Island Effects
Jon Sprouse and Sandra Villata
University of Connecticut

1. Introduction

One of the defining characteristics of human languages is the existence of long-distance dependencies: dependencies that can exist between two elements with no apparent bound on the linear distance (as measured in words) or hierarchical distance (as measured in clauses) between them. The unboundedness of long-distance dependencies is illustrated in (1) using wh-dependencies in English. The head of the dependency is the wh-word *what*; the tail of the dependency is indicated with an underscore.

(1) a. What did Lisa invent __?
     b. What did Dean think that Lisa invented __?
     c. What did Charlie say that Dean thinks that Lisa invented __?

Long-distance dependencies can involve a number of distinct items in the head position of the dependency, as illustrated in (2) for English, with the item in the head of the dependency in bold (this list is illustrative, not exhaustive, particularly for other languages).

(2) a. **What** do you think that Lisa invented __?
     b. I do not understand **the algorithm** that you think that Lisa invented __.
     c. I do not care for Pepsi, but **Coke**, I think that I like __.
     d. **Complicated** though you think the algorithm is __, you can understand it.

Though long-distance dependencies are unbounded, they are constrained: when the tail of the dependency appears within certain structures, the sentence becomes unacceptable, as illustrated in (3) for several structures in English, with square brackets around the structures that appear to be responsible for the unacceptability.

(3) a. What did you laugh [because Sam ate __ by accident]?   (adjunct island)
     b. What did you hear [the rumor that Jodie discovered __]?   (complex NP island)
     c. What did [the story about __] impress Mary?   (subject island)
     d. What did you wonder [whether Lisa invented __]?   (whether island)

Ross (1967) metaphorically named these structures *islands*; with the effect that these structures have on the acceptability called *island effects*; and the grammatical constraints that are proposed to capture these effects called *island constraints*.

   Island effects are one of the most studied phenomena in experimental syntax. There are at least two reasons for this. First, at theoretical level, island effects are a terrific case study for a number of important questions in linguistics (and cognitive science more generally): Does the grammar require complex, abstract constraints, or can the phenomena we see be explained by appeal to independently motivated features of sentence processing? How do those constraints interact with real-time sentence processing mechanisms? Are there constraints on the patterns of cross-linguistic, or cross-dependency, variation? Does the acquisition of these constraints require

innate, domain-specific mechanisms? (See Phillips 2013a and 2013b for extensive discussion of some of these issues.) Second, at a methodological level, island effects are a valuable case study for illustrating the three primary benefits of formal experiments:

(i)   Formal experiments allow (and, in fact, force) researchers to explicitly define what it means to be an effect, often in the common terminology of factorial logic, thus allowing precise testing of different theories.

(ii)   Formal experiments allow (and, in fact, force) researchers to explicitly consider the source of the effect, thus allowing precise testing of different theories of the source.

(iii)   Formal experiments can often increase precision of the data, opening up new analysis possibilities, like correlating distinct data types, measuring effect sizes, and exploring variability across languages, constructions, participants, and items.

We will organize this chapter around these three potential benefits of formal experiments: Section 2 will focus on the definition of island effects, Section 3 will focus on the source of island effects, and Section 4 will focus on the benefits of increasing the precision of the data we have about island effects. Our goals are to (i) illustrate the benefits of formal experiments for island effects, (ii) review the major empirical contributions that formal experiments have made over the past two decades, and (iii) provide readers with a relatively comprehensive list of articles that used formal experiments to explore island effects. One limitation of this approach is that we will not provide a comprehensive review of specific theories of islands, though we will point out theoretical consequences of the experimental results that we review (for a theoretically oriented review, see Szabolcsi and Lohndal 2017). Our hope is that this chapter will help researchers uncover trends in this research that will aid their own studies. To that end, Section 5 concludes with a brief discussion of some of the trends that we see in current investigations.

2. Defining Island Effects

The definition of an island effect in the syntax literature is something like this: low acceptability that (i) arises when the tail of a long-distance dependency is inside of an island structure, and (ii) cannot be explained by any other property of the construction. All experiments in syntax, whether informal or formal, require the experimenter to explicitly define the effect of interest – to think through the syntactic property or properties that will be manipulated, and the effect that the manipulation will have on the response that will be measured. Formal experiments typically make this explicit by leveraging the terminology of factorial logic. The term *factor* means a property that can be manipulated, such as some dimension of the structure of a sentence; the term *level* is used to refer to the specific values that a factor can take. Factors can be continuous or categorical. The definition of island effects in the literature can be translated into factorial logic by using two categorical factors, each with two levels: the factor DEPENDENCY, with levels manipulating the clause containing the tail of the dependency, and the factor STRUCTURE, with levels manipulating the presence or absence of an island structure. We illustrate this design here using *whether*-islands (Sprouse 2007, Sprouse et al. 2011, Sprouse et al. 2012).

(4) A 2x2 factorial design for *whether*-islands

	DEPENDENCY	STRUCTURE
a. Who __ thinks [that Lisa invented the algorithm]?	matrix	non-island

b. What do you think [that Lisa invented ___]?   embedded   non-island
c. Who ___ wonders [whether Lisa invented the algorithm]?   matrix   island
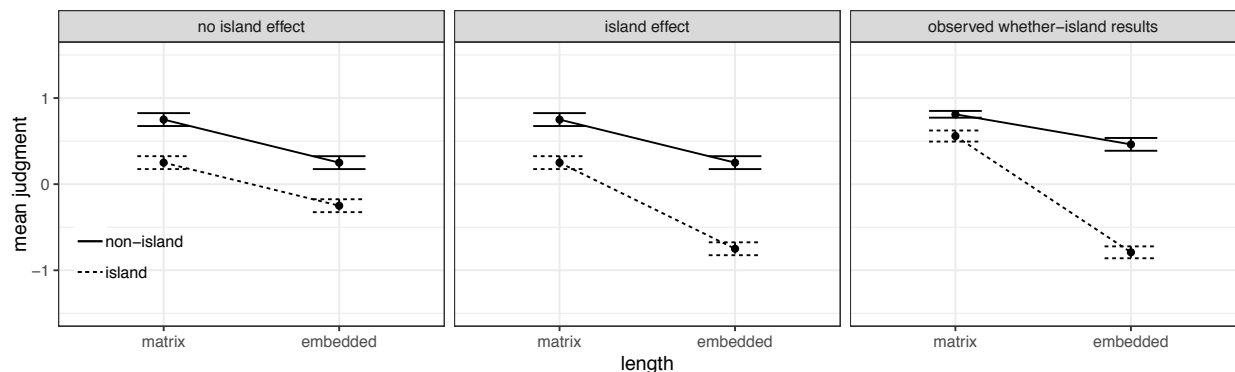d. What do you wonder [whether Lisa invented ___]?   embedded   island

This is called a 2x2 design (read "two by two") – each digit in this name represents a factor in the design, and each value of the digits represents the number of levels. The goal of factorial logic is to isolate effects using subtraction. The difference (a-b) isolates the effect of the length of the dependency (both structural and linear). The difference (a-c) isolates the effect of the island structure. Recall that the definition of island effect says that something additional happens when the tail of the dependency is inside of an island structure. In this factorial design, that means that the acceptability of (d) is more than the linear sum of the effects of dependency length and structure. We can state this mathematically as: (a-d) = (a-b) + (a-c) + X, where X is the additional effect that is not isolated by any of the factors, that is, X is the island effect. In statistical terms, this is a superadditive interaction, where the superadditive component isolates the island effect. It is sometimes useful to algebraically re-arrange the equation to isolate X, such as (b-d) – (a-c) = X. The interaction term X in this equation is called a *differences-in-differences* score (Maxwell and Delaney 2003).

There are several advantages to using a factorial definition for island effects, at both the level of experimental design, and the level of data analysis. At the level of experimental design, explicitly defining the factors in the design draws attention both to the effects that can be quantified (the length effect, the structure effect, and the island effect), and to the effects that are not being actively quantified (like the choice of wh-word). Experimenters can then easily evaluate the pros and cons of either adding additional factors for unquantified effects, holding them constant across all factors, or letting them vary freely. This plays out in a number of ways above. For example, we chose to define levels of the dependency factor as matrix-vs-embedded instead of no-dependency-vs-dependency (e.g., yes-no questions versus wh-questions) because the former allows us to isolate the effect of the length of the dependency, whereas the latter would yield a complex effect: it would capture both the effect of the presence and absence of a dependency, but also any other differences between yes-no and wh-questions. Ultimately, this would not impact the isolation of the island effect. One nice consequence of isolating the island effect in the interaction term is that the effects captured in the two factors should subtract out (unless they interact with the island effect). But the choice of the levels of the factors does influence which other properties we can quantify. In this case, we would like to explore a specific theory of island effects that posits a role for the length of the dependency (see Section 3.1 below), so this design is slightly more helpful. Similar considerations hold for the use of *who* in (a) and (c) versus *what* in (b) and (d). This difference will subtract out in the equations above as long as the choice of wh-word does not interact with island effects; but it does mean that the effect of dependency length will contain both the length manipulation and the wh-word manipulation. In short, factorial logic provides a framework for evaluating the properties of the conditions to determine precisely how they will impact the effects that can be quantified (the two main effects of the factors and the interaction).

