When do input differences matter? Using developmental computational modeling to assess input quality for syntactic islands across socio-economic status

Alandi Bates & Lisa Pearl
ajbates, lpearl@uci.edu
University of California, Irvine

Abstract

While there are observed differences in input across socio-economic status (SES), it's unclear how often these input differences are developmentally meaningful and so impact language development. We describe a way to identify developmentally-meaningful input differences that harnesses developmental computational modeling, which allows us to link children's input to predicted language development outcomes. We then apply this approach to investigate if there's developmentally-meaningful input variation across SES with respect to the complex syntactic knowledge called syntactic islands. Despite several measurable input differences for syntactic island input across SES, our model predicts no differences in the syntactic island knowledge that can be learned from that input. Interestingly, at least one key building block for syntactic island knowledge comes from a different source in low-SES children's input, but is crucially still present. This highlights a qualitative input similarity across SES. We discuss implications for linguistic development and adult syntactic knowledge variability across SES.

Key Words: socioeconomic status, linguistic development, child-directed speech, syntactic island constraints, computational cognitive modeling, quantitative approaches, input quantity, input quality, developmentally meaningful

1 Introduction

1.1 Input differences that are developmentally meaningful

There's a lot of naturally-occurring variation in children's input, including how long children are talked to every day, which people talk to them (e.g., adults, other children), what environments they experience language interaction in (e.g., home, daycare, school), and what people talk to them about, among many other types of variation. Importantly, not all this input variation is *developmentally meaningful* – that is, not all input variation impacts language development in a way that deviates significantly from a typically-developing child's trajectory. That is, while input differences may appear, the quality of the input isn't different when it comes to supporting language acquisition. However, some input variation does indeed impact language development – this variation is then developmentally meaningful; in particular, developmentally-meaningful input deficits lead to language development delays, and so indicate significant input quality differences.

Yet, how do we know if any particular observed input difference is developmentally meaningful? We know that some aspects of language development remain constant despite contextual variability that surfaces as measurable input differences (Hoff, 2006); measurable input differences in these cases don't seem to be developmentally meaningful. In contrast, when aspects of language development change in the face of measurable input differences, this suggests that those measurable input differences might be developmentally meaningful.

1.2 How do we know if a difference is developmentally meaningful?

A standard way to determine if a measurable input difference is developmentally meaningful is to observe some input difference, observe language development outcomes, and then see if the observed input difference is correlated with any observed outcome difference. If so, the language input difference *might* cause the language development outcome difference. In this case, targeting the input difference for intervention may lead to improved language development outcomes (e.g., input-based interventions allowing low-SES students to improve their language comprehension: Huttenlocher et al. 2002). If input-based intervention is indeed effective, this is more support that the language input difference caused the observed language outcome difference, and was therefore developmentally meaningful.

However, a complementary way to investigate if any observed input difference is developmentally meaningful uses developmental computational modeling (see Pearl (ress) for an overview of this technique applied to language acquisition more generally). A developmental computational model describes a specific learning mechanism that mediates between the input and a predicted language development outcome. In particular, the model implements a specific learning theory about how children use their input to acquire particular linguistic knowledge; children then use that linguistic knowledge to generate observable outcomes (e.g., determining if a question is well-formed). In this way, the developmental computational model allows us to test explicit hypotheses about the language knowledge that could be derived from the information available in children's experience (Hoff, 2006).

Because the developmental computational model concretely links children's input to a predicted language development outcome, the model will pinpoint what aspect of children's input is relevant, and predict the expected language development outcome on the basis of that relevant input (e.g., if children will believe a question is well-formed). That is, a developmental computational model identifies if any observed input difference is predicted to be developmentally meaningful.

In this way, a developmental computational model, as a concrete implementation of a theory of learning, allows us to generate testable predictions about the relationship between observed input differences and language development outcomes. These predictions can then be evaluated with targeted behavioral work that assesses the predicted development outcomes. If language outcome differences are predicted and they do indeed appear, then we have strong support that the input aspect highlighted by the developmental computational model is developmentally meaningful. In this way, developmental computational modeling connects theories of language development, empirical data on children's input, and child behavioral experiments.

