DOI: 10.1111/synt.12262

REGULAR ARTICLE

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

Rania Al-Aqarbeh[1] | **Jon Sprouse**[2]

[1]The Faculty of Arts, The Department of English Language and Literature, Mutah University, Karak, Jordan.

[2]New York University Abu Dhabi

Correspondence
Jon Sprouse, Program in Psychology, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, United Arab Emirates,
Email: jon.sprouse@nyu.edu

Abstract

This study brings 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. We report four auditory judgment studies exploring two dependency types and four island types with a gap or resumption in the tail of the dependency, yielding 16 distinct quantified effects. Our experiments identified two notable sources of variation: variation across dependency types in the sets of island effects that occur with gaps and variation across island types in amelioration by resumption. We discuss the challenges these results raise for four major classes of theories of island effects, and we 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 Jordanian Arabic. We also observe some evidence of individual variation in the availability of resumption across dependency types that could be explored in future studies.

KEYWORDS

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

1 | INTRODUCTION

In this article we report the results of four auditory judgment experiments, testing a total of 165 native speakers, that were 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, Szabolcsi & Zwarts 1993, 1997, Rizzi 2004), 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 between the set of island effects found with *wh* dependencies and the set found with relative-clause dependencies. This variation raises challenges for all four classes of theories of island effects; we 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., Choueiri 2017 and Salzmann 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 preference for gaps or resumption across *wh* dependencies and relative-clause dependencies in JA.

Taken as a whole, we believe that these experiments show that there is much to be learned about crosslinguistic variation in island effects, amelioration by resumption, and possibly even grammatical resumption through experimental-syntax studies of spoken varieties of Arabic.

We organize the article 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 or resumption in each dependency, and an exploratory analysis of individual variation in 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 for future studies of the four patterns of preferences for gaps or resumption. 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 in section 5 of the results.

2.1 | Jordanian Arabic

JA 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 and emails. As is common in Arabic-speaking countries, speakers of JA tend to use Modern Standard Arabic in formal written contexts like school assignments and 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 of Arabic (but see Tucker et al. 2019 for a formal-experimental study of island effects and resumption in Modern Standard Arabic). 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 2007, an investigation of reconstruction with resumption in JA. Malkawi & Guilliot report examples that contain adjunct-island structures (they use 'because' clauses) and *wh*-island structures (they use 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 islands 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 (in that the authors chose to use left-dislocation/topicalization dependencies for their examples), we nonetheless decided to treat this as an empirical question by testing both dependency types (with both tail types). We include analyses to account for (section 4.2) and quantify (section 4.4) preference for gaps or 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 explore the consequences of that pattern for the four dominant classes of theories: phase based, intervention based, information-structure based, and processing-complexity based. The four classes differ substantially regarding 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 and 2001, 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 head of the phase). It is easy to see how phases can give rise to something like island effects: there merely needs to be a relationship (perhaps 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 of island effects 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 & 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 should be driven by availability of a(n intermediate) landing site for movement at the edge of the phase in question. For example, Rackowski & Richards 2005 proposes 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 become closed to extraction when the final specifier of the phase is merged into it, 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 proposes tying 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 (such as in *wh* questions and relative clauses). This is because existence of a phase and availability of its edge as a landing site are both typically determined independently of dependency type. Capturing variation across dependencies would either require postulating different mechanisms underlying the dependencies, such as movement versus base generation, or require availability of edge features to somehow be determined by 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 is a *wh*-island violation like **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 intervention relationship? (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 them (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*) versus the existence 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 equivalents of *what* and *who*), since these are individual denoting but still show island effects.

Relativized Minimality defines interveners as items that carry the same syntactic feature(s) as the moved item. In the example **What did you wonder who invented __?*, both *what* and *who* might carry a + *wh* feature; therefore *who* is a potential intervener for the movement of *what*. Relativized Minimality 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, *what* c-commands *who*, and *who* c-commands the launching site of *what* (the position marked by "__"), so *who* is an intervener. The underlying insight of Relativized Minimality is that intervention blocks the dependency by causing it to violate 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. Relativized Minimality is thus part of the much broader trend in generative syntax of exploring metrics of computational efficiency in grammar.

Relativized Minimality 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 these classes to vary (e.g., Friedmann et al. 2009, Abels 2012, Belletti et al. 2012). For example, embedded *wh* questions can intervene in relative-clause dependencies, thus leading to *wh*-island effects, if the feature driving the relative-clause dependency is in the same equivalence class as the *+wh* feature in the embedded question. Similarly, relative-clause dependencies are insensitive to *wh* islands if the feature driving the relative-clause 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 Relativized Minimality and other intervention-based theories is that they tend to only cover the set of island effects associated with what are called *weak* or *selective* islands; these include *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 with bare *wh* words like *who*, and not to others, such as with 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 Nomi Erteschik-Shir's seminal 1973 dissertation. Erteschik-Shir proposed that the clauses or phrases that allow extraction are what she called semantically *dominant*, 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 either are 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* dependencies and modification through relative-clause dependencies. Nondominant clauses are presupposed or preuttered, 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 2006, 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.

In information-structure-based theories, variation in the pattern of island effects across languages reduces to a question of 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). She 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 leverages this approach to explain a difference in extractability between *wh* dependencies and relative-clause dependencies for subject islands in French and English.

One possible consequence of this approach is an expectation 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 island effects that can occur in that language, or it does not create a clash and therefore shows no island effects; we might not expect to observe two dependencies yielding two distinct patterns of island effects within the same language. However, this is something we do already observe in some languages (see section 2.3), so the all-or-nothing expectation 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., Wanner & Maratsos 1978, Deane 1991, Hawkins 1999, Hofmeister & Sag 2010, O'Grady 2010), 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. Robert Kluender's working-memory-capacity theory proposes that 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 in terms of variation in the processing dynamics of the island structures and dependencies in question (such as requiring fewer or more working-memory resources). The empirical challenge for processing-complexity-based theories of variation is to independently identify those processing dynamics (outside of island effects), such that the patterns of variation that we observe with island effects can be independently predicted. To our knowledge, there is relatively little work on processing-complexity-based theories of crosslinguistic 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 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 (see section 3 on this design). Though we intend this list to be fairly comprehensive, it may not be exhaustive, since 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.

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. We investigate both (simple) *wh* dependencies and relative-clause dependencies in JA. Third, we see that some of the most informative studies test four or more island types. We selected two strong islands (adjunct islands and complex-NP islands) and two weak islands (*wh* islands and 'whether' islands) 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 nonanimate;

TABLE 1 A summary of 15 formal-experimental studies of island effects (with gaps in the tails of the dependencies) using a factorial design. A plus sign (+) indicates an island effect, a minus sign (−) indicates no island effect, and gray shading indicates that the island type was not tested in the study. Parentheses around the plus sign indicate a subliminal island effect: an effect where the island-violating condition is rated at or above the midpoint of the scale.

