How to succeed at syntactic island acquisition without really trying: Learning about local structure

Lisa Pearl
University of California, Irvine

March 31, 2023
Workshop on Locality in Theory, Processing, and Acquisition
University of Pennsylvania
What does it mean to succeed at syntactic island acquisition?
One answer: To develop the **target behavior** we observe about **syntactic islands**...
One answer: To develop the target behavior we observe about syntactic islands...

Some example behavior:
- judgment patterns and (dis)preferences for certain utterances related to syntactic islands

What [ [ _what ] ]?

✓
One answer: To develop the target behavior we observe about syntactic islands, given the input children get and the time they have to learn.
Acquisition success for syntactic islands
“…without really trying…”

What does it mean to try?

What \[ BN1 \[ BN2 \_what \] \]?
“without really trying”

Learn about syntactic islands indirectly by learning about *wh*-dependencies more generally.
"Learning about **local structure**"

Proposal: The relevant local structure is pieces that combine to build *wh*-dependencies.
“Learning about local structure”

Proposal: Learn about syntactic islands indirectly by learning about the probabilities of the pieces that build wh-dependencies.
Part 1: Learning the **probabilities** of pre-specified pieces from the input *(Pearl & Sprouse 2013)*.

*This turns out to work pretty well.*
Part 2: Learning what the pieces are and their probabilities from the input (Dickson, Pearl & Futrell 2022, in prep).

This turns out to work even better.
But first, let’s briefly review some relevant information about the acquisition of syntactic islands.

1: probabilities of pre-specified pieces

2: what the pieces are and their probabilities
Syntactic islands involve *wh*-dependencies.

This kitty was bought as a present for someone.

Lily thinks this kitty is pretty.

Who does Lily think the kitty for is pretty?

What does Lily think is pretty, and who does she think it’s for?
Syntactic islands involve *wh*-dependencies.

There's a dependency between the *wh*-word *who* and where it's understood (the gap).

*What’s going on here?*

*Who does Lily think the kitty for ___ who is pretty?*
Syntactic islands involve *wh*-dependencies.

What’s going on here?

There’s a dependency between the *wh*-word *who* and where it’s understood (*the gap*).

Who does Lily think the kitty for *who* is pretty?

This dependency is not allowed in English.

One explanation: The dependency crosses a “syntactic island” (Ross 1967)
Syntactic islands involve *wh*-dependencies.

Subject island (Ross 1967)

*Who does Lily think the kitty for *who* is pretty?*
Syntactic islands involve *wh*-dependencies.

Who does Lily think the kitty for _who is pretty_? [Subject island]

Jack is somewhat tricksy.

He claimed he bought something.

What did Jack make the claim that he bought _what_?
Jack is somewhat tricksy.
He claimed he bought something.
Elizabeth wondered if he actually did and what it was.

What did Elizabeth wonder whether Jack bought what?
Syntactic islands involve *wh*-dependencies.

 Rpc island (Ross 1967)

Who does Lily think the kitty for ___who is pretty?  Subject island

What did Jack make the claim that he bought ___what?  Complex NP island

What did Elizabeth wonder whether Jack bought ___what?  Whether island

Jack is somewhat tricksy.

He claimed he bought something.

Elizabeth worried it was something dangerous.

What did Elizabeth worry if Jack bought ___what?
Syntactic islands involve wh-dependencies.

Who does Lily think the kitty for __ who is pretty?  (Subject island)

What did Jack make the claim that he bought __ what?  (Complex NP island)

What did Elizabeth wonder whether Jack bought __ what?  (Whether island)

What did Elizabeth worry if Jack bought __ what?  (Adjunct island)

Important: It’s not about the length of the dependency.

(Chomsky 1965, Ross 1967)
Syntactic islands involve *wh*-dependencies.

Who does Lily think the kitty for *who* is pretty?  
**Subject island**

What did Jack make the claim that he bought *what*?  
**Complex NP island**

What did Elizabeth wonder whether Jack bought *what*?  
**Whether island**

What did Elizabeth worry if Jack bought *what*?  
**Adjunct island**

Important: It’s not about the length of the dependency.
Syntactic islands involve *wh*-dependencies.

**Subject island**

Who does Lily think the kitty for *who* is pretty?

**Complex NP island**

What did Jack make the claim that he bought *what*?

**Whether island**

What did Elizabeth wonder whether Jack bought *what*?

**Adjunct island**

What did Elizabeth worry if Jack bought *what*?

Important: It’s not about the length of the dependency.

What did Elizabeth think Jack said *what*?
Syntactic islands involve *wh*-dependencies.

**Who does Lily think the kitty for ___who is pretty?**  
Subject island

**What did Jack make the claim that he bought ___what?**  
Complex NP island

**What did Elizabeth wonder whether Jack bought ___what?**  
Whether island

**What did Elizabeth worry if Jack bought ___what?**  
Adjunct island

Important: It’s not about the length of the dependency.

**What did Elizabeth think Jack said Lily saw ___what?**
Syntactic islands involve \textit{wh}-dependencies.

Who does Lily think the kitty for \_who is pretty?  \textcolor{red}{Subject island}

What did Jack make the claim that he bought \_what?  \textcolor{red}{Complex NP island}

What did Elizabeth wonder whether Jack bought \_what?  \textcolor{red}{Whether island}

What did Elizabeth worry if Jack bought \_what?  \textcolor{red}{Adjunct island}

English adults \textit{judge} these island-crossing dependencies to be \textit{far worse} than many others, including others that are very similar except that they don’t cross syntactic islands (Sprouse et al. 2012).
Syntactic islands involve *wh*-dependencies.

English-learning children strongly *disprefer* one of these island-crossing dependencies compared to others (de Villiers et al. 2008).
Syntactic islands involve wh-dependencies.

