Syntactic Islands without Universal Grammar

Lisa Pearl & Jon Sprouse
lpearl@uci.edu, jsprouse@uci.edu

Input & Syntactic Acquisition 2012
January 5th, 2012

An induction problem by any other name…

One of the most controversial claims in linguistics is that children face an induction problem:

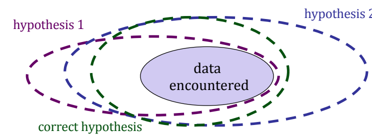
"Poverty of the Stimulus" (Chomsky 1980, Crain 1991, Lightfoot 1989, Valian 2009)

"Logical Problem of Language Acquisition" (Baker 1981, Hornstein & Lightfoot 1981)

"Plato's Problem" (Chomsky 1988, Dresher 2003)

Basic claim:

The data encountered are compatible with multiple hypotheses.

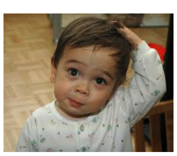


The induction problem

Extended claim:

Given this, the data are insufficient for identifying the correct hypothesis.

Big question: How do children do it?



One answer: Children come prepared

- Children are not unbiased learners.
- But if children come equipped with helpful learning biases, then what is the nature of these necessary biases?
  - Are they innate or derived from the input somehow?
  - Are they domain-specific or domain-general?
  - Are they about the hypothesis space or about the learning mechanism?

The Universal Grammar (UG) hypothesis (Chomsky 1965, Chomsky 1975):
These biases are innate and domain-specific.

The Plan

(1) Look at syntactic islands (central to UG-based syntactic theories).

(2) Explicitly define the target knowledge state, using adult acceptability judgments.

(3) Identify the data available in the input, using realistic samples.

(4) Implement a probabilistic learner that can learn about syntactic islands and see what kind of learning biases it requires.

Preview: **None** of the required biases are both innate and domain-specific (so syntactic islands don't implicate UG).

Syntactic Islands

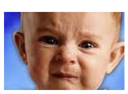
Dependencies can exist between two non-adjacent items, and these do not appear to be constrained by length (Chomsky 1965, Ross 1967).

What does Jack think __?

What does Jack think that Lily said that Sarah heard that Jareth believed __?

Syntactic Islands

However, if the gap position appears inside certain structures (called "syntactic islands" by Ross (1967)), the dependency seems to be ungrammatical.



Some example islands

Complex NP island:
*What did you make [the claim that Jack bought __]?

Subject island:
*What do you think [the joke about __] offended Jack?

Whether island:
*What do you wonder [whether Jack bought __]?

Adjunct island:
*What do you worry [if Jack buys __]?

Syntactic Islands

Predominant learning theory in generative syntax:
syntactic islands require innate, domain-specific learning biases.

Example: Subjacency
A dependency cannot cross two or more bounding nodes (Chomsky 1973, Huang 1982, Lasnik & Saito 1984).

Bounding nodes: language-specific (CP, IP, and/or NP)

Learning biases:
(1) Innate, domain-specific knowledge of hypothesis space: Exclude hypotheses that allow dependencies crossing 2+ bounding nodes.

(2) Innate, domain-specific knowledge of hypothesis space: Hypothesis space consists of bounding nodes for all languages, and the child must identify the ones applicable to her language.

The target state: Adult knowledge of syntactic islands

Sprouse et al. (2012) collected magnitude estimation judgments for four different islands, using a factorial definition that controlled for two salient properties of island-crossing dependencies:

- length of dependency (short vs. long)
- presence of an island structure (non-island vs. island)

The target state: Adult knowledge of syntactic islands

Sprouse et al. (2012) collected magnitude estimation judgments for four different islands, using a factorial definition that controlled for two salient properties of island-crossing dependencies:

- length of dependency (short vs. long)
- presence of an island structure (non-island vs. island)