Study	Language	Dependency	Adjunct	Complex NP	Subject	Relative clause	*Wh*	'Whether'
Almeida 2014	Brazilian Portuguese	*Wh* simple						(+)
Almeida 2014	Brazilian Portuguese	Topicalization						−
Lu et al. 2020	Chinese	*Wh* argument in situ				+		
Lu et al. 2020	Chinese	*Wh* adjunct 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	Relative clause	−	+	+			+
Sprouse et al. 2011	English	*Wh* argument in situ	−	−	−			−
Sprouse et al. 2016	Italian	*Wh* simple	+	+	+			+
Sprouse et al. 2016	Italian	Relative clause	+	+	−			+
Omaki et al. 2019	Japanese	NP scrambling			−			

TABLE 1 (Continued)

Study	Language	Dependency	Adjunct	Complex NP	Subject	Relative clause	*Wh*	'Whether'
Sprouse et al. 2011	Japanese	*Wh* argument in situ	−	−	−			−
Kim & Goodall 2016	Korean	*Wh* argument in situ	−					+
Kim & Goodall 2016	Korean	*Wh* argument in situ	−					+
Ko et al. 2019	Korean	NP scrambling	−			−	−	
Tucker et al. 2019	Modern Standard 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	+			+	+	

this could introduce additional differences between subject islands and other island types. We did not select relative-clause 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, subsequent works). 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, Salzmann 2017.)

To ground our first question—what is the pattern of amelioration in JA?—it may be helpful 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, Sells 1984), which do not allow resumption as a grammatical option in nonisland structures but may allow it in island structures as an exceptional device to ameliorate the island effect, and *grammatical-resumption languages*, which allow resumption as a grammatical option in nonisland structures and may also allow it as a grammatical option in island structures to ameliorate the island violation. Because JA is a grammatical-resumption language, we will focus exclusively on grammatical resumption here. (That said, there is a large and growing literature using formal experiments to explore island amelioration in intrusive-resumption languages, which we would like to cite here for readers interested in exploring that topic: 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 & Xiang 2016, Ackerman et al. 2018, A. Morgan & 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 show amelioration only with weak islands (e.g., *wh* islands and 'whether' islands; weak-set amelioration), and those that show no amelioration; see McCloskey 2017 and Salzmann 2017 for recent reviews, and see Szabolcsi & Lohndal 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 type, and *wh* islands, a weak-island type, as fully acceptable with resumptive pronouns. We will investigate this claim directly by testing two types of strong islands (adjunct islands and complex-NP islands) and two types of weak islands (*wh* islands 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 Modern Standard Arabic.

For our second question—what are the grammatical mechanisms underlying resumption in JA?—there are three classes of theories: base-generation theories, movement theories, and mixed theories (see McCloskey 2017 and Salzmann 2017 for reviews; see also Asudeh 2012 for a nontransformational 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, J. Morgan 1972, Perlmutter 1972, Givón 1973, Hayon 1973, Chomsky 1977, Bresnan & Grimshaw 1978, McCloskey 1979, Borer 1984, McCloskey 1990, subsequent works). Base-generation theories provide a straightforward explanation for languages that do not show any island effects with resumption (full-set amelioration), in that binding is generally assumed to be insensitive to islands, but they require additional assumptions to explain languages that do show island effects with resumption (weak-set amelioration, 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 in which the resumptive pronoun itself moves (possibly covertly: e.g., Demirdache 1991); there are approaches in which 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 in which the head and resumptive pronoun are merged together, with the head leaving the pronoun behind in a manner 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), in that movement is assumed to be island sensitive; but they require additional assumptions to explain languages that show no island effects (full-set amelioration) or a reduced set of island effects (e.g., weak-set amelioration).

Mixed theories allow resumption to be due to either movement or base generation plus binding (e.g., Demirdache 1991, Pesetsky 1998, Boeckx 2003, Müller 2014, 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, since learners must track resumption separately in each relevant syntactic context and must encounter evidence that indicates which mechanism underlies each separately tracked context.

As the review in Salzmann 2017 lays out, adjudicating among the three types of theories requires investigating a wide range of phenomena 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 but rather will discuss the consequences of the pattern of results that we observe for each of the three.

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

TABLE 2 The distribution of dependency types and island types across the four experiments.

	Dependency type	Island type	Participants
Experiment 1	*Wh*	*Wh*	40
	Relative clause	Complex NP	
Experiment 2	*Wh*	Complex NP	42
	Relative clause	*Wh*	
Experiment 3	*Wh*	'Whether'	43
	Relative clause	Adjunct	
Experiment 4	*Wh*	Adjunct	40
	Relative clause	'Whether'	

include multiple dependency types and multiple island types. Therefore, we decided to test both *wh* dependencies and relative-clause dependencies with 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 high-level summary of the design of the experiments; this will be followed by a detailed description, in six subsections, of each component of the experiments.

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, in all, 165 self-reported native speakers of JA from Mutah University in the Karak region of Jordan; these participants were split roughly evenly among the four experiments. Each participant completed their 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.

The experiments were auditory judgment experiments using a seven-point rating scale (from 1 to 7). The survey in each case was 42 items long: six practice items at the beginning (but not marked as such) followed by 12 experimental items and 24 filler items (eight 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 in each of the four experiments. 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 believe the factorial definition 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

TABLE 3 The 2 × 2 + 2 factorial design of the experiments, as illustrated by the examples in (1).

Example	Dependency	Structure	Tail
(1a)	Matrix	Nonisland	Gap
(1b)	Embedded	Nonisland	Gap
(1c)	Matrix	Island	Gap
(1d)	Embedded	Island	Gap
(1e)	Embedded	Nonisland	Resumption
(1f)	Embedded	Island	Resumption

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, Tail, manipulating the tail of the dependency (gap/resumptive pronoun). In principle, this should yield eight conditions in a 2 × 2 × 2 design (Dependency × Structure × 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 2 × 2 + 2 design (see also Tucker et al. 2019). We illustrate the full design with 'whether' islands and *wh* dependencies in (1); see table 3 for a key to the examples.

(1) A 2 × 2 + 2 factorial design for 'whether' islands with a *wh* dependency

a. Mi:n _ ga:l-∅ innu is-si:nama ʕaraðˤ-at filim hindi:?
who _ said-3SG.M that the-cinema presented-3SG.F film Hindi
'Who _ said that the cinema presented a Hindi film?'

b. Weiʃ ga:l-∅ Ra:mi: innu is-si:nama ʕaraðˤ-at _?
what said-3SG.M Rami that the-cinema presented-3SG.F
'What did Rami say that the cinema presented _?'

c. Mi:n _ saʔal-∅ iða is-si:nama ʕaraðˤ-at filim hindi:?
who asked-3SG.M whether the-cinema presented-3SG.F film Hindi
'Who _ asked whether the cinema presented a Hindi film?'

d. Weiʃ saʕal-∅ Ra:mi: iða is-si:nama ʕaraðˤ-at _?
what asked-3SG.M Rami whether the-cinema presented-3SG.F
Literally: 'What did Rami ask whether the cinema presented _?'

e. Weiʃ ga:l-∅ Ra:mi: innu is-si:nama ʕaraðˤ-at-**uh**?
what said.3SG.M Rami that the-cinema presented-3SG.F-it
Literally: 'What did Rami say that the cinema presented it?'

f. Weiʃ saʔal-∅ Ra:mi: iða is-si:nama ʕaraðˤ-at-**uh**?
what asked.3SG.M Rami whether the-cinema presented-3SG.F-it
Literally: 'What did Rami ask whether the cinema presented it?'