Who does Lily think the kitty for _who is pretty?  [Subject island]

What did Jack make the claim that he bought _what?  [Complex NP island]

What did Elizabeth wonder whether Jack bought _what?  [Whether island]

What did Elizabeth worry if Jack bought _what?  [Adjunct island]

These judgments and (dis)preferences are a measurable observable behavior that can signal the successful acquisition of syntactic island knowledge.
Syntactic islands involve *wh*-dependencies.

So, these judgments and (dis)preferences can serve as a target for successful acquisition — an outcome we can measure.
Syntactic islands

Adult judgments
= behavioral target outcome

Adult knowledge as measured by **acceptability judgment** behavior

Sprouse et al. 2012: **magnitude estimation judgments**
- factorial definition controlling for two salient properties of island-crossing dependencies

<table>
<thead>
<tr>
<th>Length of dependency</th>
<th>Presence of an island structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>(matrix vs. embedded)</td>
<td>(non-island vs. island)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Who</th>
<th>Who</th>
</tr>
</thead>
<tbody>
<tr>
<td>_who?</td>
<td>[non-island]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Who</th>
<th>Who</th>
</tr>
</thead>
<tbody>
<tr>
<td>[CP… _who]?</td>
<td>[island]</td>
</tr>
</tbody>
</table>
Adult knowledge as measured by **acceptability judgment** behavior

- length of dependency (matrix vs. embedded) × presence of an island structure (non-island vs. island)

**Subject island stimuli**

- Who __ thinks [the necklace is expensive]?  
  matrix | non-island
- What does Jack think [ __ is expensive]?  
  embedded | non-island
- Who __ thinks [the necklace for Lily] is expensive?  
  matrix | island
- *Who does Jack think [the necklace for __ ] is expensive?  
  embedded | island
Syntactic islands
Adult judgments
= behavioral target outcome

Adult knowledge as measured by **acceptability judgment** behavior

- Length of dependency
- Presence of an island structure
  - (matrix vs. embedded) X (non-island vs. island)

**Whether island stimuli**

- Who __ thinks [that Jack stole the necklace]?   matrix | non-island
- What does the teacher think [that Jack stole __ ]?   embedded | non-island
- Who __ wonders [whether Jack stole the necklace]?   matrix | island
- *What does the teacher wonder [whether Jack stole __ ]?   embedded | island

Sprouse et al. 2012
Syntactic islands
Adult judgments
= behavioral target outcome

Adult knowledge as measured by acceptability judgment behavior

length of dependency \( \times \) presence of an island structure
(matrix vs. embedded) \( \times \) (non-island vs. island)

Adjunct island stimuli

Who __ thinks [that Lily forgot the necklace]? \hfill \text{matrix} \ | \ \text{non-island}
What does the teacher think [that Lily forgot __ ]? \hfill \text{embedded} \ | \ \text{non-island}
Who __ worries [if Lily forgot the necklace]? \hfill \text{matrix} \ | \ \text{island}
*What does the teacher worry [if Lily forgot __ ]? \hfill \text{embedded} \ | \ \text{island}

Sprouse et al. 2012
Adult knowledge as measured by acceptability judgment behavior

- length of dependency
- presence of an island structure
  - (matrix vs. embedded) × (non-island vs. island)

Complex NP island stimuli

- Who __ claimed [that Lily forgot the necklace]? matrix | non-island
- What did the teacher claim [that Lily forgot __]? embedded | non-island
- Who __ made [the claim that Lily forgot the necklace]? matrix | island
- *What did the teacher make [the claim that Lily forgot __ ]? embedded | island

Sprouse et al. 2012
Syntactic islands
Adult judgments
= behavioral target outcome

Adult knowledge as measured by acceptability judgment behavior

length of dependency (matrix vs. embedded) \times presence of an island structure (non-island vs. island)

Syntactic island = superadditive interaction of the two factors. This is additional unacceptability that arises when the two factors — length & presence of an island structure — are combined, above and beyond the independent contribution of each factor.
Syntactic islands
Adult judgments
= behavioral target outcome

Adult knowledge as measured by acceptability judgment behavior
length of dependency (matrix vs. embedded) \( \times \) presence of an island structure (non-island vs. island)

Syntactic island = \textit{superadditive} interaction of the two factors
Syntactic islands
Adult judgments
= behavioral target outcome

Adult knowledge as measured by acceptability judgment behavior
length of dependency \( \times \) presence of an island structure
(matrix vs. embedded) \( \times \) (non-island vs. island)

Syntactic island = superadditive interaction of the two factors

Who [non-island]?
Who [island]?
Who __who? [CP... __who]?
Syntactic islands
Adult judgments
= behavioral target outcome

Adult knowledge as measured by acceptability judgment behavior
length of dependency (matrix vs. embedded) \(\times\) presence of an island structure (non-island vs. island)

Syntactic island = superadditive interaction of the two factors

Who \[\text{[non-island]}\]?
Who \[\text{[island]}\]?

Who \[\text{[non-island]}\]?
Who \[\text{[island]}\]?

Who \[\text{[CP…]}\]?
Who \[\text{[who]}\]?
Syntactic islands
Adult judgments
= behavioral target outcome

Adult knowledge as measured by acceptability judgment behavior
length of dependency (matrix vs. embedded) \times presence of an island structure (non-island vs. island)

Syntactic island = superadditive interaction of the two factors

Sprouse et al. 2012
Syntactic islands
Adult judgments
= behavioral target outcome

Adult knowledge as measured by acceptability judgment behavior
length of dependency (matrix vs. embedded) X presence of an island structure (non-island vs. island)

Syntactic island = superadditive interaction of the two factors

Who [non-island]?
Who [island]?

Who [_who]?
Who [CP... _who]?
Syntactic islands
Adult judgments
= behavioral target outcome

Adult knowledge as measured by acceptability judgment behavior

- length of dependency (matrix vs. embedded)
- presence of an island structure (non-island vs. island)

Syntactic island = superadditive interaction of the two factors

Sprouse et al. 2012
Syntactic islands

Adult judgments
= behavioral target outcome

Adult knowledge as measured by acceptability judgment behavior

length of dependency (matrix vs. embedded) \( \times \) presence of an island structure (non-island vs. island)

Syntactic island = superadditive interaction of the two factors

What [ _what ]?