Complex NP islands

Who __ claimed that Lily forgot the necklace?	short | non-island
What did the teacher claim that Lily forgot __?	long | non-island
Who __ made the claim that Lily forgot the necklace?	short | island
*What did the teacher make the claim that Lily forgot __?	long | island

The target state: Adult knowledge of syntactic islands

Sprouse et al. (2012) collected magnitude estimation judgments for four different islands, using a factorial definition that controlled for two salient properties of island-crossing dependencies:

- length of dependency (short vs. long)
- presence of an island structure (non-island vs. island)

Subject islands

Who __ thinks the necklace is expensive?	short | non-island
What does Jack think __ is expensive?	long | non-island
Who __ thinks the necklace for Lily is expensive?	short | island
*Who does Jack think the necklace for __ is expensive?	long | island

The target state: Adult knowledge of syntactic islands

Sprouse et al. (2012) collected magnitude estimation judgments for four different islands, using a factorial definition that controlled for two salient properties of island-crossing dependencies:

- length of dependency (short vs. long)
- presence of an island structure (non-island vs. island)

Whether islands

Who __ thinks that Jack stole the necklace?	short | non-island
What does the teacher think that Jack stole __ ?	long | non-island
Who __ wonders whether Jack stole the necklace?	short | island
*What does the teacher wonder whether Jack stole __ ?	long | island

The target state:
Adult knowledge of syntactic islands

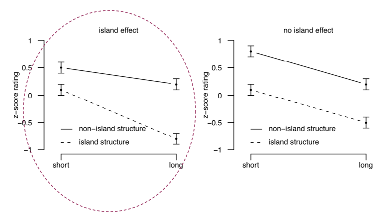
Sprouse et al. (2012) collected magnitude estimation judgments for four different islands, using a factorial definition that controlled for two salient properties of island-crossing dependencies:
- length of dependency (short vs. long)
- presence of an island structure (non-island vs. island)

Adjunct islands

Who __ thinks that Lily forgot the necklace?	short | non-island
What does the teacher think that Lily forgot __ ?	long | non-island
Who __ worries if Lily forgot the necklace?	short | island
*What does the teacher worry if Lily forgot __ ?	long | island

The target state:
Adult knowledge of syntactic islands

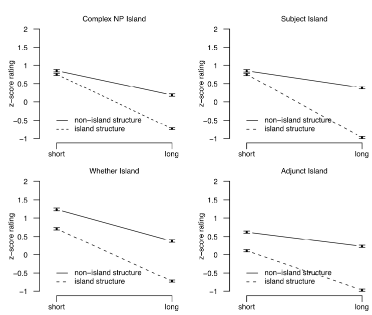
Syntactic island = superadditive interaction of the two factors (additional unacceptability that arises when the two factors are combined, above and beyond the independent contribution of each factor).



The target state:
Adult knowledge of syntactic islands

Sprouse et al. (2012)'s data on the four island types (173 subjects)

Superadditivity present for all islands tested
=
Knowledge that dependencies cannot cross these island structures is part of the adult knowledge state



The input: Induction problems

Data from three corpora of child-directed speech (Brown-Adam, Brown-Eve, Valian) from CHILDES (MacWhinney 2000): speech to 23 children between the ages of one and four years old.

Total words: 340,913

Utterances containing a *wh*-dependency: 11,308

Sprouse et al. (2012) stimuli types:

	SHORT | NON-ISLAND	LONG | NON-ISLAND	SHORT | ISLAND	*LONG | ISLAND* (*ungrammatical*)
Complex NP	4	177	0	0
Subject	4	13	0	0
Whether	4	177	0	0
Adjunct	4	177	3	0

The input: Induction problems

Data from three corpora of child-directed speech (Brown-Adam, Brown-Eve, Valian) from CHILDES (MacWhinney 2000): speech to 23 children between the ages of one and four years old.

