A new way to identify if variation in children’s input is developmentally meaningful: A look at syntactic knowledge across socio-economic status

Lisa Pearl
University of California, Irvine

November 4, 2021
EyeLands Lab
Norwegian University of Science and Technology
There’s lots of variation in children’s input
Developmentally-meaningful variation impacts language development in a measurable way.
Developmentally-meaningful input deficits can lead to language delays.
If there’s an input-based language delay and we know what the crucial input deficit is, we can intervene and fix that deficit.
Important: If a language outcome difference isn’t input-based, then “fixing” the input won’t help.
That’s why it’s **important to know** if the input is (at least partially) **causing** the language developmental issue… or not.
Input-based language delays appear across socio-economic status (SES). Lower-SES children are often behind their higher-SES peers.
Low-SES language input can differ from high-SES input in both overall quantity of speech and the quality of that speech (Hart & Risley 1995, Huttenlocher et al. 2010, Rowe 2012, Schwab & Lew-Williams 2016, Rowe et al. 2017).
Quality can be measured by different aspects of the input, like diversity of vocabulary ...
Quality can be measured by different aspects of the input, like diversity of vocabulary, diversity of syntactic constructions …
Quality can be measured by different aspects of the input, like diversity of vocabulary, diversity of syntactic constructions, and frequency of decontextualized speech.

The kitty wasn't there    Because we're going tomorrow

We saw her yesterday, didn't we?

The penguins should be at the zoo    Because the penguins were being fed.

The kitty wasn't there    Because we're going tomorrow

We'll see the kitty on Friday
How can we tell if any particular input difference is developmentally meaningful (that is, it impacts language development)?
One (standard) way:
- Notice that there’s a difference
One (standard) way:
• Notice that there's a difference
• Measure language acquisition outcomes
One (standard) way:
• Notice that there’s a difference
• Measure language acquisition outcomes
• See if that input difference correlates with any outcome differences
One (standard) way:
• Notice that there’s a difference
• Measure language acquisition outcomes
• See if that input difference correlates with any outcome differences

If so, then the input difference *might cause* the outcome difference and so be developmentally meaningful.
One (standard) way

A new (complementary) way uses developmental computational modeling.
One (standard) way

A developmental computational model implements a specific learning theory ...
One (standard) way

A developmental computational model implements a specific learning theory about how children use their input …
One (standard) way

A developmental computational model implements a specific learning theory about how children use their input to acquire the knowledge to generate their output.
One (standard) way

A developmental computational model implements a specific learning theory about how children use their input to acquire the knowledge to generate their output.

Important: the learning theory implemented by the model specifies what aspect of the input matters.
One (standard) way

A developmental computational model implements a specific learning theory about how children use their input to acquire the knowledge to generate their output.

If we know what input part matters, we can target that part for intervention if needed.
One (standard) way

So, a developmental computational model can predict the language outcome on the basis of the input.
One (standard) way

If the predicted outcomes differ, then it's because the input difference caused that outcome difference. So, the input difference is predicted to be developmentally meaningful.
One (standard) way

If the predicted outcomes differ, then it’s because the input difference caused that outcome difference. So, the input difference is predicted to be developmentally meaningful.

These outcome predictions will need to be verified, though.
One (standard) way

Bonus: Because the learning theory in the model is causal, we can predict if the input should cause similar outcomes, too.

In that case, the input difference isn’t developmentally meaningful.
Detecting if input differences are developmentally meaningful

One (standard) way

A new (complementary) way
One (standard) way

Today’s focus

A new (complementary) way

Detecting if input differences are developmentally meaningful
Detecting if input differences are developmentally meaningful

Today’s focus

A new (complementary) way

Case study:
Syntactic island acquisition

Who does...
Detecting if input differences are developmentally meaningful

One (standard) way

Today’s focus

A new (complementary) way

Case study:
Syntactic island acquisition

Who does...

Why? It’s higher-order syntactic knowledge where we don’t know much about developmentally-meaningful input differences that may exist across SES.
Syntactic island constraints 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 island constraints 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 island constraints involve *wh*-dependencies.

Who does... involve *wh*-dependencies.

Who does Lily think the kitty for _who_ is pretty?  

What’s going on here?  

Syntactic island (Ross 1967)  

Subject island
Syntactic island constraints involve *wh*-dependencies.

What’s going on here?

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?
Syntactic island constraints involve \( wh \)-dependencies.

What’s going on here? syntactic island (Ross 1967)

Who does Lily think the kitty for \( _{\text{who}} \) is pretty? Subject island

What did Jack make the claim that he bought \( _{\text{what}} \)? Complex NP island

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 \( _{\text{what}} \)?
Syntactic island constraints involve \textit{wh}-dependencies.