Who [non-island]?
Who [island]?

Who [CP… _who]?
Syntactic islands = superadditive interaction of the two factors

Sprouse et al. (2012): acceptability judgments from 173 adult subjects

Superadditivity for all four island types
**Syntactic islands**

Adult judgments
- = behavioral target outcome

Adult knowledge as measured by **acceptability judgment behavior**

- length of dependency
  - (matrix vs. embedded)
- presence of an island structure
  - (non-island vs. island)

Syntactic island = **superadditive** interaction of the two factors

Sprouse et al. (2012): acceptability judgments from 173 adult subjects

Superadditivity for all four island types

= knowledge that dependencies crossing these island structures are dispreferred.
Syntactic islands
Child judgments
= behavioral target outcome

Child knowledge as measured by preferred interpretation behavior

De Villiers et al. 2008:
How do children prefer to interpret potentially ambiguous wh-questions?
Child knowledge as measured by *preferred interpretation* behavior

How do children *prefer to interpret* potentially ambiguous *wh*-questions?
Child knowledge as measured by preferred interpretation behavior

How do children prefer to interpret potentially ambiguous *wh*-questions?

De Villiers et al. 2008
Child knowledge as measured by preferred interpretation behavior

How do children prefer to interpret potentially ambiguous *wh*-questions?

What did the boy fix the cat that was lying on the table with ___what?
Child knowledge as measured by preferred interpretation behavior

How do children prefer to interpret potentially ambiguous *wh*-questions?

What did the boy [fix the cat *that was lying on the table* [with __*what*]]?

*a needle and thread*
Child knowledge as measured by preferred interpretation behavior

How do children prefer to interpret potentially ambiguous \textit{wh}-questions?

\textbf{What} did the boy [fix [the cat [that [was [lying [on [the table [with \_	extit{what}]]]]]]]?
Child knowledge as measured by preferred interpretation behavior

How do children prefer to interpret potentially ambiguous *wh*-questions?

What did the boy fix the cat that was lying on the table with ___what?

Children strongly prefer this interpretation

*De Villiers et al. 2008*
Child knowledge as measured by preferred interpretation behavior

How do children prefer to interpret potentially ambiguous *wh*-questions?

*What* did the boy fix the cat that was lying on the table with *what*?

…and strongly disprefer this interpretation
Child knowledge as measured by preferred interpretation behavior

How do children prefer to interpret potentially ambiguous *wh*-questions?

_What_ did the boy [fix [the cat [that [was [lying [on [the table [with __what]]]]]]]]?

This means they strongly disprefer the *wh*-dependency this interpretation relies on.
Child knowledge as measured by preferred interpretation behavior.

How do children prefer to interpret potentially ambiguous *wh*-questions?

\[
\text{What did the boy } \text{fix } [\text{NP the cat } [\text{that } [\text{was } [\text{lying } \text{on } \text{the table } [\text{with } \_\text{what }]]]]]\?
\]

…which is a dependency that crosses a Complex NP.

De Villiers et al. 2008
Syntactic islands
Adult & child judgments
= behavioral target outcome

Sprouse et al. 2012

De Villiers et al. 2008
Syntactic islands

How long do children have to learn?
Syntactic islands

How long do children have to learn?

De Villiers et al. 2008: Data from four-year-olds.
Syntactic islands

How long do children have to learn?

So input through age four. (<60 months)
What input do children get?

Syntactic islands
What input do children get?

We can estimate this from samples of child-directed speech.
Syntactic islands

This is the acquisition problem
Syntactic islands

...which is where a theory of acquisition comes in.
Syntactic islands

...which is where a theory of acquisition comes in.

1: probabilities of pre-specified pieces

2: what the pieces are and their probabilities
Syntactic islands

We can evaluate a theory by implementing it concretely in a computational cognitive model.

1: probabilities of pre-specified pieces

2: what the pieces are and their probabilities
The model generates predictions that can be compared with available empirical data.

1: probabilities of pre-specified pieces

2: what the pieces are and their probabilities
Syntactic islands

And then we can look inside it to see what makes it work (or not work).

1: probabilities of pre-specified pieces

2: what the pieces are and their probabilities
Syntactic islands

So let’s do this for our two theories.

1: probabilities of pre-specified pieces

2: what the pieces are and their probabilities
Intuition:
- Learn what you can from the *wh*-dependencies you observe in the input over time
Intuition:

- Learn what you can from the *wh*-dependencies you observe in the input over time

- Apply it to generate behavior for *wh*-dependencies you haven’t seen before, like those crossing syntactic islands (or other longer *wh*-dependencies).

Pearl & Sprouse 2013
Dickson, Pearl, & Futrell 2022, in prep.
View *wh*-dependencies in terms of their **building blocks** and track those building blocks in the input.
What phrases contain the gap (but not the *wh*-word)?

Dependencies represented as a sequence of container nodes.

*Pearl & Sprouse 2013*  
Dickson, Pearl, & Futrell 2022, in prep.
What phrases contain the gap (but not the *wh*-word)?

What did you see __?  
= What did [IP you [VP see __]]?  
= start-IP-VP-end

Dependencies represented as a sequence of container nodes

Pearl & Sprouse 2013  
Dickson, Pearl, & Futrell 2022, in prep.
Dependencies represented as a sequence of container nodes.

What phrases contain the gap (but not the *wh*-word)?

What did you see _?  
= What did \[_{IP} \text{ you } _{VP} \text{ see } _{__}\]?  
= start-IP-VP-end

What _ happened?  
= What \[_{IP} _{__} \text{ happened}\]?  
= start-IP-end

Pearl & Sprouse 2013  
Dickson, Pearl, & Futrell 2022, in prep.
What phrases contain the gap (but not the wh-word)?