Total words: 340,913

Utterances containing a *wh*-dependency: 11,308

Sprouse et al. (2012) stimuli types: *ungrammatical*

	SHORT | NON-ISLAND	LONG | NON-ISLAND	SHORT | ISLAND	*LONG | ISLAND*
Complex NP	4	177	0	0
Subject	4	13	0	0
Whether	4	177	0	0
Adjunct	4	177	3	0

These kinds of utterances are fairly rare in general - the most frequent appears about 0.016% of the time (177 of 11,308.)

The input: Induction problems

Data from three corpora of child-directed speech (Brown-Adam, Brown-Eve, Valian) from CHILDES (MacWhinney 2000): speech to 23 children between the ages of one and four years old.

Total words: 340,913

Utterances containing a *wh*-dependency: 11,308

Sprouse et al. (2012) stimuli types: *ungrammatical*

	SHORT | NON-ISLAND	LONG | NON-ISLAND	SHORT | ISLAND	*LONG | ISLAND*
Complex NP	4	177	0	0
Subject	4	13	0	0
Whether	4	177	0	0
Adjunct	4	177	3	0

Being grammatical doesn't necessarily mean an utterance will appear in the input at all.

The input: Induction problems

Data from three corpora of child-directed speech (Brown-Adam, Brown-Eve, Valian) from CHILDES (MacWhinney 2000): speech to 23 children between the ages of one and four years old.

Total words: 340,913

Utterances containing a *wh*-dependency: 11,308

Sprouse et al. (2012) stimuli types: *ungrammatical*

	SHORT | NON-ISLAND	LONG | NON-ISLAND	SHORT | ISLAND	*LONG | ISLAND*
Complex NP	4	177	0	0
Subject	4	13	0	0
Whether	4	177	0	0
Adjunct	4	177	3	0

Unless the child is sensitive to very small frequencies, it's difficult to tell the difference between grammatical and ungrammatical dependencies sometimes…

The input: Induction problems

Data from three corpora of child-directed speech (Brown-Adam, Brown-Eve, Valian) from CHILDES (MacWhinney 2000): speech to 23 children between the ages of one and four years old.

Total words: 340,913

Utterances containing a *wh*-dependency: 11,308

Sprouse et al. (2012) stimuli types: *ungrammatical*

	SHORT | NON-ISLAND	LONG | NON-ISLAND	SHORT | ISLAND	*LONG | ISLAND*
Complex NP	4	177	0	0
Subject	4	13	0	0
Whether	4	177	0	0
Adjunct	4	177	3	0

…and impossible to tell no matter what the rest of the time. This looks like an induction problem for the language learner.

Building a computational learner: Proposed learning biases

Learning Bias: Children track the occurrence of structures that can be derived from phrase structure trees - container nodes.

$[_{CP}$ Who did $[_{IP}$ she $[_{VP}$ like __]]]?
IP VP

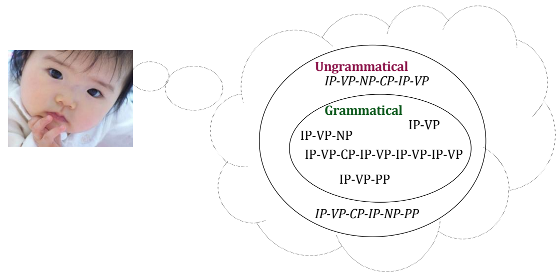
Container node sequence: IP-VP

$[_{CP}$ Who did $[_{IP}$ she $[_{VP}$ think $[_{CP}$ $[_{IP}$ $[_{NP}$ the gift] $[_{VP}$ was $[_{PP}$ from __]]]]]]]]?
IP VP CP IP VP PP

Container node sequence: IP-VP-CP-IP-VP-PP

Building a computational learner: Proposed learning biases

Children's hypotheses are about what container node sequences are grammatical for dependencies in the language.



Building a computational learner: Proposed learning biases

Children's hypotheses are about what container node sequences are grammatical for dependencies in the language.