Developmentally-meaningful input differences can then be targeted for intervention, with the strong possibility of positively impacting language development outcomes. Importantly, because a developmental computational model describes exactly how the input causes the predicted developmental outcome, the model can also determine if an observed input difference is predicted *not* to be developmentally meaningful (because the predicted outcome isn't qualitatively different). That is, the model can identify contextual variation surfacing in children's input that's predicted not to impact language development (Hoff, 2006). In this case, we would expect an input-based intervention to be ineffective at improving children's language development.

1.3 Language development across socio-economic status

Language development delays appear across socio-economic status (**SES**), with lower-SES children behind their higher-SES peers for different aspects of language development (e.g., vocabulary development: Hart and Risley 1995; Hoff 2003, language processing: Fernald et al. 2013; Weisleder and Fernald 2013). Yet, there are also aspects of language development that don't appear to be delayed (e.g., complex syntactic knowledge: de Villiers et al. 2008), and there are many aspects where we simply don't know. It's often unclear which input differences matter for the development of different types of linguistic knowledge. Certainly, there are observed input quantity and input quality differences across SES (though also within SES).

For instance, when it comes to input quantity at the word-level, some studies have found that lower-SES children may encounter 30 million fewer words of caretaker speech than their higher-SES peers (Hart and Risley, 1995; Schwab and Lew-Williams, 2016); other studies have found greater differences in input quantity within SES rather than across SES (Blum, 2015; Sperry et al., 2018). At the clause-level however, there appear to be fewer observed input quantity differences across SES; for instance, caretakers across SES produce approximately the same number of multi-clause utterances (e.g., *[He gave the book to the girl [who lived down the street]]* = 2 clauses) in their child-directed speech (Huttenlocher et al., 2002).

For input quality, differences across SES have been observed at the lexical and foundational syntactic levels (Huttenlocher et al., 2010; Rowe, 2012; Rowe et al., 2017). These differences include the relative frequency of word types, word tokens, and rare words; the diversity of syntactic constructions; and the relative frequency of decontexualized utterances like explanations (*Oh, we can't put them in the bus because the bus is full of blocks*), pretend (*I'll save you from the wicked sister*), and narrations (*He is going to look in your nose and your throat and your ears*).

Again, what's often unclear is whether a specific measurable input difference matters for developing a specific type of linguistic knowledge. Developmental computational modeling, by implementing a learning theory that links input to the development of language knowledge, can identify when a measurable input difference is predicted to matter – that is, when a difference is developmentally meaningful.

1.4 Using developmental computational modeling to predict whether syntactic input differences across SES are developmentally meaningful

Here, we harness developmental computational modeling to do precisely this: identify if input differences across SES for certain aspects of complex syntax are predicted to impact development of that knowledge and so be developmentally meaningful. We focus on a certain type of complex syntactic knowledge known as *syntactic islands* that concerns *wh*-questions (e.g., the grammatical *Who did Lily think the pretty kitty was for?* vs. the ungrammatical *Who did Lily think the kitty for was pretty?*). More specifically, syntactic islands are constraints on the permitted forms of *wh*-questions (among other linguistic forms). Knowledge of syntactic islands thus allows children to know which *wh*-questions are well-formed and which aren't; so, even if children have never heard a particular *wh*-question before, they can tell if it's acceptable or not. This means that children who have knowledge of syntactic islands know something quite sophisticated about the syntax of *wh*-questions – not simply how to understand *wh*-questions that occur in their language, but also which ones aren't going to occur at all because those *wh*-questions are ill-formed.

We first briefly review what's currently known about the development of complex syntactic knowledge across SES, focusing on knowledge related to *wh*-questions in general and syntactic islands in particular. We then discuss the complex syntactic knowledge that syntactic islands involve, and describe the particular syntactic islands we focus on; we selected these due to the available empirical data on the behavior that signals successful acquisition of this knowledge.

