Reassessing a model of syntactic island acquisition

Avni Gulrajani
Department of Linguistics
1401 Marie Mount Hall
University of Maryland
College Park, MD 20742
avni@terpmail.umd.edu

Jeffrey Lidz
Department of Linguistics
1401 Marie Mount Hall
University of Maryland
College Park, MD 20742
jlidz@umd.edu

Abstract

This paper examines the limits of the learning model for syntactic islands from Pearl and Sprouse (2013), which challenges linguistic nativist perspectives by suggesting that island effects can be learned from language input and domain-general or learned abilities. Our investigations focus on sentences that would be ambiguous if there were no island constraints, where one conceivable interpretation violates an island constraint. A learner without any knowledge of islands could incorrectly treat the island-violating parses of such sentences as grammatical. We conducted simulations introducing these sentences in the model's input and also analyzed their frequency in the child-directed speech corpora used as the model's input. The results show that a small number of potentially island-violating sentences in the model's input impairs its ability to exhibit island effects, and potential island violations occur frequently enough in children's input to degrade the model's performance.

1 Introduction

Island effects have played a central role in controversies around nativism in linguistics. While many linguists have argued that they are entirely a consequence of innate linguistic knowledge, Pearl and Sprouse (2013) offer a different viewpoint. They developed a computational model that suggests that these effects can be learned through language input and various abilities which might be learned or domain-general, such as parsing sentences and calculating probabilities. This model warrants thorough scrutiny as it represents the first serious attempt to explain how knowledge of islands could possibly be learned. Understanding the limitations of this model could be helpful in developing improved models of the acquisition of islands, potentially leading to a more comprehensive understanding of islands overall.

This paper explores the limits of Pearl and Sprouse's model through computational simulations and an examination of children's linguistic input. It specifically focuses on how different assumptions about the learner's intake might affect the model's performance. Originally, Pearl and Sprouse tested their model with adult-like parses of sentences. Our analysis considers the possibility that learners could misparse sentences that would be ambiguous if there were no island constraints.

To understand this issue, consider the sentences in (1). Sentence (1a) is ambiguous because the wh-phrase could relate to either verb, leading to different interpretations about thinking or smiling. In contrast, sentence (1b) only allows the interpretation where "why" is associated with "wonder" because an island structure blocks the alternative interpretation. But a learner without any knowledge of islands might not know this about sentence (1b) and could misparse it in a way where the wh-phrase relates to the verb inside the island.

(1) a. Why does Leo think that Meredith smiles?
b. Why does Leo wonder whether Meredith smiles?

Throughout the rest of this paper, we will use the term "potential island violation" for a sentence like (1b), which is unambiguous in English but would be ambiguous if English had no islands.

Our results indicate that a very small number of potential island violations in the model's input hinders its ability to display island effects, and children's input contains a large enough number of these sentences to degrade the model's performance.

2 A description of syntactic islands

Languages allow certain dependencies to extend over any number of words or phrases; however, these dependencies can still be restricted by particular structures. The examples in (2) demonstrate that wh-dependencies can span many clauses, but example (3) shows that the same type of dependency cannot cross even a single wh-clause. In these examples and subsequent examples, the underscore represents the position associated with the wh-phrase (called the gap position).

(2) a. What does Meredith like ___?
b. What does Leo think that Meredith likes ___?
c. What does the teacher believe... that Leo thinks that Meredith likes ___?

(3) * What does Leo wonder why Meredith likes ___?

The structures that constrain these dependencies are called syntactic islands (Ross, 1967). Many types of structures have been identified as islands, including complex noun phrases, subjects, coordinate structures, adjuncts, and wh-clauses. Examples of these are shown in (4)-(8). The island structure in each example is shown in brackets.

(4) Complex NP: * What did he make [the claim that the teacher celebrated ___]?

(5) Subject: * What do [pictures of ___] make you happy?

(6) Coordinate structure: * What did she see [the elephant and ___]?

(7) Adjunct: * What did you smile [after she said ___]?

(8) Wh-clause * What did you ask [why she said ___]?

While these examples focus on wh-dependencies, islands also affect other kinds of dependencies, including tough movement, relative clauses, comparative deletion, and clefting. (Chomsky, 1977; Bresnan, 1975).