What did you see __?
= What did [IP you [VP see __]]?
= start-IP-VP-end

What __ happened?
= What [IP __ happened]?
= start-IP-end

What did she want to do __ ?
= What did [IP she [VP want [IP to [VP do __]]]]?
= start-IP-VP-IP-VP-end

Both theories

Pearl & Sprouse 2013
Dickson, Pearl, & Futrell 2022, in prep.
Both theories

What happened?
= What [IP you [VP see __]]?
= start-IP-VP-end

What she want to do __?
= What did [IP she [VP want [IP to [VP do __]]]]?
= start-IP-VP-IP-VP-end

What you see __?
= What did [IP you [VP see __]]?
= start-IP-VP-end

(Much) less acceptable dependencies have low probability segments

start-IP-VP-CP-that-IP-NP-PP-end

Pearl & Sprouse 2013
Dickson, Pearl, & Futrell 2022, in prep.
So if children break these dependencies into smaller building blocks, they can identify if a dependency has bad segments (made up of one or more low probability building blocks).

Pearl & Sprouse 2013
Dickson, Pearl, & Futrell 2022, in prep.
1: probabilities of pre-specified pieces

The building blocks: trigrams of container nodes
1: probabilities of pre-specified pieces

The building blocks: trigrams of container nodes

start-IP-VP-end
start-IP-VP
IP-VP-end
start-IP-VP-IP-VP-end
start-IP-VP-CP_that_IP-NP-PP-end
start-IP-end
syntactic trigrams
The building blocks: **trigrams of container nodes**

1: **probabilities of pre-specified pieces**

- start-IP-VP-end
- start-IP-VP-CP_{that}-IP-NP-PP-end
- start-IP-end
- start-IP-VP-IP-VP-end

**syntactic trigrams**

Pearl & Sprouse 2013
1: probabilities of pre-specified pieces

The building blocks: trigrams of container nodes

start-IP-VP-end  start-IP-VP-CP(that)-IP-NP-PP-end
start-IP-end   start-IP-VP-IP-VP-end

syntactic trigrams

start-IP-VP
IP-VP-IP
VP-IP-VP
IP-VP-end

start-IP-end
start-IP-end
1: probabilities of pre-specified pieces

The building blocks: trigrams of container nodes
Learning: Track the relative frequency of the syntactic trigrams in the input.
1: probabilities of pre-specified pieces

Some of them are common and some of them aren’t.
Some of them are common and some of them aren’t.
(And some never occur at all.)
1: probabilities of pre-specified pieces

Relative syntactic trigram frequency:

\[ p(t) \approx \frac{\# \text{trigram}}{\text{total} \# \text{trigrams}} \]
Any *wh*-dependency can then be constructed from its syntactic trigram building blocks.
1: probabilities of pre-specified pieces

\[ \prod_{t \in \text{trigrams}} p(t) \]

\( \text{start-IP-end} \)

\( \text{start-IP-VP} \)

\( \text{IP-VP-IP} \)

\( \text{VP-IP-VP} \)

\( \text{IP-VP-PP} \)

\( \text{VP-PP-end} \)

\( \text{IP-VP-CP}_{\text{that}} \)

\( \text{IP-NP-PP} \)

\( \text{NP-PP-end} \)

\( \text{start-IP-VP} \)

\( \text{start-IP-VP-end} \)

\( \text{start-IP-VP-CP}_{\text{that}} \)
1: probabilities of pre-specified pieces

\[ \Pi_{t \in \text{trigrams}} p(t) \]
A wh-dependency’s probability can stand in for its predicted acceptability or preference.
Lower probability dependencies are predicted to be less acceptable (dispreferred), compared to higher probability dependencies.
If we learn from the input children get the way this theory specifies, can this theory output the behavior children (should) produce?
Both theories

Evaluating the theory

What's the input look like?

Pearl & Sprouse 2013, Bates & Pearl 2019, Pearl & Bates in press
Evaluating the theory

Both theories

102K utterances ($\approx$21K wh-dependencies) from the CHILDES Treebank (Pearl & Sprouse 2013) of speech directed at 25 children between the ages of 1 and 5 years old.

Pearl & Sprouse 2013, Bates & Pearl 2019, Pearl & Bates in press
This lets us estimate which *wh*-dependencies children hear and how often they hear them (the *wh*-dependency distribution).
Both theories

Evaluating the theory

We can then estimate how many wh-dependencies children hear during the learning period.

(<60 months)

Pearl & Sprouse 2013, Bates & Pearl 2019, Pearl & Bates in press
Children begin to represent the full structure of *wh*-dependencies (e.g., *wh*-questions and relative clauses) around 20 months: Seidl et al. 2003, Gagliardi et al. 2016, Perkins & Lidz 2020.
Evaluating the theory

wh-dependency distribution

(20 months ≤ age < 60 months)

Educated guess: This is when children can start processing *wh*-dependencies reliably from their input.
Both theories

Evaluating the theory

How many minutes is this? In particular, children are awake for only a certain portion of the day at different ages (Davis et al. 2004).

Pearl & Sprouse 2013, Bates & Pearl 2019, Pearl & Bates in press
Evaluating the theory

Both theories

wh-dependency distribution

(20 months ≤ age < 60 months)

<table>
<thead>
<tr>
<th>age</th>
<th>age range</th>
<th>waking</th>
<th>total waking hours</th>
<th>cumulative waking hrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>one</td>
<td>20-23 months</td>
<td>10</td>
<td>11 hrs/day * 365 days/yr * 4/12 = 1216.67</td>
<td>1216.67</td>
</tr>
<tr>
<td>two</td>
<td>24-35 months</td>
<td>11</td>
<td>11 hrs/day * 365 days/yr = 4015</td>
<td>5231.67</td>
</tr>
<tr>
<td>three</td>
<td>36-47 months</td>
<td>12</td>
<td>12 hrs/day * 365 days/yr = 4380</td>
<td>9611.67</td>
</tr>
<tr>
<td>four</td>
<td>48-59 months</td>
<td>12.5</td>
<td>12.5 hrs/day * 365 days/yr = 4562.5</td>
<td>14174.17</td>
</tr>
</tbody>
</table>

cumulative waking mins
14174.17 * 60 min/hour

850450.2

Pearl & Sprouse 2013, Bates & Pearl 2019, Pearl & Bates in press
Evaluating the theory

Both theories

wh-dependency distribution

(≈850450 minutes)

How many wh-dependencies is this?