We then review a developmental computational model for learning syntactic islands that connects children's input to language development (Pearl and Sprouse, 2013); this model implements a specific learning theory for how children use their input to acquire knowledge of syntactic islands. The learning theory pinpoints that the relevant aspect of children's input for learning syntactic islands involves *wh*-dependencies, which rely on "*wh*-words" like *what* and *who* (among others).

We then investigate input variation for learning syntactic islands, looking at the distributions of *wh*-dependencies in American English child-directed speech (**CDS**) between high-SES populations and low-SES populations. In particular, we provide both a descriptive corpus analysis and a quantitative analysis comparing high-SES to low-SES input. We then assess input quantity differences, and derive realistic estimates of the quantity of *wh*-dependencies that high-SES vs. low-SES children would hear by age four, when one of the syntactic islands we investigate has been acquired (de Villiers et al., 2008). With realistic estimates of the input data to high-SES and low-SES children, we then provide a developmental computational modeling analysis of the input quality; in particular, the model predicts the syntactic island knowledge that high-SES and low-SES children would be able to acquire on the basis of their *wh*-dependency input.

We find that the low-SES input, in terms of *wh*-dependency distribution and the syntactic building blocks needed for syntactic islands, is similar in several key respects to the high-SES CDS distribution. More specifically, our modeling results predict that low-SES input can support acquisition of all the investigated syntactic islands by age four as well as high-SES input does. Thus, our results predict that input quality for syntactic islands is the same across SES – there are no developmentally-meaningful differences across SES coming from children's input, with respect to acquiring this complex syntactic knowledge. This result accords with known developmental trajectory evidence for one type of syntactic island knowledge, and predicts additional trajectory similarities for the other types we investigate here.

Interestingly, a syntactic building block involving complementizer *that* (e.g., *that* in *Who do* *you think that Lily likes?*) is predicted to be crucial for successful knowledge development and comes from a different *wh*-dependency type in low-SES CDS, compared with high-SES CDS; this difference highlights that surface input quality differences may mask deeper input quality similarities. More generally, our results suggest that the quality of the input for learning about syntactic islands doesn't fundamentally differ across SES. We discuss implications for linguistic development across SES and potential adult syntactic knowledge variation.

2 The development of *wh*-dependency knowledge across SES

Currently, less is known about the development of complex syntactic knowledge across SES (especially with respect to *wh*-dependencies) than about the development of lexical and foundational syntactic knowledge. Still, we do know about the development of some *wh*-dependency knowledge across SES and a little about the *wh*-dependency input.

In terms of *wh*-dependency knowledge, high-SES English-learning children at 20 months seem to represent the full structure of *wh*-dependencies in *wh*-questions (e.g., *Which cat did the dog bump?*) and relative clauses (e.g., *Show me the dog [who the cat bumped]*), rather than relying on vocabulary-based heuristics to understand these *wh*-dependencies (Seidl et al., 2003; Gagliardi et al., 2016; Perkins and Lidz, 2020). High-SES children are also able to correctly repeat back well-formed *wh*-questions like *Who can Falkor save?* and generate new well-formed *wh*-questions by two and a half to three years old (Valian and Casey, 2003).

By age four, we see similar knowledge across SES about several aspects of *wh*-dependencies (see de Villiers et al. (2008) for empirical data across SES, as well as a review of prior empirical data from high-SES children). This knowledge includes sensitivity to allowed interpretations of *wh*-dependencies – that is, constraints on which interpretations are allowed because those interpretations depend on which *wh*-dependencies are allowed.

For instance, four-year-olds (like adults) can interpret *wh*-dependencies like "*How did the boy say he hurt himself?*" with *how* modifying the embedded clause verb *hurt*; so, the *wh*-question can be interpreted as asking about how the boy hurt himself. Children as young as four are also sensitive to the difference between the possible interpretations of "*How did the mom learn what to bake?*" The allowed interpretation has *how* modifying the main clause verb *learn* (e.g., a possible answer is "from a recipe book"); the disallowed interpretation has *how* modifying the embedded clause verb *bake* (e.g., a possible answer would be "in a glass dish").