What's going on here?

Who does Lily think the kitty for \underline{who} is pretty? \textbf{Subject island}

What did Jack make the claim that he bought \underline{what}? \textbf{Complex NP island}

What did Elizabeth wonder whether Jack bought \underline{what}? \textbf{Whether island}

Who does Lily think the kitty for \underline{who} is pretty?

What did Elizabeth wonder whether Jack bought \underline{what}?

Jack is somewhat tricksy.

He claimed he bought something.

Elizabeth worried it was something dangerous.
Syntactic island constraints involve *wh*-dependencies.

Who does Lily think the kitty for—who is pretty?  
What did Jack make the claim that he bought—who?  
What did Elizabeth wonder whether Jack bought—who?  
What did Elizabeth worry if Jack bought—who?

Subject island  Complex NP island  Whether island  Adjunct island

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

(Chomsky 1965, Ross 1967)
Syntactic island constraints 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 island constraints involve wh-dependencies.

What’s going on here? syntactic 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

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 ___what?
Syntactic island constraints involve wh-dependencies.

What’s going on here? syntactic 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**
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 island constraints 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**

High-SES adults *judge* these dependencies to be *far worse* than many others, including others that are very similar except that they don’t cross syntactic islands (Sprouse et al. 2012).

These judgments are a measurable observable behavior that signal the successful acquisition of syntactic island knowledge.
Syntactic island constraints 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**

High-SES adults *judge* these dependencies to be *far worse* than many others, including others that are very similar except that they don’t cross syntactic islands (Sprouse et al. 2012).

So, these judgments can serve as a target for successful acquisition — an outcome we can measure.
Syntactic island constraints

High-SES 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

length of dependency (matrix vs. embedded)

Who [CP... _who]?

Who _who?

presence of an island structure (non-island vs. island)

Who [non-island ]?

Who [island ]?
Syntactic island constraints

High-SES adult judgments
= behavioral target outcome

Adult knowledge as measured by acceptability judgment behavior

<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>

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 island constraints

High-SES 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)

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

Also, 4-year-olds across SES strongly disprefer wh-dependencies like this one (De Villiers et al. 2008).
Syntactic island constraints

High-SES 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)

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

Sprouse et al. 2012
Syntactic island constraints

High-SES adult judgments = behavioral target outcome

Adult knowledge as measured by acceptability judgment behavior

- length of dependency
- presence of an island structure

(matrix vs. embedded) \(\times\) (non-island vs. island)

Whether island stimuli

Who ___ thinks [that Jack stole the necklace]?\nWhat does the teacher think [that Jack stole ___]?\nWho ___ wonders [whether Jack stole the necklace]?\n*What does the teacher wonder [whether Jack stole ___]?
Syntactic island constraints

High-SES 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)

Adjunct island stimuli

<table>
<thead>
<tr>
<th>Who __ thinks [that Lily forgot the necklace]?</th>
<th>matrix</th>
<th>non-island</th>
</tr>
</thead>
<tbody>
<tr>
<td>What does the teacher think [that Lily forgot __ ]?</td>
<td>embedded</td>
<td>non-island</td>
</tr>
<tr>
<td>Who __ worries [if Lily forgot the necklace]?</td>
<td>matrix</td>
<td>island</td>
</tr>
<tr>
<td>*What does the teacher worry [if Lily forgot __ ]?</td>
<td>embedded</td>
<td>island</td>
</tr>
</tbody>
</table>

Sprouse et al. 2012
Syntactic island constraints

High-SES adult judgments
= behavioral target outcome

Adult knowledge as measured by acceptability judgment behavior

\[ \text{length of dependency (matrix vs. embedded)} \times \text{presence of an island structure (non-island vs. island)} \]

Syntactic island = superadditive interaction of the two factors (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 island constraints

High-SES 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

\[ \begin{align*}
\text{Who} & \quad \text{[non-island]} \quad ? \\
\text{Who} & \quad \text{[island]} \\
\end{align*} \]

Who \[ \text{[CP…} \quad \_{\text{who}}? \]

Sprouse et al. 2012
Syntactic island constraints

High-SES 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)

Syntactic island = superadditive interaction of the two factors

Sprouse et al. 2012

Who does...
Syntactic island constraints

High-SES 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
Syntactic island constraints

High-SES 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): acceptability judgments from 173 adult subjects

Superadditivity for all four island types = knowledge that dependencies can’t cross these island structures.

Sprouse et al. 2012
Okay, so what's the **relevant input** for learning this **target knowledge**?
That depends on **how** we think children learn it.
That depends on how we think children learn it.

