classify clitics as *weak resumptives*, which contrast with *strong resumptives* such as non-clitic pronouns or epithets. Both types of resumptives lead to amelioration of island effects – in fact, the examples that Malkawi & Guilliot 2007 provide for amelioration by resumption in adjunct and wh-islands both involve weak resumptives. Weak and strong resumptives differ in two ways: only weak resumptives are possible in non-subject position (Aoun et al. 2010), and only weak resumptives allow

reconstruction within an island (Malkawi & Guilliot 2007). The strong/weak resumptive distinction will therefore likely matter for a complete theory of resumption. But for the narrow goals of this project, we believe either type is appropriate. We chose weak resumptives (in non-subject positions) to better match the existing literature (e.g., Malkawi & Guilliot 2007).

3.2 The choice of auditory modality

The linguistic context in Jordan is diglossic. JA is the most common native language, and the language of daily interactions. It is primarily a spoken language, as discussed in section 2.1. Modern Standard Arabic (MSA) is the language of formal contexts, including government business (it is the official language of the country), education, and the vast majority of print and broadcast media. MSA is rarely, if ever, spoken as a native language, but it is taught extensively in the Jordanian educational system because of its role in formal contexts. Our focus in this study is on JA exclusively. To that end, we 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 JA. To aid in this, the first author spoke to the participants in JA during the laboratory visit. Furthermore, the instructions for the experiment directed the participants to imagine conversing with a friend in JA, and to judge whether a native speaker of JA could produce these sentences. Though nothing can ensure that participants do not allow their judgments about MSA to influence their judgments of JA, we do believe that the auditory presentation was successful in eliciting judgments that are distinct from MSA, because the pattern of results that we observed differ from the results of study of MSA in Tucker et al. 2019.

3.3 Materials and survey construction

Each participant completed a survey that consisted of 42 items: 6 items at the beginning of the survey to help familiarize participants with the task (but not marked as distinct from the main task), followed by 12 experimental items and 24 filler items pseudorandomized to avoid related experimental items appearing in succession. The 12 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 to achieve a filler-to-target ratio of 2:1 while still maintaining a reasonable experiment length given the increased cognitive burden and time requirements of the auditory judgment task. 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 such 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, attempting to produce all items with a natural intonational contour for information-seeking questions in JA. 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. Though we believe it is unlikely that a prosodic artifact could be responsible for the complex pattern of results that we observe here (varying by dependency, island type, structure, and tail type), in the interest of full transparency (and reproducibility), the full set of recorded materials are available on the authors' websites for any readers who wish to explore the prosodic properties of the materials as a possible explanation of the results.

3.4 Participants

All participants were students at Mutah University in Jordan. They were all self-reported native speakers of JA (from the Karak region). They received course credit for their participation. For the four experiments, we recruited 40, 42, 43, and 40 participants, respectively. Participants completed the experiment during a visit to the first author's laboratory.

3.5 Presentation

Participants were instructed (in spoken JA) to rate the acceptability of the sentences in the experiment. We defined acceptability in terms of possibility: we instructed participants to imagine conversing with a friend in JA, and to judge whether a native speaker of JA could produce these sentences. The auditory experiments were implemented using PennController for IBEX (Drummond 2019; Zehr & Schwarz 2018). Each sentence received its own presentation screen with a 1 غير مقبول اطلاقًا (totally unacceptable) to 7 مقبول تمامًا (perfectly acceptable) scale. Each 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.

3.6 Statistical analyses

We chose our target sample sizes to be around 40 participants per experiment based on the empirical estimates of statistical power for 7-point acceptability judgment tasks in Sprouse & Almeida 2017 and Marty et al. 2020. Those studies found that sample sizes of 40-43 (with one judgment per participant) yield nearly 100% statistical power for large effect sizes, which is the typical size of island effects in the experimental literature (see the studies in Table 1). These samples sizes also yield over 80% statistical power for medium effect sizes. Given that 80% power is the target level for best practices in experimental psychology (Cohen 1988/2013), we concluded that even if the island effects in JA are smaller than island effects in other languages, our study would still be well-powered to detect them.

We ran two sets of statistical tests as our primary analyses for the presence or absence of island effects. 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 (Kuznetsova et al. 2017), which uses the Satterthwaite approximation for degrees of freedom to derive an F test from the linear mixed effects model. The full set of statistical results for the mixed effects models are in the appendix. For ease of exposition, we have added the interaction term *p*-value to each cell of the interaction plots in section 4.

In the second set of analyses, we derived Bayes factors using the BayesFactor package (Morey & Rouder 2018) for the interaction of DEPENDENCY and STRUCTURE (the island effect) by comparing a fixed effects model with the interaction term to a fixed effects model without the interaction term. 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 (H1) to the likelihood of the data under the null hypothesis (H0). 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. For example, a BF_{10} of .33 would indicate that the data is 3x more likely under the null hypothesis than the experimental hypothesis. This helps to distinguish between null results that are evidence for the absence of an island effect and null results that are inclusive.

For the primary question of the presence versus absence of island effects, we will therefore look for three patterns in the statistical results: a *p*-value less than the conventional threshold of .05 and a BF_{10} greater than the conventional threshold of 3 (Jeffries 1961), which is indicative of an island effect; a *p*-value greater than .05 and a BF_{10} less than 0.33, which is indicative of no island effect; 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. This last possibility would suggest that there is no classic (medium or large) island effect because our experiments have high statistical power to detect medium and large effects. But it would be unable to distinguish between a true null effect and a very small effect (for which the experiments do not have high statistical power).

As an anonymous reviewer notes, one potential drawback of Bayes factors is that their magnitude can be dependent on the choice of priors (and in particular the width of the prior probability distribution). The BayesFactor package implements uninformative priors within the "objective" or "default" framework that has been explored by Rouder, Morey, and colleagues over the past several years (the specific priors for the regression models implemented here are based on the work of Liang et al. 2008 as implemented by Rouder and Morey 2012). To explore the stability of the Bayes factors for our results, in the appendix we report Bayes factors for the three preset prior distribution widths made available by the BayesFactor package (medium, wide, and ultrawide). Because the Bayes factors that we observe are remarkably stable, we report only one value (the medium width) in the plots in the main text to avoid visual clutter.

The exploration of some of the patterns in our results also required two ancillary statistical tests. The first is an analysis to show that the elimination of the interaction effect that is indicative of amelioration by resumption is driven by an increase in acceptability in the island violating condition (island/long) with resumption, and not by changes in the other three conditions. We test this using a pairwise comparison between the island violating condition (island/long) for the gap and resumption tail types. The second is an analysis to show that the interaction effect is smaller with resumption than with gaps in situations where amelioration due to resumption is partial. This would typically be achieved by looking at the three-way interaction in our 2x2x2 design (DEPENDENCY x STRUCTURE x TAIL); however, given that the short conditions are identical for both tail types, this more properly reduces to a 2x2 interaction of STRUCTURE x TAIL in the long conditions. We report both of these analyses in full in the appendix, and report the results as needed in line in the text.

For readers interested in other statistical analyses (including Bayesian analyses with priors that go beyond the three presets of the BayesFactor package), the raw data is available for download and re-analysis on the authors' websites.

4. The results of the experiments

In this section we describe the results of the experiments. Section 4.1 presents our primary analysis of the data set. Section 4.2 presents an additional analysis to evaluate the reliability of these results, taking into consideration participants' preferences for gaps or resumption. The

additional analysis reveals the same general pattern of results as the primary analysis, suggesting that the results are reliable. Section 4.3 presents an exploratory analysis of individual variation in the preferences for gaps or resumption.

We z-score transformed the results for each participant prior to the analyses presented in this section to reduce the impact of common forms of scale bias. Though we believe this is the most appropriate way to analyze judgment results (e.g., Schütze & Sprouse 2014), we also note that there are no differences in the pattern of results using the raw judgments directly. In fact, the mid-point (0) of the z-score scale corresponds nearly perfectly to the absolute mid-point of the raw judgment scale (4), which suggests that the filler items succeeded in minimizing scale bias directly. We will therefore report the z-score transformed results here. The raw data is available on the authors' websites for readers to reproduce these analyses or explore others.