What [__what__]?

Pearl & Sprouse 2013, Bates & Pearl 2019, Pearl & Bates in press
Hoff-Ginsberg (1998) and Rowe (2012): Estimates of utterances per minute in speech directed at children from different backgrounds.

Pearl & Sprouse 2013, Bates & Pearl 2019, Pearl & Bates in press
Both theories

Evaluating the theory

from our own corpus samples: rate of \textit{wh-dep} / utterance

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
 & \textbf{utt/min} & \textbf{min} & \textbf{wh-dep/utt} & \textbf{total wh-dep} \\
\hline
\text{- 2 s.d.} & 7.4 & \text{*} 850,450.2 & \text{*} 20,932/101,838 & \text{=} 1,293,545 \\
\text{- 1 s.d.} & 11.6 & & & \text{=} 2,027,719 \\
\text{average} & 15.8 & & & \text{=} 2,761,893 \\
\text{+ 1 s.d.} & 20.0 & & & \text{=} 3,496,067 \\
\text{+ 2 s.d.} & 24.2 & & & \text{=} 4,230,241 \\
\hline
\end{tabular}
\end{table}

Pearl & Sprouse 2013, Bates & Pearl 2019, Pearl & Bates in press
Both theories

Evaluating the theory

Can the modeled learner produce the appropriate observable behavior?

wh-dependency distribution

≈ 1.3 million - 4.2 million wh-dependencies

Pearl & Sprouse 2013, Bates & Pearl 2019, Pearl & Bates in press
Evaluating the theory

Reminder: Target behavior

- Subject island
- Complex NP island
- Whether island
- Adjunct island

Both theories

Sprouse et al. 2012

De Villiers et al. 2008

What [NP [CP _what]]?

Pearl & Sprouse 2013, Bates & Pearl 2019, Pearl & Bates in press
Both theories

Evaluating the theory

Reminder: Target behavior

Subject island
Complex NP island
Whether island
Adjunct island

Whether island X

Complex NP island

Subject island
Adjunct island

Who [non-island] [island]?

Looking for superadditivity in selected judgments as the sign of syntactic islands knowledge

Who [CP... who]?
Evaluating the theory

Reminder: Target behavior

Each set of island stimuli from Sprouse et al. 2012...

Complex NP island stimuli

- Who ___ claimed [that Lily forgot the necklace]?  
  matrix | non-island
- What did the teacher claim [that Lily forgot ___]?  
  embedded | non-island
- Who ___ made [the claim that Lily forgot the necklace]?  
  matrix | island
- *What did the teacher make [the claim that Lily forgot ___]?  
  embedded | island
Each *wh*-dependency from the island stimuli of Sprouse et al. 2012
• can be transformed into container node sequences

\[
\begin{align*}
\text{start-IP-end} & \quad \text{matrix} \mid \text{non-island} \\
\text{start-IP-VP-CP}_{\text{that}}\text{-IP-VP-end} & \quad \text{embedded} \mid \text{non-island} \\
\text{start-IP-end} & \quad \text{matrix} \mid \text{island} \\
\text{start-IP-VP-NP-CP}_{\text{that}}\text{-IP-VP-end} & \quad \text{embedded} \mid \text{island}
\end{align*}
\]
Evaluating the theory

Reminder: Target behavior

Both theories

- Subject island
- Complex NP island
- Whether island
- Adjunct island

Each *wh*-dependency from the island stimuli of Sprouse et al. 2012
- can be transformed into container node sequences
- can be broken into *syntactic trigram building blocks* and have its *probability calculated*

**Complex NP island stimuli**

<table>
<thead>
<tr>
<th>Structure</th>
<th>Start-IP-End</th>
<th>Start-IP-VP-CP&lt;sub&gt;that&lt;/sub&gt;-IP-VP-end</th>
<th>Start-IP-End</th>
<th>Start-IP-VP-NP-CP&lt;sub&gt;that&lt;/sub&gt;-IP-VP-end</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>matrix</td>
<td>non-island</td>
<td>embedded</td>
<td>non-island</td>
</tr>
</tbody>
</table>

\[ \prod_{t \in \text{trigrams}} p(t) \]

Pearl & Sprouse 2013, Bates & Pearl 2019, Pearl & Bates in press
Both theories

Evaluating the theory

Reminder: Target behavior

Subject island

Complex NP island

Whether island

Adjunct island

These probabilities can then be plotted to see if *superadditivity* is present in the predicted acceptability judgments.

**Complex NP island stimuli**

<table>
<thead>
<tr>
<th>start-IP-end</th>
<th>matrix</th>
<th>non-island</th>
</tr>
</thead>
<tbody>
<tr>
<td>start-IP-VP-CP$_{\text{that-IP-VP-end}}$</td>
<td>embedded</td>
<td>non-island</td>
</tr>
<tr>
<td>start-IP-end</td>
<td>matrix</td>
<td>island</td>
</tr>
<tr>
<td>start-IP-VP-NP-CP$_{\text{that-IP-VP-end}}$</td>
<td>embedded</td>
<td>island</td>
</tr>
</tbody>
</table>

\[ \prod_{t \in \text{trigrams}} p(t) \]
Both theories

Evaluating the theory

Reminder: Target behavior

If so, then we predict the modeled child has syntactic island knowledge that allows the same judgment pattern as adults, learned from the building blocks in children’s input.