Many attempts have been made to create general theories explaining a variety of island effects. These theories vary, with some attributing islands to grammatical knowledge and others to factors like pragmatics of questions or sentence processing difficulties. Among the grammatical theories, one particularly noteworthy example is the Subjacency Condition (Chomsky, 1973), which restricts dependencies to positions separated by no more than one bounding node. A paraphrased version of its original definition is given in (9).

(9) Subjacency Condition:
No rule can involve X and Y in the structure:
$\ldots X \ldots [_a \ldots [_b \ldots Y \ldots] \ldots] \ldots$
where a and b are bounding nodes.

Often, island phenomena are used to support linguistic nativist perspectives because comprehensive theories of islands are stated in terms of highly abstract linguistic properties which are not directly observable to learners. Island structures and island-sensitive dependencies vary widely in their surface-level characteristics, which makes them difficult to explain using directly observable properties. However, a potential concern with such abstract theories is the learning puzzle they present. Learners must somehow converge on the same abstract representations even though many representations can be compatible with their experience (cf. Chomsky, 1975, Goodman, 1955). Nativist theories address the puzzle of acquiring such abstract knowledge by considering it a component of an innate language faculty.

3 Pearl and Sprouse's model

Contrasting with theories that attribute island effects mostly or entirely to innate linguistic knowledge, Pearl and Sprouse (2013) suggest that a substantial portion of the knowledge resulting in island effects can be learned through experience. Instead of relying on innate linguistic knowledge, their model requires several biases that are possibly either learned and domain-specific, or innate and domain-general. Since linguistic nativism depends on biases that are both innate and domain-specific at once, their model could possibly challenge this perspective.

3.1 The learning process

At the beginning of the learning process, the learner is able to identify wh-dependencies, which means knowing that a wh-phrase must correspond to a gap elsewhere in the sentence. When a sentence with a wh-dependency is encountered, the learner parses the sentence into a phrase structure tree and extracts a sequence of "container nodes," which are phrasal nodes in the tree that contain the gap but not the wh-phrase. While parsing sentences, CP nodes are subcategorized according to the lexical item that introduces the CP. Next, the sequence of container nodes is broken into smaller sequences of three container nodes, called trigrams. The learner records the individual frequencies of trigrams and the total number of trigrams observed throughout a period of time. A small smoothing constant of 0.5 is added to all trigram frequencies, so even unobserved trigrams have a frequency of 0.5.

A "grammaticality preference" for a sentence is calculated by multiplying the probabilities of all trigrams in its container node sequence. The probability of a trigram is estimated by dividing its frequency by the total number of observed trigrams.

Below is a walk-through of the process of learning and calculating a grammaticality preference, demonstrated with a specific sentence example.[1]

(10) Sentence: What do you think she saw?
Parsed sentence:
[$_{CP}$ What do [$_{IP}$ [$_{NP}$ you] [$_{VP}$ think [$_{CP}$ [$_{IP}$ [$_{NP}$ she] [$_{VP}$ saw ___]]]]]]
Container node sequence:
IP–VP–CP_{null}–IP–VP
Trigrams:
start–IP–VP
IP–VP–CP_{null}
VP–CP_{null}–IP
CP_{null}–IP–VP
IP–VP–end
Updating trigram counts:
add 1 each trigram count
add 5 to the number of trigrams observed
Calculating a grammaticality preference:
Grammaticality preference =
P(start–IP–VP) × P(IP–VP–CP_{null}) × P(VP–CP_{null}–IP) × P(CP_{null}–IP–VP) × P(IP–VP–end)

These learning biases enable the learner to generalize beyond the input while still avoiding ungrammatical sentences. Focusing exclusively on wh-dependencies and container node sequences ensures that the learner avoids learning from irrelevant information. Subcategorizing CPs is a necessary step in distinguishing certain island violations from grammatical sentences. Without this information, *whether* and adjunct island violations, which are characterized by $CP_{whether}$ and CP_{if} nodes, would be indistinguishable from grammatical dependencies that include CP_{that} or CP_{null} nodes. Keeping track of trigram probabilities and calculating the grammaticality of a dependency from the probabilities of its trigrams allows the learner's knowledge to extend beyond the specific sentences that have been observed. If a new sentence has a dependency containing frequent trigrams, it is perceived as grammatical even if the whole sentence or container node sequence has never been encountered before. Pearl and Sprouse note that although these biases are conducive to learning, some of them have no other obvious motivation. It's not obvious that a learner would know to pay close attention to small sequences of nodes involved in wh-dependencies without any prior knowledge that islands exist. Still, this model is important because it appears to demonstrate the possibility of acquiring knowledge of islands without innate island constraints.