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, using this design we can look for these interactions for both gaps and resumption—to determine if an island effect that is present with gaps is eliminated with resumption. We can also look for both total amelioration, which will result in no interaction as in the left panel, and partial amelioration, which will result in a smaller interaction, as illustrated in the right panel.

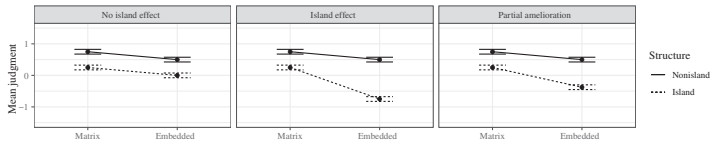

FIGURE 1 Possible outcomes for the factorial design.

In (2) we illustrate just the island-violating conditions for each of the eight combinations of islands and dependency types. See section 3.3 for more on the materials used in the experiments; the full list is available on the authors' websites.

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

a. Weiʃ saʔal-∅ Ra:mi: iða is-si:nama ʕaraðˁ-at __?
what asked-3SG.M Rami whether the-cinema presented-3SG.F
Literally: 'What did Rami ask whether the cinema presented __?'
= (1d)

b. Weiʃ saʔal-∅ Baha:ʔ leiʃ Ama:ni: ʃatˁab-at __?
what asked-3SG.M Baha' why Amani crashed-3SG.F
Literally: 'What did Baha' ask why Amani crashed __?'

c. Weiʃ naʃar-∅ Sa:mi: il-iʃaʕah innu Intisˁa:r iʃtar-at __?
what spread-3SG.M Sami the-rumor that Intisar bought-3SG.F
Literally: 'What did Sami spread the rumor that Intisar bought __?'

d. Weiʃ ziʕil-∅ Na:sˁir laʔinnu il-maktabih tˁabaʕ-at __?
what got-angry-3SG.M Naser because the-press printed-3SG.F
Literally: '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 __.
know.1SG the-manager who the-secretary asked-3SG.F whether the-board chose-3SG.F
Literally: 'I know the manager who the secretary asked whether the board chose __.'

f. ʃif-it il-asatðeh illi sˁa:ħab-ak saʔal-∅ leiʃ itˁ-tˁulla:b waddaʕ-u __.
saw.1SG the-teachers who friend-your asked.3SG.M why the-students said.goodbye.to-3PL.M
Literally: '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-∅ __
know.1SG the-secretary who the-clerk heard-3SG.F the-rumor that the-principal married-3SG.M
Literally: '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-∅ __.
know.1SG the-manager who the-girl felt-happy-3SG.F because neighbor-our invited-3SG.M
Literally: '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

TABLE 4 Key to the examples in (2) illustrating the island-violating conditions.

Example	Dependency type	Island type
(2a)	*Wh*	'Whether'
(2b)	*Wh*	*Wh*
(2c)	*Wh*	Complex NP
(2d)	*Wh*	Adjunct
(2e)	Relative clause	'Whether'
(2f)	Relative clause	*Wh*
(2 g)	Relative clause	Complex NP
(2 h)	Relative clause	Adjunct

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 classifies clitics as *weak resumptives*, which contrast with *strong resumptives* such as nonclitic pronouns and epithets. Could our choice of weak resumptives have an impact on the results? Both types of resumptives lead to amelioration of island effects. Indeed, the examples that Malkawi & Guilliot provide for amelioration by resumption in adjunct and *wh* islands involve weak resumptives. Weak and strong resumptives differ in two ways: only weak resumptives are possible in nonsubject position (Aoun et al. 2010), and only weak resumptives allow reconstruction within an island (Malkawi & Guilliot 2007). The strong–weak distinction in resumptives 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 nonsubject 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 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. Modern Standard Arabic 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 of the participants applying Modern Standard Arabic grammatical rules to their judgments and ideally to maximize the possibility of their engaging 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 Modern Standard Arabic to influence their judgments of JA, we do believe that the auditory presentation was successful in eliciting judgments that are distinct from

Modern Standard Arabic, because the pattern of results that we observed differs from the results in Tucker et al. 2019 for Modern Standard Arabic.

3.3 | Materials and survey construction

Each participant completed a survey that consisted of 42 items: six items at the beginning of the survey to help familiarize them 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 one token of each of the six conditions for each of the two island–dependency combinations in the 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 six 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 eight sentence types expected to receive low ratings, eight expected to receive ratings in the middle of the scale, and eight 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 50 ms 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 type, island type, and tail type), in the interest of full transparency (and reproducibility), the full set of recorded materials is 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 (Zehr & Schwarz 2018, Drummond 2019). Each sentence received its own presentation screen with a scale of 1 *ʕeir maqbu:l ɪʔtˤla:qan* (written in the Arabic abjad) (totally unacceptable) to 7 *maqbu:l tama:man* (perfectly acceptable). 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 in Sprouse & Almeida 2017 and Marty et al. 2020 of statistical power for seven-point acceptability-judgment tasks. 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 sample 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), 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 is in the supporting information for this article. 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, using the BayesFactor package (Morey & Rouder 2018) we derived Bayes factors 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 three times 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 three times more likely under the null hypothesis than under 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 inconclusive.

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 (Jeffreys 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, indicating 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 not then be possible 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 Jeffrey N. Rouder, Richard D. Morey, and colleagues over the past several years (the specific priors for the regression models implemented here are based on the work in Liang et al. 2008, as implemented in Rouder & Morey 2012). To

explore the stability of the Bayes factors for our results, in the supporting materials 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 this article, 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/embedded) with resumption and not by changes in the other three conditions. We test this using a pairwise comparison between the island-violating condition (island/embedded) for the gap tail type and the same condition for the resumption tail type. 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 $2 \times 2 \times 2$ design (Dependency $\times$ Structure $\times$ Tail); however, given that the matrix conditions are identical for both tail types, this more properly reduces to a 2×2 interaction of Structure $\times$ Tail in the embedded conditions. We report both of these analyses in full in the supporting materials, and we report the results as needed in the text of the article.