As another example, four-year-olds across SES are sensitive to the difference between the possible interpretations of "*What is Jane drawing a monkey that is drinking milk with?*"; the allowed interpretation has *what* linked to a position outside the relative clause ("*What is Jane drawing [a monkey that is drinking milk] with* $\_\_{what}$*?*"), with a possible answer of what Jane is drawing with (e.g., "a pencil"); the disallowed interpretation has *what* linked to a position inside the relative clause ("*What is Jane drawing [a monkey that is drinking milk with* $\_\_{what}$*]?*"), with a possible answer of what the monkey is drinking with (e.g., "a straw").

So, there appear to significant similarities in the developmental outcomes by age four across SES with respect to allowing and disallowing possible interpretations for *wh*-questions; these interpretations rest on children being sensitive to several constraints on allowed *wh*-dependencies. These developmental outcome similarities suggest that input differences across SES for these types of *wh*-dependency knowledge shouldn't be developmentally meaningful. Yet, we know much less about any input differences there might be for *wh*-dependencies, let alone how children's input leads to the development of *wh*-dependency knowledge despite any input variation that might be present. We do know that the use of *wh*-questions in input to low-SES two-year-olds helps build their vocabulary and reasoning skills more generally (Rowe et al., 2017). However, it's unclear how the *wh*-questions in the input impact the development of complex *wh*-dependency knowledge (such as constraints on *wh*-dependencies). More generally, much remains unknown about the input variation present across SES for learning about constraints on *wh*-dependencies, how the input scaffolds the development of this complex syntactic knowledge, and whether any input variation present is developmentally meaningful for other types of constraints on *wh*-dependencies that have yet to be assessed in children across SES.

3 Syntactic islands

A key component of human syntactic knowledge is the ability to have long-distance dependencies, where there's a relationship between two words that aren't adjacent to each other. Long-distance dependencies, such as the dependencies between the *wh*-word *what* and *eat* in (1), can be arbitrarily long (Chomsky, 1965; Ross, 1967; Chomsky, 1973). In (1), we can see that this dependency can stretch across one, two, three, or four clauses. In each case, *what* is understood as the thing Falkor ate, despite *what* not being adjacent to *eat*.

(1)
a. What did Falkor eat $\_\_{what}$?
b. What did Atreyu see Falkor eat $\_\_{what}$?
c. What did the Childlike Empress say Atreyu saw Falkor eat $\_\_{what}$?
d. What did Bastian hear the Childlike Empress say Atreyu saw Falkor eat $\_\_{what}$?

However, there are constraints on the *wh*-dependencies that are allowed. These constraints have been discussed as specific syntactic structures that long-distance dependencies can't cross, called *syntactic islands* (Chomsky, 1965; Ross, 1967; Chomsky, 1973). Four examples of syntactic islands are in (2), with * indicating ungrammaticality and [...] highlighting the proposed island structure that a *wh*-dependency can't cross in English.

(2)
a. **Complex NP island**
*What did Falkor make [the claim [that Atreyu fought $\_\_{what}$]]?
b. **Subject island**
*What did Falkor think [[the joke about $\_\_{what}$] was hilarious]?
c. **Whether island**
*What did Falkor wonder [whether Atreyu bought $\_\_{what}$]?
d. **Adjunct island**
*What did Falkor worry [if Atreyu buys $\_\_{what}$]?

During language development, children must infer and internalize the constraints on long-distance *wh*-dependencies (i.e., syntactic island constraints) that allow them to recognize that the questions in (2) are not allowed, while the questions in (1) are fine. We note that this recognition is the measurable behavior of children's internalized knowledge – that is, distinguishing grammatical questions like (1) from ungrammatical questions like (2) is one way to indicate knowledge of the relevant syntactic island constraints (whatever form that knowledge may take).