[1] Grammaticality preferences are not necessarily calculated after each sentence observation, but the calculation process is included here for clarity.

3.2 The model's input

The input for the model consists of 200,000 container node sequences, randomly selected from a frequency distribution that represents approximately 21,000 wh-dependencies from four child-directed speech corpora: the Adam and Eve corpora from the Brown dataset (Brown, 1973), the Valian corpus (Valian, 1991), and the Suppes corpus (Suppes, 1974). The number 200,000 is Pearl and Sprouse's estimate of the number of wh-dependencies a child would encounter between the ages of 2 and 5. According to Pearl and Sprouse, this period spans the time from when children start recognizing wh-dependencies to when they exhibit knowledge of islands.

3.3 Measuring the success of the model

Pearl and Sprouse compared the model's grammaticality preferences to adult acceptability judgements in experiments from Sprouse et al. (2012). Here, island effects were defined as superadditive interactions between two factors: gap position (MATRIX or EMBEDDED) and structure (ISLAND or NON-ISLAND). Example (11) includes different combinations of gap position and structure for *whether* islands. The interaction is measured using the differences-in-differences score, which is calculated by subtracting the difference in the MATRIX conditions from the difference in the EMBEDDED conditions.

(11) a. MATRIX | NON-ISLAND: Who ___ thinks that Leo plays piano?
b. EMBEDDED | NON-ISLAND: What does Meredith think that Leo plays ___?
c. MATRIX | ISLAND: Who ___ wonders whether Leo plays piano?
d. EMBEDDED | ISLAND: * What does Meredith wonder whether Leo plays ___?

In addition to *whether* islands, Sprouse et al. also tested complex NP islands, subject islands, and adjunct islands. The results of these experiments show superadditive interactions for all four

Island type	MATRIX | NON-ISLAND	EMBEDDED | NON-ISLAND	MATRIX | ISLAND	EMBEDDED | ISLAND	Differences-in-differences
Subject	-1.21	-7.89	-1.21	-20.17	12.28
Complex NP	-1.21	-13.84	-1.21	-19.81	5.97
Whether	-1.21	-13.84	-1.21	-18.54	4.7
Adjunct	-1.21	-13.84	-1.21	-18.54	4.7

Table 1: Model's grammaticality preferences and differences-in-differences for four island types. To maintain consistency with Pearl and Sprouse's reported results, all values in this table are presented as log probabilities.

island types. Similarly, Pearl and Sprouse tested their model on the same sentence types and found superadditive patterns in the model's grammaticality preference scores for all island types tested. These scores and their differences-in-differences are shown in Table 1.

4 Interpreting the model's results

Before examining the model's response to potential island violations, it is important to clarify the extent of its success to begin with. Although the original results demonstrate the model's success at displaying island effects for four specific island types, it remains unclear whether it achieves a true separation of island structures from all other structures.

The model's probability-based grammaticality preference scores are used as replacements for both acceptability and grammaticality at once, although the exact relationships between these concepts are not straightforward (see Phillips, 2013 for discussion). Since the model is not designed to encompass all aspects of acceptability judgements, there are noticeable differences between its scores and true acceptability judgements. For example, experiments from (Sprouse et al., 2012) show that the presence of an island structure outside a wh-dependency affects acceptability, but the model does not display this pattern because it ignores all properties of a sentence other than the nodes in its dependency. This might be appropriate if the model is only supposed to detect differences in grammaticality; however, the model also seems to capture some acceptability judgement patterns that go beyond grammaticality alone, such as the effect of a dependency's length. In general, the model assigns lower scores to longer dependencies because it involves multiplying many probabilities between 0 and 1.[2]

Since it is unclear which exact components of acceptability judgements the model's scores are supposed to represent, it could be more productive to focus on the broader idea that learning to identify islands involves separating them from all other structures in some way. If there is a detectable pattern in the input that distinguishes islands from non-island structures, then the model's scores should reflect this distinction somehow, regardless of how exactly they relate to acceptability and grammaticality. According to Pearl and Sprouse, the definition of an island effect is a superadditive pattern. So, islands should be associated with stronger superadditive patterns than non-island structures if the model is successful.