The factorial definition of island effects also provides benefits at the level of data analysis. For one, the factorial definition provides a straightforward graphical prediction for the presence or absence of an island effect: when plotted in an interaction plot as in Figure 10.1, the absence of an island effect will appear as parallel lines (i.e., no interaction) as in the leftmost panel of Figure 10.1, and the presence of an island effect will appear as the "alligator mouth"

pattern indicative of a monotonic superadditive interaction as in the center panel of Figure 10.1. The rightmost panel of Figure 10.1 shows the results of an experiment from Sprouse and Messick (2015) using the factorial design for *whether*-islands, which show the characteristic superadditive interaction.

Figure 10.1: The graphical predictions of the 2x2 design for *whether*-islands. The left panel is the prediction for the absence of an island effect (in the presence of two main effects of dependency length and structure); the center panel is the prediction for the presence of an island effect (a monotonic superadditive interaction); the right panel is what we observe for a real experiment in English. For the real data, judgments were z-score transformed prior to analysis, therefore the y-axis is the mean of z-scores.

A closely related benefit is that the factorial definition provides a straightforward statistical definition for the presence or absence of an island effect: the presence or absence of a statistically significant interaction between the two factors in the design. Finally, the factorial definition provides a method for quantifying the size of the island effect: the size of the interaction term (or differences-in-differences score). Though effect sizes are rarely used in linguistic theory (or cognitive science more broadly), there are a number of questions about island effects for which effect sizes may yield relevant information; therefore effect sizes will arise throughout the discussions below.

Before leaving this section, there are three more advanced topics related to factorial designs that are worth mentioning. The first is that factorial logic is not just for island effects – factorial designs ultimately underlie all effects in the syntax literature (and cognitive science more generally). This means that all of the benefits that factorial designs provide for the investigation of island effects are in principle available for other phenomena. Our impression is that factorial designs have played a more explicit role in the island effects literature for the very same reasons that island effects have played such a large role in the experimental syntax literature – because island effects are an excellent case study for exploring questions about the source of effects, the complexity of the grammar, the interaction of the grammar and sentence processing, etc. We expect factorial designs to play a more central role as more phenomena are studied within the experimental syntax literature.

The second issue is that the benefits of isolating effects using superadditive monotonic interactions come with a minor methodological cost – superadditive monotonic interactions can be caused by non-linearity in the response scale (specifically, larger intervals at one end, and smaller intervals at the other). Unfortunately, for most cognitive measures, including acceptability judgments, there is no way to independently verify that the response scale is linear.

The explicitly linearly spaced numbers of the scale could map to a non-linear underlying scale in the minds of participants (that we therefore cannot observe). This means that superadditive monotonic interactions can arise even when there is no true interaction present. For this reason, such interactions are sometimes called *removable interactions* in the statistics literature (Loftus 1978, Wagenmakers et al. 2012). This is an issue that all users of factorial designs must keep in mind. We do not have space here to review this issue in detail, but we would like to note that we do not believe that island effects are the result of an underlying non-linearity in the scale of acceptability. One piece of evidence for this is empirical: superadditive monotonic interactions do not appear to arise spuriously for phenomena that should not yield them, contrary to what we might expect if the underlying scale of acceptability were non-linear. Another piece of evidence is statistical: the interactions that indicate the presence of island effects survive even when researchers use statistical analysis techniques that are designed to circumvent the problem raised by removable interactions (such as signal detection theory, and cultural consensus theory).

The third issue is that factorial designs only instantiate one of the two components of the traditional definition of island effects – that the acceptability effect cannot be explained by other known factors. They do not enforce the second component – that the acceptability of the sentence containing the island effect is relatively low in the scale. Keller (2000) and Featherston (2005) both observed that acceptability judgment effects can be present in formal experiments without leading to low acceptability. Almeida (2014) was the first to observe this for island effects, demonstrating that *whether*-islands in Brazilian Portuguese show a small superadditive effect with all four conditions rated above the mid-point in the acceptability scale. Almeida calls this a *subliminal* island effect to reflect the fact that there is an acceptability effect (albeit a small one) that is unexplained in the factorial design, but that speakers may not report the critical sentence as unacceptable. As Almeida points out, subliminal island effects raise difficult questions about what it means to be an island effect, what the source of island effects are, and ultimately, what the overall architecture of the language faculty is. These questions were recently underscored by Keshev and Meltzer-Asscher (2019), who observe a subliminal island effect for wh-islands in Hebrew for both wh-dependencies, which is unexpected because Hebrew has been claimed to lack wh-islands (Reinhart 1981), and for backward binding dependencies, which is unexpected because binding dependencies are not typically claimed to show island effects. We will discuss research on the source of island effects in detail in the next section. The important point here is that formalizing the definition of island effects using a factorial design helps to bring these issues into sharp relief, thereby opening a number of new avenues of research.

3. The Source of Island Effects

One of the driving questions in linguistics is whether the grammar requires complex, abstract constraints, or whether the phenomena that we see in human languages can be explained by appealing to other, perhaps independently motivated, aspects of cognition. Island effects present a classic case study for this, as the components involved in island effects, such as long-distance dependencies and complex syntactic structures, raise the possibility that island effects could be reduced to an independently motivated consequence of sentence processing complexity (e.g., Kluender and Kutas 1993, Hofmeister and Sag 2010). In this section, we will primarily focus on the debate between approaches that postulate complex, abstract grammatical constraints to explain island effects, and approaches that seek to reduce island effects to independently motivated aspects of sentence processing. We see this as a first cut in the space of theories. After

this cut, one could then explore different approaches within a tradition; for example, within the grammatical tradition, one could then explore the semantic approaches of Szabolcsi and Zwarts (1993) or Abrusán (2014), or the pragmatic approaches of Erteschik-Shir (1973) or Goldberg (2006). But for space reasons we will focus on the first cut in this chapter. We will review multiple strands of experimental research dedicated to this question, including studies on working memory capacity (Section 3.1), studies on the sentence processing dynamics of filler-gap dependencies (Section 3.2), studies on the sentence processing dynamics of binding dependencies (Section 3.3), and studies on judgment satiation (Section 3.4). Though the results of these studies do tend to support grammatical approaches to islands over sentence processing approaches, our focus will be on the way that formal experiments can be used to gather this kind of evidence.

3.1 Working Memory Capacity

Perhaps the most prominent sentence-processing based theory of island effects is the working memory capacity theory proposed by Kluender and Kutas (1993) (recently advocated by Hofmeister and Sag 2010). Kluender and Kutas observe that there are two sources of sentence processing complexity in the critical sentences of island effects (the (d) condition): the long-distance dependency, and the complex syntactic structure that we call an island. These sources of sentence processing complexity provide two potential explanations for the unacceptability that we see in island effects. One explanation could be called the *simple reductionist approach*. This approach argues that the long-distance dependency and the complex island structure each decrease acceptability (presumably due to their sentence processing costs), and that these decreases sum linearly as in the left panel of Figure 10.1. In this way, these two costs completely determine the acceptability of the critical sentence without any need for a grammatical constraint. The second type of explanation is a *complex reductionist approach*, in which independently motivated sentence processing factors lead to the superadditive interaction. Kluender and Kutas propose just such a theory. Under their theory, the processing of the long-distance dependency and the processing of the complex island structure each draw from the same limited pool of working memory resources. There are not enough resources in the pool to deploy both sets of processes simultaneously. This leads the parse to crash, yielding the perception of unacceptability. This working memory capacity theory can explain the superadditive pattern of acceptability in the center panel of Figure 10.1 (an interaction between dependency length and structure) without the need for a grammatical constraint. This means that, for the working memory capacity theory, the critical (d) sentences for island effects are grammatical, but are perceived as unacceptable because the parser does not have access to the working memory resources that it needs. These two approaches to reductionism lead to the following logic: the linearly additive pattern in the left panel of Figure 10.1 is unambiguously evidence of a simple reductionist theory of island effects, whereby the unacceptability is completely explained by the costs associated with the two factors of the factorial design; the superadditive pattern in the center panel is ambiguous –it shows that there is a mystery that needs to be explained, with the space of possible explanations including a grammatical constraint, the working memory capacity theory, or some other complex reductionist theory that has not yet been proposed.