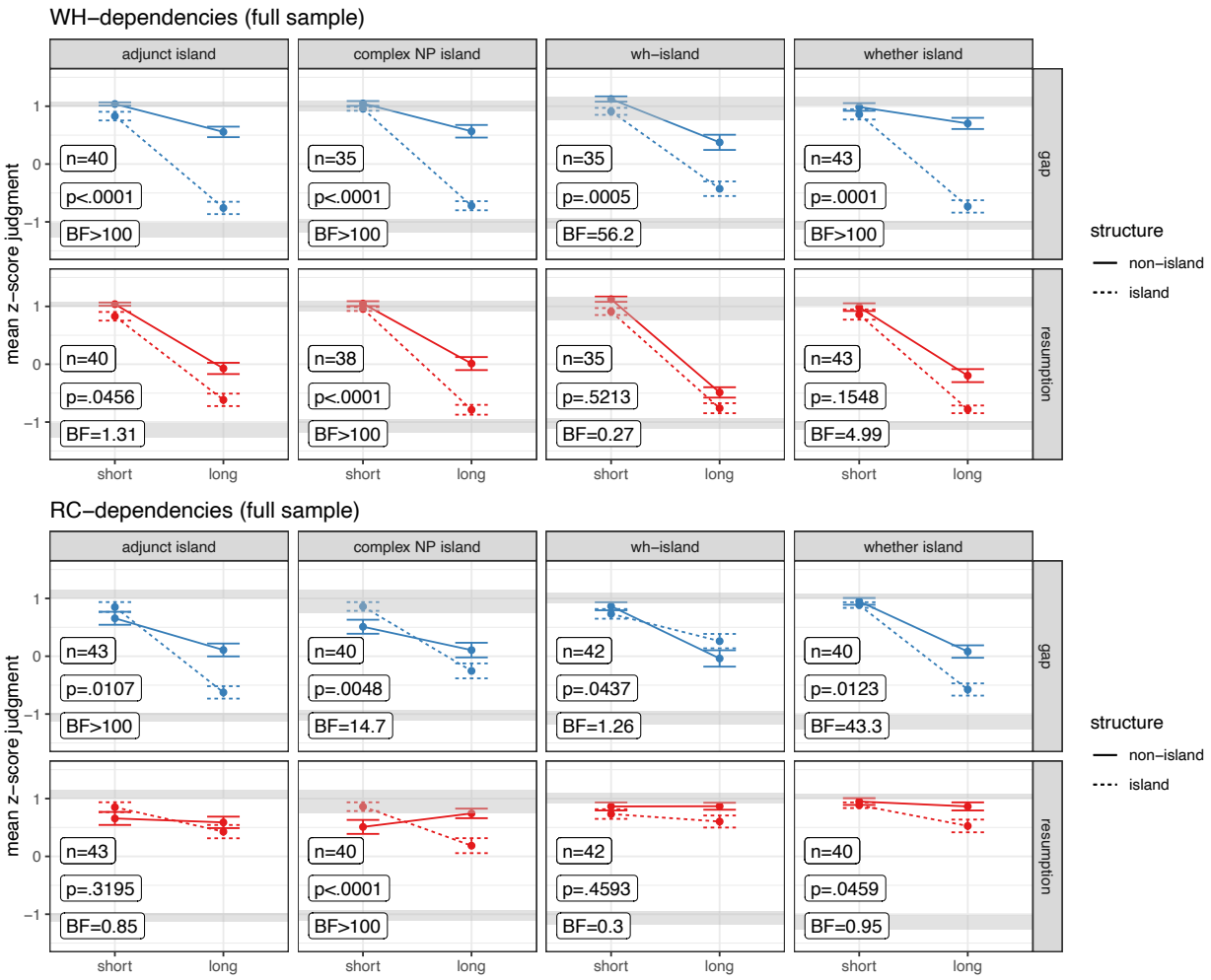
4.1 Island effects and amelioration by resumption for the full sample

Figure 2 reports the means and standard errors arranged in an interaction plot, with wh-dependencies in the top half of the plot and rc-dependencies in the bottom half of the plot. Within each dependency, the top row reports the results for gap conditions, and the bottom row reports the results for resumption conditions. Because of the incomplete factorial design, we use the same two short conditions in both the gap and resumption rows. The columns report the four island types (adjunct, complex NP, wh, and whether islands). The gray bars represent the range of mean ratings for the eight unacceptable and eight acceptable filler types in order to look for evidence of floor or ceiling effects that could influence the results (we do not see any, so they will not be commented on further). The counts in each facet represent the number of observations for the island/long condition.

For wh-dependencies, in the gap conditions (top row), we see large superadditive interactions for all four island types 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.

The resumption conditions (bottom row) do not show amelioration. The amelioration pattern would be an increase in the acceptability of the island/long condition, as compared to the gap version, that eliminates the superadditive interaction. But that is not what we see. The island/long condition is not statistically different for adjunct, np, and whether islands (*p*-values above .05 and Bayes factors near or below 0.33; see tables A4 and A5 in the appendix), and is statistically inconclusive for wh-islands (p=.046 and BF_{10}=1.31), with a trend in the opposite direction (resumption is rated lower than a gap). What we see instead is that the non-island/long condition is rated lower with resumption than with gaps – near the mid-point (0) for adjunct, complex NP, and whether islands, and substantially lower for wh-islands. This suggests a dispreference for resumption with wh-dependencies that we will attempt to control for in section 4.2 and explore directly in section 4.3. This dispreference leads to a smaller interaction (*p*-values below .05 and BF_{10} above 3 for all four islands; see tables A6 and A7 in the appendix). But, crucially, that smaller interaction is not amelioration – it is the consequence of simultaneously combining two independent effects (the island effect and the dispreference effect) in one paradigm.

Figure 2: Interaction plots for island effects and amelioration by resumption. 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 counts represent the number of observations for the island/long condition. The horizontal gray bars represent the range of mean ratings for the eight unacceptable and eight acceptable filler types.



For rc-dependencies, in the gap conditions (top row), we see island effects for adjunct, complex NP, and whether islands, but not wh-islands. For adjunct islands and whether islands, we see the typical superadditivity pattern, and it is confirmed by both statistical tests. For complex NP islands, we see a non-monotonic interaction (i.e., crossing lines) because the island/short condition is rated higher than the non-island/short condition. This is still an island effect in that the island/long condition is less acceptable than one would expect given the acceptability of the other three conditions. And the interaction is confirmed by both statistical tests. The question of why the island/short condition appears to be relatively more acceptable likely lies beyond the theory of syntax (perhaps in theories of semantic plausibility or even sentence processing), therefore we do not explore it here. We also note that the shift upward in acceptability in the island conditions means that the island/long condition is rated relatively high

(-0.25) compared to the other island effects (adjunct is -0.63 and whether is -0.58). 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 either the hypothesis that there is a reverse interaction or the hypothesis that there is no interaction. Though this appears inconclusive, we note that there is no evidence of a trend toward a true island effect, so we take this as evidence against the presence of wh-islands. We thus conclude that there is strong evidence of adjunct, complex NP, and whether islands, and evidence against wh-islands.

For the resumption conditions (bottom row), we see a form of amelioration for all three of the islands that were present with gaps. We will discuss each island type turn.

Adjunct islands show a trend toward classic amelioration: the visual pattern shows a large increase in acceptability for the island/long condition. The *p*-value is greater than the conventional threshold, suggesting no evidence of an interaction; however, the Bayes factor is 0.8, which suggests that the data is roughly equally likely under both hypotheses. We are inclined to interpret this as amelioration because, even if a substantially larger experiment were to detect an interaction, it would be an extremely small interaction with all four conditions in the upper half of the range of acceptability. In other words, this hypothetical effect would be a small subliminal island – that is, a superadditive interaction with the island/long condition rated above the mid-point of the scale (Almeida 2014, Kush et al. 2018, Keshev & Meltzer-Asscher 2019) – and not a classic island effect.

For complex NP islands, we once again find an interaction that is confirmed by both of our statistical tests, but with the atypical, non-monotonic form. However, both the non-island/long and island/long conditions have shifted higher in acceptability compared to the corresponding gap conditions. This suggests a general preference for resumption over gaps in rc-dependencies. We will explore this further in sections 4.2 and 4.3. Given the shift into the mid-range of the acceptability scale one might wonder whether the complex NP island with resumption is a subliminal island effect. We would argue that it is not a subliminal island based on the size of the effect: 0.9 on the z-score scale. This is two to three times larger than the subliminal islands previously reported in the literature (Almeida 2014, Kush et al. 2018, and Keshev & Meltzer-Asscher 2019), and roughly the same size as the island effect with gaps (the statistical comparison of the interaction effects between gaps and resumption in tables A6 and A7 suggest that the effect may even be larger with resumption, as BF_{10} is slightly above 3). We will therefore treat the complex NP island as a classic island effect.

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/long condition is in the upper half of the acceptability scale, which means that even if a substantially larger experiment were to detect an interaction, it would again be a very small subliminal island. We thus conclude that adjunct islands show the clearest pattern of amelioration, whether islands show amelioration that could possibly involve a very small subliminal island (though our experiment was not sensitive enough to detect effect sizes that small), and complex NP islands do not show amelioration (but rather a non-monotonic island effect for both gaps and resumption). We do not discuss wh-islands in detail because they

do not show island effects with gaps, so the lack of island effect with resumption is to be expected.

Table 3 summarizes the results of these experiments. We have categorized the effects as islands only when both statistical analyses (null hypothesis testing and Bayes factors) converge, and as none or amelioration otherwise.