Pearl & Sprouse 2013 intuition for a strategy:
- Learn what you can from the dependencies you do actually observe in the input
- Apply it to make a judgment about the dependencies you haven’t seen before, like those crossing syntactic islands (and maybe other longer dependencies, too).
The learning theory = a concrete learning strategy (Pearl & Sprouse 2013): View *wh*-dependencies in terms of their **building blocks** and **track** those building blocks in the input.
A strategy for learning syntactic islands

Dependencies represented as a sequence of container nodes

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

```
CP
  NP₁
    What
  did
    IP
      NP
        Pro
        you
      VP
        V
        see
        NP₁
          ...
```
A strategy for learning syntactic islands

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 you [VP see __]]?
= *start-IP-VP-end*
A strategy for learning syntactic islands

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

What __ happened?
= What [IP __ happened]?
= start-IP-end
A strategy for learning syntactic islands

Dependencies represented as a sequence of container nodes

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

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

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

What did she want to do __?
= What did \([_{IP} \text{ she } [_{VP} \text{ want } [_{IP} \text{ to } [_{VP} \text{ do } __]]]]\)\)?
= \(\text{start-IP-VP-IP-VP-end}\)
A strategy for learning syntactic islands

Who does...

What __ happened?
= What \[ IP \__ happened \]\?
= \textit{start-IP-VP-end}

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

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

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

(Much) less acceptable dependencies have low probability segments

\[
\begin{array}{l}
\text{[CP Who \quad did \quad [IP Lily \quad [VP \quad think \quad [CP-that \quad [IP \quad [NP \quad the \quad kitty \quad [PP \quad for \quad __ ] \quad was \quad pretty \quad ?]]]]]} \\
\text{\textit{start-IP-VP-CP-that-IP-NP-PP-end}}
\end{array}
\]
A strategy for learning syntactic islands

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

[CP Who did [IP Lily [VP think [CP-that [IP [NP the kitty [PP for __ ] was pretty ?]]]]]

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).
A strategy for learning syntactic islands

The building blocks: trigrams of container nodes

start-IP-VP-end

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

start-IP-VP-VP-end

start-IP-VP-IP-VP-end

Pearl & Sprouse 2013
A strategy for learning syntactic islands

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-end
start-IP-VP-CP_{that}-IP-NP-PP-end
syntactic trigrams
A strategy for learning syntactic islands

The building blocks: trigrams of container nodes

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

start-IP-VP-CP_{that}-IP-NP-PP-end
A strategy for learning syntactic islands

The building blocks: trigrams of container nodes

syntactic trigrams

start-IP-VP-end

start-IP-VP-CP\text{that}-IP-NP-PP-end

start-IP-VP-IP-VP-end

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

start-IP-VP

IP-VP-IP

VP-IP-VP

IP-VP-end

start-IP-end

start-IP-end

Who does...
A strategy for learning syntactic islands

The building blocks: trigrams of container nodes

start-IP-VP-end

start-IP-end

start-IP-VP-IP-VP-end

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

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

start-IP-VP

IP-VP-CP_{that}

VP-CP_{that}-IP

CP_{that}-IP-NP

IP-NP-PP

NP-PP-end

Pearl & Sprouse 2013
A strategy for learning syntactic islands

The strategy: Track the relative frequency of the syntactic trigrams in your input.
A strategy for learning syntactic islands

Some of them are common and some of them aren’t.

Who does…
A strategy for learning syntactic islands

Some of them are common and some of them aren’t.

(And some never occur at all.)
A strategy for learning syntactic islands

Relative syntactic trigram frequency:

\[ p(t) \approx \frac{\text{# trigram}}{\text{total \# trigrams}} \]
A strategy for learning syntactic islands

Any wh-dependency can then be constructed from its syntactic trigram building blocks.
A strategy for learning syntactic islands

\[ \prod_{t \in \text{trigrams}} p(t) \]
A strategy for learning syntactic islands

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

Who does…

VP-PP-end

start-IP-VP-end

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

start-IP-VP-IP-VP-end

start-IP-VP

\[ \text{IP-VP-IP} \]

\[ \text{VP-IP-VP} \]

\[ \text{IP-VP-PP} \]

\[ \text{VP-PP-end} \]

\[ \text{IP-VP-CP_{\text{that}}} \]

\[ \text{IP-NP-PP} \]

\[ \text{VP-CP_{\text{that}}-IP} \]
A strategy for learning syntactic islands

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

Pearl & Sprouse 2013
A strategy for learning syntactic islands

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

A *wh*-dependency’s probability can stand in for its predicted acceptability.
A strategy for learning syntactic islands