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 reanalysis 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 performed a *z*-score transformation on 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 (see, 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 midpoint (0) of the *z*-score scale corresponds nearly perfectly to the absolute midpoint 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 relative-clause dependencies in the bottom half. For 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

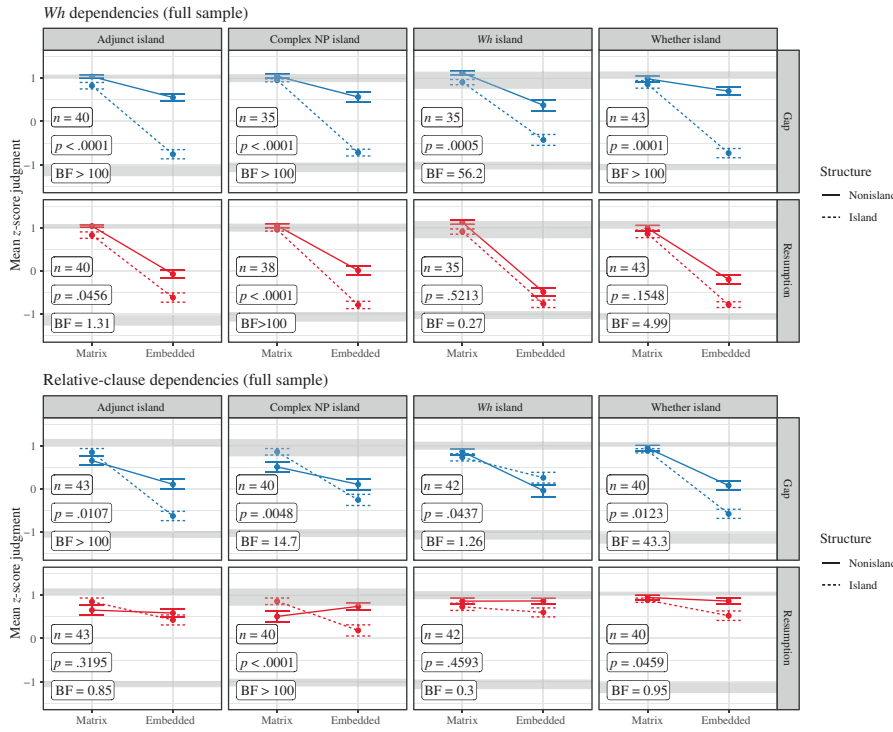

FIGURE 2 Interaction plots for island effects and amelioration by resumption. Points are condition means. Error bars represent estimated standard error. For space reasons, *p* values are rounded to a floor of 0.0001 and Bayes factors are rounded to a ceiling of 100. The counts represent the number of observations for the island/embedded condition. The horizontal gray bars represent the range of mean ratings for the eight unacceptable and eight acceptable filler types. [Colour figure can be viewed at wileyonlinelibrary.com]

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/embedded 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/embedded) 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 both by 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 for *wh* dependencies (bottom row) do not show amelioration. The amelioration pattern would be an increase in the acceptability of the island/embedded condition, as compared to the gap version, that eliminates the superadditive interaction. But that is not what we see. The island/embedded condition is not statistically different for adjunct, complex-NP, and 'whether' islands (*p* values above .05 and Bayes factors near or below 0.33; see tables A4 and A5 in the supporting materials) and is statistically inconclusive for *wh* islands ($p = 0.046$ 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 nonisland/embedded condition is rated lower with resumption than

with gaps: near the midpoint (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 supporting materials). 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.

For relative-clause 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 nonmonotonic interaction (i.e., crossing lines) because the island/matrix condition is rated higher than the nonisland/matrix condition. This is still an island effect in that the island/embedded 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/matrix 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/embedded condition is rated relatively high (−0.25) compared to the other island effects (adjunct islands are −0.63 and 'whether' islands are −0.58). Finally, *wh* islands show an interaction in a direction that is opposite to the direction predicted—the island/embedded condition is rated higher than the nonisland/embedded condition. This is not an island effect. It is also not clear if this interaction is robust, since 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 the hypothesis that there is a reverse interaction and 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 for relative-clause dependencies (bottom row), we see a form of amelioration for all three of the island effects that occurred with gaps. We will discuss each island type in turn.

Adjunct islands show a trend toward classic amelioration: the visual pattern shows a large increase in acceptability for the island/embedded 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 effect—that is, a superadditive interaction with the island/embedded condition rated above the midpoint 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, nonmonotonic form. However, both the nonisland/embedded and island/embedded conditions have shifted higher in acceptability compared to the corresponding gap conditions. This suggests a general preference for resumption over gaps in relative-clause dependencies. We will explore this further in sections 4.2 and 4.3. Given the shift into the midrange of the acceptability scale, one might wonder whether the complex-NP island with

TABLE 5 Summary of the results of the experiments.

		Adjunct island	Complex-NP island	*Wh* island	'Whether' island
Wh dependencies	Gap	Island effect	Island effect	Island effect	Island effect
	Resumption	Island effect	Island effect	Island effect	Island effect
Relative-clause dependencies	Gap	Island effect	Island effect	None	Island effect
	Resumption	Amelioration	Island effect	None	Amelioration

resumption is a subliminal island effect. We would argue that it is not a subliminal island effect based on its size: .9 on the *z*-score scale. This is two to three times larger than the subliminal island effects previously reported in the literature (Almeida 2014, Kush et al. 2018, Keshev & Meltzer-Asscher 2019) and roughly the same size as the island effect with gaps. We will therefore treat what we see with complex-NP islands 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 0.05, but the Bayes factor is 0.99, which suggests that the data is equally likely under both hypotheses. Furthermore, the island/embedded 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 effect (though our experiment was not sensitive enough to detect effect sizes that small), and complex-NP islands do not show amelioration (but rather a nonmonotonic 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 effects with resumption is to be expected.

Table 5 summarizes the results of the experiments. We have categorized the effects as island effects only when both statistical analyses (null-hypothesis testing and Bayes factors) converge, and as "none" or "amelioration" otherwise.

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 relative-clause dependencies. This is a common pattern in grammatical-resumption languages (see Salzmann 2017 and Choueiri 2017 for reviews). However, for our full-sample analysis such preferences are a potential confound that could add noise to the island analyses in figure 2. In figure 3, we reanalyze the data in an attempt to eliminate this potential confound. For the gap rows of the plot, we only include participants who rated gaps in the nonisland/embedded condition above the midpoint (0) on the *z*-score scale. For the resumption rows of the plot, we only include participants who rated resumption in the nonisland/embedded condition above the midpoint. Filtering based on the ratings of the (by hypothesis, grammatical) nonisland/embedded 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