Table 3: Summary of the results of the experiments.

		adjunct	np	wh	whether
wh-dependencies	gap	island	island	island	island
	resumption	island	island	island	island
rc-dependencies	gap	island	island	none	island
	resumption	amelioration	island	none	amelioration

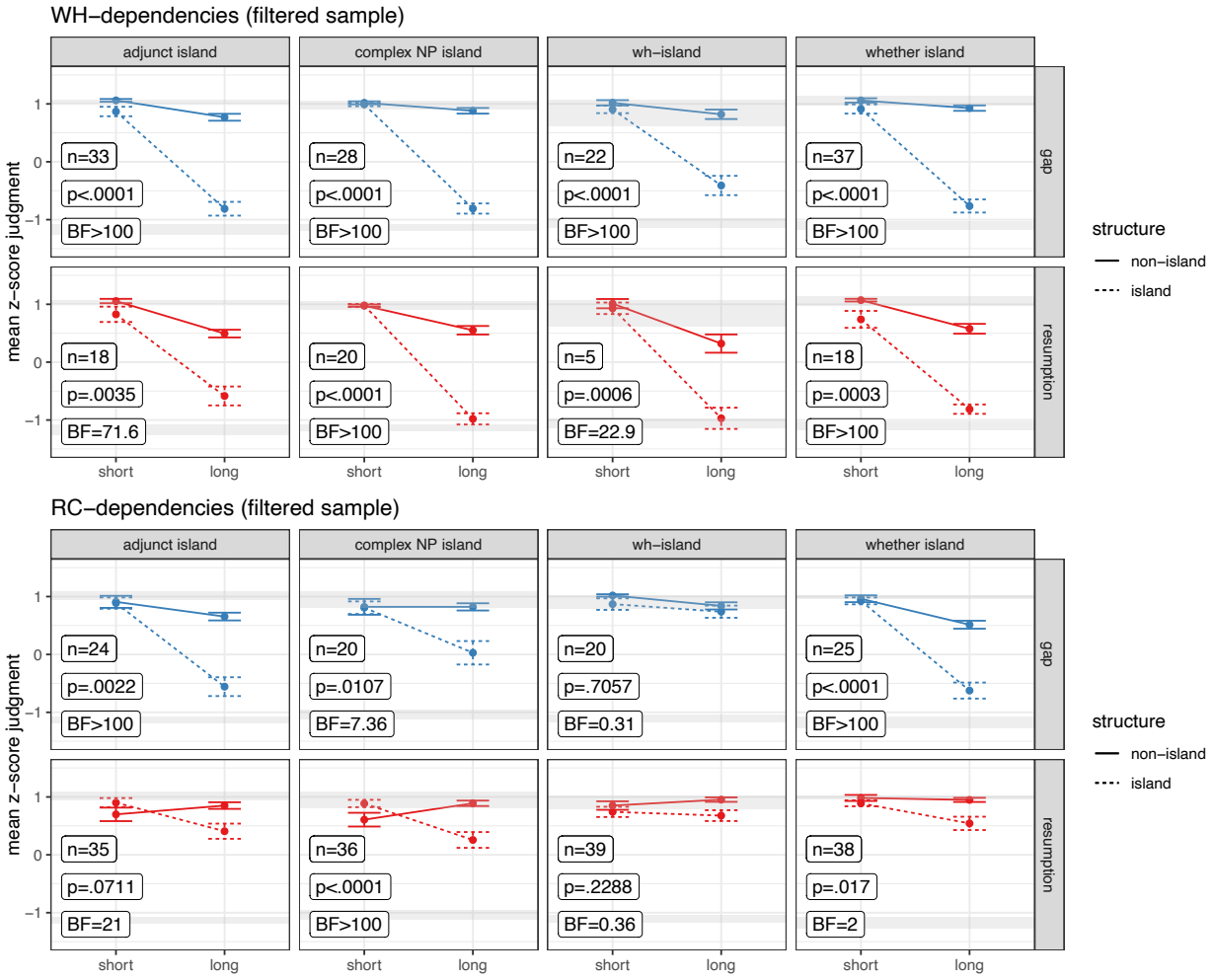
4.2 Island effects and amelioration by resumption while controlling for preferences for gaps or resumption

We saw indications in the full sample analysis in Figure 2 that some participants may have preferred gaps with wh-dependencies and some may have preferred resumption with rc-dependencies. This is a common pattern in grammatical resumption languages (see Salzmann 2017 and Choueiri 2017 for reviews). However, for our full sample analysis these preferences are a potential confound that could add noise to the island analyses in Figure 2. In Figure 3 below, we re-analyze the data in an attempt to eliminate this potential confound. For the gap rows of the plot, we only included participants who rated gaps in the non-island/long condition above the mid-point (0) on the z-score scale. For the resumption rows of the plot, we only included participants who rated resumption in the non-island/long condition above the mid-point. Filtering based on the ratings of the (by hypothesis, grammatical) non-island/long condition should eliminate any possibility that a dispreference for a specific tail type would contaminate the island analyses. The one limitation of this analysis is that the filtered subsets will be smaller than the full sample, thus reducing the statistical power of our analyses. To track this reduction, each cell contains the number of participants included in the analysis. The critical question is whether the pattern in the filtered subsets differs meaningfully from the pattern in the full sample analysis.

The overall pattern of results that we see in Figure 3 is qualitatively identical to Figure 2 – we see the same island effects, the same amelioration effects, and the same absence of island effects. There are some minor quantitative differences, but they are generally in line with what we might expect given the way that we defined the subsets. For wh-dependencies and gap conditions, we still see large island effects for all four island types. The only change appears to be a higher rating for the non-island/long conditions (in effect making the island effects larger), presumably due to removing the influence of participants who do not allow gaps with wh-dependencies. For the resumption conditions, we now see patterns consistent with relatively large island effects, presumably due to removing the influence of participants who do not allow resumption with wh-dependencies. In short, just as with Figure 2, Figure 3 shows classic island effects with gaps, and no amelioration by resumption. For rc-dependencies and gap conditions, we see large classic island effects for adjunct and whether islands. These are larger than the

island effects in Figure 2, with the increase driven by an increase in the non-island/long condition, presumably due to removing the influence of participants who do not allow gaps. The complex NP pattern is now that of a monotonic interaction, not the non-monotonic pattern in Figure 2. The mean rating of the island/long condition is still near the midpoint of the scale, similar to Figure 2, albeit a bit higher, suggesting either a classic island effect or a nearly subliminal island effect. And we still see no wh-island effect. Finally, for rc-dependencies and resumption, we see a subliminal island effect for adjunct islands, a non-monotonic interaction for complex NP islands that is still a subliminal island (but only based on five participants), no island effect for wh-islands, and either no island effect or a very small subliminal island for whether islands. In short, for rc-dependencies we see amelioration patterns similar to Figure 2.

Figure 3 – Interaction plots for island effects and amelioration by resumption, allowing for variation in preferences for gaps and resumption. Details are the same as Figure 2, except that gap rows only include participants who accept gaps in the non-island/long condition, and resumption rows only include participants who accept resumption in the non-island/long condition.



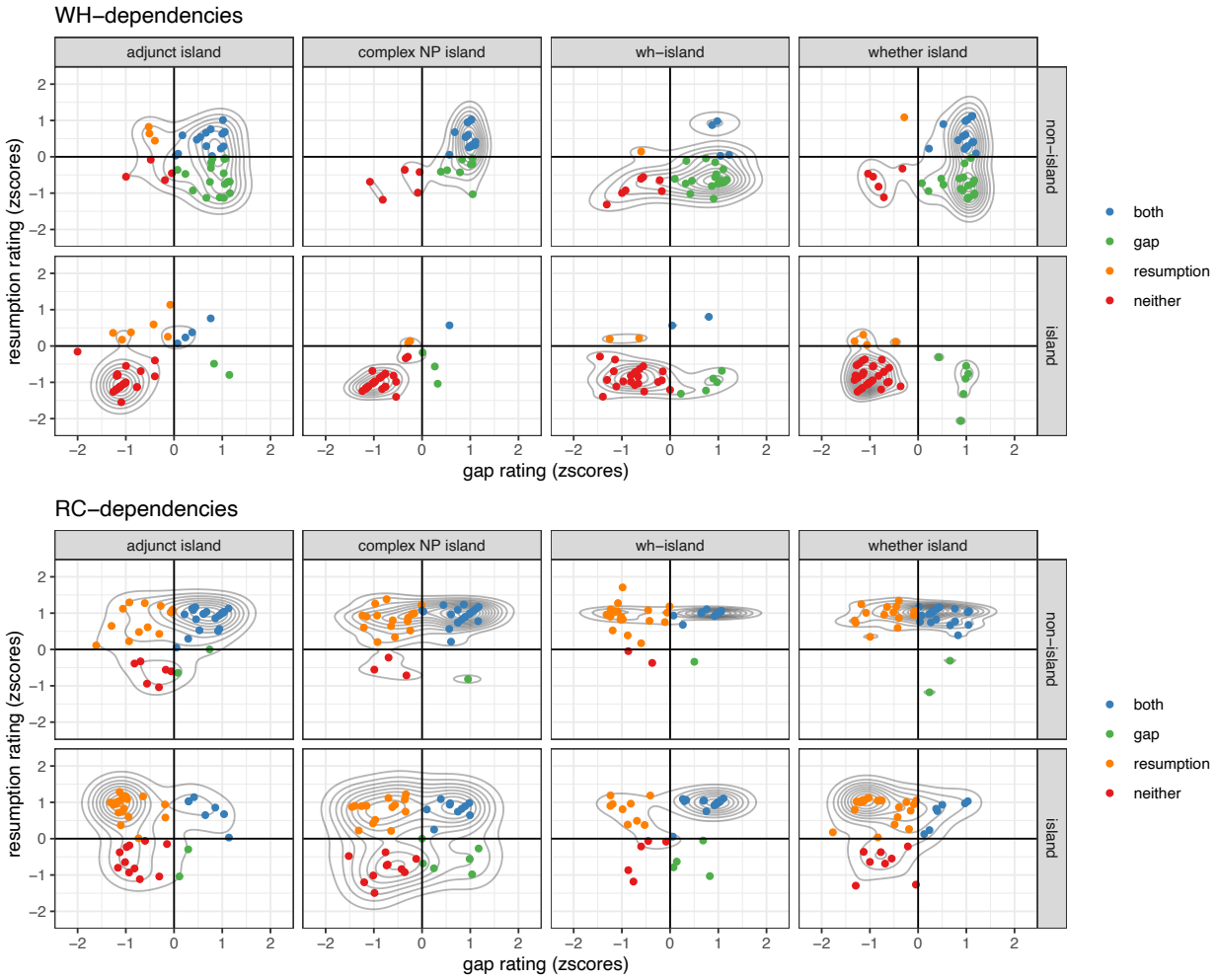
The possibility that the amelioration by resumption effect for adjunct and whether islands could give rise to subliminal island effects is potentially novel. Unfortunately, because our study was not explicitly designed to detect subliminal islands (which are, by definition, very small effects), it is difficult to distinguish full amelioration from subliminal island effects with high confidence. We see this in the fact that the two statistical tests rarely converge for the potential subliminal islands. We therefore note the possibility that adjunct and whether islands with resumptive pronouns in JA may yield subliminal island effects as a hypothesis to be explored with a dedicated study in the future (specifically one with higher statistical power for small effect sizes, and perhaps more observations per condition per participant to make it possible to explore individual variation).