Using this definition, the model does not achieve a perfect separation of islands from other structures. Superadditive patterns appear even when comparing sentences without island violations, suggesting that this measure is susceptible to false positives. Because the model is unaffected by island structures outside of wh-dependencies, the differences-in-differences measurement effectively reduces to a single difference, and a superadditive pattern appears with any difference at all between two probabilities. Since the model prefers shorter dependencies, and trigram probabilities naturally vary widely, these differences appear in nearly any pair of sentences compared. Table 2 presents a variety of similarly acceptable sentence pairs whose differences in grammaticality preference scores exceed those associated with island violations.[3] Although it might be impractically difficult to create a complete model of acceptability judgements, verifying the model's success still requires an explanation of why its superadditive effects are relevant in situations involving island violations but not in others. Without this explanation, it seems that the model

[2]The model's preference for shorter dependencies might initially seem desirable, since acceptability judgements share this pattern. However, the underlying reasons for these preferences are quite different. Long dependencies are rated as less acceptable because of parsing difficulties that are unrelated to probability (Gibson, 1998; Sprouse, 2020). By attributing these low ratings entirely to probability, the model possibly overestimates the impact of probability on acceptability.

[3]Although we haven't run experiments showing that these sentences are similar in acceptability, it seems unlikely that they would show differences as large as true island effects.

Sentence 1	Sentence 2	Difference in grammaticality preferences
What did she think he saw? IP–VP–CP_{null}–IP–VP	What did she think that he saw? IP–VP–CP_{that}–IP–VP	6.61
What did she think about? IP–VP–PP	What did she think about seeing? IP–VP–PP–IP–VP	10.43
What did she see? IP–VP	What did she see a picture of? IP–VP–NP–PP	9.93
What was she hoping to see? IP–VP–IP–VP	What was she happy to see? IP–VP–AdjP–IP–VP	14.77
What did she want him to see? IP–VP–IP–VP	What did she hope for him to see? IP–VP–CP_{for}–IP–VP	11.08
What did she allow him to see? IP–VP–IP–VP	What did she give him a chance to see? IP–VP–NP–IP–VP	11.64
What did she think he saw? IP–VP–CP_{null}–IP–VP	What did she feel like he saw? IP–VP–PP–CP_{null}–IP–VP	7.83

Table 2: Differences in log probabilities of similarly acceptable sentences. Below each sentence is its container node sequence.

cannot easily distinguish between these.

4.1 Unobserved trigrams

It is possible that slight adjustments to the learning procedure could result in a clearer separation of islands from other structures. Pearl and Sprouse mention an important distinction between island violations and grammatical dependencies: island violations always contain at least one trigram that has never been observed, whereas grammatical dependencies consist of trigrams that have been observed previously, even if infrequently. To differentiate these cases, they suggest calculating grammaticality preferences in ways that penalize unseen trigrams more strongly. For example, instead of taking the product of trigram probabilities, grammaticality preferences could be calculated using the geometric mean instead, which moderates the impact of multiplying many probabilities. Another possible solution is to lower the smoothing constant to a much smaller number, which further decreases the probabilities of unobserved trigrams, and consequently any dependencies containing these trigrams. A third idea is that the learner could "simply note the presence of a very low-probability trigram," instead of aggregating trigram probabilities.

However, a potential remaining problem with all of these suggestions is that they all depend on island violations containing unobserved trigrams. If the model's input includes even a single island violation, the model could still fail to differentiate the island violation from other rare grammatical dependencies even after employing these strategies. This is particularly likely if potentially island-violating sentences are parsed incorrectly. The next section focuses on this issue.