4 Assessing knowledge of syntactic islands

Previous work assessing children's knowledge has focused on the interpretations of *wh*-dependencies that are allowed, rather than the grammaticality of the *wh*-dependencies directly (Otsu, 1981; De Villiers et al., 1990; Roeper and Seymour, 1994; de Villiers and Roeper, 1995; McDaniel et al., 1995; Vainikka and Roeper, 1995; De Villiers and Pyers, 2002; Coles-White et al., 2004; de Villiers et al., 2008). The idea was that it's easier to ask children if they can allow a particular interpretation that relies on a certain *wh*-dependency (something more similar to naturalistic communication) rather than asking children directly if that *wh*-dependency is grammatical (something more meta-linguistic that requires reasoning about language forms). If children don't allow a certain interpretation (e.g., "*What is Jane drawing a monkey that is drinking milk with?*" with *what* interpreted as "the straw"), this can be interpreted as children not allowing the *wh*-dependency that the interpretation relies on (e.g., "*What is Jane drawing [a monkey that is drinking milk with* $\_\__{what}$*]?*"); so, this behavior can then be interpreted as children knowing the syntactic island that disallows that *wh*-dependency (e.g., a Complex NP island, since "a monkey that is..." is an NP with a relative clause, which is a type of complex NP). In this way, children disallowing a particular interpretation indirectly indicates their knowledge of a particular syntactic island – specifically, the syntactic island that disallows the *wh*-dependency that the disallowed interpretation relies on.

A more direct way to assess syntactic island knowledge is with the less-natural task of directly judging how acceptable a *wh*-dependency is (e.g., Sprouse et al. 2012). When the stimuli are carefully designed (as discussed below), relative differences in judged acceptability can be used to infer whether a particular dependency is allowed (i.e., grammatical). In particular, island-crossing *wh*-dependencies can be compared against *wh*-dependencies that don't cross islands, yet are similar in other important ways to the island-crossing ones. When the island-crossing *wh*-dependencies are still judged as far more unacceptable, this signals knowledge of the relevant constraint on *wh*-dependencies captured by syntactic islands. We follow Pearl and Sprouse (2013), and use acceptability judgment data to indicate knowledge of syntactic islands. In particular, the developmental models we implement will attempt to replicate the appropriate acceptability judgment pattern found by Sprouse et al. (2012) that indicates syntactic island knowledge, as this is the target knowledge for development.

Sprouse et al. (2012) investigated the four islands from (2); a sample set for each island type is shown in (3)-(6), where island structures are indicated with [...]. These stimuli were designed using a 2x2 factorial design, involving two factors deemed important for judging acceptability: *wh*-dependency length (matrix vs. embedded) and presence of an island structure in the utterance (non-island vs. island). Each island stimuli set therefore had four *wh*-dependency types: matrix+non-island, embedded+non-island, matrix+island, and embedded+island. The embedded+island stimulus in each case involved an island-crossing *wh*-dependency, and so was ungrammatical.

(3) Sample Complex NP island stimuli

a. matrix+non-island
Who $\_\__{who}$ claimed that Atreyu fought the goblin?

b. embedded+non-island

Who did Falkor claim that Atreyu fought $\_\_{who}$?

c. matrix+island:
Who $\_\_{who}$ made [the claim that Atreyu fought the goblin]?

d. embedded+island:
*Who did Falkor make [the claim that Atreyu fought $\_\_{who}$]?

(4) Sample Subject island stimuli

a. matrix+non-island:
Who $\_\_{who}$ thinks the joke is hilarious?

b. embedded+non-island:
What does Falkor think $\_\_{what}$ is hilarious?

c. matrix+island:
Who $\_\_{who}$ thinks the joke about Atreyu is hilarious?

d. embedded+island:
*Who did Falkor think [[the joke about $\_\_{who}$] was hilarious]?