4.3 The preference for gaps versus resumption

The preceding group-level analyses of island effects and amelioration by resumption suggest that there may be variation across participants in their preferences for gaps versus resumption. Though our experiments were not designed to study individual variation, we can provide an exploratory analysis of the individual variation in our results to reveal potential hypotheses to be explored in future studies. We provide such an analysis in this subsection, with two caveats. The first is that our study only collected one observation per condition per participant, thus potentially leading to variability across participants due to other, theoretically uninteresting, factors (e.g., item-level effects, fatigue/attention effects). The second is that individual variation itself has proven relatively unreliable, at least with measures of (real-time) sentence processing (e.g., James et al. 2018, Staub 2021); it is currently unclear if this reliability issue extends to (offline) acceptability judgments.

Figure 4 plots each participant's rating for the gap version of a condition along the x-axis against that same participant's rating for the resumption version of a condition along the y-axis. The figure is divided as before into wh-dependencies in the top panel and rc-dependencies in the bottom. The columns represent island types, and the rows represent the specific condition – either the grammatical non-island/long condition or the potentially ungrammatical island/long condition. We divided each plot into four quadrants based on the mid-point of the z-score scale (0): a point in the top right quadrant represents a participant who rated both the gap and resumption version of the condition in the upper (positive) half of the z-score scale, and therefore appears to allow both tail types (labeled "both"); a point in the top left quadrant represents a participant who rated the gap condition low and the resumption condition high, and therefore appears to prefer resumption only (labeled "resumption"); the bottom left quadrant represents a participant who rated both conditions low, and therefore appears to allow neither tail type (labeled "neither"); and the bottom right represents a participant who rated gaps high and resumption low, and therefore appears to prefer gaps only (labeled "gap"). We have added two features to make the plot more informative: 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 (i.e., number of participants) in each quadrant. Similar to a topographic map, concentric circles that are closer together represent higher density (i.e., more participants). Though the dividing lines and colors create clear distinctions among participants, we should note, in line with our caveats above, that the potential for noise in these (single) ratings could mean that participants nearer to the dividing lines may be misclassified.

Figure 4: Scatterplots identifying preferences for gaps and resumption for each participant. 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.



To determine the general preferences among our sample of speakers, we can look at the non-island conditions (the top row in each panel). The non-island structure for all four island types is the same – a declarative CP. Therefore, the four columns are simply four replicated experiments testing the same CP structure, albeit with distinct lexicalizations (lexically matching the island conditions for the island indicated in the column label). We can then count how many participants appear in each of the quadrants across all four columns in the top rows for each dependency. Table 4 provides those counts, setting aside the neither (red dots) classification as participants who may have provided noisy judgments in these trials.

Table 4: The number of participants reporting each tail preference based on the classification in Figure 4. The cells for the four most common patterns are shaded in gray.

		wh-preference			
		both	gap	resumption	total
rc-preference	both	34	35	2	71
	gap	2	2	0	4
	resumption	15	21	3	39
	total	51	58	5	114

We can first look at each dependency in isolation. For wh-dependencies, these preferences are captured in the row labeled total. We see that most participants tended to prefer either both gaps and resumption (blue dots, 51 total participants), or only gaps (green dots, 58 participants). Very few participants preferred resumption only (orange dots, 5). For rc-dependencies, these preferences are captured in the column labeled total. We see that most participants tended to prefer either both gaps and resumption (blue dots, 71 total participants), or only resumption (orange dots). Very few participants accepted gap only (green dots, 4). This exploratory analysis suggests two preference patterns for each dependency type, which, to our knowledge, has not been reported previously in the resumption literature. We can also look at how the preferences for the two dependencies combine. This is captured in the 9 interior cells. We have shaded the 4 combined patterns formed by the most frequent preferences. As an anonymous reviewer notes, the counts for each of the four combined patterns appear to be what we would expect if the preferences for each dependency combined independently, suggesting that the preference in one dependency does not appear to influence the preference in the other. A contingency table Bayes factor test from the BayesFactor package yields a BF_{01} of 3.36 for the four shaded patterns, suggesting that the data is a bit more than 3x more probable under the null hypothesis that the two preferences are independent (and a typical chi-square test yields a *p*-value of .59), thus corroborating this observation. We discuss the potential research questions raised by our observation that there are four preference patterns in more detail in section 5.3.

We turn next to the bottom rows of the two panels in Figure 4, which show participants' ratings of the island/long conditions. In effect, these rows show us the classification of participants according to island effects and amelioration by resumption if we used a single condition definition for these effects, rather than the factorial definition, such that a low rating with a gap indicates the presence of an island effect and a high rating with resumption indicates the presence of amelioration. Though the theoretical literature often reports judgments for individual sentences, we believe syntacticians typically have the factorial definition in mind as discussed in section 3.1. Nonetheless, we can explore this single-sentence approach for completeness. For both dependency types, what we find is a distribution of judgments that matches the factorial approach taken in sections 4.1-4.3. For wh-dependencies (top panel), we see almost all participants are in the bottom left quadrant (red), indicating that they rated both

gaps and resumption low. This suggests that nearly all participants show an island effect and no amelioration effect. For rc-dependencies (bottom panel), we see a similar convergence in analyses. For adjunct and whether islands, we see a large group of participants in the top left quadrant (orange), indicating an island effect for gaps and amelioration with resumption, and a smaller group of participants in the bottom left quadrant (red), indicating an island effect without amelioration. For complex NP islands we see three groups: those showing amelioration (orange), those showing no amelioration (red), and those showing no island effect at all (blue). This three-way split, coupled with patterns that we saw in the group results, suggests that the status of complex NP islands may be a substantial point of variation among speakers of JA. Finally, for wh-islands, the largest group of participants show no island effect (blue), with a small group showing amelioration (orange). Though we believe that the factorial definition best matches the logic of the theoretical syntax literature, it is reassuring that a single-sentence approach to these effects would yield a substantially similar pattern.

5. Discussion

We conducted four auditory judgment experiments exploring the presence of island effects and amelioration effects in JA, for two dependency types, four island types, and two tail types, using a factorial definition for both island effects and amelioration by resumption. Within these experiments, we were also able to explore the preferences for gaps and resumption that our participants displayed. We found four patterns formed by fully crossing gaps or both for wh-dependencies and resumption or both for rc-dependencies. We discuss the theoretical consequences of these results for island effects, amelioration, and theories of the distribution of gaps and resumption in the rest of this section.

5.1 Consequences for theories of island effects

The major finding for theories of island effects is that the pattern of island effects varies between wh-dependencies and rc-dependencies. As mentioned in section 2, variation across dependencies raises challenges for all four classes of theories of island effects. The specific pattern of variation that we observed in JA is also relatively rare – we know of no previous study that observed variation in complex NP islands across dependencies (though in JA it is just a subset of participants who show the variation), and we know of no previous study that observed the simultaneous absence of wh-islands and presence of whether islands within a single dependency (the closest is Pañeda & Kush 2021, which found the opposite pattern in Spanish – the presence of wh-islands and a subliminal effect for whether islands). In this subsection, we discuss the challenges that this pattern of variation raises for each of the four classes of theories, and the possible avenues available for modifying existing theories to capture this pattern.

The challenge for phase-based theories is that both phases and their edges are defined independently of the dependency types. This is a conceptual strength of the theory – phases and their edges are grounded in deep architectural principles of the system (see Müller 2021). But it comes with a steep empirical cost. We can see three general options for modifying phase-based theories to accommodate the variation we observed in JA. The first, and most radical, would be to abandon the phase-based approach to island effects, and instead capture island effects using less-grounded grammatical constraints (similar to previous approaches such as individual island constraints, Subjacency, or the CED). Though the link between phases and locality is

conceptually appealing, it has been noted since the earliest days of the phase-based approach that the only types that can be explained by phase impenetrability without additional assumptions are wh-islands and whether islands (not unlike Chomsky's 1986 Barriers approach; see Boeckx 2013 for a review). The variation that has been observed in JA and other languages may thus be seen as additional evidence that phases may not be the source of island effects. If island effects are due to (ungrounded) syntactic constraints, then it is possible to define those constraints separately for each dependency type. Similarly, the pattern of variation that we observed can be captured with the standard analytic tools of syntax. For example, the difference between wh-islands (no island effect) and whether islands (island effect) with rc-dependencies could suggest that *whether* occupies a position that is to the left of a non-phase-edge position occupied by *why* (perhaps spec of an interrogative phrase and spec of a focus phrase, respectively) as in Rizzi 2001. Similarly, the absence of complex NP islands with rc-dependencies for some participants could be analyzed parallel to the absence of complex NP islands in some languages (e.g., Japanese; Haig 1976, Yano 2019).