5 Addressing potential island violations

We explored the impacts of potential island violations in the model's input using two approaches. First, we conducted simulations where we incorporated varying numbers of possibly island-violating sentences in the model's input, regardless of their presence in children's actual input. The purpose of these simulations was to assess the model's capacity to handle potential island violations and identify the number of potential island violations that would cause it to be unable to display island effects. Second, we searched through the four child-directed speech corpora used as input for potential island violations and included their island-violating parses in the model's input. This analysis was intended to determine the frequency of potential island violations in children's input and whether a model with limited tolerance for island violations could still succeed.

In both of these investigations, it was necessary to modify the model's process for selecting input container node sequences so that it could accommodate ambiguous sentences. Originally, each sentence was represented by a single container node sequence, and 200,000 sequences were randomly chosen one at a time from this collection. In our new setup, sentences are represented as groups of container node sequences, and the selection procedure involves selecting a sentence and one of its possible container node sequences randomly, meaning each parse for a particular sentence has an equal chance of being selected. This might overestimate the chance that a learner would misparse potential island violations, but we want to consider the worst-case scenario to understand the full range of possibilities (contrasting with Pearl and Sprouse, who focused on the best-case scenario). In the absolute worst case, learners would consistently choose island-violating parses, but this situation seems unlikely. Instead, we are considering a more realistic worst-case scenario where learners are completely unbiased.

5.1 Simulations

For each island type, we attempted to include potential island violations with the exact EMBEDDED | ISLAND container node sequences used by Pearl and Sprouse. This regime required sentences with adjunct wh-phrases and verbs inside island structures. Consequently, it was possible to find such sentences for all island types except subject islands, which are typically nominal. Examples of the types of potential island violations identified are shown in (12), (13), and (14), along with the container node sequences of the island-violating parses.

(12) Complex NP island:
Why did Meredith make the claim that Leo plays piano?
Grammatical: IP–VP
Island-violating: IP–VP–NP–CP_{that}–IP–VP

(13) Whether island:
Why does Meredith wonder whether Leo plays piano?
Grammatical: IP–VP
Island-violating: IP–VP–$CP_{whether}$–IP–VP

(14) Adjunct island:
How does Meredith smile if Leo plays piano?
Grammatical: IP–VP
Island-violating: IP–VP–CP_{if}–IP–VP

We conducted a separate simulation for each island type and examined the model's grammaticality preference scores for each pair of EMBEDDED | ISLAND and EMBEDDED | NON-ISLAND sentences after including different numbers of island-violating parses. We ignored the MATRIX gap position conditions because the model always rates them as the same. The EMBEDDED | NON-ISLAND baseline for these three island types is IP-VP-CP_{that}-IP-VP. The results are displayed in Figure 1 and explained below.