Though the superadditive pattern is itself ambiguous, the factorial definition does allow us to potentially test unique predictions of the competing theories of the superadditive interaction. For example, Sprouse et al. (2012) argue that one potential prediction of the

Kluender and Kutas (1993) working memory capacity theory is that variability in working memory capacity should lead to variability in the size of the island effect (based on discussion in Kluender and Kutas (1993) that the additional unacceptability may be driven by a mechanism in the memory system that penalizes the parse based on how much the processes exceed the pool of resources). Sprouse et al. then test this prediction by looking for correlations between the interaction term (the differences-in-differences score) from the factorial definition of island effects as a measure of the size of island effects, and both serial recall scores and performance in an *n*-back task as measures of working memory capacity. They find no evidence of a relationship between the two across two experiments, each testing four island types (whether, complex NP, subject, and adjunct islands), using both measures of working memory capacity, with relatively large samples sizes. These results were replicated in Michel (2014) using a reading span tasks as the measure of working memory capacity. These results suggest that the most straightforward approach to working memory capacity is unlikely to be sufficient to explain island effects. However, these results do leave open the possibility of adopting more complex working memory capacity theories (e.g., Hofmeister, Staum, Casasanto, and Sag (2012) suggest that a step-like or sigmoidal relationship between working memory capacity and acceptability might account for these results). This kind of theory revision is the normal course for science. The primary point here is that formal experiments, and explicit factorial definitions of effects, can be leveraged to explore the source of island effects (and indeed all acceptability judgment effects), as long as the competing theories make unique, testable predictions. The ability of formal experiments to quantify effect sizes can be particularly useful if the predictions involve relationships with other types of measures, like working memory measures.

3.2 The Sentence Processing Dynamics of Long-Distance Dependencies

The sentence processing dynamics of long-distance dependencies is a complex topic in its own right; and, as such, we cannot hope to do it justice in one sub-section (see Chapters 24 and 25 in this volume). Therefore here we will focus specifically on the ways in which this topic could potentially interact with the question of the source of island effects. The critical issue involves *gap-filling* – a cover term for the set of process that the parser deploys to identify the tail of a long-distance dependency (the *gap*), retrieve the displaced element (the *filler*) from memory, and integrate it into the structure. One of the driving questions in the literature has been whether gap-filling is *passive*, such that the parser waits for unambiguous evidence for a gap (e.g., a sequence of words that could not appear together without a gap in between) to deploy gap-filling processes, or whether gap-filling is *active*, such that that the parser attempts to deploy gap-filling processes before encountering unambiguous evidence for the gap. The consensus in the literature is that gap-filling is active: Crain and Fodor (1985) and Stowe (1986) first demonstrated this with an experimental diagnostic called the filed-gap effect, illustrated in (5) using examples from Stowe (1986).

(5)   a.   My brother wanted to know **who** Ruth will bring <u>us</u> home to __ at Christmas.
        b.   My brother wanted to know **if** Ruth will bring <u>us</u> home to Mom at Christmas.

In a self-paced reading task, Stowe found that reading times at *us* are slower in a sentence with a long-distance dependency (5a) then in a sentence without a long-distance dependency (5b). This slow-down can be explained if the parser (actively) engages gap-filling processes prior to *us*,

integrating the filler as the object of the verb, such that the parser must reanalyze the structure when it encounters the "filled-gap" represented by *us*. In addition to the filled-gap effect, Garnsey et al. (1989) demonstrated that active gap-filling can be revealed by manipulating the plausibility of a filler relative to the verb that potentially selects it, and later work, like Traxler and Pickering (1996) discussed below, showed that the plausibility manipulation also triggers a reading time slow-down at the verb. Building on results like these (and many others), Frazier and Flores d'Arcais (1989) formalized the idea of active gap-filling as the *Active Filler Strategy*, which states that the parser attempts to complete filler-gap dependencies at the first possible location. The "first possible location" has traditionally been taken to be the first gap-selecting category (such as a verb or preposition), but Omaki et al. (2015) have recently shown that gap-filling may in fact be *hyperactive*, such that the parser predicts the gap-selecting category (such as a transitive verb) before encountering it within sentences with filler-gap dependencies.

The active (or hyperactive) nature of gap-filling raises an interesting question for island effects – does the parser attempt to actively engage gap-filling inside of islands? Stowe (1986) investigated this question in the second experiment of her seminal paper with the paradigm in (6):

(6)   a.   The teacher asked **what** the team laughed about <u>Greg's</u> older brother fumbling __.
          b.   The teacher asked **if** the team laughed about <u>Greg's</u> older brother fumbling the ball.
          c.   The teacher asked **what** the silly story about <u>Greg's</u> older brother was supposed to mean __.
          d.   The teacher asked **if** the silly story about <u>Greg's</u> older brother was supposed to mean anything.

She found the classic filled-gap effect in (6a) versus (6b) at *Greg's*, as expected. However, in (6c) and (6d), which contain a subject island structure, she found no filled-gap effect at *Greg's*, suggesting that active gap-filling is suppressed within subject islands. This close alignment between active gap-filling and island effects is tantalizing – it raises the possibility that active gap-filling may reveal information about island effects that we may not get from acceptability judgments alone. To be absolutely clear, as far as we can tell, the alignment of active gap-filling with island effects is logically independent of the source of island effects: active gap-filling could in principle align or not with island effects caused by grammatical constraint, and could in principle align or not with island effects caused by sentence-processing complexity. To our minds, the theoretical value of exploring the alignment of active gap-filling is that it can help to refine the space of possible theories of island effects for both types of sources.

Nearly all of the published studies on active gap-filling within islands using reading time measures are studies of English, and all test islands in subject position. The focus on subject position is methodological: investigating an island in subject position means that there is a potential acceptable continuation of the dependency after the island (typically in the form of a post-verbal gap). If there is no acceptable continuation of the sentence, it is possible that the parser might recognize the unacceptability of the sentence at the leading edge of the island, making it impossible to interpret reading time results within the island (see Phillips 2006 for a review of studies that find reading and ERP effects at the leading edge of islands). The focus on English is perhaps a consequence of the fact that English has subject island effects, and is generally overrepresented in linguistics. What this means in practice is that there is quite a bit of

information that we do not yet have – information on languages other than English, and on islands that cannot occur in subject position (like wh-islands and adjunct islands). What we do know is that subject islands in English appear to reliably suppress active gap-filling for wh-dependencies (Stowe 1986, Pickering et al. 1994), and that relative clauses in subject position (so, perhaps a type of double island effect) in English appear to reliably suppress active gap-filling for relative-clause dependencies (Traxler and Pickering 1996, Omaki et al. 2015). The one complication to this pattern was reported by Pickering et al. 1994 for wh-dependencies out of relative clauses in subject position, which did show evidence of active gap-filling (in contrast to the relative clause dependencies tested by Traxler and Pickering 1996). Setting aside the (still unexplained) results of Pickering et al. (1994), the general consensus in the literature is that subject islands suppress active gap-filling in English (see also Freedman and Forster 1985 for evidence using the sentence matching task, and Clifton and Frazier 1989 and Kurtzman and Crawford 1991 for evidence using speeded grammaticality).

Phillips (2006) investigated active gap-filling within subject islands that can host a parasitic gap. A parasitic gap is an acceptable gap that appears inside of an island when there is a second gap outside of the island; in other words, the gap inside the island is parasitic on the gap outside of the island (Engdahl 1983; see Culicover and Postal 2001 for a review of the conditions on parasitic gaps). As a concrete example, (7a) is unacceptable in English, giving rise to a subject island effect. However, (7b) is acceptable, apparently because of the second gap added to the post-verbal object position. The ability of the subject in (7b) to host a parasitic gap appears to be tied to the finiteness of the relative clause; a finite relative clause is unacceptable regardless of the presence of the second gap (8a-b). All of these facts were corroborated by Phillips (2006) in a formal acceptability experiment.

(7)   a.   *The outspoken environmentalist worked to investigate what the local campaign to preserve __ had harmed the annual migration.
      b.   The outspoken environmentalist worked to investigate what the local campaign to preserve __ had harmed __.

(8)   a.   *The outspoken environmentalist worked to investigate what the local campaign that preserved __ had harmed the annual migration.
      b.   *The outspoken environmentalist worked to investigate what the local campaign that preserved __ had harmed __.

Phillips then used the plausibility mismatch paradigm in a self-paced reading study to show that non-finite subject relative clauses, like those in (7), show evidence of active gap-filling within the subject; whereas finite subject relative clauses, like those in (8), show no evidence of active gap-filling within the subject. The descriptive fact seems to be that active gap-filling is very tightly aligned with the ability to host an acceptable gap – subjects that can potentially host a parasitic gap (an acceptable gap) show evidence of active gap-filling, and subjects that cannot host a parasitic gap show no active gap-filling. The theoretical import of this is quite stunning. The critical fact is that the parser has no way of knowing whether the gap in the subject in (7) is licit or illicit at the time of active gap-filling. The parser simply deploys gap-filling because the subject could potentially host a gap if the right conditions hold later in the sentence. This means that the unacceptability that arises for sentences like (7a) cannot be caused by the inability of the parser to fill the gap in the subject. The parser can, and does, fill that gap during the first pass of

the parse. The unacceptability in (7a) must occur later, after the parser has realized that there is no second gap to license the parasitic gap. The consequence of this is that any theory of the source of the island effect in (7a) must <u>not</u> prevent gap-filling wholesale. It must allow gap-filling, and then yield unacceptability later when the licensing condition is not met. As Phillips notes, a check of licensing is something that grammatical approaches to islands are well-equipped to handle; it is less clear how this can be achieved with the sentence processing approaches to island effects that are currently in the literature. (For additional evidence that the parser is aware of the licensing of parasitic gaps, see Wagers and Phillips 2009.)