The second path forward for phase-based approaches would be to link the variation in JA to the strong/weak island distinction, perhaps by positing that strong islands are explained by phases, while weak islands are explained by intervention-based theories like RM. This approach would potentially maintain the groundedness of phases; however, it is not clear how to motivate the separation of strong and weak islands into two distinct theories of locality. This has long been an issue in the literature, as phase-based and intervention-based theories partially overlap precisely in the case of weak islands. That overlap has generally been unresolvable because weak islands are the core case for phase impenetrability – CPs are the canonical example of a phase, and weak islands typically involve an item occupying the edge of the CP phase. Relatedly, strong islands tend to involve a number of additional assumptions built upon the core mechanisms of phase impenetrability. For this approach to be viable, the field would have to identify a new theory of phases that somehow eliminates this superset/subset relationship of the mechanisms underlying strong and weak islands.

The final path forward for phase-based theories would be to link the dependency type to the availability of a phase-edge position, either through additional positions (as in the multiple specifier approach of Nyvad et al. 2017) or through the availability of edge features (as in Müller 2010). We do not know of any existing proposals in the literature that would accomplish this, as variation in island effects across dependencies has not historically been discussed in the phase-based literature. But this seems like a promising avenue for researchers who wish to maintain the groundedness of phases and the coverage of both strong and weak islands.

Intervention-based theories like RM are not intended to explain variation in strong islands like we observed for complex NP islands in JA; however, they can potentially explain variation in weak islands by leveraging featural differences between the dependencies and featural differences between wh-islands and whether islands. Given that rc-dependencies show whether islands but not wh-islands, the challenge is to either find a feature that is shared between relative clause heads and whether clauses but not shared by the why questions in wh-islands, or to establish a hierarchy of features that groups a feature of relative clause heads and a feature of whether clauses together to the exclusion of the features of why questions. We know of no existing feature or hierarchy of features that would accomplish this. In fact, the standard features in RM analyses, such as +Q for questions or +focus for focus constructions, would potentially predict the opposite pattern: if whether clauses are +Q, and why questions are +Q and +focus (under an analysis in which wh-words occupy a focus position), and relativization involves

+focus, then rc-dependencies would be expected to show wh-islands but not whether islands, in contrast to the pattern that we observe in JA. Thus, even granting the restricted empirical scope of intervention-based theories, the pattern we observe in JA will require either new featural analyses of rc-dependencies and weak islands, or new proposals for hierarchies of features in JA.

Information-structure-based theories can capture variation across dependency types by postulating distinct information-structure properties for each dependency. But, as discussed in section 2, the effects would likely hold for all island types in all-or-nothing patterns, making the complex patterns observed here difficult to capture. Recognizing this issue in her initial investigation of variation between Danish and English, Erteschik-Shir (1973) proposes that the basic information-structure clash defines the set of possible island effects, and that the input that children receive determines the subset of these possible island effects that become actual island effects. Though originally proposed for cross-linguistic variation, this mechanism could be extended to variation between dependencies. Erteschik-Shir (1973) left the details of this mechanism to future work in language acquisition, as it obviously requires a theory of the inferences that can be drawn from the input that children receive. Here we will simply note two things about this work. First, there have been a number of advances in theories of language acquisition that may make this work more tractable today (see Pearl *in press* for a review). Second, the general idea pursued by Erteschik-Shir – that innate knowledge of the set of possible island effects combines with experience to shape the final set of island effects – can be found in many theories of island effects that arose after her seminal work (possibly in all but the independently grounded theories like phase-based or processing-complexity-based theories). Therefore, there seems to be quite a bit of potential for additional work exploring how mechanisms of language acquisition could give rise to island variation through variation in the input that children receive (e.g., Pearl & Sprouse 2013), in both JA and other languages that show variation in island effects (e.g., Table 1).

Processing-complexity-based theories are relatively underdeveloped when it comes to variation in island effects. This is because the only mechanism available to explain variation is an interaction between the processing dynamics of the dependencies (e.g., encoding and retrieval from memory) and the processing requirements of the individual island structures. Though it is possible in principle to construct a theory of these interactions, there is no such theory yet. Furthermore, the specific pattern that we observed in JA, with rc-dependencies showing whether islands but not wh-islands, appears to run contrary to the processing-complexity findings in Kluender & Kutas 1993. In an event-related potential study, they report larger N400-like effects for embedded wh-questions compared to embedded polar-questions. They interpret this as evidence that embedded wh-questions require greater processing resources than polar-questions. This in turn appears to predict that if one of the two islands were to be absent, it would be whether islands, but that is opposite to what we observe in JA. This suggests that the specific challenge for processing-complexity-based theories raised by the variation that we observe in JA is to identify new processing dynamics for whether and wh-islands that pattern in the correct direction.

We have attempted to keep the preceding discussion as objective as possible so that the results of the current study can be useful to researchers working in all four major theories of island effects. That said, the pattern of cross-linguistic variation that has been emerging in the experimental syntax literature (e.g., Table 1), coupled with the results here, suggests that island effects are more variable across both island types and dependency types than previously believed. Given that, our personal subjective beliefs are that the most profitable approach to

island effects moving forward will be one that allows island types and dependency types to vary independently, such as one that postulates distinct syntactic constraints for each island type, and allows the acquisition process to track input for each dependency type separately. We believe that the challenges posed by the variation we observe across languages and studies are, at least at present, unlikely to be resolved for the other theories without significant changes to their architectural assumptions.

5.2 Consequences for theories of amelioration by resumption

We begin with wh-dependencies, which showed a fairly uniform pattern, and therefore license the strongest conclusions. We observed no amelioration by resumption for wh-dependencies. It is tempting to attribute this to the relatively common pattern in which grammaticalized resumption languages disallow all resumption with wh-dependencies, even within non-island structures (see Demidarche 1991 for a theory that predicts this pattern). But in our exploratory analysis of individual variation, nearly half of our participants reported judgments that could be interpreted as allowing resumption with wh-dependencies in non-island structures, and even these participants show no amelioration by resumption in island structures. This suggests the lack of amelioration by resumption with wh-dependencies is truly a fact about amelioration, and therefore that JA is a no amelioration language with respect to wh-dependencies. This in turn suggests that wh-dependencies with resumption are most likely generated by syntactic movement rather than base generation, as it is generally assumed that movement is potentially sensitive to islands, while base generation is always insensitive to islands. Our observation that wh-dependencies in JA show no amelioration by resumption accords well with analyses that posit that wh-dependencies must involve movement, such as Tellier 1991 and Merchant 2004 (though each motivates this restriction differently).

For rc-dependencies and resumption, we see amelioration for adjunct and whether islands, but not for complex NP islands. Though this is a mixed result, we can see why JA has previously been reported to be a full set amelioration language (e.g., Malkawi & Guilliot 2007). First, the amelioration spans both a canonical strong island (adjunct) and a canonical weak island (whether), indicating full set amelioration. Second, the effect for complex NP islands yields a rating near the middle of the scale for the island violating sentence, which could potentially be interpreted as no island effect in an informal judgment study. That said the pattern that we observed suggests that resumption in JA is a mix of both movement (for complex NP islands) and base generation (for adjunct and whether islands). This accords well with the overall pattern of cross-linguistic variation in resumption that has emerged from large scale reviews (e.g., Salzmann 2017) and from targeted reviews of varieties of Arabic (e.g., Choueiri 2017). This also accords well with the literature on reconstruction in varieties of Arabic, which also postulates a mixed source for resumption even within a single language (e.g., Aoun et al. 2001, Malkawi & Guilliot 2007, Aoun et al. 2010, Rouveret 2011). Future studies could establish a theoretically meaningful connection between variation in island amelioration and variation in reconstruction by testing both phenomena within the same set of participants.

Before moving on, it is also worth noting the implications of mixed theories of resumption for language acquisition. Mixed theories must assume that a set of innate biases will combine with evidence in the input available to help children make the correct inference about the mechanism, movement or base generation, underlying each instance of resumption (where “instance” in this case means something like distinct structure containing the resumptive

pronoun). One possibility is that learners are innately biased to look for island constraints of some sort (see Roeper & de Villiers 2011 for a review), and also innately biased to associate movement with sensitivity to islands and to associate base generation with insensitivity to islands (as is commonly assumed in syntactic theory). With these biases, learners could use the presence or absence of dependencies that span islands in the input as evidence to make an inference about the underlying mechanism. Investigating the acquisition of resumption is far beyond the scope of this project; however, we note that the speaker variation that we observed here could be taken to predict that the frequency of island-spanning rc-dependencies with resumption will be quite small in child-directed JA, such that some learners might never be exposed to them during acquisition (leading to no amelioration in their grammars). This prediction could be explored in future work through systematic corpus studies of child-directed speech in JA.

5.3 Consequences for theories of the preference for gaps and resumption

Though it was not an initial goal of these experiments, our results also revealed both expected and unexpected variation in the preference for gaps and resumption in JA. Our analyses replicated the frequent observation among varieties of Arabic that, of the two dependency types, wh-dependencies are more likely to reveal a gap-only pattern, and rc-dependencies are more likely to reveal a resumption-only pattern (see, for example the review in Choueiri 2017). With the caveat that our analysis of preferences is only exploratory, we also found evidence that some speakers of JA accept both gaps and resumption for wh-dependencies and that some speakers of JA accept both gaps and resumption for rc-dependencies. To our knowledge, the possibility of two patterns of preferences for each dependency type has not been previously reported in the literature on spoken varieties of Arabic. Table 5 below extends a table from Choueiri 2017 to include these new observations for JA. The data in the table comes from Choueiri 2017 (Lebanese); Wahba 1984, Brustad 2000, Aoun et al. 2010, and Soltan 2011 (Egyptian); and Nouhi 1996 (Moroccan). Given that our experiments were not explicitly designed to investigate individual differences, we present these new observations as potential hypotheses to be tested explicitly in future studies.