Classification of learning bias:

Identifying container nodes
- applies to language data: domain-specific
- requires child to represent the hypothesis space a certain way
- derived from ability to parse utterances

Parsing utterances
- requires chunking data into cohesive units: likely to be innate and domain-general
- units being chunked are domain-specific phrasal units: derived from distributional data

Building a computational learner: Proposed learning biases

Learning Bias: Implicitly assign a probability to a container node sequence by tracking trigrams of container nodes. A sequence's probability is the smoothed product of its trigrams.

$[_{CP}$ Who did $[_{IP}$ she $[_{VP}$ like __]]]?
IP VP
start-IP-VP-end =
start-IP-VP
IP-VP-end

Probability(IP-VP) = p(start-IP-VP-end)
= p(start-IP-VP) * p(IP-VP-end)

Building a computational learner: Proposed learning biases

Learning Bias: Implicitly assign a probability to a container node sequence by tracking trigrams of container nodes. A sequence's probability is the smoothed product of its trigrams.

[$_{CP}$ Who did [$_{IP}$ she [$_{VP}$ think [$_{CP}$ [$_{IP}$ [$_{NP}$ the gift] [$_{VP}$ was [$_{PP}$ from __]]]]]]]]?
IP VP CP IP VP PP

start-IP-VP-CP-IP-VP-PP-end =
start-IP-VP
IP-VP-CP
VP-CP-IP
CP-IP-VP
IP-VP-PP
VP-PP-end

Probability(IP-VP-CP-IP-VP-PP) = p(start-IP-VP-CP-IP-VP-PP-end)
= p(start-IP-VP) * p(IP-VP-CP)*p(VP-CP-IP)*p(CP-IP-VP)
*p(IP-VP-PP)*p(VP-PP-end)

Building a computational learner: Proposed learning biases

Learning Bias: Implicitly assign a probability to a container node sequence by tracking trigrams of container nodes. A sequence's probability is the smoothed product of its trigrams.

What this does:
- longer dependencies are less probable than shorter dependencies, all other things being equal

- individual trigram frequency matters: short dependencies made of infrequent trigrams will be less probable than longer dependencies made of frequent trigrams

Effect: the frequencies observed in the input temper the detrimental effect of dependency length.

Building a computational learner: Proposed learning biases

Learning Bias: Implicitly assign a probability to a container node sequence by tracking trigrams of container nodes. A sequence's probability is the smoothed product of its trigrams.

Classification of learning bias:
- have enough memory to hold the utterance and its dependency in mind: innate and domain-general

- have enough memory to hold three units in mind (Mintz 2006, Wang & Mintz 2008, Saffran et al. 1996, Aslin et al. 1996, Saffran et al. 1999, Graf Estes et al. 2007, Saffran et al. 2008, Pelucchi et al. 2009a, 2009b): innate and domain-general

- track trigrams of units: innate, domain-general, learning mechanism

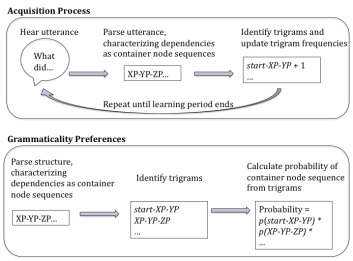
Building a computational learner: Proposed learning biases

None of the proposed learning biases are innate and domain-specific.

Description of process	Domain-specific	Domain-general	Innate	Derived
Parse utterance & identify dependencies	*			*
Identify container nodes	*			*
Extract trigram sequences		*	*	
Update probability of each trigram		*	*	
Calculate probability of utterance's dependency		*	*	

Building a computational learner: Proposed learning biases

Learning biases operate together to generate grammaticality preferences



Building a computational learner: Empirical grounding

Child-directed speech (Brown-Adam, Brown-Eve, Valian) from CHILDES: If we want to model child learners.