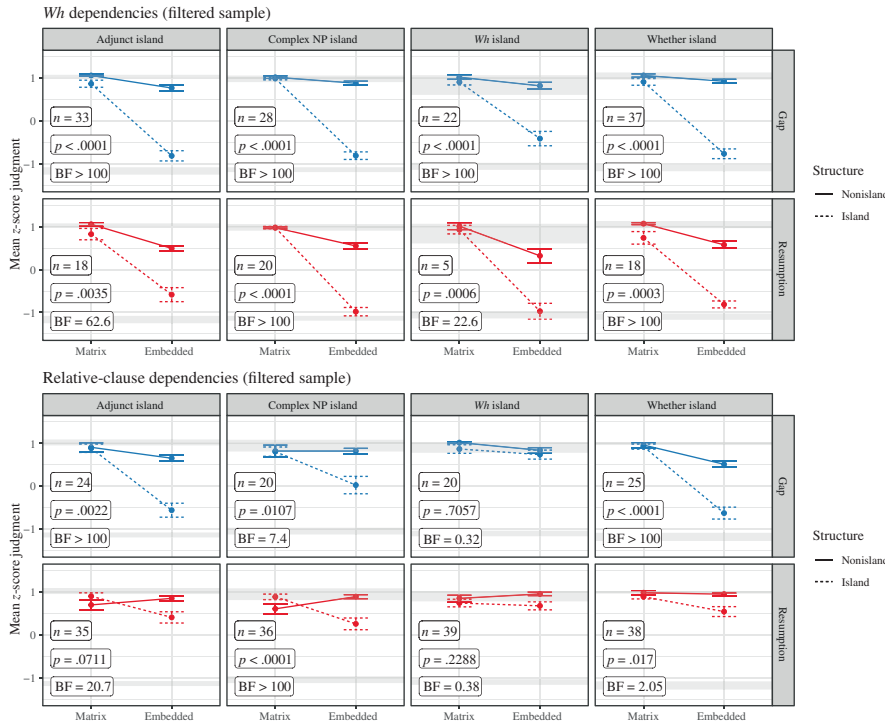

FIGURE 3 Interaction plots for island effects and amelioration by resumption, allowing for variation in preferences for gaps and resumption. Details are the same as figure 2, except that gap rows only include participants who accept gaps in the nonisland/embedded condition, and resumption rows only include participants who accept resumption in the nonisland/embedded condition. [Colour figure can be viewed at wileyonlinelibrary.com]

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 absences 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*-dependency gap conditions, we still see large island effects for all four island types. The only change appears to be a higher rating for the nonisland/embedded 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 *wh*-dependency 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 in figure 2, figure 3 shows classic island effects with gaps and no amelioration by resumption for this dependency type. For relative-clause-dependency gap conditions, we see large classic island effects for adjunct and 'whether' islands in figure 3. These are larger than those seen in figure 2, with the increase driven by an increase in the nonisland/embedded 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 nonmonotonic pattern in figure 2. The mean rating of the island/embedded

condition is still near the midpoint of the scale, much as in 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 relative-clause dependency resumption conditions, we see a subliminal island effect for adjunct islands, a nonmonotonic interaction for complex-NP islands that is still a subliminal island effect (but only based on five participants), no island effect for *wh* islands, and either no island effect or a very small subliminal island effect for 'whether' islands. In short, for relative-clause dependencies we see amelioration patterns similar to those in figure 2.

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 island effects (which are, by definition, very small), 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 that the possibility that adjunct and 'whether' islands with resumptive pronouns in JA yield subliminal island effects is a hypothesis to be explored with a dedicated study in the future (specifically one with higher statistical power for small effect sizes and perhaps with 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 or 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 explore 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 (see, e.g., James et al. 2018, Staub 2021); it is currently unclear if this reliability issue extends to (offline) acceptability judgments.

Each scatterplot in figure 4 plots, along the *x* axis, each participant's rating of the gap version of a condition against, along the *y* axis, that same participant's rating of the resumption version of the same condition. The figure is divided as before between *wh* dependencies in the top panel and relative-clause dependencies on the bottom. The columns represent island types, and the rows represent the specific condition: either the grammatical nonisland/embedded condition or the potentially ungrammatical island/embedded condition. We divided each plot into four quadrants based on the midpoint of the *z*-score scale (0): a point in the top right quadrant represents a participant who rated both the gap and resumption versions of the condition in the upper (positive) half of the *z*-score scale and who 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 who therefore appears to prefer resumption (labeled "resumption"); the bottom left quadrant represents participants who rated both conditions low and therefore appear to allow neither tail type (labeled "neither"); and the bottom right represents participants who rated gaps high and resumption low and therefore appear to prefer gaps (labeled "gap"). We have added two features to make the plot more informative: unique colors for the points in each of the

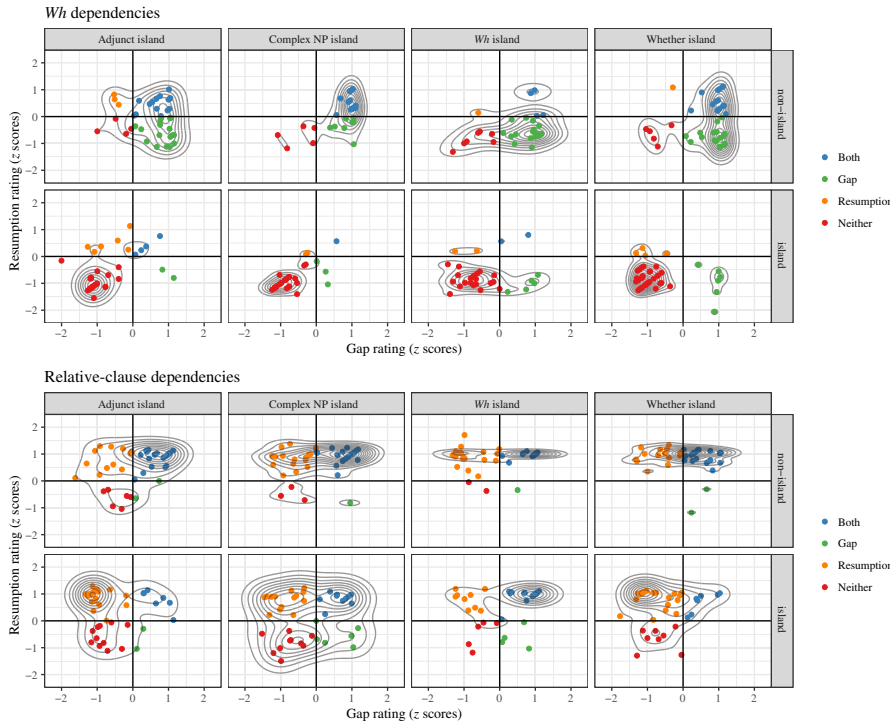

FIGURE 4 Scatterplots identifying preferences for gaps and resumption for each participant. The lines represent two-dimensional (joint) probability-density estimates. The points are colored according to the quadrants defined by the midpoint (0) of the *z*-score scales. [Colour figure can be viewed at wileyonlinelibrary.com]

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. Much as in 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 in section 4.2, that the potential for noise in these (single) ratings means that participants nearer to the dividing lines may be misclassified.

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 nonisland 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 6 provides those counts, setting aside the "neither" classification (red dots) as participants who may have provided noisy judgments in these trials.

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 either allow both gaps and resumption (blue dots, 51 total participants) or prefer gaps (green dots, 58 participants). Very few participants preferred resumption (orange dots, five participants). For relative-clause dependencies, these preferences are captured in the column labeled "total." We see that most participants tended to either allow both gaps and resumption (blue dots, 71 total participants) or prefer resumption (orange dots). Very few participants accepted only gaps (green dots, four

TABLE 6 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**
Relative-clause preference	Both	34	35	2	71
	Gap	2	2	0	4
	Resumption	15	21	3	39
	Total	51	58	5	114

participants). 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 nine interior cells. We have shaded the four 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 three times 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. In section 5.3 we discuss in more detail the potential research questions raised by our observation that there are four preference patterns 3.