Table 5: The preference for gaps or resumption in four varieties of Arabic, adapted from Choueiri 2017 and including the results of the current study.

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

To our knowledge there is no overarching theory of the distribution of preferences for gaps and resumption. It is primarily presented as a descriptive generalization in the literature, therefore we cannot comment on the consequences for the theory should our observations of variability within JA be corroborated by future experiments. However, these results do suggest that there may be more variability in resumption than has previously been reported, both for island amelioration and preferences in non-island structures. This variability must be taken into account to gain a clear picture of the properties of both gap and resumption dependencies,

therefore future studies may profit from designs that provide a high level of confidence in the preferences of individual participants, and analyses that filter participants based on those preferences. An anonymous reviewer points out two potentially interesting follow-up studies to better explore this variation. The first is to elicit judgments for both JA and MSA from the same set of participants to systematically quantify the effect of diglossia at both the group and individual level. The second is to systematically vary the type of wh-items at the head of a resumption dependency. Our experiments only tested *weiʃ* ('what') with resumption, never *miːn* ('who') or d-linked wh-phrases. Following Aoun & Choueiri 1999 and Shlonsky 2002, the reviewer notes that there may be variability across wh-dependency types, and that this variability may be linked to the ability of the wh-item to be d-linked.

6. Conclusion

Our broad goal in this study was to bring evidence from Jordanian Arabic, a primarily spoken variety of Arabic, into the (formal experimental) empirical base of both theories of island effects and theories of island amelioration by resumption. To that end, we ran four auditory judgment studies exploring two dependency types, four island types, and both gaps and resumption. Our experiments identified three sources of variation that raise challenges for existing theories: variation across dependency types in the sets of islands present with gaps, variation across island types in the presence of amelioration by resumption, and potentially variation across participants in the preferences for gaps versus resumption. The variation across dependency types suggests that each of the four major classes of theories of island effects – phase-based, intervention-based, information-structure-based, and processing-complexity-based – require substantial modification. For each, we discussed specific paths forward that theoretical work could pursue. The variation across island types for amelioration by resumption suggests a mixed source of resumption dependencies in JA – i.e., both movement and base generation. The variation across participants also suggests that future studies of resumption in JA, and perhaps other grammatical resumption languages, will benefit from experimental designs that quantify individual variation, both in the preference for gaps or resumption in non-island structures, and in the amelioration effect across island types. The variation also suggests that there is a profitable path forward for systematic studies of the input that children receive during the acquisition of JA and perhaps the acquisition of other grammatical resumption languages. Taken as a whole, we believe that these experiments show that spoken varieties of Arabic have much to offer the experimental syntax literature, both in terms of the impact that their patterns of variation will have on theories, and in terms of the impact that their patterns of variation will have on the types of studies that the field explores moving forward.

Data availability statement

Both the materials and data that support the findings of this study are openly available on the authors' websites.

Acknowledgments

We would like to thank three anonymous reviewers and the editors at Syntax for helpful comments and suggestions throughout the review process. We would also like to thank audiences at Aarhus University, New York University, and the University of Delaware for helpful feedback on this project at earlier stages of its development. All errors remain our own.

References

Abeillé, Anne, Barbara Hemforth, Elodie Winckel, and Edward Gibson. 2020. Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition* 204.104293.

Abels, Klaus. The Italian left periphery: A view from locality. *Linguistic Inquiry* 43.229–254.

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. and Adele Goldberg. 2008. The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics* 19.357–389.

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.

Aoun, Joseph, and Lina Choueiri. 1999. Modes of interrogation. In Elabbas Benmamoun, editor. *Perspectives on Arabic Linguistics XII: Papers from the annual symposia on Arabic Linguistics*. Amsterdam: John Benjamins. 7-26.

Asudeh, Ash. 2012. *The logic of pronominal resumption*. Oxford: Oxford University Press

Bates, Douglas, Martin Mächler, Ben Bolker and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67.1-48.

Belletti, Adriana. 2006. Extending doubling to non-local domains: Complete vs. partial copying + deletion and related reconstruction issues. In Patrick Brandt and Eric Fuß, editors. *Form, structure, and grammar*. Berlin: Akademie. 129–136.

Belletti, Adriana, Naama Friedmann, Dominique Brunato, and Luigi Rizzi. 2012. Does gender make a difference? Comparing the effect of gender on children's comprehension of relative clauses in Hebrew and Italian. *Lingua* 122.1053–1069

Beltrama, Andrea and Ming Xiang. 2016. Unacceptable but comprehensible: The facilitation effect of resumptive pronouns. *Glossa* 1.1.

Bianchi, Valentina. 2004. Resumptive relatives and LF chains. In Luigi Rizzi, editor. *The Cartography of syntactic Structures, Volume 2*. Oxford: Oxford University Press. 76–114.

Boeckx, Cedric. 2003. *Islands and chains. Resumption as stranding*. Amsterdam: Benjamins.

Boeckx, Cedric. 2013. *Syntactic Islands*. Cambridge: Cambridge University Press.

Boeckx, Cedric and Norbert Hornstein. 2008. Superiority, reconstruction, and islands. In Robert Freidin, Carlos P. Otero, and Maria Luisa Zubizarreta, editors. *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*. Cambridge, MA: MIT Press. 197–225.

Boersma, Paul. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5.341–345.

Borer, Hagit. 1984. Restrictive relatives in Modern Hebrew. *Natural Language and Linguistic Theory* 2:209–261.

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.

Chacón, Dustin. A. 2019. Minding the gap?: Mechanisms underlying resumption in English. *Glossa* 4.68.

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

Chidambaram, Vrinda S. 2013. *On resumptive pronouns in Slavic: Towards a generalized theory of pronominal structure and interpretation*. Doctoral thesis. Princeton, NJ: Princeton University.

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

Chomsky, Noam. 1977. On Wh-movement. In Peter Culicover, Thomas Wasow, and Adrian Akmajian, editors. *Formal Syntax*. New York: Academic Press. 71–132.

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

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In *Step by Step*, ed. R. Martin, D. Michaels and J. Uriagereka, 89–155. Cambridge, MA: MIT Press.

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

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

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.

Citko, Barbara. 2014. *Phase Theory: An introduction*. Cambridge: Cambridge University Press.

Cohen, Jacob. 1988/2013. *Statistical power analysis for the behavioral sciences*. New York: Academic press.

Deane, Paul D. 1991. Limits to attention: a cognitive theory of island phenomena. *Cognitive Linguistics* 2.1–63.

Demirdache, Hamida. 1991. *Resumptive Chains in Restrictive Relatives, Appositives and Dislocation Structures*, Doctoral thesis. Cambridge, MA: Massachusetts Institute of Technology.

Dickey, Michael. W. 1996. Constraints on the sentence processor and the distribution of

resumptive pronouns. In Michael W. Dickey and Susanne Tunstall, editors. *Linguistics in the Laboratory*. Amherst: Graduate Linguistics Student Association. 157–192.
Donati, Caterina and Carlo Cecchetto. 2011. Relabeling heads: A unified account for relativization structures. *Linguistic Inquiry* 42.519–560.
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*. Doctoral thesis. Cambridge, MA: Massachusetts Institute of Technology.
Farby, Shira, Gabi Danon, Joel Walters and Michael Ben-Shachar. 2010. The acceptability of resumptive pronouns in Hebrew. In Yehuda N. Falk, editor. *The Proceedings of the Israel Association for Theoretical Linguistics (IATL)* 26. Ramat Gan: Bar Ilan University. 1–16.
Featherston, Sam. 2005. The Decathlon Model of empirical syntax. In Marga Reis and Stephen Kepser, editors. *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*. Berlin: Mouton de Gruyter. 187–208.
Ferreira, Fernanda and Benjamin Swets. 2005. The production and comprehension of resumptive pronouns in relative clause "island" contexts. In Anne Cutler, editor *Twenty-first century psycholinguistics: Four cornerstones*. Mahwah, New Jersey: Laurence Erlbaum Associates. 263–278.
Frazier, L. and C. Clifton Jr. 2002. Processing 'd-linked' phrases. *Journal of Psycholinguistic Research* 31.633–660.
Friedmann, Naama, Adriana Belletti and Luigi Rizzi. 2009. Relativized relatives: Types of intervention in the acquisition of A′-dependencies. *Lingua* 119.67–88.
Givón, Talmy. 1973. Complex NPs, word order and resumptive pronouns in Hebrew. In Claudia Corum, Cedrik Smith-Stark, and Ann Weiser, editors. *I'll Take the High Node and You Take the Low Node*: *Papers from the Comparative Syntax Festival*. Chicago: Chicago Linguistic Society. 135–146.
Goldberg, Adele. 2007. *Constructions at work*. London: Oxford University Press.
Haig, John H. 1976. Shadow pronoun deletion in Japanese. *Linguistic Inquiry* 7.363–371.
Hawkins, John. A. 1999. Processing complexity and filler-gap dependencies across grammars. *Language* 75.244-285.
Hayon, Yehie. 1973. *Relativization in Hebrew*. The Hague: Mouton.
Heestand, Dustin, Ming Xiang, and Maria Polinsky. 2011. Resumption still does not rescue islands. *Linguistic Inquiry* 42.138–152.
Hladnik, M. 2015. *Mind the Gap: Resumption in Slavic Relative Clauses*. Utrecht: LOT.
Hofmeister, Philip. and Ivan A. Sag. 2010. Cognitive constraints and island effects. *Language* 86.366–415.
Hornstein, Norbert. 2001. *Move! A minimalist theory of construal*. Malden, MA: Blackwell.
Huang, C-T James. 1982. *Logical relations in Chinese and the theory of grammar*. Doctoral thesis. Cambridge, MA: Massachusetts Institute of Technology.
Jeffreys, H. 1961. *Theory of probability*. Oxford: Oxford University Press.
Keffala, Bethany. 2013. Resumption and gaps in English relative clauses: Relative acceptability creates an illusion of 'saving'. In Chundra Catchart, I-Hsuan Chen, Greg Finley, Shinae Kang, Clare S. Sandy and Elise Stickles, editors. *Proceedings of the 37th annual meeting of the Berkeley Linguistics Society*. Berkeley, CA: Berkeley Linguistics Society. 140–154.