For completeness we should mention that there are a few EEG studies of island effects. Neville et al. (1991) found a P600 response at the verb for subject islands (*_What was a sketch of admired by the man_). Kluender and Kutas (1993) found an N400 at the clause boundary for a wh-island (*_What do you wonder <u>who</u> they caught at <u>by</u> accident?_), as well as a LAN at the word after the gap. McKinnon and Osterhout (1996) found a positive component at the first word of an adjunct clause forming an adjunct island (*_I wonder which of his staff members the candidate was annoyed <u>when</u> his son was questioned by?_). These results all show that island effects are detected during online processing, but it is less clear how these results relate to the question of active gap-filling within islands. One problem that arises when trying to link these results to the active gap-filling literature is that there is no EEG response uniquely associated with active gap-filling. The two responses associated with gap-filling, a P600 at the verb (Kaan et al. 2000) and a LAN at the word after the gap (Kluender and Kutas 1993), are both later than the reading time effects that indicate active gap-filling. Another problem is that only subject islands allow for an acceptable continuation, making it difficult to interpret effects that might arise inside of other island types. Therefore, for now, we simply note that EEG can be used to detect island effects during real-time sentence processing, and that there is still quite a bit of potential to make advances in this literature.

3.3 The Sentence Processing Dynamics of Binding Dependencies

The binding dependencies that exist between the pronoun and the noun that it is co-referent with do not appear to give rise to island effects. Therefore it may seem odd to have a section dedicated to binding dependencies in a chapter about island effects. But this is in fact a strength. Binding dependencies share a number of sentence processing features with the long-distance dependencies that give rise to island effects: binding dependencies can be infinite in length; when the pronoun or anaphor appears before its coreferent noun in a configuration called _backward binding_, the parser engages in an _active_ search for that noun (Van Gompel and Liversedge 2000) similar to active gap-filling; the active search for the coreferent noun is constrained by the licensing conditions on binding dependencies (Sturt 2003, Kazanina et al. 2007) similar to the way active gap-filling is constrained by island effects; and both long-distance dependencies and backward binding dependencies appear to increase activation in the left inferior frontal gyrus (Matchin et al. 2014). These similarities suggest that binding dependencies and long-distance dependencies may share substantially similar sentence processing mechanisms. The fact that binding dependencies share these sentence processing mechanisms but do not share island effects means that (backward) binding dependencies can serve as an interesting minimal pair with long-distance dependencies in studies designed to explore the role of sentence processing mechanisms in island effects.

Yoshida et al. 2014 demonstrated that the active search for a coreferent noun in backward binding dependencies is not suppressed within relative clauses in subject position (a combination of a relative clause island and a subject island), but is suppressed by Binding Condition C. They did this using the *gender mismatch effect* – a slow-down in reading times that occurs when the gender of a pronoun and the gender of its potential coreferent NP do not match. A gender mismatch effect is typically interpreted as evidence that the parser has attempted to create a binding dependency between the pronoun and the NP. Yoshida et al. created two pairs of conditions, as in (9) and (10).

(9)   a.   **His managers** revealed that the studio that notified <u>Jeffrey Stewart</u> about the new film selected a novel for the script, but Annie did not seem to be interested in this information.

        b.   **Her managers** revealed that the studio that notified <u>Jeffrey Stewart</u> about the new film selected a novel for the script, but Annie did not seem to be interested in this information.

(10)   a.   **He** revealed that the studio that notified <u>Jeffrey Stewart</u> about the new film selected a novel for the script, but Annie did not seem to be interested in this information.

        b.   **She** revealed that the studio that notified <u>Jeffrey Stewart</u> about the new film selected a novel for the script, but Annie did not seem to be interested in this information.

The pair of conditions in (9) instantiates a gender mismatch paradigm between a possessive pronoun (*his/her managers*) and a proper noun inside of a relative clause (*Jeffrey Stewart*). This pair tests whether a binding dependency is actively constructed within relative clause islands in subject position. The pair in (10) is identical, except that the pronoun is no longer possessive (*he/she*). The pronouns in (10) now c-command the critical proper noun, violating Binding Condition C (crucially, the possessive pronouns in (9) do not c-command the critical proper noun, causing no Binding Condition C violation). In this way (10), replicates the design of previous studies on backward binding and Binding Condition C (e.g., Kazanina et al. 2007). Yoshida et al. found a gender mismatch effect in (9), but not in (10). This suggests that the active search for a coreferent NP is not suppressed within islands, but is suppressed when Binding Condition C would be violated by a c-command relationship between the pronoun and its coreferent NP. Yoshida et al. interpret these results to indicate that island effects cannot be the result of sentence processing complexity driven by the sentence processing mechanisms that backward binding and long-distance dependencies share, otherwise we would expect islands to suppress binding dependencies, contrary to fact.

A recent sentence acceptability experiment by Keshev and Meltzer-Asscher (2019) raises a potential complication for the relationship between binding dependencies and island effects. Keshev and Meltzer-Asscher tested wh-islands in Hebrew using both wh-dependencies and backward binding dependencies. Hebrew is typically claimed to not have wh-island effects (Reinhart 1981). However, Keshev and Meltzer-Asscher find small superadditive interactions for both wh-dependencies and backward binding dependencies. Keshev and Meltzer-Asscher point out that the small superadditive effect for wh-dependencies looks like a subliminal island effect in the sense of Almeida 2014 (see Section 2). They further observe that the existence of a similar

small interaction effect for backward binding dependencies could be taken to suggest that superadditive interactions can potentially be caused by something other than a grammatical island constraint, under the assumption that backward binding would not be subject to a grammatical island constraint. This means that it is possible that the superadditive interactions for both wh-dependencies and binding dependencies in Hebrew could be driven by a set of shared sentence processing mechanisms (though the judgment experiments were not designed to isolate those mechanisms). No such interactions have been demonstrated in English (a language that has true wh-islands), so these results do not directly complicate the Yoshida et al. 2013 results. Nonetheless, these results add a layer of complication to the assumption that binding dependencies never give rise to effects that look like island effects. These results underscore the need for more research on subliminal island effects with both long-distance dependencies and binding dependencies across languages.

3.4 Satiation of Judgments

Satiation is the term that syntacticians tend to use to describe an increase in the perception of acceptability after repeated exposures to the same sentence or the same structure. As a phenomenon, satiation raises a number of interesting questions, such as whether it is related to the phenomenon of syntactic priming in production and reading time studies (see Do and Kaiser 2017), and whether satiation as measured in the laboratory is the same phenomenon that professional linguists report after working on one phenomenon for an extended period of time (see Dabrowska 2010 for some investigations of potential differences between linguists and non-linguists). As such, satiation is the topic of its own chapter in this volume (Chapter 7). However, in this chapter, we would like to focus on the claim, to our knowledge first proposed by Snyder 2000, that satiation could be used to distinguish between unacceptability that arises due to the violation of a grammatical constraint and unacceptability that arises due to sentence processing issues, particularly with respect to island effects.

Table 10.1 presents our attempt to summarize the acceptability judgment satiation literature for four island types (we apologize for any studies that we inadvertently left out of this summary).