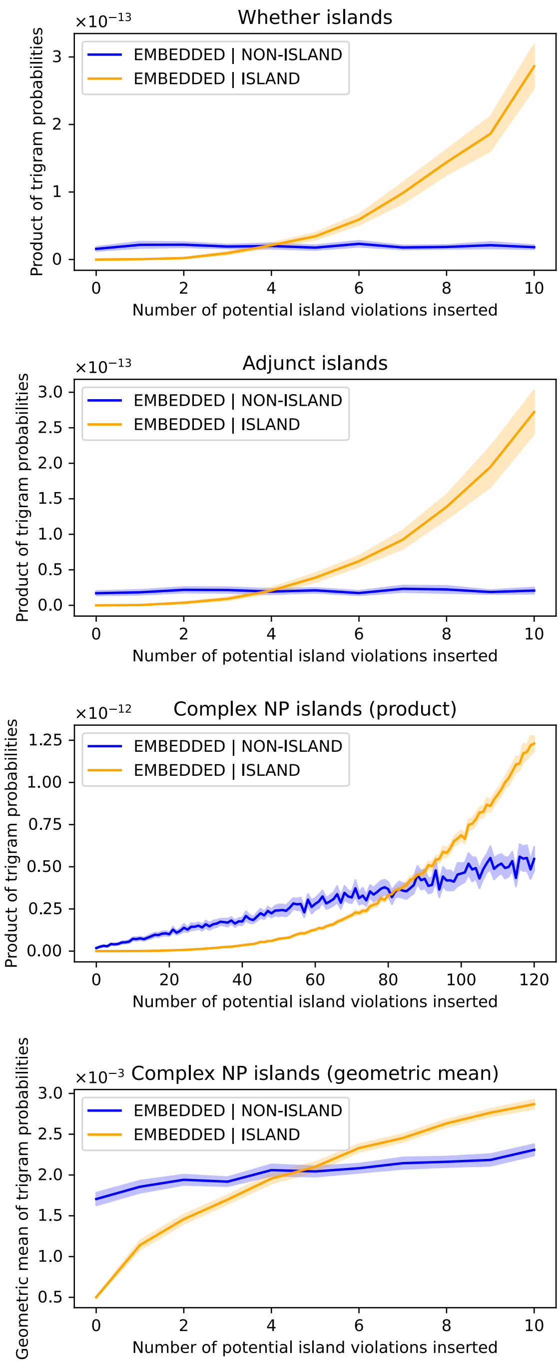

Figure 1: Grammaticality preferences with varying numbers of island violations in the input. Each pair of points represents the average of 50 repetitions of the model. Colored areas represent 95% confidence intervals. These charts display raw probabilities instead of log probabilities for clearer visualization.

For *whether* islands and adjunct islands, including just five island violations of each type results in higher scores for island-violating sentences than the grammatical baseline. Complex NP island effects might better withstand island violations in the input for two reasons. First, the grammatical sentence has an advantage because of its shorter container node sequence. Second, the two container node sequences share many trigrams, so observing island violations actually increases the score of the grammatical sentence. As a result of these two issues, this island effect is quite persistent; it remains until approximately 90 potential island violations are inserted. However, when using alternative measurements that do not favor shorter dependencies, like the geometric mean, the complex NP island effect disappears with just five potential island violations.

While these simulations clearly demonstrate that *whether* and adjunct island effects disappear with a small number of island violations in the input, interpreting the results for complex NP islands depends heavily on the exact method used to calculate grammaticality preferences. Using Pearl and Sprouse's original approach, it might seem like a small number of island violations has no serious impact on the complex NP island effect. However, as explained in Section 4, this approach leads to difficulties in differentiating between true island violations and uncommon grammatical dependencies. To achieve a clearer separation, several solutions were suggested which all focus on penalizing unseen trigrams, since this is the only unique characteristic of island violations that this model can detect. Because these solutions depend on island violations containing unseen trigrams, introducing even very few potential island violations lands us back at the original problem. For this reason, even a small number of island violations in the input might present problems for the model overall.

The reason only five potential island violations are required to eliminate these island effects is because the baseline grammatical sequence includes a rare container node, CP_{that}, which only appears twice in the entire input corpus. We selected this container node sequence to remain consistent with Pearl and Sprouse's original tests, but it's worth considering what might have happened if we had used a more common baseline, such as one with CP_{null}. In this situation, more potential island violations would be required to undo the island effects, but the challenge of distinguishing island violations from rare grammatical dependencies would remain the same.

5.2 Children's input

After examining the child-directed speech corpus used as the model's input, we found several different types of potential island violations, presented in Table 3.

We included the island-violating container node sequences for each potential island violation and retested the model with this revised input. We tested various island types, including two of the four types tested by Pearl and Sprouse, excluding subject and *whether* islands because of their ab-

Island type	Example sentence	Count
Complex NP	Adam, how would I know that those are the wheels that go on here?	24
Adjunct	How can he sit comfortably if you take all the pillows off?	69
Wh	How do you know what we find at the carnival?	35
Extraction from NP	What do you build a ship with?	68
Coordinate structure	How can the tiger be so healthy and fly like a kite?	151

Table 3: Types and frequencies of potential island violations in children's input, with examples from the input corpus