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

Keshev, Mayaan and Aya Meltzer-Asscher. 2017. Active dependency formation in islands: How grammatical resumption affects sentence processing. *Language* 93.549–68.

Keshev, Mayaan and Aya Meltzer-Asscher. 2019. A processing-based account of subliminal wh-island effects. *Natural Language and 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.

Klein, Timo. 2014. Deriving a typology of resumption. In Anke Assmann, Sebastian Bank, Doreen Georgi, Timo Klein, Philipp Weisser, and Eva Zimmermann, editors. *Topics at Infl*. Leipzig: Leipzig University. 329–346.

Klein, Timo. 2016. *Patterns of resumption – towards a derivational account*. Doctoral thesis. Leigzig, Germany: Leipzig University.

Kluender, Robert. 1991. *Cognitive constraint on variables in syntax*. Doctoral thesis. University of California, San Diego.

Kluender, Robert. 1998. On the distinction between strong and weak islands: a processing perspective. *Syntax and Semantics* 29.241–279.

Kluender, Robert. 2004. Are subject islands subject to a processing account? In, Vineeta Chand, Ann Kelleher, Angelo Rodríguez, Benjamin Schmeiser, editors. *The Proceedings of the WCCFL 23*. Somerville, MA: Cascadilla Press. 101–125.

Kluender, Robert and Simone Gieselman. 2013. What's negative about negative islands? A re-evaluation of extraction from weak island contexts. In Jon Sprouse and Norbert Hornstein, editors. *Experimental Syntax and Island Effects*. Cambridge: Cambridge University Press. 186–207.

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 and Information Society* 37.

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

Kroch, Anthony. 1981. On the role of resumptive pronouns in amnestying island violations. In Roberta A. Hendrick, Carrie S. Masek and Mary Frances Miller, editors. *Papers from the 17th Regional Meeting of the Chicago Linguistic Society*. Chicago, Illinois: Chicago Linguistic Society. 125-135.

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 and Rune H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82.1–26.

Lasnik, Howard and Mamoru Saito. 1984. On the nature of proper government. *Linguistic Inquiry* 15:235–89.

Liang, Feng, Rui Paulo, German Molina, Merlise A. Clyde and Jim O. Berger. 2008. Mixtures of

g Priors for Bayesian Variable Selection, Journal of the American Statistical Association 103.410-423.
Lu, Jiayi, Cynthia K. 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 Mustafa Mughazy, editor. *Perpectives on Arabic linguistics XX: Papers from the 20th Annual Symposium on Arabic linguistics*. Amsterdam: John Benjamins. 87-104.
Marty, Paul, Emmanuel Chemla, and Jon Sprouse. 2020. The effect of three basic task features on the sensitivity of acceptability judgment tasks. *Glossa* 5.1–23.
Maxwell, Scott E., Harold D. Delaney and K. Kelley. 2017. *Designing experiments and analyzing data: A model comparison perspective*. Abingdon: Routledge.
McCloskey, Jim. 1979. *Transformational syntax and model theoretic semantics: A case-study in Modern Irish*. Dordrecht: D. Reidel Publishing Company.
McCloskey, Jim. 1990. Resumptive pronouns, A'-binding and levels of representation in Irish. In Randall Hendrick, editor. *Syntax of the modern Celtic languages*, *vol. 23*, *Syntax and semantics*. New York: Academic Press. 199–248.
McCloskey, Jim. 2017. Resumption. In Martin Everaert and Henk van Riemsdijk, editors. *The Blackwell companion to syntax*. Malden, MA: Blackwell. 84–117.
McDaniel, Dana. and Wayne Cowart. 1999. Experimental evidence for a minimalist account of English resumptive pronouns. *Cognition* 70.B15–B24.
Merchant, Jason. 2004. Resumptivity and non-movement. *Studies in Greek linguistics* 24.471–481.
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 Gloria C. Peranteau, Paul M. Levi, and Judith N. Phares, editors. *The Chicago which hunt: Papers from the relative clause festival*. Chicago, IL: Chicago Linguistic Society. 63-72.
Morgan, Adam and Matthew Wagers. 2018. English resumptive pronouns are more common where gaps are less acceptable. *Linguistic Inquiry* 49.861–876.
Müller, Gereon. 2010. On deriving CED effects from the PIC. *Linguistic Inquiry* 41.35–82.
Müller, Gereon. 2014. *Syntactic buffers: A local-derivational approach to improper movement, remnant movement, and resumptive movement*. Leipzig: Institut für Linguistik, University of Leipzig.
Müller, Gereon. 2021. Constraints on grammatical dependencies. In Nicholas Allott, Terje Lohndal, and Georges Rey, editors. *A companion to Chomsky*. Hoboken: Wiley. 190-209.
Nouhi, Youssef. 1996. Wh-constructions in Moroccan Arabic. Masters thesis. Ottawa, Canada: University of Ottawa.
Nyvad, Anne Mette, Christensen, Ken Ramshøj, and Sten Vikner. 2017. CP-Recursion in Danish: A cP/CP-analysis. *The Linguistic Review* 34:449–477.
O'Grady, Wayne. 2010. An emergentist approach to syntax. In Bernd Heine and Heiko Narrog and, editors. *The Oxford handbook of linguistic analysis*. Oxford: Oxford University Press. 257–83.
Omaki, Akira and Chizuru Nakao. 2010. Does English resumption really help to repair island violations? *Snippets* 21.11–12.
Omaki, Akira, Shin Fukuda, Chizuru Nakao and Maria Polinsky. 2019. Subextraction in

Japanese and subject-object symmetry. *Natural Language and Linguistic Theory* 38.627–669.
Pañeda, Claudia, Sol Lago, Elena Vares, João Veríssimo, and Claudia Felser. 2020. Island effects in Spanish comprehension. *Glossa* 5.21.
Pañeda, Claudia and Dave Kush. 2021. Spanish embedded question island effects revisited: An experimental study. *Linguistics*.
Pearl, Lisa. 2022. Modeling syntactic acquisition. In Jon Sprouse, editor. *The Oxford handbook of experimental syntax*. Oxford: Oxford University Press.
Pearl, Lisa and Jon Sprouse. 2013. Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition* 20.23–68.
Perlmutter, David. 1972. Evidence for shadow pronouns in French relativization. In Gloria C. Peranteau, Paul M. Levi, and Judith N. Phares, editors. *The Chicago which hunt: Papers from the relative clause festival*. Chicago, IL: Chicago Linguistic Society. 73–105.
Pesetsky, David. 1998. Some optimality principles of sentence pronunciation. In Pilar Barbosa, Danny Fox, Martha McGinnis, and David Pesetsky, editors. *Is the best good enough? Optimality and competition in syntax*. Cambridge, MA: MIT Press. 337–383.
Rackowski, Andrea and Norvin Richards. 2005. Phase edge and extraction: a Tagalog case study. *Linguistic Inquiry* 36:565–599.
Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge, MA: MIT Press.
Rizzi, Luigi. 2001. On the position "Int(errogative)" in the left periphery of the clause. In Guglielmo Cinque and Giampaolo Salvi, editors. *Current studies in Italian syntax: Essays offered to Lorenzo Renzi*. New York: Elsevier. 287–296.
Rizzi, Luigi. 2004. Locality and left periphery. In *Structures and beyond*, ed. A. Belletti, 223–251. Oxford: Oxford University Press.
Rizzi, Luigi. 2013. Locality. *Lingua* 130.169–186.
Roeper, Tom and Jill de Villiers. 2011. The acquisition path for wh-questions. In Jill de Villiers and Tom Roeper, editoirs. *The Handbook of Generative Approaches to Language Acquisition*. Springer: New York. 189–246.
Ross, John Robert. 1967. *Constraints on variables in syntax*. Doctoral thesis. Cambridge, MA: Massachusetts Institute of Technology.
Rouder, Jeffrey N., and Richard D. Morey. 2012. Default Bayes Factors for Model Selection in Regression, Multivariate Behavioral Research 47.877–903.
Rouveret, Alain. 2011. Some issues in the theory of resumption: A perspective on early and recent research. In Alain Rouveret, editor. *Resumptive pronouns at the interfaces*, vol. 5. *Language faculty and beyond: Internal and external variation in linguistics*. Philadelphia: John Benjamins Publishing Company. 1–62.
Salzmann, Martin. 2017. *Reconstruction and resumption in indirect A'-dependencies: On the syntax of prolepsis and relativization in (Swiss) German and beyond*. Berlin: Walter de Gruyter.
Schütze, Carson. and Jon Sprouse. 2014. Judgment Data. In Robert Podesva and Devyani Sharma, editors. *Research methods in linguistics*. Cambridge: Cambridge University Press. 27–51.
Sells, Peter. 1984. Syntax and semantics of resumptive pronouns. Doctoral thesis. University of Massachusetts, Amherst.
Shlonsky, Ur. 1992. Resumptive pronouns as last resort. *Linguistic Inquiry* 23.443–468.