Adult-directed speech (Treebank-3-Switchboard corpus: Marcus et al. 1999) and text (Treebank-3-Brown corpus: Marcus et al. 1999): If we want to model adult learners (since we have adult data).

	Child-directed: speech	Adult-directed: speech	Adult-directed: text
total utterances	65932	74576	24243
total *wh*-dependencies	11308	8508	4230

Note: Child-directed speech and adult-directed speech are qualitatively similar in being mostly IP-VP and IP dependencies, with many more IP-VP dependencies.

Building a computational learner: Empirical grounding

Hart & Risley 1995: Children hear approximately 1 million utterances in their first three years.

Assumption: learning period for modeled learners is 3 years (ex: between 2 and 5 years old for modeling children's acquisition), so they would hear one million utterances.

Total learning period: 175,000 *wh*-dependency data points (rest of utterances heard do not contain *wh*-dependencies)

Success metrics

Compare learned grammaticality preferences to Sprouse et al. (2012) judgment data.

To do this, we need to identify the container node sequences for each stimuli for each island type.

Complex NP islands

IP	short | non-island
IP-VP-CP-IP-VP	long | non-island
IP	short | island
*IP-VP-NP-CP-IP-VP	long | island

Success metrics

Compare learned grammaticality preferences to Sprouse et al. (2012) judgment data.

To do this, we need to identify the container node sequences for each stimuli for each island type.

Subject islands

IP	short | non-island
IP-VP-CP-IP	long | non-island
IP	short | island
*IP-VP-CP-IP-NP-PP	long | island

Success metrics

Compare learned grammaticality preferences to Sprouse et al. (2012) judgment data.

To do this, we need to identify the container node sequences for each stimuli for each island type.

Whether islands

IP	short | non-island
IP-VP-CP-IP-VP	long | non-island
IP	short | island
*IP-VP-CP-IP-VP	long | island

Success metrics

Compare learned grammaticality preferences to Sprouse et al. (2012) judgment data.

To do this, we need to identify the container node sequences for each stimuli for each island type.

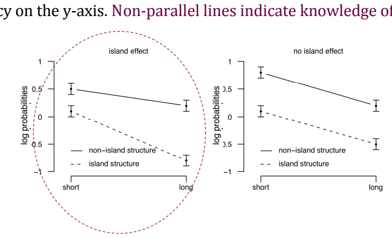
Adjunt islands

IP	short | non-island
IP-VP-CP-IP-VP	long | non-island
IP	short | island
*IP-VP-CP-IP-VP	long | island

Success metrics

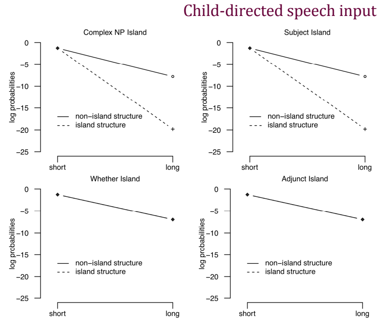
Compare learned grammaticality preferences to Sprouse et al. (2012) judgment data.

Then, for each island, we plot the predicted grammaticality preferences from the modeled learner on an interaction plot, using log probability of the dependency on the y-axis. Non-parallel lines indicate knowledge of islands.