Table 10.1: A (potentially incomplete) summary of the acceptability judgment satiation literature for four island types. The first column identifies the study. The next four columns summarize information about the experiments: the experiment number, the task - yes/no (YN), magnitude estimation (ME), and five and seven point scales (5P and 7P), the number of repetitions of each condition, and whether the items were presented with context sentences. All tested English, except for Christensen et al. (2013), which tested Danish. A plus (+) indicates evidence of satiation according to some statistical test; a dash (–) indicates no evidence of satiation; and a blank cell indicates that the island was not tested. The size column reports the slope of the line of best fit for repetition for those studies that used linear regression as part of the analysis.

study	exp	task	reps	context	adjunct	complex np	subject	whether	size
Snyder 2000	1	YN	5	yes	–	+	–	+	
Hiramatsu 2000	1	YN	7	yes	–	–	+	–	
Hiramatsu 2000	2	YN	7	yes	–	–	–	+	
Sprouse 2009	1-3	YN	5	yes	–	–	–	–	
Sprouse 2009	4-7	ME	14	no	–	–	–	–	

Francom 2009	1	YN	5	no	–	–	+	+	
Francom 2009	2	YN	5	no	–	–	–	+	
Goodall 2011	1	YN	5	yes	–	+	–		
Crawford 2012	1	7P	7	no	–		–	+	
Christensen et al. 2013	1	5P	16	no				+	.12
Christensen et al. 2013	2	5P	12	no				–	
Chaves & Dery 2014	1	7P	20	no			+		.02
Chaves & Dery 2014	2	7P	14	no			+		.10
Do & Kaiser 2017	1	5P	1	no		+	–		.10
Do & Kaiser 2017	2	5P	1	no		–	–		

The first pattern that emerges is that there is no island that consistently shows satiation: adjunct islands do not show satiation in any of these studies, complex NP islands show satiation in 3 out of 10 studies, subject islands show satiation in 4 out of 13, and whether islands show satiation in 6 out of 10. The second pattern that emerges is that none of the obvious properties of these experiments can explain this variability: the results that show satiation span different tasks, different numbers of repetitions, and both the presence and absence of context sentences. The third pattern that emerges is that, for those studies that used a numerical scale and found a significant satiation effect, the size of the effect of satiation is very small: repetition increases the acceptability rating by between .02 and .12 units on the scale (i.e., a 5 point or 7 point scale) per repetition. For the largest effect, it would take 8 repetitions to increase acceptability by approximately one point on the scale. (The reason that the YN tasks show satiation despite these small effects and the relatively low power of the YN task (Sprouse and Almeida 2017) is that the definition of satiation in these studies focuses on a comparison of first blocks to last blocks, and in some cases, ignores data from participants who show no changes in their judgments – two choices that likely increase the chances of detecting small effects.) Taken together, these three patterns suggest that, at a purely empirical level, satiation is unlikely to yield the kind of reliable results that we would need to make the strong claim that different island types should be divided into distinct classes.

We would also like to note that there is a deeper theoretical challenge facing satiation studies – there is no explicit theory of the mechanisms underlying satiation. To be clear, there a number of phenomena in language studies that satiation could be related to, such as implicit learning (see Luka and Barsalou 2005 for a discussion) or syntactic priming (see Do and Kaiser 2017 for a discussion). But these phenomena labels do not tell us what the underlying mechanisms are. Without a theory of those mechanisms, we cannot evaluate the proposal that satiation will affect constructions differently depending on the source of the violation (e.g., grammar versus sentence processing). It is clear that there is much work to be done to better understand satiation of island effects, but given that the effects appear to be very small and relatively fragile, we recommend caution to researchers deciding whether to invest significant time or resources into this topic.

4. Precision in the Data

The third major benefit of formal experimental work (for all phenomena) is that it allows us to increase the precision of our data, thereby potentially increasing the precision of our theories. In this section we will look at four topics in the islands literature for which experimental work is beginning to refine the data set in potentially theoretically interesting ways: cross-linguistic

variation, cross-dependency variation, the effect of complex wh-phrases, and the effect of resumptive pronouns.

4.1 Cross-Linguistic Variation

It is empirically valuable to establish the extent, and pattern, of cross-linguistic variation for any phenomenon because constraints on variation can help to refine the space of viable theories for that phenomenon. Island effects are no different. Much of the literature on island effects has sought to establish what, if any, constraints exist on their variation. When looking at the contribution of formal experimental studies, we think there are two questions that we can profitably ask. The first is whether the pattern that emerges using formal experiments differs in any meaningful way from the pattern that emerges using traditional informal experiments. Table 10.2 below presents a list of formal experimental studies on island effects, organized first by language, then by dependency type. The data is reported by island: adjunct, complex NP, subject, relative clause, and wh-islands. For space reasons, we have collapsed all types of adjunct islands into one column, and have collapsed wh-islands and whether islands into one column. We restricted our attention to published studies using the (2x2) factorial definition of island effects discussed in Section 2 (we apologize for any studies that we have missed). If an island effect was found, we have listed the size of the island effect (the differences-in-differences score) in that cell. The scale column indicates the scale of the effect size – either a z-score scale or a raw judgment scale, with the number of points on the scale indicated in parentheses.

Table 10.2: A (potentially incomplete) summary of formal experimental studies on island effects. The table is organized by language. The four island columns represent adjunct, complex NP, subject, relative clause, and wh-islands. For space reasons, we collapsed all adjuncts into one column, and both whether islands and wh-islands into one column. The numbers in the island columns indicate the size of the superadditive interaction as measured on the scale listed in the scale column (either a z-score scale or a raw judgment scale with the number of points in parentheses). Most of the effect sizes are reported directly in the articles, but some we estimated from plots (and rounded to two significant digits). A dash (–) indicates no evidence of an island effect; and a blank cell indicates that the island was not tested.

study	language	dependency	adj	np	sub	rc	wh	scale
Almeida 2014	Br. Portuguese	bare wh					0.6	z-score
Almeida 2014	Br. Portuguese	topicalization					–	z-score
Lu et al. 2019	Chinese	wh-arg-in-situ				1.5		raw (7)
Lu et al. 2019	Chinese	wh-adj-in-situ				1.6		raw (7)
Christensen et al. 2013	Danish	bare wh					1.2	raw (5)
Poulsen 2008	Danish	topicalization	4.0					raw (5)
Sprouse et al. 2016	English	bare wh	0.7	1.1	0.6		1.2	z-score
Sprouse et al. 2016	English	complex wh	0.8	0.5	0.5		0.6	z-score
Sprouse et al. 2016	English	rel. clause	–	0.5	0.5		0.4	z-score
Almeida 2014	English	topicalization					–	z-score
Sprouse et al. 2011	English	wh-arg-in-situ	–	–	–		–	z-score
Sprouse et al. 2016	Italian	bare wh	1.3	0.9	1.4		1.7	z-score
Sprouse et al. 2016	Italian	rel. clause	1.1	0.6	–		0.7	z-score
Omaki et al. 2019	Japanese	np scrambling			–			z-score

Sprouse et al. 2011	Japanese	wh-arg-in-situ	–	–	–		–	z-score
Kim & Goodall 2016	Korean	wh-arg-in-situ	–				0.3	z-score
Kim & Goodall 2016	Korean	wh scrambling	–				0.7	z-score
Ko et al. 2019	Korean	np scrambling	–			–	–	z-score
Tucker et al. 2019	MS Arabic	complex wh	0.8	0.5			0.4	z-score
Kush et al. 2018	Norwegian	bare wh	1.1	1.7	1.3	1.4	0.4	z-score
Kush et al. 2018	Norwegian	complex wh	1.3	1.2	1.2	1.4	0.3	z-score
Kush et al. 2019	Norwegian	topicalization	0.2	0.5	1.7	0.7	–	z-score
Stepanov et al. 2018	Slovenian	bare wh			0.6		–	z-score

From this table, readers can evaluate any specific claims about cross-linguistic variation that they might have seen in the literature. For example, the Subjacency approach to island effects (explicit in Chomsky 1986, building on work by Rizzi 1982 and Torrego 1984) makes the strong claim that subject islands and wh-islands will covary together (either both present or both absent). This is true for most of the languages in Table 10.2, but we also see that Italian provides evidence of a language with wh-islands but not subject islands, and Slovenian provides evidence of a language with subject islands but not wh-islands. The Subjacency versus ECP approach to wh-in-situ (building on work by Huang 1982) makes the strong claim that wh-arguments in-situ will not show island effects, but wh-adjuncts in-situ will. Again, this is true for Japanese, but we also see that Korean shows wh-island effects with wh-arguments in-situ, and that Chinese shows relative clause islands with wh-arguments in-situ. The second question we can ask is whether there are any new patterns that emerge in these studies that could be used to constraint theories of island effects. A full evaluation of this question is beyond the scope of this chapter, but we can at least note that there are no obvious universal correlations: there are no two islands that are either always present together or always absent together. It is possible that more complex patterns may exist, or that there may be some sort of implicational hierarchy among island effects. It will take a larger sample of languages to explore these more complex hypotheses.

4.2 Cross-Dependency Variation and the Uniformity Hypothesis

Another dimension of variation in Table 10.2 above is cross-dependency variation: how island effects vary based on the type of long-distance dependency tested. At times there appears to be an unspoken assumption in the literature that all long-distance dependencies will behave the same with respect to island effects, at least within specific languages. This can be seen in any study that tests one specific kind of long-distance dependency but makes claims about island effects more generally in that language. We will call this assumption the *uniformity hypothesis* so that we can refer to it efficiently. The uniformity hypothesis likely has its roots in considerations of language acquisition – the acquisition process would be more complicated if children had to learn the constraints on each type of long-distance dependency separately. But the uniformity hypothesis appears to be false. Six of the languages in Table 10.2 show variation in island effects based on the dependency type tested. It is, of course, possible that future studies will find confounds in these studies that may explain the variability; but for now, both syntacticians and language acquisition researchers should consider the possibility that the uniformity hypothesis may be false.