Pearl & Sprouse 2013, Bates & Pearl 2019, Pearl & Bates in press
Evaluating the theory

1: probabilities of pre-specified pieces

---

Pearl & Sprouse 2013, Bates & Pearl 2019, Pearl & Bates in press
Evaluating the theory

1: probabilities of pre-specified pieces

Superadditivity predicted for judgments of all four island types.

Pearl & Sprouse 2013, Bates & Pearl 2019, Pearl & Bates in press
Evaluating the theory

1: probabilities of pre-specified pieces

Reminder: Target behavior

Children prefer this interpretation.

Sprouse et al. 2012

De Villiers et al. 2008
Evaluating the theory

1: probabilities of pre-specified pieces

\[ \prod_{t \in \text{trigrams}} p(t) \]

The \textit{wh}-dependency this interpretation relies on is \(10^{18}\) times more probable than the other one.

Subject island
Complex NP island
Whether island
Adjunct island

\[ \text{matrix} \quad \text{embedded} \]

\[ \text{Sprouse et al. 2012} \]

\[ \text{De Villiers et al. 2008} \]
Evaluating the theory

1: probabilities of pre-specified pieces

\[ \prod_{t \in \text{trigrams}} p(t) \]

So, the modeled child prefers it.

What

[complex NP [what]]

\[ \text{What} \textunderscore \text{what?} \]

De Villiers et al. 2008

Sprouse et al. 2012
Takeaway: This theory can work for learning knowledge about syntactic islands.

1: probabilities of pre-specified pieces
Key idea: Learning about the building blocks of \textit{wh}-dependencies leads to knowledge about syntactic islands.
What about the other theory that learns what the building blocks of wh-dependencies are at the same time as it learns their probabilities?
2: what the pieces are and their probabilities

Dickson et al. 2022, in prep.
2: what the pieces are and their probabilities

The building blocks from Pearl & Sprouse (2013) were pre-specified. The modeled child already knew to look for syntactic trigrams of a certain kind.

- start-IP-VP
- IP-VP-end
- IP-VP-IP
- VP-IP-VP
- start-IP-end
- IP-VP-CP_{that}
- VP-CP_{that}-IP
- IP-VP-PP
- VP-PP-end
- CP_{that}-IP-NP
- IP-NP-PP
- NP-PP-end

Dickson et al. 2022, in prep.
In particular:
(1) Look for groups of three units
(2) If the unit is a CP, include the lexical item

Dickson et al. 2022, in prep.
Empirical motivation for the CP lexical item:
Two of the islands (Whether and Adjunct) only differ from more acceptable wh-dependencies by the complementizer used.

What does the teacher think \([\text{that Lily forgot } \_\_]\)?

non-island

*What does the teacher wonder \([\text{whether Lily forgot } \_\_]\)?

island

*What does the teacher worry \([\text{if Lily forgot } \_\_]\)?

island
Empirical motivation for the CP lexical item: Two of the islands (Whether and Adjunct) only differ from more acceptable wh-dependencies by the complementizer used.

\[\text{Empirical motivation for the CP lexical item: Two of the islands (Whether and Adjunct) only differ from more acceptable } \text{wh-dependencies by the complementizer used.}\]

\[\text{start-IP-VP-CP}_{\text{that}}\text{-IP-VP-end}\]

\[\text{start-IP-VP-CP}_{\text{whether}}\text{-IP-VP-end}\]

\[\text{start-IP-VP-CP}_{\text{if}}\text{-IP-VP-end}\]

\[\text{embedded } | \text{ non-island}\]

\[\text{embedded } | \text{ island}\]

\[\text{embedded } | \text{ island}\]

Dickson et al. 2022, in prep.
So the building blocks need to include this lexical item type.

\[ \text{Whether Adjunct} \]

\[ \text{* } \text{start-IP-VP-CP}_{\text{that}} \text{- IP-VP-end} \]

\[ \text{embedded | non-island} \]

\[ \text{embedded | island} \]

\[ \text{embedded | island} \]
2: what the pieces are and their probabilities

Is this the only one needed?

\[\text{start-IP-VP-CP}_{\text{that}}\text{-IP-VP-end}\]
\[\text{start-IP-VP-CP}_{\text{whether}}\text{-IP-VP-end}\]
\[\text{start-IP-VP-CP}_{\text{if}}\text{-IP-VP-end}\]

embedded | non-island
embedded | island
embedded | island

Dickson et al. 2022, in prep.
Liu et al 2019: Acceptability of *wh*-dependencies can depend on the *lexical item in the main verb*.

Dickson et al. 2022, in prep.

What did she *think* [that he saw ___]?
What did she *say* [that he saw ___]?

What did she *whine* [that he saw ___]?
What did she *mumble* [that he saw ___]?
Liu et al. 2019: Acceptability of *wh*-dependencies can depend on the lexical item in the main verb.
Bigger question: Are there other lexical item types the building blocks need to include?

2: what the pieces are and their probabilities
Theory: The child tries to learn what the “best” building blocks are at the same time she learns about their distributions in the input.
the best building blocks

Before:
(1) Look for groups of three units
(2) If the unit is a CP, include the **lexical item**

- **start-IP-VP**
- **IP-VP-end**
- **IP-VP-CP_{that}**
- **VP-CP_{that}-IP**
2: what the pieces are and their probabilities

the best building blocks

(1) Look for groups of three units
(2) If the unit is a CP, include the lexical item

Maybe the best size is sometimes bigger than three and sometimes smaller.
2: what the pieces are and their probabilities

the best building blocks

(1) Look for groups of three units
(2) If the unit is a CP, include the lexical item
the best building blocks

(1) Look for groups of three units
(2) If the unit is a CP, include the lexical item
2: what the pieces are and their probabilities

the best building blocks

(1) Look for the best-sized units
(2) If the unit is a CP, include the lexical item

Maybe the lexical item is needed sometimes… but sometimes not.

\[\text{start-IP-VP}_{\text{think-CP}}_{\text{that}}\]

\[\text{start-IP-VP}_{\text{say-CP}}_{\text{that}}\]
2: what the pieces are and their probabilities

the best building blocks

(1) Look for the best-sized units
(2) Sometimes include the lexical item

Maybe the lexical item is needed sometimes…but sometimes not.
How can the child learn what the best building blocks are?