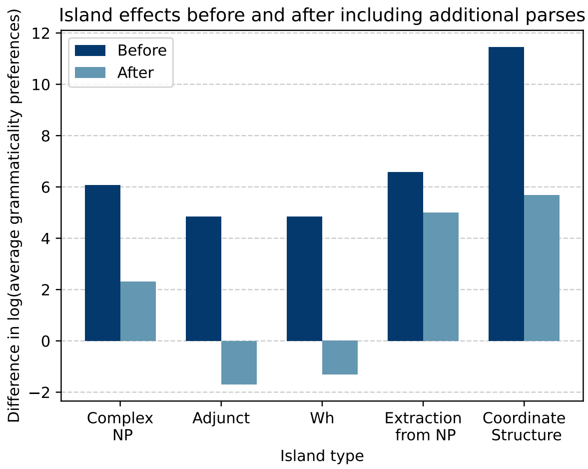

Figure 2: Differences in average grammaticality preferences (transformed to log probabilities) before and after inserting island-violating parses from children's input. Each bar represents the result from 1,000 model runs.

sence in the input corpus. Each test involved a single comparison of an island violation and a similar grammatical sentence. The complete list of test sentences is shown in Table 4. Figure 2 displays the model's grammaticality preferences for these test sentences before and after inserting the island-violating parses. These results indicate that the potential island violations in children's input can impair the model's ability to recognize several island types, although some island effects remain.

The results for complex NP and adjunct islands are consistent with the simulation results presented earlier. The island effect for adjuncts beginning with "if" disappears entirely because the input contains many instances of these. Adjuncts beginning with "when," "while," and "so" are similarly af-

Island type	Non-island sentence	Island sentence
Complex NP	What did he claim that she saw? IP-VP-CP_{that}-IP-VP	What did he make the claim that she saw? IP-VP-NP-CP_{that}-IP-VP
Adjunct	What did he think that she saw? IP-VP-CP_{that}-IP-VP	What did he worry if she saw? IP-VP-CP_{if}-IP-VP
Wh	What did he think that she saw? IP-VP-CP_{that}-IP-VP	What did he wonder when she saw? IP-VP-CP_{when}-IP-VP
Extraction from NP	What did he see with? IP-VP-PP	What did he see the elephant with? IP-VP-NP-PP
Coordinate structure	What did he see? IP-VP	What did he see an elephant and hear? IP-VP-VP

Table 4: Test sentences and container node sequences

fected. Other adjuncts exhibit small island effects with score differences less than 2. The complex NP island effect only partially remains. The score difference decreases to 2.31, which is smaller than many differences found between grammatical sentences.

Similar to adjuncts, wh-islands are also affected, but not uniformly. Because CPs are subcategorized by their initial words, certain wh-words form islands while others do not.[4] Our test sentence contains an embedded clause beginning with "when," which appears often enough in the input that the model does not consider it an island. However, embedded "why" questions are rare, so the model still treats these as islands.

Some extractions from NPs are grammatical while others are not, as shown by the examples in (15), and linguists have not conclusively determined the underlying distinctions between these (Davies and Dubinsky, 2003).

(15) a. What did you see [a picture of ___]?
b. * What did you see [the elephant with ___]?

Since they share identical container node sequences, the model is unable to differentiate between grammatical and ungrammatical extractions from NPs and instead generally rates them low because they contain uncommon trigrams. This effect partially remains despite many potential ungrammatical extractions from NPs in the input, although its size is smaller than some differences between grammatical sentences. If the model were enhanced in such a way that it could differentiate between grammatical and ungrammatical extractions from NPs, potential island violations could become problematic. There are 68 potential ungrammatical extractions from NPs compared to only 8 grammatical ones. The larger number of ungrammatical extractions suggests that they could interfere with learning.

The model's ability to recognize coordinate structure island violations is uncertain to begin with. Although our test shows a large difference for this island type, the probability of a coordinate structure island violation even before adding island violations to the input is higher than that of many grammatical dependencies, such as two-clause dependencies. The difference in our test sentences partially remains after inserting island violations, probably because the baseline container node sequence is shorter and overlaps with the island-violating sequence, similar to complex NP islands and extractions from NPs. However, its size diminishes to a value smaller than some grammatical sentences display. Every grammatical sentence pair in Table 2 from Section 4 exhibits a larger score difference.