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

Lower probability dependencies are predicted to be less acceptable (dispreferred), compared to higher probability dependencies.
A strategy for learning syntactic islands

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

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

Complex NP island stimuli

Who __ claimed [that Lily forgot the necklace]?  
What did the teacher claim [that Lily forgot __]?
Who __ made [the claim that Lily forgot the necklace]?
*What did the teacher make [the claim that Lily forgot __]?

matrix | non-island  
embedded | non-island  
matrix | island  
embedded | island

Pearl & Sprouse 2013
A strategy for learning syntactic islands

Each *wh*-dependency from the island stimuli of Sprouse et al. 2012
- can be transformed into container node sequences

Complex NP island stimuli

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

Pearl & Sprouse 2013
A strategy for learning syntactic islands

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>Start-IP-End</th>
<th>Start-IP-VP-CP_{that}-IP-VP-End</th>
<th>Start-IP-VP-NP-CP_{that}-IP-VP-End</th>
</tr>
</thead>
<tbody>
<tr>
<td>matrix</td>
<td>embedded</td>
<td>embedded</td>
</tr>
<tr>
<td>non-island</td>
<td>non-island</td>
<td>island</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A strategy for learning syntactic islands

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

Complex NP island stimuli

\[
\begin{align*}
\text{start-IP-end} & \quad \text{matrix} \mid \text{non-island} \\
\text{start-IP-VP-CP_{that-IP-VP-end}} & \quad \text{embedded} \mid \text{non-island} \\
\text{start-IP-end} & \quad \text{matrix} \mid \text{island} \\
\text{start-IP-VP-NP-CP_{that-IP-VP-end}} & \quad \text{embedded} \mid \text{island}
\end{align*}
\]
A strategy for learning syntactic islands

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

Complex NP island stimuli

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

\( \prod_{t \in \text{trigrams}} p(t) \)
This strategy works for high-SES children’s input

Judgments from a modeled child learning from the same amount of data as high-SES children seem to, with those data having the same composition as high-SES child-directed speech data (Pearl & Sprouse 2013).
This strategy works for high-SES children’s input.

Judgments from a modeled child learning from the same amount of data as high-SES children seem to, with those data having the same composition as high-SES child-directed speech data (Pearl & Sprouse 2013).

Superadditivity for all four islands.
This strategy works for high-SES children’s input

Implication: High-SES child input can support the acquisition of syntactic islands, using this learning strategy that depends on a certain part of the input.

Judgments from a modeled child learning from the same amount of data as high-SES children seem to, with those data having the same composition as high-SES child-directed speech data (Pearl & Sprouse 2013).
This strategy works for high-SES children’s input.

Judgments from a modeled child learning from the same amount of data as high-SES children seem to, with those data having the same composition as high-SES child-directed speech data (Pearl & Sprouse 2013).

Implication:
High-SES child input can support the acquisition of syntactic islands, using this learning strategy that depends on a certain part of the input.

In line with high-SES 4-year-olds who disprefer the Complex NP island-crossing dependency (De Villiers et al. 2008).
This strategy works for high-SES children’s input. Judgments from a modeled child learning from the same amount of data as high-SES children seem to, with those data having the same composition as high-SES child-directed speech data (Pearl & Sprouse 2013).

That input part is the *wh*-dependencies, and their building blocks (the syntactic trigrams).
$\Pi_{t \in \text{trigrams}} p(t)$

Are there meaningful differences across SES in this part of the input (the *wh*-dependencies and syntactic trigrams)?
First, what are the differences? That is, how different does this input part look across SES?

Let's measure the distribution of the relevant parts: the *wh*-dependencies and the syntactic trigrams.

*Bates & Pearl 2019, submitted*
who does $\prod_{t \in \text{trigrams}} p(t)$?

One way to measure differences in distribution: the Jensen-Shannon divergence ($\text{JSDiv}$) (Endres & Schindelin 2003).

$0 \leq \text{JSDiv} \leq 1$

identical distributions $= \neq$ dissimilar distributions

Bates & Pearl 2019, submitted
Measurable input differences

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

\[ = \neq \]

\[ 0 \leq \text{JSDiv} \leq 1 \]

102K utterances (21K wh-dependencies) from the CHILDES Treebank (Pearl & Sprouse 2013) of speech directed at 25 high-SES children between the ages of 1 and 5 years old.
31.8K utterances (3.9K wh-dependencies) from a subpart of the HSLD corpus (Dickinson & Tabors 2001) in the CHILDES Treebank (Pearl & Sprouse 2013) of speech directed at 78 low-SES children between the ages of 3 and 5.
Measurable input differences

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

\[ 0 \leq JSDiv \leq 1 \]

Note: SES was defined by the creators of the HSLLD corpus according to maternal education (6 years to some post-high school education) and annual income (70% reported < $20K/year).
Measurable input differences

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

\[ 0 \leq \text{JSDiv} \leq 1 \]

High-SES child-directed

\( \geq 21K \) \text{wh-dependencies}

Low-SES child-directed

\( \leq 3.9K \) \text{wh-dependencies}

High-SES adult-directed

74.6K utterances (8.5K \textit{wh-dependencies}) from the Switchboard corpus (Marcus et al. 1999) of adults speaking to each other over the phone.