(1) Look for the best-sized units
(2) Sometimes include the lexical item
How can the child learn what the best building blocks are?

1. Look for the best-sized units
2. Sometimes include the lexical item

Theory: Look for an “efficient” set of building blocks.

Dickson et al. 2022, in prep.
2: what the pieces are and their probabilities

How can the child learn what the best building blocks are?

(1) Look for the best-sized units
(2) Sometimes include the lexical item

Efficient building blocks allow the representation of current and future *wh*-dependencies to be more probable.
2: what the pieces are and their probabilities

How can the child learn what the best building blocks are?

(1) Look for the best-sized units
(2) Sometimes include the lexical item

Efficient building blocks allow the representation of current and future wh-dependencies to be more probable.

Why? One idea: Higher probability wh-dependencies are faster to process (comprehending or producing).

Dickson et al. 2022, in prep.
How? Look for building blocks that are a balance between (1) how big they are (2) how fast they are to put together to make a wh-dependency
a balance between
(1) how big they are
(2) how fast they are to put together to make a wh-dependency

What did she say that he saw __ ?
2: what the pieces are and their probabilities

learning efficient building blocks

a balance between
(1) how big they are
(2) how fast they are to put together to make a *wh*-dependency

\[
\text{start-IP}_\text{past-VP}_\text{say-CP}_\text{that-IP}_\text{past-VP}_\text{see-end}
\]
2: what the pieces are and their probabilities

learning efficient building blocks

a balance between
(1) how big they are
(2) how fast they are to put together to make a *wh*-dependency

\[
\text{start-IP}_{\text{past}}\text{-VP}_{\text{say}}\text{-CP}_{\text{that}}\text{-IP}_{\text{past}}\text{-VP}_{\text{see}}\text{-end}
\]

Pieces can be small, so that many of them make up a *wh*-dependency
2: what the pieces are and their probabilities

learning efficient building blocks

a balance between
(1) how big they are
(2) how fast they are to put together to make a *wh*-dependency

\[ \text{start-IP} \_\text{past-VP} \_\text{say-CP} \_\text{that-IP} \_\text{past-VP} \_\text{see-end} \]

It may be slower to put together many small pieces.

Dickson et al. 2022, in prep.
2: what the pieces are and their probabilities

learning efficient building blocks

a balance between
(1) how big they are
(2) how fast they are to put together to make a wh-dependency

\[ \text{start-IP}_{\text{past}}\text{-VP}_{\text{say}-\text{CP}}\text{that-IP}_{\text{past}}\text{-VP}_{\text{see}-\text{end}} \]

many smaller

slower because many

But these pieces may get reused, so that makes them faster to put together.
2: what the pieces are and their probabilities

learning efficient building blocks

(a balance between
(1) how big they are
(2) how fast they are to put together to make a *wh*-dependency

\[ \text{start-IP}_{\text{past}}-\text{VP}_{\text{say-CP}}\text{that-IP}_{\text{past}}-\text{VP}_{\text{see-end}} \]

many smaller

slower because many

But these pieces may get reused, so that makes them faster to put together.
2: what the pieces are and their probabilities

learning efficient building blocks

a balance between
(1) how big they are
(2) how fast they are to put together to make a *wh*-dependency

\[ \text{start-IP}_{\text{past}}-\text{VP}_{\text{say}}-\text{CP}_{\text{that}}-\text{IP}_{\text{past}}-\text{VP}_{\text{see}}-\text{end} \]

Pieces can be big, so that only one makes up a *wh*-dependency

many smaller many reused

Dickson et al. 2022, in prep.
2: what the pieces are and their probabilities

learning efficient building blocks

a balance between
(1) how big they are
(2) how fast they are to put together to make a *wh*-dependency

\[ \text{start-IP}_{\text{past}}-\text{VP}_{\text{say}}-\text{CP}_{\text{that}}-\text{IP}_{\text{past}}-\text{VP}_{\text{see}}-\text{end} \]

many smaller
many reused

It may be faster to put together one big piece.

Dickson et al. 2022, in prep.
2: what the pieces are and their probabilities

learning efficient building blocks

a balance between

(1) how big they are
(2) how fast they are to put together to make a *wh*-dependency

\[ \text{start-IP}_\text{past}-\text{VP}_\text{say}-\text{CP}_\text{that}-\text{IP}_\text{past}-\text{VP}_\text{see}-\text{end} \]

many smaller

many reused

one big

faster because one

It may be slower if the piece is used rarely.

Dickson et al. 2022, in prep.
2: what the pieces are and their probabilities

learning efficient building blocks

a balance between
(1) how big they are
(2) how fast they are to put together to make a *wh*-dependency

\[ \text{start-IP}_{\text{past}}-\text{VP}_{\text{say}}-\text{CP}_{'that'}-\text{IP}_{\text{past}}-\text{VP}_{\text{see}}-\text{end} \]

many smaller many reused

one big faster because one

It may be slower if the piece is used rarely.

Dickson et al. 2022, in prep.
What did she say that he saw...

...start past VP say CP that IP past VP see end

The most efficient option is probably a balance of bigger and smaller blocks that collectively are faster to put together.