In summary, the impact of incorporating island-violating parses varies: certain adjunct and wh-island effects disappear entirely; complex NP, coordinate structure, extraction from NP, and other adjunct and wh-island effects are substantially reduced; and subjects and whether-clauses continue to display island effects. It is important to recognize that these tests were conducted using Pearl and Sprouse's original method for calculating grammaticality preferences. If we had used alternative approaches, particularly ones that focus on unobserved trigrams, any potential island violations would have removed the island effects entirely. In this situation, only subject and *whether* island effects would remain, because only these islands contain unobserved trigrams.

[4] It's not entirely clear that this is how CPs are subcategorized. According to Pearl and Sprouse, the category depends on the word that "introduces" the CP, which could mean either the complementizer or the first word. However, using the complementizer would cause the model to fail to recognize all wh-islands because wh-words are not complementizers.

6 Conclusion

This paper has concentrated on exploring the limits of Pearl and Sprouse's model, focusing on sentences with potential island violations. Undertaking this analysis is important because their model represents a serious effort to explain how knowledge of islands could be learned from experience. Two potential problems have been identified here: the challenge of distinguishing true island violations from grammatical dependencies with low probabilities, and the possibility that sentences with potential island violations could be misparsed. Resolving these issues is important for a comprehensive understanding of island acquisition.

Of course, our simulations reflect a kind of worst case scenario by treating each potential island violation as though each parse had an equal chance of being selected. It may be that the impact of these sentences could be reduced by semantic and pragmatic factors. For example, we can imagine a learning scenario in which the child uses the discourse context to estimate the intended interpretation independent of the parse. Such a child could then use that interpretive estimate as a factor in deciding on a parse, possibly lessening the impact of the potential island violations.

It is also worth noting that the majority of the potential island violations come from adjunct questions, where there is not an independent source (such as argument structure) to identify the extraction site. It could be that learners down-weight evidence from adjunct questions precisely because they lack an independent means of verifying the extraction site. We can also imagine an enriched version of the Pearl and Sprouse model that tracks extraction paths separately for argument wh-phrases and adjunct wh-phrases. Such a model could also down-weight evidence from extraction paths that only occur for adjunct wh-phrases, on the assumption that the locality domains for adjunct wh-phrase should not be less restrictive than the locality domains for argument wh-phrases. Of course, such a model would be quite distinct in spirit from the original Pearl and Sprouse model.

Acknowledgements

We are grateful to Lisa Pearl for providing the code for the model and for her invaluable comments and feedback on this project.

References

Joan W. Bresnan. 1975. Comparative deletion and constraints on transformations. *University of Massachusetts Occasional Papers in Linguistics*, 1:Article 4.

Roger Brown. 1973. *A First Language: The Early Stages*. Harvard University Press, Cambridge, MA.

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

Noam Chomsky. 1975. *Reflections on Language*. Pantheon Books, New York.

Noam Chomsky. 1977. On wh-movement. In Peter W. Cullicover, Thomas Wasow, and Adrian Akmajian, editors, *Formal Syntax*, pages 71–132. Academic Press, New York.

William D. Davies and Stanley Dubinsky. 2003. On extraction from NPs. *Natural Language & Linguistic Theory*, 21(1):1–37.

Edward Gibson. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68:1–76.

Nelson Goodman. 1955. *Fact, Fiction, and Forecast*. Harvard University Press, Cambridge, MA.

Lisa Pearl 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.

Colin Phillips. 2013. On the nature of island constraints ii: Language learning and innateness. In Jon Sprouse and Norbert Hornstein, editors, *Experimental syntax and island effects*, pages 132–158. Cambridge University Press, Cambridge.

John Robert Ross. 1967. *Constraints on Variables in Syntax*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.

Jon Sprouse. 2020. A user's view of the validity of acceptability judgments as evidence for syntactic theories. In Samuel Schindler, Anna Drożdżowicz, and Karen Brøcker, editors, *Linguistic Intuitions: Evidence and Method*. Oxford University Press, Oxford.

Jon Sprouse, Matthew Wagers, and Colin Phillips. 2012. A test of the relation between working-memory capacity and syntactic island effects. *Language*, 88:82–123.

Patrick Suppes. 1974. The semantics of children's language. *American Psychologist*, 29:103–114.

Virginia Valian. 1991. Syntactic subjects in the early speech of American and Italian children. *Cognition*, 40:21–81.