We focus next on the bottom rows of the two panels in figure 4, which show participants' ratings of the island/embedded 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, bottom row), we see that 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 relative-clause dependencies (bottom panel, bottom row), 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 shows 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 that our participants displayed regarding gaps versus resumption. We found that for *wh* dependencies, speakers either preferred gaps only or allowed both gaps and resumption (never preferring only resumption); and we found that for relative-clause dependencies, speakers either preferred resumption only or allowed both gaps and resumption (never preferring only gaps). This results in four speaker patterns formed by fully crossing the two observed *wh*-dependency preferences with the two observed relative-clause-dependency preferences. Here we discuss the theoretical consequences of these results for island effects, amelioration, and theories of the distribution of gaps and resumption.

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 relative-clause 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 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 this variation), and we know of no previous study that observed the simultaneous absence of *wh*-island effects and presence of 'whether'-island effects for a single dependency (the closest is Pañeda & Kush 2021, which found the opposite pattern in Spanish—*wh*-island effects 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 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 them using less-grounded grammatical constraints (similar to previous approaches involving, for example, individual island constraints, Subjacency, or the Condition on Extraction Domains). 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 island types that can be explained by phase impenetrability without additional assumptions are *wh* islands and 'whether' islands (not unlike the Barriers approach of Chomsky 1986; 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 attributed 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 relative-clause dependencies could suggest that 'whether' occupies a position that is to the left of a non-phase-edge position occupied by 'why' (perhaps the specifier of an interrogative phrase, versus the specifier of a focus phrase), as in Rizzi 2001. Similarly, the absence of complex-NP island effects with relative-clause dependencies for some participants could be analyzed in parallel to the absence of complex-NP island effects in some languages (e.g., scrambling in 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 distinction for islands, perhaps by positing that strong islands are explained by phases while weak islands are explained by intervention-based theories like Relativized Minimality. 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, since 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, since 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 Relativized Minimality are not intended to explain variation in strong islands like that 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 between *wh* islands and 'whether' islands. Given that relative-clause dependencies show 'whether' islands but not *wh* islands in JA, the challenge is to either find a feature that is shared by relative-clause heads and 'whether' clauses but not 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 Relativized Minimality analyses, such as +Q for questions and + focus for focus constructions, would potentially predict the opposite pattern: if 'whether' clauses are +Q and 'why' questions are both +Q and + focus (under an analysis in which *wh* words occupy a focus position) and relativization involves +focus, then relative-clause 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 relative-clause 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 1973 investigation of variation between Danish and English, Nomi Erteschik-Shir proposed 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 crosslinguistic 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, since 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 2022 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 and 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 (see, e.g., Pearl & Sprouse 2013), in both JA and other languages that show variation in island effects (see, 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 relative-clause dependencies showing 'whether' island effects but not *wh* island effects, appears to run contrary to the processing-complexity findings in Kluender & Kutas 1993. In an event-related-potential study, Kluender & Kutas report larger N400-like effects for embedded *wh* questions than for 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 island effects were to be absent, it would be the 'whether'-island effect, 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 crosslinguistic 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 belief is 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 nonisland structures (see Demirdache 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 nonisland structures, and even these participants show no amelioration by resumption in island structures. This suggests that 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, since 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 relative-clause 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 crosslinguistic 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). It also accords well with the literature on reconstruction in varieties of Arabic, which postulates mixed sources 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 available input 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 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 relative-clause dependencies with resumption will be quite small in

TABLE 7 Preference for gaps or resumption in four varieties of Arabic, adapted from Choueiri 2017 and including the results of the current study.

	***Wh* dependencies**	**Relative-clause dependencies**
Egyptian Arabic	Resumption	Resumption
Lebanese Arabic	Both	Resumption
Moroccan Arabic	Gap	Both
JA	Gap or both	Resumption or both

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 preferences for gaps and resumption

Though it was not an initial goal of these experiments, our results also revealed both expected and unexpected variation in preferences for gaps and resumption in JA. Our analyses replicated the observation frequently made with respect to different varieties of Arabic that, of the two dependency types, *wh* dependencies are more likely to reveal a gap-only pattern and relative-clause dependencies are more likely to reveal a resumption-only pattern (see, for example the review in Choueiri 2017). We also found evidence—with the caveat that our analysis of preferences is only exploratory—that some speakers of JA accept both gaps and resumption for *wh* dependencies and that some accept both for relative-clause 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 7 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.

To our knowledge there is no overarching theory of the distribution of preferences for gaps and resumption. The distribution of these preferences is primarily presented as a descriptive generalization in the literature; therefore we cannot comment on the theoretical consequences should our observations of variability within JA be corroborated by future experiments. However, these results do suggest that there may be more variability with respect to resumption than has previously been reported, both for island amelioration and for preferences in nonisland 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 from 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 Modern Standard Arabic 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 *mi:n* ('who') or d-linked *wh* phrases. Citing 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 JA, 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 island effects that occur with gaps, variation across island types in the presence of amelioration by resumption, and potentially variation across participants in preferences for gaps versus resumption. The variation across dependency types suggests that all 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 mixed sources for resumption dependencies in JA: both movement and base generation. The variation across participants also suggests that future studies of resumption in JA and perhaps in other grammatical-resumption languages will benefit from experimental designs that quantify individual variation, both in preference for gaps or resumption in nonisland 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, in terms of the impact that their patterns of variation will have both on theories and on the types of studies that the field explores moving forward.

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.

DATA AVAILABILITY STATEMENT

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

ORCID

Jon Sprouse https://orcid.org/0000-0002-8875-3204

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. 2012. The Italian left periphery: A view from locality. *Linguistic Inquiry* 43.2.229–254.

Ackerman, Lauren, Michael Frazier, and Masaya Yoshida. 2018. Resumptive pronouns can ameliorate illicit island extractions. *Linguistic Inquiry* 49.4.847–859.

Alexopoulou, Theodora, and Frank Keller. 2007. Locality, cyclicity, and resumption: At the interface between grammar and the human sentence processor. *Language* 83.1.110–160.

Almeida, Diogo. 2014. Subliminal wh-islands in Brazilian Portuguese and the consequences for Syntactic theory. *Revista da Abralin* 13.2.55–93.

Ambridge, Ben, and Adele Goldberg. 2008. The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics* 19.3.357–389.

Ambridge, Ben, Julian M. Pine, and Elena V. M. Lieven. 2014. Child language acquisition: Why universal grammar doesn't help. *Language* 90.3.e53–e90.

Aoun, Joseph, Elabbas Benmamoun, and Lina Choueiri. 2010. *The syntax of Arabic*. Cambridge, the United Kingdom: 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, the Netherlands: John Benjamins Publishing Company. 7–26.

Aoun, Joseph, Lina Choueiri, and Norbert Hornstein. 2001. Resumption, movement, and derivational economy. *Linguistic Inquiry* 32.3.371–403.

Asudeh, Ash. 2012. *The logic of pronominal resumption*. Oxford, the United Kingdom: 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.

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, Germany: Akademie Verlag. 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* 122.10.1053–1069.