\textit{Bates & Pearl 2019, submitted}
Who does…

$$\prod_{t \in \text{trigrams}} p(t)$$

$$= 0 \leq JSDiv \leq 1$$

So what do we find?

In particular, is high-SES child-directed speech more like low-SES child-directed speech or more like high-SES adult-directed speech?

High-SES adult-directed

High-SES child-directed

Low-SES child-directed

8.5K wh-dependencies

21K wh-dependencies

3.9K wh-dependencies

Bates & Pearl 2019, submitted
Measurable input differences

\[ \prod_{t \in \text{trigrams}} p(t) = 0 \leq \text{JSDiv} \leq 1 \]

If high-SES child-directed speech is more like low-SES child-directed speech, then SES differences matter less than who the speech is directed at.

High-SES

- adult-directed
  - 8.5K wh-dependencies

High-SES

- child-directed
  - 21K wh-dependencies
  - SES differences

Low-SES

- child-directed
  - 3.9K wh-dependencies

Bates & Pearl 2019, submitted
\[ \prod_{t \in \text{trigrams}} p(t) = 0 \leq JSDiv \leq 1 \]

If high-SES child-directed speech is more like high-SES adult-directed speech, then SES differences matter more than who the speech is directed at.
Whether we look at *wh*-dependencies or syntactic trigrams, we find the same pattern: high-SES and low-SES child-directed speech are more similar than high-SES child-directed and high-SES adult-directed speech.
Who does $\prod_{t \in \text{trigrams}} p(t)$?

$0 \leq JSDiv \leq 1$

For $wh$-dependencies, high-SES child-directed speech is twice as similar to low-SES child-directed speech as it is to high-SES adult-directed speech.
Measurable input differences

\[ \prod_{t \in \text{trigrams}} p(t) = 0 \leq JSDiv \leq 1 \]

For syntactic trigrams, high-SES child-directed speech is twice as similar to low-SES child-directed speech as it is to high-SES adult-directed speech.

- High-SES adult-directed: 8.5K wh-dependencies
- High-SES child-directed: 21K wh-dependencies
- Low-SES child-directed: 3.9K wh-dependencies

Bates & Pearl 2019, submitted
Measurable input differences

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

=  \neq

0 \leq JSDiv \leq 1

Takeaway: This part of the input looks pretty similar across SES — more similar than child-directed vs. adult-directed speech within SES.

High-SES
adult-directed

Directed at who differences

High-SES
child-directed

SES differences

Low-SES
child-directed

8.5K \textit{wh}-dependencies

21K \textit{wh}-dependencies

3.9K \textit{wh}-dependencies

Bates & Pearl 2019, submitted
\[ \prod_{t \in \text{trigrams}} p(t) \]

\[ 0 \leq \text{JSDiv} \leq 1 \]

Measurable input differences

But this is just a (quantitative) way to describe the input differences…

High-SES
- adult-directed
  - directed at who differences
  - 8.5K wh-dependencies

High-SES
- child-directed

Low-SES
- child-directed
  - SES differences
  - \( \approx \)
  - 3.9K wh-dependencies

Bates & Pearl 2019, submitted
Who does \( \prod_{t \in \text{trigrams}} p(t) \) mean?

Meaningful input differences

Does this part of the input act differently across SES? That is, are any differences (even if they’re smaller) developmentally meaningful?

They might be — small differences in the input distribution might snowball into learning outcome differences.

*Bates & Pearl 2019, submitted*
Does this part of the input act differently across SES? That is, are any differences (even if they’re smaller) developmentally meaningful?

They might be — small differences in the input distribution might snowball into learning outcome differences.

<table>
<thead>
<tr>
<th>Dependency Type</th>
<th>Percentage</th>
<th>Example</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>start-IP-VP-end</td>
<td>76.7%</td>
<td>What did Lily read __what?</td>
<td>75.5%</td>
</tr>
<tr>
<td>start-IP-end</td>
<td>10.3%</td>
<td>What __what happened?</td>
<td>12.8%</td>
</tr>
</tbody>
</table>

Bates & Pearl 2019, submitted
Who does \( \prod_{t \in \text{trigrams}} p(t) \)?