Dickson et al. 2022, in prep.
2: what the pieces are and their probabilities

learning efficient building blocks

a balance between
(1) how big they are
(2) how fast they are to put together to make a *wh*-dependency

\[ \text{start-IP}_{\text{past}}-\text{VP}_{\text{say}}-\text{CP}_{\text{that}}-\text{IP}_{\text{past}}-\text{VP}_{\text{see}}-\text{end} \]

many smaller many reused

one big one rare

Dickson et al. 2022, in prep.
2: what the pieces are and their probabilities

learning efficient building blocks

How can children find the best balance?

many smaller many reused

many reused

one big one rare

Dickson et al. 2022, in prep.
Use Bayesian inference to search through the hypothesis space of all possible building blocks (O’Donnell 2015) and find an efficient set for children’s input.
2: what the pieces are and their probabilities

So that's what the modeled child will do
There's additional target behavior about *wh*-dependencies we'd like to capture.
2: what the pieces are and their probabilities

Before:
Adult judgments + child preferences of certain wh-dependencies

Dickson et al. 2022, in prep.
Whether island

Adjunct island

Before: Target behavior

Children prefer this interpretation.

Sprouse et al. 2012

Bates et al. in prep.

De Villiers et al. 2008

Dickson et al. 2022, in prep.
2: what the pieces are and their probabilities

- certain *wh*-dependencies
  - Subject island
  - Complex NP island
  - Whether island
  - Adjunct island

+ additional target behavior with *wh*-dependencies that vary main verb frequency

- What did she think [that he saw ___]? [Liu et al. 2019]
- What did she say [that he saw ___]?
- What did she whine [that he saw ___]? [Liu et al. 2019]
- What did she mumble [that he saw ___]?
2: what the pieces are and their probabilities

certain *wh*-dependencies

+ additional target behavior with *wh*-dependencies that vary main verb frequency

What did she VERB [that he saw __ ]?

*Liu et al. 2019*
2: what the pieces are and their probabilities

- Certain *wh*-dependencies
  - Subject island
  - Complex NP island
  - Whether island
  - Adjunct island

+ Additional target behavior with *wh*-dependencies that vary main verb frequency

What did she VERB [that he saw __ ]?

Liu et al. 2019

Important pattern: Positive correlation between main verb frequency and judged acceptability.

Dickson et al. 2022, in prep.
2: what the pieces are and their probabilities

Before: Target behavior

Children prefer this interpretation.

What ___what?

What [NP [CP ___what]]?

De Villiers et al. 2008

Dickson et al. 2022, in prep.
2: what the pieces are and their probabilities

+ additional target behavior with other wh-dependencies

Who did the little sister ask how to see?
Who did the boy ask what to bring?
How did the mom learn what to bake?
How did the girl ask where to ride?
How did the boy who sneezed drink the milk?
2: what the pieces are and their probabilities

Subject island
Complex NP island
Whether island
Adjunct island

Who
Who [CP-how _who]?
Who [CP-what _who]?
How [CP-what _how]?
How [CP-where _how]?
How [NP [CP-who _how]]?

De Villiers et al. 2008

Dickson et al. 2022, in prep.
So what does the modeled child do?

2: what the pieces are and their probabilities

wh-dependency distribution
Superadditivity predicted for judgments of all four island types.

Dickson et al. 2022, in prep.
2: what the pieces are and their probabilities

Positive correlation predicted with verb frequency for judgments of this *wh*-dependency.

Dickson et al. 2022, in prep.
2: what the pieces are and their probabilities

Children prefer this interpretation.

Dickson et al. 2022, in prep.
2: what the pieces are and their probabilities

The *wh*-dependency this interpretation relies on is $10^5$ times more probable than the other one.

Dickson et al. 2022, in prep.
2: what the pieces are and their probabilities

So, the modeled child prefers it.

Dickson et al. 2022, in prep.
2: what the pieces are and their probabilities

The modeled child also prefers the child-preferred ones for the other wh-dependencies.
What do the learned building blocks that lead to this behavior look like?

2: what the pieces are and their probabilities

Dickson et al. 2022, in prep.
Different sizes that the modeled child learned

2: what the pieces are and their probabilities

learned building blocks
Some lexicalization based on frequency: more frequent lexical items are used. The frequency threshold is learned by the modeled child per node type (IP, VP, CP, etc.).

Dickson et al. 2022, in prep.
Takeaway:
This theory can work (even better) for learning knowledge about syntactic islands.
Key idea (again): Learning about the building blocks of wh-dependencies leads to knowledge about syntactic islands, even when there’s less knowledge built in.
Key idea: This strategy works when the child’s goal is finding efficient building blocks.

2: what the pieces are and their probabilities

wh-dependency distribution

Dickson et al. 2022, in prep.
One way to succeed at learning about constraints on *wh*-dependencies (syntactic islands) is to learn them indirectly.

Pearl & Sprouse 2013, Bates & Pearl 2019, Dickson et al. 2022, Pearl & Bates in press, Dickson et al. in prep.
In particular, learn how to represent *wh*-dependencies efficiently using pieces that can be composed locally to represent any *wh*-dependency.

Pearl & Sprouse 2013, Bates & Pearl 2019, Dickson et al. 2022, Pearl & Bates in press, Dickson et al. in prep.
The big picture

The efficient pieces that are learned, with their associated probabilities, allow the constraints to emerge automatically.

Pearl & Sprouse 2013, Bates & Pearl 2019, Dickson et al. 2022, Pearl & Bates in press, Dickson et al. in prep.
The big picture

So this is one way **locality** could play a key role in the **acquisition** of complex syntactic knowledge.

*Pearl & Sprouse 2013, Bates & Pearl 2019, Dickson et al. 2022, Pearl & Bates in press, Dickson et al. in prep.*
Thank you!

Jon Sprouse  Alandi Bates  Richard Futrell  Niels Dickson

BUCLD 2018  UCSD Linguistics 2020  ForMA Group 2020
UMD Linguistics 2020  BUCLD 2021  SCiL 2022
UArizona Linguistics 2022  UChicago LEAP 2022
UCI QuantLang Collective

Lisa S. Pearl
Professor
Department of Language Science
SSPB 2219
University of California, Irvine
lpearl@uci.edu