Beltrama, Andrea, and Ming Xiang. 2016. Unacceptable but comprehensible: The facilitation effect of resumptive pronouns. *Glossa* 1.29.

Bianchi, Valentina. 2004. Resumptive relatives and LF chains. In Luigi Rizzi, editor. *The structure of CP and IP. (The Cartography of Syntactic Structures 2.)* Oxford, the United Kingdom: Oxford University Press. 76–114.

Boeckx, Cedric. 2003. *Islands and chains: Resumption as stranding*. Amsterdam, the Netherlands: John Benjamins Publishing Company.

Boeckx, Cedric. 2013. *Syntactic islands*. Cambridge, the United Kingdom: 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* 5.9–10.341–345.

Borer, Hagit. 1984. Restrictive relatives in Modern Hebrew. *Natural Language and Linguistic Theory* 2.2.209–261.

Bresnan, Joan, and Jane Grimshaw. 1978. The syntax of free relatives in English. *Linguistic Inquiry* 9.3.331–391.

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 the North East Linguistic Society 13*. Amherst, MA: Graduate Linguistics Student Association, University of Massachusetts Amherst. 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, NY: 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, NY: 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:** Roger Martin, David Michaels, Juan Cambridge, the United Kingdomka, and Samuel Jay Keyser, editors. *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*. Cambridge, MA: MIT Press. 89–155.

Chomsky, Noam. 2001. Derivation by phase. **In:** Michael Kenstowicz, editor. *Ken Hale: A life in language*. Cambridge, MA: 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, the United Kingdom: 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* 42.1.51–70.

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

Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences*, second edition. New York, NY: Academic Press.

Deane, Paul D. 1991. Limits to attention: A cognitive theory of island phenomena. *Cognitive Linguistics* 2.1.1–63.

Demirdache, Hamida. 1991. *Resumptive chains in restrictive relatives, appositives and dislocation structures*. Doctoral thesis. Cambridge, MA: Massachusetts Institute of Technology.

de Vries, Mark. 2002. *The syntax of relativization*. Utrecht, the Netherlands: LOT Publications.

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, MA: Graduate Linguistics Student Association, University of Massachusetts Amherst. 157–192.

Donati, Caterina, and Carlo Cecchetto. 2011. Relabeling heads: A unified account for relativization structures. *Linguistic Inquiry* 42.4.519–560.

Drummond, Alex. 2019. Ibex farm. *Software.* http://spellout.net/ibexfarm.

Engdahl, Elisabet. 1985. Parasitic gaps, resumptive pronouns, and subject extractions. *Linguistics* 23.1.3–44.

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. *Proceedings of IATL 26*. Jerusalem, Israel: Israel Association for Theoretical Linguistics. https://www.iatl.org.il/?page_id=46.

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, NJ: Lawrence Erlbaum Associates, Publishers. 263–278.

Frazier, Lyn, and Charles Clifton Jr. 2002. Processing "d-linked" phrases. *Journal of Psycholinguistic Research* 31.6.633–660.

Friedmann, Naama, Adriana Belletti, and Luigi Rizzi. 2009. Relativized relatives: Types of intervention in the acquisition of A′-dependencies. *Lingua* 119.1.67–88.

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, IL: Chicago Linguistic Society. 135–146.

Goldberg, Adele. 2006. *Constructions at work: The nature of generalization in language*. London, the United Kingdom: Oxford University Press.

Haig, John H. 1976. Shadow pronoun deletion in Japanese. *Linguistic Inquiry* 7.2.363–371.

Hawkins, John A. 1999. Processing complexity and filler-gap dependencies across grammars. *Language* 75.2.244–285.

Hayon, Yehie. 1973. *Relativization in Hebrew*. The Hague, the Netherlands: Mouton.

Heestand, Dustin, Ming Xiang, and Maria Polinsky. 2011. Resumption still does not rescue islands. *Linguistic Inquiry* 42.1.138–152.

Hladnik, Marko. 2015. *Mind the gap: Resumption in Slavic relative clauses*. Utrecht, the Netherlands: LOT Publications.

Hofmeister, Philip, and Ivan A. Sag. 2010. Cognitive constraints and island effects. *Language* 86.2.366–415.

Hornstein, Norbert. 2001. *Move! A minimalist theory of construal*. Malden, MA: Blackwell Publishers.

Huang, C.-T. James. 1982. *Logical relations in Chinese and the theory of grammar*. Doctoral thesis. Cambridge, MA: Massachusetts Institute of Technology.

James, Ariel N., Scott H. Fraundorf, Eun-Kyung Lee, and Duane G. Watson. 2018. Individual differences in syntactic processing: Is there evidence for reader-text interactions? *Journal of Memory and Language* 102.155–181.

Jeffreys, Harold. 1961. *The theory of probability*, third edition. Oxford, the United Kingdom: Oxford University Press.

Keffala, Bethany. 2013. Resumption and gaps in English relative clauses: Relative acceptability creates an illusion of "saving." **In:** Chundra Cathcart, I-Hsuan Chen, Greg Finley, Shinae Kang, Clare S. Sandy, and Elise Stickles, editors. *Proceedings of the thirty-seventh annual meeting of the Berkeley Linguistics Society: General session and parasession on language, gender, and sexuality*. Berkeley, CA: Berkeley Linguistics Society. 140–154.

Keshev, Mayaan, and Aya Meltzer-Asscher. 2017. Active dependency formation in islands: How grammatical resumption affects sentence processing. *Language* 93.3.549–568.

Keshev, Mayaan, and Aya Meltzer-Asscher. 2019. A processing-based account of subliminal wh-island effects. *Natural Language and Linguistic Theory* 37.2.621–657.

Kim, Boyoung, and Grant Goodall. 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, Germany: Leipzig University. 329–346.

Klein, Timo. 2016. *Patterns of resumption—Towards a derivational account*. Doctoral thesis. Leipzig, Germany: Leipzig University.

Kluender, Robert. 1991. *Cognitive constraints 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. **In:** Peter Culicover and Louise McNally, editors. *The limits of syntax. (Syntax and Semantics 29.)* San Diego, CA: Academic Press. 241–279.

Kluender, Robert. 2004. Are subject islands subject to a processing account? **In:** Vineeta Chand, Ann Kelleher, Angelo Rodríguez, and Benjamin Schmeiser, editors. *Proceedings of the 23rd West Coast Conference on Formal Linguistics*. 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, the United Kingdom: Cambridge University Press. 186–207.

Kluender, Robert, and Marta Kutas. 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8.4.573–633.

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* 37.287–323.

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 seventeenth regional meeting of the Chicago Linguistic Society*. Chicago, IL: 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* 36.3.743–779.

Kush, Dave, Terje Lohndal, and Jon Sprouse. 2019. On the island sensitivity of topicalization in Norwegian: An experimental investigation. *Language* 95.3.393–420.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82.13.

Lasnik, Howard, and Mamoru Saito. 1984. On the nature of proper government. *Linguistic Inquiry* 15.2.235–289.

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* 51.3.611–623.

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, the Netherlands: John Benjamins Publishing Company. 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* 5.72.