Meaningful input differences

Does this part of the input act differently across SES? That is, are any differences (even if they’re smaller) developmentally meaningful?

They might be — small differences in the input distribution might snowball into learning outcome differences.

<table>
<thead>
<tr>
<th>Syntactic Trigrams</th>
<th>Start-IP-VP</th>
<th>IP-VP-end</th>
<th>Start-IP-end</th>
</tr>
</thead>
<tbody>
<tr>
<td>41.4%</td>
<td>41.8%</td>
<td>40.0%</td>
<td>6.1%</td>
</tr>
</tbody>
</table>

Bates & Pearl 2019, submitted
Meaningful input differences

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

Does this part of the input act differently across SES? That is, are any differences (even if they’re smaller) developmentally meaningful?

But…we already have some evidence (De Villiers et al. 2008) that low-SES children disprefer \textit{wh}-dependencies that cross one of the islands we’re looking at (Complex NP), just like high-SES children.
Who does

Meaningful input differences

Does this part of the input act differently across SES? That is, are any differences (even if they’re smaller) developmentally meaningful?

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

Checkpoint: So at least for Complex NP islands, if our learning theory is right, we shouldn’t predict learning outcome differences across SES by age four. The input differences shouldn’t be developmentally meaningful.
Does this part of the input act differently across SES? That is, are any differences (even if they’re smaller) developmentally meaningful?

But what about the other three islands types (Subject, Whether, and Adjunct)? What do we predict for that knowledge at four years old? Are any input differences predicted to be developmentally meaningful for these islands?
Meaningful input differences

Let's use developmental computational modeling to find out.

Bates & Pearl 2019, submitted
First, how much *wh*-dependency input do low-SES children hear?
We can estimate how much by

• identifying the potential learning period for these syntactic islands (how much time)

• calculating how many *wh*-dependencies low-SES children would hear during this amount of time (how much input in that time).
Children (both high-SES and low-SES) seem to know about Complex NP islands by four years old (<60 months).

End point: Just before that (59 months).
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.

Educated guess: This is when they can start extracting *wh*-dependencies reliably from their input.
How much time
20 months → 59 months

How many hours is this? In particular, children are awake for only a certain portion of the day at different ages (Davis et al. 2004).
<table>
<thead>
<tr>
<th>age</th>
<th>age range</th>
<th>waking</th>
<th>total waking hours</th>
<th>cumulative waking</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>

How much time
20 months → 59 months

How many hours
Who does...?

How much time
20 months $\rightarrow$ 59 months
$\approx 14174$ hours

How much input (in particular, how many *wh*-dependencies) do children encounter during that time?
How much time
20 months $\rightarrow$ 59 months
$\approx$ 14174 hours

How many \textit{wh}-dependencies

Hart & Risley 1995: Estimates of \textit{utterances per hour}
for high-SES and low-SES households

301  
487

\textit{Bates & Pearl} 2019, submitted
Who does...

How much time
20 months $\longrightarrow$ 59 months
$\approx$ 14174 hours

How many *wh*-dependencies

utterances per hour

301 487

We can use our own corpus samples to estimate the rate of *wh*-dependencies/utterance.
How much time
20 months \( \rightarrow \) 59 months
\( \approx 14174 \) hours

How many \textit{wh}-dependencies

\textbf{utterances per hour}

\begin{tabular}{|c|c|c|c|c|}
\hline
SES & hours & \( \times \) & utt/hour & \( \times \) & \( \textit{wh}-\text{dep}/\text{utt} \) & = & \textit{total \textit{wh}-dep} \\
\hline
high-SES & 14174 & \( \times \) & 487 & \( \times \) & 20932/101838 & = & 1,418,193 \\
low-SES & 14174 & \( \times \) & 301 & \( \times \) & 3904/31875 & = & 522,539 \\
\hline
\end{tabular}
Who does…

How many *wh*-dependencies

<table>
<thead>
<tr>
<th>Low-SES children</th>
<th>High-SES children</th>
</tr>
</thead>
<tbody>
<tr>
<td>522,539</td>
<td>1,418,193</td>
</tr>
</tbody>
</table>

Low-SES children hear about a third the *wh*-dependencies that high-SES children do!
Who does X

How many *wh*-dependencies

522,539  1,418,193

But does the learning theory implemented by our developmental model predict that quantitative difference to matter by age four?
Let’s look first at Complex NP islands (where it shouldn’t matter if the learning theory is in fact right)
Who does \( \prod_{t \in \text{trigrams}} p(t) \)?