Shlonsky Ur. 2002. Constituent Questions in Palestinian Arabic. In Jamal Ouhalla and Ur Shlonsky, editors. *Themes in Arabic and Hebrew Syntax*. Studies in Natural Language and Linguistic Theory 53. Dordrecht: Springer. 37-159.

Sichel, Ivy. 2014. Resumptive pronouns and competition. *Linguistic Inquiry* 45.655–693.

Smolensky, Paul and Geraldine 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*. Doctoral thesis. University of Maryland, College Park.

Sprouse, Jon, Shin Fukuda, Hajime Ono, and Robert Kluender. 2011. Reverse island effects and the backward search for a licensor 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.

Sprouse, Jon and Diogo Almeida. 2017. Design sensitivity and statistical power in acceptability judgment experiments. *Glossa* 2.14

Stepanov, Artur, Manca Mušič and 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.

Szabolcsi, Anna and Frans Zwarts. 1997. Weak islands and an algebraic semantics for scope taking. In Anna Szabolcsi, editor. *Ways of scope taking*. Dordrecht: Kluwer. 217–262.

Szabolcsi, Anna. and Terje Lohndal. 2017. Strong vs. weak islands. In Martin Everaert and Henk van Riemsdijk *The Wiley Blackwell companion to syntax*. Berlin: Wiley.

Tellier, Christine. 1991. *Licensing theory and French parasitic gaps*. Dordrecht: Springer.

Tucker, Matthew, Ali Idrissi, Jon Sprouse and Diogo Almeida. 2019. Resumption ameliorates different islands differentially: Acceptability data from Modern Standard Arabic. In Amel Khalfaoui and Matthew A. Tucker, editors. *Perspectives on Arabic Linguistics XXX: Papers from the annual symposia on Arabic Linguistics*. Amsterdam: John Benjamins. 159–93.

Uriagereka, Juan. 1999. Multiple spell-out. In Samuel D. Epstein and Norbert Hornstein, editors. *Working minimalism*. Cambridge: MIT Press. 251–282.

de Vries, Mark. 2002. *The Syntax of Relativization*. Utrecht, The Netherlands: LOT.

Wahba, Wafaa. 1984. Wh-constructions in Egyptian Arabic. Doctoral thesis. University of Illinois, Urbana-Champaign.

Wanner, Eric. and Michael Maratsos: 1978. An ATN approach to sentence comprehension. In Morris Halle, Joan Bresnan, and George A. Miller, editors. *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press. 119–160.

Yano, Masataka. 2019. On the nature of the discourse effect on extraction in Japanese. *Glossa* 4.90.

Zaenen, A., E. Engdahl and J. M. Maling. 1981. Resumptive pronouns can be syntactically bound. *Linguistic Inquiry* 12:679–682.

Zehr, Jeremy. and Florian Schwarz. 2018. PennController for internet-based experiments (PCIBEX). https://doc.pcibex.net/.

Appendix

Table A1: Results of 2x2 linear mixed effects models for the full sample of participants, with 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	
			β	*p*	β	*p*	β	*p*	β	*p*
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

Table A2: Results of 2x2 linear mixed effects models for the sample of participants filtered by preferences for gaps and resumption, with 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	
			β	*p*	β	*p*	β	*p*	β	*p*
WH	gap	intercept	1.06	.001	1.02	.001	1.02	.001	1.07	.001
		distance	-0.30	.042	-0.14	.074	-0.20	.161	-0.15	.419
		structure	-0.20	.163	-0.03	.708	-0.11	.432	-0.21	.247
		dist x struct	-1.37	.001	-1.66	.001	-1.12	.001	-1.47	.001
	resumption	intercept	1.05	.001	0.98	.001	1.03	.001	1.07	.001
		distance	-0.56	.005	-0.43	.001	-0.70	.003	-0.49	.015
		structure	-0.23	.195	0.00	.978	-0.09	.685	-0.27	.159
		dist x struct	-0.84	.004	-1.53	.001	-1.22	.001	-1.12	.001
RC	gap	intercept	0.89	.001	0.83	.001	1.02	.001	0.96	.001
		distance	-0.25	.200	-0.01	.950	-0.18	.053	-0.45	.004
		structure	-0.01	.944	-0.03	.894	-0.15	.106	-0.03	.819
		dist x struct	-1.17	.002	-0.79	.011	0.05	.706	-1.12	.001
	resumption	intercept	0.68	.001	0.61	.001	0.85	.001	0.98	.001
		distance	0.18	.466	0.28	.037	0.10	.297	-0.03	.716
		structure	0.23	.354	0.28	.040	-0.11	.276	-0.09	.311
		dist x struct	-0.65	.071	-0.91	.001	-0.17	.229	-0.31	.017

Table A3: Bayes factors for the interaction term in 2x2 linear models for the full sample of participants, with DISTANCE x STRUCTURE for both dependency types and tail types (separately), using the BayesFactor package. Bayes factors were calculated with each of the three preset values for the width of the prior distribution in the BayesFactor package to explore the stability of the resulting Bayes factors.

dependency	tail	prior width	adjunct	np	wh	whether
WH	gap	medium	>100	>100	56.24	>100
		wide	>100	>100	52.72	>100
		ultrawide	>100	>100	51.64	>100
	resumption	medium	1.31	>100	0.27	4.99
		wide	1.15	>100	0.28	4.78
		ultrawide	1.37	>100	0.28	4.64
RC	gap	medium	>100	14.66	1.26	43.30
		wide	>100	13.18	1.27	43.45
		ultrawide	>100	14.90	1.23	44.06
	resumption	medium	0.85	>100	0.30	0.95
		wide	0.84	>100	0.29	0.90
		ultrawide	1.65	>100	0.30	1.00

Table A4: Results of pairwise linear mixed effects models comparing the two tail types (gap vs resumption) for the island | long condition using the lme4 and lmerTest packages, and treatment coding (with gap as a reference level).

dependency	model factor	adjunct		np		wh		whether	
		β	*p*	β	*p*	β	*p*	β	*p*
WH	intercept	-0.76	.001	-0.71	.001	-0.46	.001	-0.73	.001
	resumption	0.14	.401	-0.11	.234	-0.30	.046	-0.04	.745
RC	intercept	-0.63	.003	-0.27	.087	0.26	.070	-0.58	.001
	resumption	1.07	.001	0.46	.042	0.35	.067	1.10	.001

Table A5: Bayes factors for pairwise comparisons of the two tail types (gap vs resumption) for the island | long condition using the BayesFactor package. Bayes factors were calculated with each of the three preset values for the width of the prior distribution in the BayesFactor package to explore the stability of the resulting Bayes factors.

dependency	prior width	adjunct	np	wh	whether
WH	medium	>100	2.73	1.55	>100
	wide	>100	2.73	1.55	>100
	ultrawide	>100	2.73	1.55	>100
RC	medium	0.34	0.29	1.31	0.24
	wide	0.34	0.29	1.31	0.24
	ultrawide	0.34	0.29	1.31	0.24

Table A6: Results of 2x2 linear mixed effects models comparing the long conditions with STRUCTURE x RESUMPTION as fixed factors for each dependency type, using the lme4 and lmerTest packages, and treatment coding (with non-island and gap as reference levels).

dependency	model factor	adjunct		np		wh		whether	
		β	*p*	β	*p*	β	*p*	β	*p*
WH	intercept	0.56	.001	0.54	.001	0.38	.001	0.70	.001
	resumption	-1.31	.001	-1.26	.001	-0.84	.001	-1.44	.001
	structure	-0.63	.001	-0.56	.001	-0.88	.001	-0.90	.001
	struct x res.	0.77	.001	0.48	.035	0.59	.006	0.85	.001
RC	intercept	0.09	.567	0.11	.409	-0.04	.717	0.07	.536
	resumption	-0.72	.003	-0.38	.042	0.31	.047	-0.65	.001
	structure	0.52	.024	0.63	.002	0.91	.001	0.79	.001
	struct x res.	0.55	.084	-0.17	.484	-0.56	.013	0.32	.175

Table A7: Bayes factors for the interaction term in 2x2 linear models comparing the long conditions with STRUCTURE x RESUMPTION as fixed factors for each dependency type, using the BayesFactor package. Bayes factors were calculated with each of the three preset values for the width of the prior distribution in the BayesFactor package to explore the stability of the resulting Bayes factors.

dependency	prior width	adjunct	np	wh	whether
WH	medium	117.88	3.32	6.04	824.21
	wide	114.32	3.45	6.16	790.22
	ultrawide	117.77	3.55	6.21	848.48
RC	medium	5.09	0.29	3.63	0.70
	wide	4.74	0.31	3.72	0.68
	ultrawide	4.78	0.32	3.64	0.68