McCloskey, Jim. 1979. *Transformational syntax and model theoretic semantics: A case-study in Modern Irish*. Dordrecht, the Netherlands: 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. (Syntax and Semantics 23.)* New York, NY: Academic Press. 199–248.

McCloskey, Jim. 2017. Resumption. **In:** Martin Everaert and Henk van Riemsdijk, editors. *The Blackwell companion to syntax*. Malden, MA: Blackwell Publishers. 84–117.

McDaniel, Dana, and Wayne Cowart. 1999. Experimental evidence for a minimalist account of English resumptive pronouns. *Cognition* 70.2.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. *Software.* http://CRAN.R-project.org/package=BayesFactor.

Morgan, Adam, and Matthew Wagers. 2018. English resumptive pronouns are more common where gaps are less acceptable. *Linguistic Inquiry* 49.4.861–876.

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.

Müller, Gereon. 2010. On deriving CED effects from the PIC. *Linguistic Inquiry* 41.1.35–82.

Müller, Gereon. 2014. *Syntactic buffers: A local-derivational approach to improper movement, remnant movement, and resumptive movement*. Leipzig, Germany: Institute of Linguistics, University of Leipzig.

Müller, Gereon. 2021. Constraints on grammatical dependencies. **In:** Nicholas Allott, Terje Lohndal, and Georges Rey, editors. *A companion to Chomsky*. Hoboken, NJ: Wiley Blackwell. 190–209.

Nouhi, Youssef. 1996. *Wh-constructions in Moroccan Arabic*. Master's thesis. Ottawa, Canada: University of Ottawa.

Nyvad, Anne Mette, Ken Ramshøj Christensen, and Sten Vikner. 2017. CP-Recursion in Danish: A cP/CP-analysis. *Linguistic Review* 34.3.449–477.

O'Grady, Wayne. 2010. An emergentist approach to syntax. **In:** Bernd Heine and Heiko Narrog, editors. *The Oxford handbook of linguistic analysis*. Oxford, the United Kingdom: Oxford University Press. 257–283.

Omaki, Akira, Shin Fukuda, Chizuru Nakao, and Maria Polinsky. 2019. Subextraction in Japanese and subject-object symmetry. *Natural Language and Linguistic Theory* 38.2.627–669.

Omaki, Akira, and Chizuru Nakao. 2010. Does English resumption really help to repair island violations? *Snippets* 21.11–12.

Pañeda, Claudia, and Dave Kush. 2021. Spanish embedded question island effects revisited: An experimental study. *Linguistics* 60.2.463–504.

Pañeda, Claudia, Sol Lago, Elena Vares, João Veríssimo, and Claudia Felser. 2020. Island effects in Spanish comprehension. *Glossa* 5.21.

Pearl, Lisa. 2022. Modeling syntactic acquisition. **In:** Jon Sprouse, editor. *The Oxford handbook of experimental syntax*. Oxford, the United Kingdom: 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* 20.1.23–68.

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.

Poulsen Mads. 2008. Acceptability and processing of long-distance dependencies in Danish. *Nordic Journal of Linguistics* 31.1.73–107.

Rackowski, Andrea, and Norvin Richards. 2005. Phase edge and extraction: A Tagalog case study. *Linguistic Inquiry* 36.4.565–599.

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, NY: Elsevier. 287–296.

Rizzi, Luigi. 2004. Locality and left periphery. **In:** Adriana Belletti, editor. *Structures and beyond*. Oxford, the United Kingdom: Oxford University Press. 223–251.
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, editors. *The handbook of generative approaches to language acquisition*. New York, NY: Springer. 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* 47.6.877–903.
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. (Language Faculty and Beyond: Internal and External Variation in Linguistics 5.)* Philadelphia, PA: 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, Germany: Walter de Gruyter.
Schütze, Carson, and Jon Sprouse. 2014. Judgment data. **In:** Robert Podesva and Devyani Sharma, editors. *Research methods in linguistics*. Cambridge, the United Kingdom: Cambridge University Press. 27–51.
Sells, Peter. 1984. *Syntax and semantics of resumptive pronouns*. Doctoral thesis. Amherst, MA: University of Massachusetts Amherst.
Shlonsky, Ur. 2002. Constituent questions in Palestinian Arabic. **In:** Jamal Ouhalla and Ur Shlonsky, editors. *Themes in Arabic and Hebrew syntax*. Dordrecht, the Netherlands: Springer. 37–159.
Sichel, Ivy. 2014. Resumptive pronouns and competition. *Linguistic Inquiry* 45.4.655–693.
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* 17.1.24.215–224.
Sprouse, Jon. 2007. *A program for experimental syntax*. Doctoral thesis. College Park, MD: University of Maryland, College Park.
Sprouse, Jon, and Diogo Almeida. 2017. Design sensitivity and statistical power in acceptability judgment experiments. *Glossa* 2.14.
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* 34.1.307–344.
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* 14.2.179–203.
Sprouse, Jon, Matthew Wagers, and Colin Phillips. 2012. A test of the relation between working memory capacity and syntactic island effects. *Language* 88.1.82–123.
Staub, Adrian. 2021. How reliable are individual differences in eye movements in reading? *Journal of Memory and Language* 116.104190.
Stepanov, Artur, Manca Mušič, and Penka Stateva. 2018. Two (non-) islands in Slovenian: A study in experimental syntax. *Linguistics* 56.3.435–476.
Szabolcsi, Anna, and Terje Lohndal. 2017. Strong vs. weak islands. **In:** Martin Everaert and Henk van Riemsdijk, editors. *The Wiley Blackwell companion to syntax*, second edition. Berlin, Germany: John Wiley and Sons. https://doi.org/10.1002/9781118358733.wbsyncom008.
Szabolcsi, Anna, and Frans Zwarts. 1993. Weak islands and an algebraic semantics for scope taking. *Natural Language Semantics* 1.3.235–284.
Szabolcsi, Anna, and Frans Zwarts. 1997. Weak islands and an algebraic semantics for scope taking. **In:** Anna Szabolcsi, editor. *Ways of scope taking*. Dordrecht, the Netherlands: Kluwer Academic Publishers. 217–262.
Tellier, Christine. 1991. *Licensing theory and French parasitic gaps*. Dordrecht, the Netherlands: 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, the Netherlands: John Benjamins Publishing Company. 159–193.
Uriagereka, Juan. 1999. Multiple spell-out. **In:** Samuel D. Epstein and Norbert Hornstein, editors. *Working minimalism*. Cambridge, MA: MIT Press. 251–282.
Wahba, Wafaa. 1984. *Wh-constructions in Egyptian Arabic*. Doctoral thesis. Urbana, IL: 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, Annie, Elisabet Engdahl, and Joan M. Maling. 1981. Resumptive pronouns can be syntactically bound. *Linguistic Inquiry* 12.4.679–682.

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

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Al-Aqarbeh, Rania, and Jon Sprouse. 2023. Island effects and amelioration by resumption in Jordanian Arabic: An auditory acceptability-judgment study. *Syntax* 1–37. https://doi.org/10.1111/synt.12262