Reminder: Judgment behavior

Looking for superadditivity as the sign of syntactic islands knowledge

Bates & Pearl 2019, submitted
Complex NP islands: Superadditivity predicted to be present for low-SES children too! (Low-SES children don’t like dependencies crossing Complex NP islands either.)
Predicted judgment behavior

Complex NP islands: **Superadditivity** predicted to be present for low-SES children too! (Low-SES children don’t like dependencies crossing Complex NP islands either.)

No difference from high-SES behavior, despite measurable input differences. The learning theory has passed this checkpoint!

Now, what does the learning theory predict for the other three island types (Subject, Adjunct, and Whether)?
Meaningful input differences

The other islands: Superadditivity predicted to be present for low-SES children too! (Low-SES children shouldn’t like dependencies crossing these islands either.)
Who does \( \prod_{t \in \text{trigrams}} p(t) \) represent?

Meaningful input differences

This means low-SES input is predicted to support the same learning outcome knowledge (of these four syntactic islands) as high-SES input.

We know this is true for Complex NP islands.

Future work: Collect child judgments for the other island types.

Bates & Pearl 2019, submitted
So, our developmental computational model predicts no meaningful input differences across SES when it comes to learning this syntactic island knowledge from this part of the input.
No meaningful input differences predicted

Useful: Because the learning theory specifies how the input is predicted to cause the knowledge to develop, we know which building blocks are particularly important.

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

\[ \approx \]

\[ \approx \]

Bates & Pearl 2019, submitted
Key building blocks for success involve complementizer that (CP\textsubscript{that}) - this is because two of the islands (Whether and Adjunct) only differ from grammatical dependencies by the complementizer used.

What does the teacher think [that Lily forgot __ ]?  
*What does the teacher wonder [whether Lily forgot __ ]?  
*What does the teacher worry [if Lily forgot __ ]?

\[ \prod_{t \in \text{trigrams}} p(t) \]
Key building blocks for success involve complementizer *that* (CP_{that}) - this is because two of the islands (Whether and Adjunct) only differ from grammatical dependencies by the complementizer used.

\[ \Pi_{t \in trigrams} p(t) \]

Key building blocks are there

\[ \begin{align*}
\text{start-IP-VP-CP}_{\text{that}} & \quad \text{IP-VP-end} \\
\text{adjunct} & \quad \text{*start-IP-VP-CP}_{\text{whether}} \quad \text{IP-VP-end} \\
\text{whether} & \quad \text{*start-IP-VP-CP}_{\text{ir}} \quad \text{IP-VP-end}
\end{align*} \]

Embedded | Non-island
----------|----------------
Embedded   | Island
Embedded   | Island

* Bates & Pearl 2019, submitted
Key building blocks are there

So, children need to encounter grammatical *wh*-dependencies that involve $\text{CP}_{\text{that}}$. These are actually pretty rare in child-directed speech.

Low-SES child-directed

2 instances of 3.9K (=.05%)

High-SES child-directed

2 instances of 21K (<.01%)

What do you think *that* ___ what happens?  
What do you think *that* Jack read ___ what?

*Bates & Pearl 2019, submitted*
Key building blocks are there

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

But with enough input (over several years), even these rare cases are predicted to support learning.

Low-SES child-directed

2 instances of 3.9K (=.05%)

522,539 * 2/3904 \approx 268

What do you think \textbf{that} ___what happens?

High-SES child-directed

2 instances of 21K (<.01%)

1,418,193 * 2/20932 \approx 136

What do you think \textbf{that} Jack read ___what?

\textit{Bates & Pearl 2019, submitted}
Key building blocks are there

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

But with enough input (over several years), even these rare cases are predicted to support learning.

**Low-SES child-directed**
- 2 instances of 3.9K (=.05%)
- 8.45/month

**High-SES child-directed**
- 2 instances of 21K (<.01%)
- 3.39/month

What do you think that __what happens?

What do you think that Jack read __what?

*Bates & Pearl 2019, submitted*
Key building blocks are there

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

And in fact, if the samples are reasonably accurate, low-SES children actually see this building block more often.

Low-SES child-directed

- 2 instances of 3.9K (≈0.05%) 8.45/month

What do you think that \_\_\_ happens?

High-SES child-directed

- 2 instances of 21K (<0.01%) 3.39/month

What do you think that Jack read \_\_\_ what?
Key building blocks are there

[Diagram showing trigrams and dependency relations]

Interesting: The low-SES *wh*-dependency from the input sample with this building block is typically judged to be ungrammatical in the high-SES dialect (a *that*-trace violation).