4.3 Complex Wh-Phrases and Selective Islands

One proposal in the literature is that island types can be divided into two classes: *unselective islands*, which block extraction of all types of wh-items (and presumably all types of long-distance dependencies), and *selective islands*, which block certain types of wh-items, and allow others to pass. (Sometimes the terms *strong* and *weak* are used for *unselective* and *selective*, respectively; but these terms are also sometimes misinterpreted as labels for how large the island effect is, so we will use the terms *unselective* and *selective* in this chapter to avoid any ambiguity.) There is quite a bit of debate in the literature about (i) which islands constitute unselective and selective islands, and (ii) which wh-items, and heads of other long-distance dependencies, are blocked by selective islands. Despite these debates, there are some recurring claims in the literature. One recurring claim is that complex wh-phrases of the form *which* NP, *what* NP, or *which of the* NP are not blocked by selective islands. (Complex wh-phrases are also sometimes called d(iscourse)-linked wh-phrases in the literature, referring to a specific analysis of their properties (Pesetsky 1987). For this chapter, we prefer the theoretically neutral term *complex wh-phrase*.) Another recurring claim is that wh-islands and whether islands are selective islands. These two claims are easily characterized using the factorial definition of island effects: a selective island should either show no superadditive interaction for complex wh-phrases, or show a smaller superadditive interaction compared to bare wh-words. This is another example of formal experiments opening the door to productive investigations of some of the more enduring topics in the islands literature.

Two languages in Table 10.2 were tested using both complex wh-phrases and bare wh-words: English and Norwegian. In the Sprouse et al. (2016) experiments, English shows superadditive interactions for all four island types tested, with both complex wh-phrases and bare wh-words. This suggests that, if selectivity is defined as absolute elimination of island effects, there are no selective islands in this group of four island types in English. Though there were significant interactions for all four island types, whether islands and complex NP islands showed effect sizes with complex wh-phrases that are approximately one half of the size of the effect sizes with bare wh-words. This suggests that whether islands and complex NP islands could be considered selective islands in English, as long as selectivity is defined as smaller island effect sizes, and not complete elimination of the island effect. (We should note that Goodall 2015 found a main effect of complex wh-phrases that impacted both islands and non-island controls; this would imply that in a full factorial design, the size of the interaction would stay the same between bare wh-words and complex wh-phrases. The reason for this difference with the Sprouse et al. results remains a mystery). In the Kush et al. (2018) experiments, Norwegian shows superadditive interactions for all five island types tested, with both complex wh-phrases and bare wh-words. Only complex NP islands show a reduced effect size with complex wh-phrases; however, the resulting effect size is still as large or larger than effect sizes with bare wh-words in other languages. Therefore it is not clear whether complex NP islands should be considered selective islands in Norwegian. If we broaden the definition of selectivity to include other long-distance dependencies that involve NPs, such as relative clauses and topicalization, we see a number of additional effects. In English, whether islands and complex NP islands show reduced effect sizes with relative clauses; adjunct islands disappear completely with relative clauses (Sprouse et al. 2016). In both English and Brazilian Portuguese, whether islands disappear completely with topicalization (Almeida 2014). In Italian, relative clauses cause subject islands to disappear completely, wh-islands to be substantially smaller, and adjunct and complex NP islands to be a bit smaller (Sprouse et al. 2016). And, in Norwegian, topicalization

causes whether islands to disappear completely, while adjunct, complex NP, and relative clause islands reduce substantially in size (Kush et al. 2019).

The broadened definition of selectivity entertained at the end of the preceding paragraph raises the question of whether selectivity in its classic, constrained form can be maintained or not. In its classic form, it refers to an invariable set of dependency types (e.g., complex wh-phrases), and is a binary phenomenon (an island is either selective or unselective). The broadened definition in the previous paragraph opens the door to considering different sets of dependencies for different islands, and consequently expanding beyond a binary classification (to different types of selective islands). This question interacts directly with the uniformity hypothesis discussed in the previous subsection. Selectivity is a departure from the uniformity hypothesis, as it means that two or more dependencies are behaving differently with respect to one island. Selectivity and the uniformity hypothesis can coexist if selectivity is constrained – for example, there is a long literature starting with Pesetsky (1987) that attempts to explain why it is that complex wh-phrases behave differently within selective islands. But once selectivity is no longer constrained, not only does the term selectivity lose meaning, but so too does the uniformity hypothesis.

Before leaving this section, it is important to note that there is a second strand of research on complex wh-phrases in the literature that explores their consequences for the *relativized minimality* approach to wh-islands (building on work beginning with Rizzi 1990). Relativized minimality is a general configurational constraint that says no syntactic dependency can hold between two items if a third item of the same type intervenes between them, where *same type* is typically defined in terms of features, such as a wh-feature or noun-feature, and *intervenes* is typically defined in terms of c-command, such that the head of the dependency c-commands the intervener, and the intervener c-commands the tail of the dependency (see Rizzi 2013 for a review). Relativized minimality provides a potential analysis for wh-islands that does not rely on a specific island constraint: the wh-word at the edge of the island structure (the embedded wh-clause) intervenes between the wh-word at the head of the dependency and the tail of the dependency. Complex wh-phrases and bare wh-words potentially stand in a complex featural relationship with one another: they both involve a wh-feature, but complex wh-phrases likely involve an additional feature or features due to the extra specification of the noun. This raises the possibility that the size of the wh-island may vary based on the precise featural relationships between the head of the dependency and the intervener: the two can be identical, as in (wh… wh… __) or (which NP… which NP… __), and the two can partially match, as in (wh… which NP… __) or (which NP… wh… __). Atkinson et al. (2015) (English) and Villata et al. (2016) (French) explore these featural relationships using formal judgment experiments. We do not attempt to summarize the results here because *relativized minimality* is only an analysis for wh-islands (not for other island types), and because the results are fairly complex (simultaneously confirming and falsifying various facets of the featural relativized minimality account). But we refer readers interested in relativized minimality and island effects to these articles for discussion of these issues.

4.4 Resumptive Pronouns

Resumptive pronouns have the formal shape of typical pronouns, but unlike typical pronouns, they appear in the tail position of long-distance dependencies, and are obligatorily co-referent with the item in the head of the long-distance dependency (see McCloskey 2006 for a review).

Though there is quite a bit of debate surrounding the correct analysis of resumptive pronouns in the world's languages, many analyses divide languages into (at least) two types: those that allow resumptive pronouns as a completely grammatical option (e.g., Arabic, Hebrew, and Irish), and those that do not allow resumptive pronouns as a grammatical option (e.g., English; often called *intrusive resumption* following Sells 1984). There are a number of interesting questions about resumptive pronouns, and their properties in different types of languages; as such resumptive pronouns are the topic of their own chapter in this volume (Chapter 9). In this chapter, we would like to focus on the claim that resumptive pronouns can be used to eliminate island effects, both in languages that allow resumptive pronouns as a grammatical option, and in languages that do not allow resumptive pronouns as a grammatical option (Ross 1967, Kroch 1981, Sells 1984, Engdahl 1985, and much subsequent work).

The claim that resumptive pronouns eliminate or reduce island effects can be directly translated into the factorial definition of island effects: resumptive pronouns should either completely eliminate the superadditive interaction, or substantially decrease the size of the interaction relative to the size of the interaction with gaps. For languages that allow resumptive pronouns as a grammatical option, this would require a 2x2x2 design that uses the typical distance and structure factors, and adds a third factor with gap and resumption as levels (this design often devolves into a 2x2 because resumption languages typically do not allow resumptive pronouns in the matrix subject position, thus eliminating the short level from the distance factor). Tucker et al. (2019) used precisely such a design to study three island types (adjunct, complex NP, and wh-islands) with complex wh-phrases in Modern Standard Arabic. They found no reduction in the size of the superadditive interactions for resumptive pronouns versus gaps for any of the islands, contrary to the claim in the literature. For languages that do not allow resumptive pronouns as a grammatical option, the full factorial design cannot be used, because resumptive pronouns cannot appear in non-island structures (so there would be a non-monotonic interaction). Therefore most studies in these languages focus on the relative difference between resumptive pronouns and gaps within islands (sometimes with additional conditions showing that resumptive pronouns are unacceptable in non-island structures). In Table 10.3 below, we list formal experimental studies on resumptive pronouns, organized first by language, and then by dependency type (and again we apologize for any studies that we inadvertently left off). This table includes both types of languages: languages without grammatical resumption (intrusive resumption) appear first, and languages with grammatical resumption appear second (below the solid horizontal line). For all languages without grammatical resumption and for the Keshev and Meltzer-Asscher (2017) Hebrew results, the data in the cells reports the relative difference between resumptive pronouns and gaps within island structures. For the Tucker et al. (2019) Modern Standard Arabic results, the cells report the change in the size of the differences-in-differences score in a factorial design.