The non-UG learner

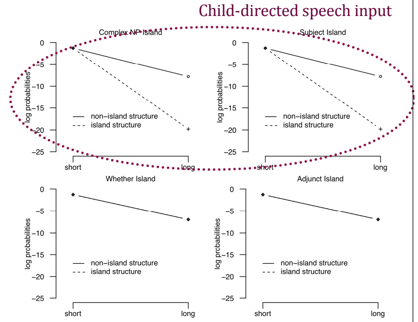
Child-directed speech input



The non-UG learner

Complex NP and Subject islands have the correct superadditive behavior…

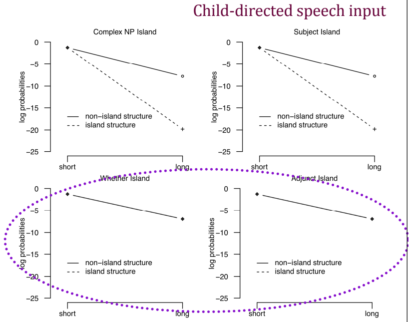
Child-directed speech input



The non-UG learner

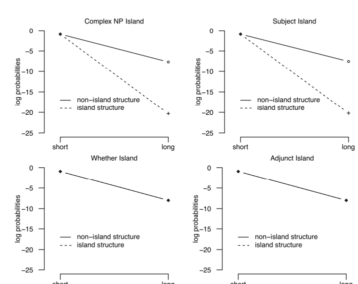
But Whether and Adjunct islands don't. In fact, the lines are overlapping - the learner thinks the grammatical long | non-island stimuli and ungrammatical long | island stimuli are equally good.

Child-directed speech input



The non-UG learner

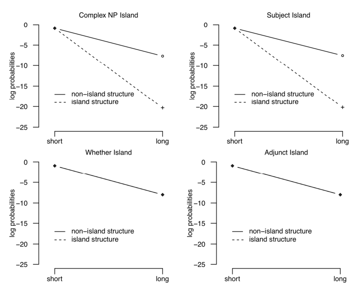
Adult-directed speech & text input



The non-UG learner

Adult-directed speech & text input

The same is true for adult-directed input: the learner has the correct preferences for Complex NP islands and Subject islands, but has the incorrect preferences for Whether and Adjunct islands.



The non-UG learner

Why do we see this behavior?

The learner does not distinguish between grammatical structures with the sequence IP-VP-$CP_{null/that}$-IP-VP

What did he think (that) she saw?

and structures with the ungrammatical sequence IP-VP-$CP_{whether/if}$-IP-VP

* What did he wonder whether/if she saw?

This means that Whether and Adjunct island violations, which contain specific types of CPs ($CP_{whether}$ and CP_{if}), are treated identically to grammatical utterances containing CP_{null} or CP_{that}.

The non-UG learner

Solution:
Have CP container nodes be more specified for the learner: CP_{null}, CP_{that}, $CP_{whether}$, CP_{if}, etc.

The learner can then distinguish between these structures:

IP-VP-$CP_{null/that}$-IP-VP
IP-VP-$CP_{whether/if}$-IP-VP

Does CP specification require UG?

Not necessarily:
• uncontroversial to assume that children learn to distinguish different types of CPs since the lexical content of CPs has substantial consequences for the semantics of a sentence (e.g., declaratives versus interrogatives)

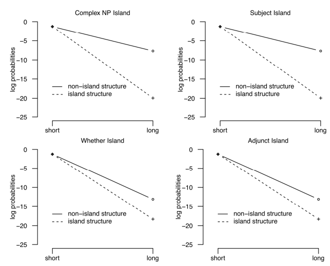
• adult speakers are sensitive to the distribution of *that* versus null complementizers (Jaeger 2010)

Likely a derived, domain-specific learning bias about the representation of the hypothesis space.

The non-UG learner

Using finer-grained container nodes: include CP specification
- ex: use CP_{null}, CP_{that}, etc.

Child-directed speech input

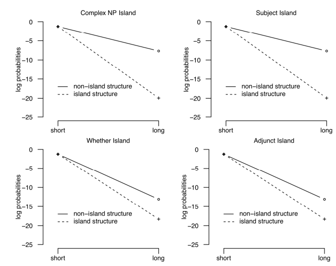


The non-UG learner

Using finer-grained container nodes: include CP specification
- ex: use CP_{null}, CP_{that}, etc.

Child-directed speech input

Problem solved! Superadditivity observed for all four island types.