Low-SES child-directed

2 instances of 3.9K (=.05%)

[Image of child with speech bubble]

What do you think *that* ___what happens?

High-SES child-directed

2 instances of 21K (=.01%)

[Image of child with speech bubble]

What do you think *that* Jack read ___what?

*Bates & Pearl 2019, submitted*
Who does \[ \prod_{t \in \text{trigrams}} p(t) \]

Key building blocks are there

\[ \approx \]

Upshot: Low-SES children are \textit{predicted} to achieve the same learning outcome as high-SES children by \textit{leveraging key building blocks from sources a high-SES child wouldn’t hear} (because those sources are ungrammatical for high-SES speakers).

What do you think \textbf{that} \underline{\text{what}} happens?

\textbf{Bates & Pearl 2019, submitted}
Key building blocks are there

Takeaway: This is one reason why measurable differences in the input might not be developmentally meaningful. Under this learning theory, the building blocks may show up in different places, but they’re still present in the input.

What do you think that ___ happens?
Who does ... $\prod_{t \in \text{trigrams}} p(t)$

No meaningful input differences predicted

So now what?
We should measure the learning outcomes in children across SES to see if in fact there are any learning outcome differences.

We already know there’s no difference for Complex NP islands, despite measurable input differences.
Who does …

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

No meaningful input differences predicted

One caveat: If there are in fact differences, it could be due to other factors besides input differences.

Example factor: Language processing ability is known to differ across SES, with low-SES children sometimes slower compared to their high-SES counterparts (Fernald et al. 2013, Weisleder & Fernald 2013). If low-SES children are less able to harness the information in their input (even if it’s there), they might be delayed in acquiring syntactic island knowledge.
\[ \Pi_{t \in \text{trigrams}} p(t) \]

No meaningful input differences predicted

But, if there aren’t outcome differences (perhaps after any language processing ability differences have resolved), then this supports syntactic island input quality being the same across SES (and also that the learning theory is plausible).
Who does... 

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

Building block origins

Low-SES child-directed

What do you think that ___what happens?

Remember that key building blocks involving CP that are predicted to come from a particular wh-dependency in low-SES child-directed speech that’s ungrammatical in the high-SES dialect.

Bates & Pearl 2019, submitted
Who does $\prod_{t \in \text{trigrams}} p(t)$?

Building block origins

Low-SES child-directed

What do you think that ___what happens?

This means low-SES adults are predicted to view this wh-dependency as grammatical if we expect low-SES children to hear it and harness those crucial CP that building blocks from it.

We can test this.

Bates & Pearl 2019, submitted
The big picture

One (standard) way

Developmental computational modeling complements existing techniques for identifying if input differences may be developmentally meaningful.

A new (complementary) way
The big picture

One (standard) way

≠

✓ ≠

A new (complementary) way

Developmental computational modeling complements existing techniques for identifying if input differences may be developmentally meaningful.

Who does…

We demonstrated this approach for syntactic island knowledge, and predicted no meaningful input differences across SES for knowledge by age four.
The big picture

One (standard) way

Developmental computational modeling complements existing techniques for identifying if input differences may be developmentally meaningful.

A new (complementary) way

This means we predict that input-based interventions wouldn’t be impactful if there actually are any differences in the acquisition of these syntactic islands across SES.
The big picture

One (standard) way

≠

✓

A new (complementary) way

≠

✓

Something useful: This technique can provide a causal explanation (using the learning theory implemented in the model) for how input differences could affect learning outcomes.
The big picture

Something useful: This technique can provide a causal explanation (using the learning theory implemented in the model) for how input differences could affect learning outcomes.

A new (complementary) way

For syntactic islands, the building blocks needed for this knowledge don’t seem to differ enough across SES to matter by age four.
The big picture

One (standard) way

Something else useful: This technique can make predictions about differences we might expect in both child outcomes and eventual adult knowledge.

A new (complementary) way

What do you think that __what happens?
The big picture

One (standard) way

Something important: Any predicted differences still need to be measured to see if they’re actually present. But at least we have a better idea what to look for.

A new (complementary) way

What do you think that __what happens?
The big picture

One (standard) way

Bonus: Modeling is often faster (and cheaper to do) than behavioral work. So it can be very useful as a first-pass input-quality assessment.

A new (complementary) way

Extra bonus: Possible to do in pandemic times.

What do you think that __what happens?
So let's use developmental computational modeling when we want to identify and understand if input variation could be developmentally meaningful!
Thank you!

BUCLD 2018
UCSD Linguistics 2020
ForMA Group 2020
UMD Linguistics 2020
UCI QuantLang Collective

Computation of Language Laboratory
UC Irvine

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