Table 10.3: A (potentially incomplete) summary of studies on resumptive pronouns using formal acceptability judgment methods. The scale column indicates the task: forced choice between two sentences (FC), comprehensibility (compr.) with the number of points on the scale indicated in parentheses, and acceptability (accept.) with the number of points on the scale indicated in parentheses. Most of the effect sizes are reported directly in the articles, but some we estimated from plots (and rounded to two significant digits). A dash (–) indicates no effect of resumption; and a blank cell indicates that the island was not tested.

study	language	dependency	adj	np	sub	rc	wh	scale
Alexopoulou & Keller 2007	English	bare wh				−	−	
Ackerman et al. 2018	English	complex wh	0.2			0.4	0.5	FC
Omaki & Nakao 2010	English	complex wh				−		
Beltrama & Xiang 2016	English	rel. clause				0.4		compr. (7)
Beltrama & Xiang 2016	English	rel. clause				−		accept. (7)
Heestand et al. 2011	English	rel. clause	−	−		−		
Keffala 2013	English	rel. clause				0.7	0.6	accept. (11)
Morgan & Wagers 2018	English	rel. clause	−	−			−	
Alexopoulou & Keller 2007	German	bare wh				−	−	
Alexopoulou & Keller 2007	Greek	bare wh				−	−	
Beltrama & Xiang 2016	Italian	rel. clause				0.4		compr. (7)
Keshev & Meltzer-Asscher 2017	Hebrew	rel. clause		1.2				accept. (7)
Tucker et al. 2019	MS Arabic	bare wh	−	−			0.1	z-score
Tucker et al. 2019	MS Arabic	complex wh	0.5	−			−	z-score

Three patterns emerge in Table 10.3. The first is that resumption effects are not particularly reliable. There is no island that shows a resumption effect every time that it is tested. The second is that the effect appears to be difficult to detect with rating-scale acceptability judgment tasks. Only one rating-scale sentence acceptability study detected an effect in a language that does not allow grammatical resumption (Keffala 2013). The other positive results are either for languages that allow resumption as a grammatical option, or for other tasks, such as the forced-choice task of Ackerman et al. (2018) or the comprehensibility task of Beltrama and Xiang (2016). The difficulty of detecting an effect in judgments, coupled with the relative ease of eliciting resumptive pronouns in production tasks (Ferreira and Swets 2005, Morgan and Wagers 2018), could itself be a piece of evidence about the nature of resumptive pronouns in languages that do not allow resumption. The third pattern is that the resumption effect is very small in languages that do not allow resumption as a grammatical option. In the Beltrama and Xiang (2016) and Keffala (2013) results, the resumption effect is only about 6% of the size of the scale of the task. In languages that do allow resumption as a grammatical option, the effect is about twice as large (Keshev and Meltzer-Asscher 2017, Tucker et al. 2019). Taken together, these results suggest that, while there is certainly some empirical support for the claim that resumptive pronouns increase the acceptability of island effects, the effect is far less robust, and far smaller, than traditional studies had suggested. In many ways, resumption resembles satiation – the variability across studies suggests that more research is necessary to understand the phenomenon; but that variability also suggests that researchers should exercise caution in deciding to invest time and resources into this topic.

5. Moving Forward

In this chapter we focused on three benefits of formal experiments, and how they have helped to expand our knowledge of island effects, from exploring the definition of island effects, to probing the source of island effects, to increasing the precision of the data on the various properties of island effects. In this last section, we would like to briefly mention some of the patterns that are emerging in recent work on island effects in the hopes that it may help researchers formulate their own studies. One major pattern that is emerging is the role that

differences in effect sizes may play in the theory. We see this both in fundamental questions such as whether subliminal island effects should be considered the same kind of effect as traditional island effects, and in higher order questions such as whether the smaller effect sizes that we see with complex wh-phrases should be interpreted as selectivity or not. Exploring these questions will require both a concerted effort to quantify effect sizes in a comparable way across languages and dependency types, and a concerted effort to create a theory of the factors that can influence effect sizes. This is no small challenge – we know of no domain of cognitive science that has a substantive theory of effect sizes. Another major pattern that is emerging is that cross-linguistic and cross-dependency variation is similar to what has been reported using traditional informal experiments, but not quite identical. This suggests that there is real empirical value in systematically re-testing languages for island effects, both to establish the range of variation across languages, and the range of variation within long-distance dependency types. Finally, it is apparent that much more work is needed to explore the space of possible sources of island effects, both within sentence processing and within the domains of grammatical theory. This work will likely require creative thinking about the types of predictions that these theories make beyond basic acceptability judgment patterns, as all of these theories can typically explain these basic facts. The studies that push the field forward will likely combine sentence acceptability studies with other data types, such as working memory measures, reading times, EEG, and semantic or pragmatic tasks.

References

Abrusán, M. (2014). *Weak island semantics*. Oxford University Press: Oxford.

Ackerman, L., Frazier, M. & Yoshida, M. (2018). Resumptive pronouns can ameliorate illicit island extractions. *Linguistic Inquiry*, 49: 847-859.

Alexopoulou, T. & Keller, F. (2007). Locality, cyclicity, and resumption: At the interface between the grammar and the human sentence processor. *Language* 83: 110-160.

Almeida, D. (2014). Subliminal wh-islands in Brazilian Portuguese and the consequences for syntactic theory. *Revista da ABRALIN*, 13.55–93. DOI: 10.5380/rabl.v13i2 .39611.

Atkinson, E., Apple, A. Rawlins, K. & Omaki, A. (2015). Similarity of wh-Phrases and Acceptability Variation in wh-Islands. *Frontiers in psychology* 6: 2048.

Beltrama, A. & Xiang, M. (2016). Unacceptable but comprehensible: the facilitation effect of resumptive pronouns. *Glossa* 1: 1.

Chaves, R. P. & Dery, J. E. (2014). Which subject islands will the acceptability of improve with repeated exposure? Proceedings of the 31st West Coast Conference on Formal Linguistics, ed. Robert E. Santana-La Barge, 96-106. Somerville, MA: Cascadilla Proceedings Project.

Christensen, K. R., Kizach, J. & Nyvad, A. M. (2013). Escape from the island: Grammaticality and (reduced) acceptability of wh-island violations in Danish. *Journal of Psycholinguistic Research* 42: 51-70.

Clifton C. & Frazier L. (1989) Comprehending Sentences with Long-Distance Dependencies. In Carlson G. N., Tanenhaus M. K., eds., *Linguistic Structure in Language Processing. Studies in Theoretical Psycholinguistics*, vol 7. Springer, Dordrecht, pp. 273-317.

Crain, S. & Fodor, J. D. (1985). How can grammars help parsers? In D. Dowty, L. Kartuunen, and A. Zwicky, eds., *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives*. Cambridge University Press: Cambridge.

Crawford, J. (2012). Using syntactic satiation to investigate subject islands. Proceedings of the 29th West Coast Conference on Formal Linguistics, eds., Jaehoon Choi et al.. Somerville, MA: Cascadilla Proceedings Project, pp. 38-45.

Culicover, P. W. & Postal, P. M. (Eds.). (2001). *Parasitic gaps* (Vol. 35). MIT Press.

Dąbrowska, E. (2010). Naive v. expert intuitions: An empirical study of acceptability judgments. *The linguistic review*, 27(1), 1-23.

Do, M. L. & Kaiser, E. (2017). The relationship between syntactic satiation and syntactic priming: A first look. *Frontiers in Psychology,* 8:1851.

Engdahl, E. (1983). Parasitic gaps. *Linguistic Inquiry*, 6, 5–34.

Engdahl, E. (1985). Parasitic gaps, resumptive pronouns, and subject extractions. *Linguistics*, 23, 3-44.

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

Featherston, S. (2005). Magnitude estimation and what it can do for your syntax: Some wh constraints in German. *Lingua,* 115, 1525–1550

Ferreira, F. & Swets, B. (2005). The production and comprehension of resumptive pronouns in relative clause "island" contexts. *Twenty-first century psycholinguistics: Four cornerstones*, pp. 263-278.

Francom, J. C. (2009). Experimental Syntax: exploring the effect of repeated exposure to anomalous syntactic structure--evidence from rating and reading tasks. PhD Dissertation, University of Arizona.

Frazier, L., & d'Arcais, G. B. F. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of memory and language*, 28(3), 331-344.

Freedman, S. E. & Forster, K. I. (1985). The psychological status of overgenerated sentences. *Cognition*, *19*(2), 101-131.

Garnsey, S., Tanenhaus, M. & Chapman, R. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research,* 18:51–60.

Goldberg, A. (2006). *Constructions at work*. London: Oxford University Press.

Goodall, G. (2011). Syntactic satiation and the inversion effect in English and Spanish wh-questions. *Syntax*, 14(1), 29-47.

Goodall, G. (2015). The D-linking effect on extraction from islands and non-islands. *Frontiers in Psychology,* 5: 1493.

Heestand, D., Xiang, M. & Polinsky, M. (2011). Resumption still does not rescue islands. *Linguistic Inquiry*, 42(1):138–152.

Hiramatsu, K. (2000). Accessing linguistic competence: Evidence from children's and adults' acceptability judgments. PhD dissertation, University of Connecticut, Storrs.