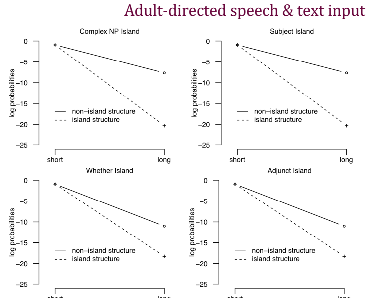


The non-UG learner

Using finer-grained container nodes: include CP specification
- ex: use CP_{null}, CP_{that}, etc.

Adult-directed speech & text input

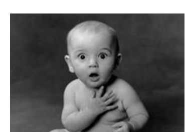
Same for adult-directed data: superadditivity observed for the ungrammatical island dependency.



Main implication of this learner

A learner using no biases that would traditionally be considered part of UG (i.e., both innate and domain-specific) was able to learn the correct grammaticality preferences for dependencies over four different island types. This suggests that adult knowledge of these syntactic islands does not implicate UG.

Though there appears to be an induction problem, it does not require UG to solve it.



Other implications & open questions

- It may be useful for children to have complex learning biases comprised of simpler learning biases.

- If children use a strategy similar to this learner's, predictions can be made about the acquisition trajectory of different islands.

- What about other more complex dependencies like parasitic gaps?

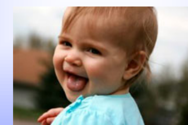
Thank You!

Jessica Lee
Uma Patel
Christine Thrasher
Cort Werner
Alec Finnell

Amy Rothermund
Sarah Hunt
Courtney Smith
Shannon Stanton
Galina Tucker

Caroline Wagenaar
Lyna Armendariz
Kristen Byer
Brian Toms
Joshua Bancroft

The National Science Foundation
EU PASCAL2 Network of Excellence



Implications of this learner

Something useful for children to have: Complex learning biases that are made up of simpler biases. (So, perhaps a bias to combine existing biases.)

Ex: Tracking trigrams of container nodes
- basic unit is container node (derived, domain-specific, hypothesis space)
- tracking 3 unit sequences (innate, domain-general, learning mechanism)

A developmental prediction

If children begin with only a basic specific of container nodes (CP instead of CP_{that}), we may expect a period of time when they recognize Complex NP and Subject islands but view dependencies spanning Whether and Adjunct islands as grammatical. Once they allow CP specification, they will recognize Whether and Adjunct islands as well.

Stage 1
* Complex NP island
* Subject island
Whether island
Adjunct island

→

Stage 2
* Complex NP island
* Subject island
* Whether island
* Adjunct island

de Villiers & Roeper (1995) suggest that children as young as 3 years old may view dependencies spanning *wh*-islands (such as *whether* islands) as ungrammatical. If they recognize *whether* islands as well, this suggests Stage 2 would be complete by this age.

A remaining issue

This learner can't handle parasitic gaps, which are dependencies that span an island (and so should be ungrammatical) but which are somehow rescued by another dependency in the utterance.

*Which book did you laugh [before reading __]?
Which book did you judge $\_\_{true}$ [before reading $\_\_{parasitic}$]?

Adjunct island

*What did [the attempt to repair __] ultimately damage the car?
What did [the attempt to repair $\_\_{parasitic}$] ultimately damage $\_\_{true}$?

Complex NP island

A remaining issue

Why not? The current learner would judge the parasitic gap as ungrammatical since it is inside an island, irrespective of what other dependencies are in the utterance.

*Which book did you laugh [before reading __]?
Which book did you judge $\_\_{true}$ [before reading $\_\_{parasitic}$]?

Adjunct island

*What did [the attempt to repair __] ultimately damage the car?
What did [the attempt to repair $\_\_{parasitic}$] ultimately damage $\_\_{true}$?

Complex NP island

This may be able to be addressed in a learner that is able to combine information from multiple dependencies in an utterance (perhaps because the learner has observed multiple dependencies resolved in utterances in the input).