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

Hofmeister, P., Casasanto, L. S. & Sag, I. A. (2012). How do individual cognitive differences relate to acceptability judgments? A reply to Sprouse, Wagers, and Phillips. *Language,* 390-400.

Huang, C.T. J. (1982). Logical relations in Chinese and the theory of grammar. Ph.D. thesis, MIT.

Kaan, E., Harris, A., Gibson, E. & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes,* 15,159–201.

Kazanina, N., Lau, E. F., Lieberman, M., Yoshida, M., & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language,* 56: 384–409.

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

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

Keshev, M. & Meltzer-Asscher, A. (2017). Active dependency formation in islands: How grammatical resumption affects sentence processing. *Language,* 93, 549–68.

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

Kim, B. & Grant, G. (2016). Islands and non-islands in native and heritage Korean. *Frontiers in psychology*, *7*, 134.

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

Ko, H., Chung, H-B., Kim, K. & Sprouse, J. An Experimental Study on Scrambling out of Islands: To the Left and to the Right. *Language & Information Society* 37.

Kroch, A. (1981). On the role of resumptive pronouns in amnestying island constraint violations. In Hendrick, R. Maseh, C. & Miller, M., eds., *Chicago Linguistics Society* 17. University of Chicago Press: Chicago, pp.125-135.

Kurtzman, H. S. & Crawford, L. F. (1991). Processing parasitic gaps. Proceedings of the 21st Annual Meeting of the North Eastern Linguistics Society, ed. by Tim Sherer,. Amherst, MA: LSA Publications., pp. 217–231.

Kush, D., Lohndal, T. & Sprouse, J. (2018). Investigating variation in island effects: A case study of Norwegian wh-extraction. *Natural Language and Linguistic Theory,* 36, 743–79.

Kush, D., Lohndal, T. & Sprouse, J. (2019). On the Island Sensitivity of Topicalization in Norwegian: An experimental investigation. *Language 95*(3), 393-420.

Loftus, G. R. (1978). On interpretation of interactions. *Memory & Cognition,* 6, 312–319.

Lu, J., Thompson, C. & Yoshida, M. (2019). Chinese Wh-in-situ and Islands: A Formal Judgment Study. *Linguistic Inquiry*, 1-20.

Luka, B. J. & Barsalou, L. W. (2005). Structural facilitation: Mere exposure effects for grammatical acceptability as evidence for syntactic priming in comprehension. *Journal of Memory and Language*, 52, 436-459.

Matchin, W., Sprouse, J. & Hickok, G. (2014). A structural distance effect for backward anaphora in Broca's area: An fMRI study. *Brain and language*, 138, 1-11.

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2017). *Designing experiments and analyzing data: A model comparison perspective*. Routledge.

McCloskey, J. (2006). Resumption. In Everaert, M. and van Riemsdijk, H., eds.,
The Blackwell Companion to Syntax. Oxford: Blackwell Publishing, pp. 94–117.

McKinnon, R. & Osterhout, L. (1996). Event-related potentials and sentence processing:

Evidence for the status of constraints on movement phenomena. *Language and Cognitive Processes* 11, 495–523.

Michel, D. (2014). Individual cognitive measures and working memory accounts of syntactic island phenomena. PhD diss., University of California, San Diego.

Morgan, A. M. & Wagers, M. W. (2018). English resumptive pronouns are more common where gaps are less acceptable. *Linguistic Inquiry*, 49(4), 861-876

Neville, H. J., Nicol, J., Barss, A., Forster, K. I. & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience* 3: 151–165.

Omaki, A. & Nakao, C. (2010). Does English resumption really help to repair island violations? *Snippets* 21: 11–12.

Omaki, A., Lau, E.F., Davidson White, I., Dakan, M. L., Apple, A. & Phillips, C. (2015). Hyper-active gap filling. *Frontiers in psychology*, 6, 384.

Omaki, A., Fukuda, S., Nakao, C. & Polinsky, M. (2019). Subextraction in Japanese and subject-object symmetry. *Natural Language & Linguistic Theory*, 1-43.

Pesetsky, D. (1987). Wh-In-situ: movement and unselective binding. In The representation of (in)definiteness, eds., Eric Reuland and Alice ter Meulen. Cambridge: MIT Press, pp. 98–129.

Phillips, C. (2006). The real-time status of island constraints. *Language* 82: 795–823.

Phillips, C. (2013a). On the nature of island constraints. I: Language processing and reductionist accounts. Sprouse, Jon; Hornstein, Norbert, eds., *Experimental syntax and island effects*. Cambridge University Press: Cambridge, pp. 64-108.

Phillips, C. (2013b). On the nature of island constraints. II: Language learning and innateness. Sprouse, Jon; Hornstein, Norbert, eds., *Experimental syntax and island effects*. Cambridge University Press: Cambridge, pp. 132-157.

Pickering, M. J., Barton, S. & Shillcock, R. (1994). Unbounded dependencies, island constraints, and processing complexity. In C. Clifton, Jr., L. Frazier, and K. Rayner, eds., Perspectives on Sentence Processing. London: Erlbaum, pp. 199–224.

Poulsen, M. (2008). Acceptability and processing of long-distance dependencies in Danish. Nordic Journal of Linguistics 31: 73–107.

Reinhart, T. (1981). A second COMP position. In A. Belletti, L. Brandi, and L. Rizzi, eds., Theory of markedness in generative grammar. Pisa: Scuola Normale Superiore, pp. 517-557.

Rizzi, L. (1982). Violations of the wh-island constraint and the subjacency condition. In Issues in Italian syntax, ed. Luigi Rizzi. Dordrecht: Foris, pp. 49–76.

Rizzi, L. (1990). Relativized minimality. Cambridge: MIT Press.

Rizzi, L. (2013). Locality. Lingua 130, 169–186.

Ross, J. R. (1967). Constraints on variables in syntax. Cambridge, MA: MIT dissertation. [Published as Infinite syntax!, Norwood, NJ: Ablex, 1986.]

Sells, P. (1984). Syntax and semantics of resumptive pronouns. Amherst, MA: University of Massachusetts dissertation.

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

Sprouse, J. (2007). A program for experimental syntax. Ph.D. thesis, University of Maryland.

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

Sprouse, J., Fukuda, S., Ono, H. & Kluender, R. (2011). Reverse island effects and the backward search for a licensor in multiple wh-questions. *Syntax,* 14: 179–203.

Sprouse, J., Wagers, M. & Phillips, C. (2012). A test of the relation between working memory capacity and syntactic island effects. *Language* 88: 82–123.

Sprouse, J., Caponigro, I., Greco, C. & Cecchetto, C. (2016). Experimental syntax and the variation of island effects in English and Italian. *Natural Language and Linguistic Theory* 34.307–44.

Sprouse, J. & Almeida, D. (2017). Design sensitivity and statistical power in acceptability judgment experiments. *Glossa*, 2(1), 1.

Stepanov, A., Mušič, M. & Stateva, P. (2018). Two (non-) islands in Slovenian: A study in experimental syntax. *Linguistics*, 56(3), 435-476.

Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes* 3: 227–245.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48(3), 542-562.

Szabolcsi, A. & Zwarts, F. (1993). Weak islands and an algebraic semantics of scope taking. *Natural Language Semantics* 1: 235-284.

Szabolcsi, A. & Lohndal, T. (2017). Strong vs. Weak Islands. In The Wiley Blackwell Companion to Syntax, Second Edition, M. Everaert and H. C. Riemsdij, eds. doi:10.1002/9781118358733.wbsyncom008.

Torrego, E. (1984). On inversion in Spanish and some of its effects. *Linguistic Inquiry,* 15: 103–129.

Traxler, M. J. & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language,* 35: 454–475.

Tucker, M. A., Idrissi, A., Sprouse, J. & Almeida, D. (2019). Resumption ameliorates different islands differentially: Acceptability data from Modern Standard Arabic. Perspectives on Arabic Linguistics XXX: Papers from the annual symposia on Arabic Linguistics, Stony Brook, New York, 2016 and Norman, Oklahoma, 2017, ed. by Amel Khalfaoui and Matthew A. Tucker, 159–93. Amsterdam: John Benjamins.

Van Gompel, R. P. & Liversedge, S. P. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(1), 128.

Villata, S., Rizzi, L. & Franck, J. (2016). Intervention effects and relativized minimality: New experimental evidence from graded judgments. *Lingua,* 179: 76–96.

Wagenmakers, E., Krypotos, A., Criss, A. H. & Iverson, G. (2012). On the interpretation of removable interactions: A survey of the field 33 years after Loftus. *Memory and Cognition* 40: 145-160. https://doi.org/10.3758/s13421-011-0158-0

Wagers, M. W. & Phillips, C. (2009). Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics* 45: 395–433.

Yoshida, M., Kazanina, N., Pablos, L. & Sturt, P. (2014). On the origin of islands. *Language, Cognition and Neuroscience* 29: 761–770.