(5) Sample Whether island stimuli

a. matrix+non-island:
Who $\_\_{who}$ thinks Atreyu bought the medallion?

b. embedded+non-island:
What does Falkor think Atreyu bought $\_\_{what}$?

c. matrix+island:
Who $\_\_{who}$ wonders if Atreyu bought the medallion?

d. embedded+island:
*What did Falkor wonder [whether Atreyu bought $\_\_{what}$]?

(6) Sample Adjunct island stimuli

a. matrix+non-island:
Who $\_\_{who}$ thinks Atreyu bought the medallion?

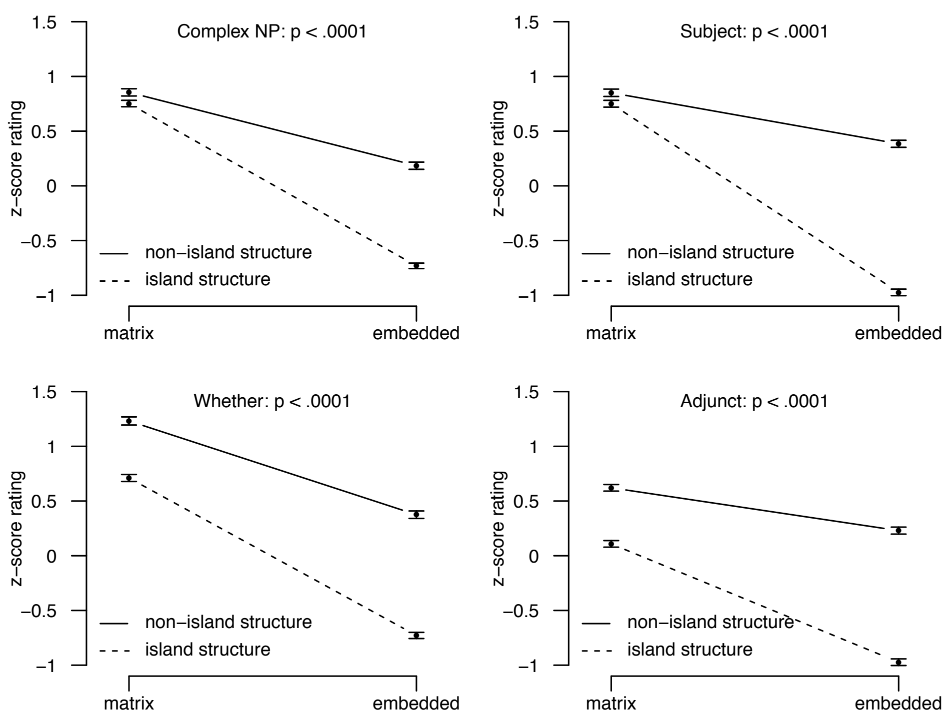
b. embedded+non-island:
What does Falkor think that Atreyu bought $\_\_{what}$?

c. matrix+island:
Who $\_\_{who}$ worries if Atreyu bought the medallion?

d. embedded+island:
*What did Falkor worry [if Atreyu buys $\_\_{what}$]?

This design allows syntactic island knowledge to surface as a superadditive interaction of acceptability judgments; this superadditivity appears as non-parallel lines in an interaction plot, such as those in Figure 1, which come from the judgments of high-SES adults tested by Sprouse et al. (2012). In particular, if we consider the Complex NP plot in the top row, there are four acceptability judgments, one for each of the stimuli in (3). The matrix+non-island dependency of (3a) has a certain acceptability score – this is the top-lefthand point. There is a (slight) drop in acceptability when the matrix+island dependency of (3c) is judged in comparison to (3a) – this is the lower-lefthand point. We can interpret this as the unacceptability associated with simply having an island structure in the utterance. There's also a drop in acceptability when the embedded+non-island dependency of (3b) is judged in comparison to (3a) – this is the upper-righthand point. We can interpret this as the unacceptability associated with simply having an embedded *wh*-dependency. If the unacceptability of the embedded+island dependency of (3d) were simply the result of those two unacceptabilities (having an island structure in the utterance and having an embedded *wh*-dependency), the drop in unacceptability would be additive and the lower-righthand point would be just below the upper-righthand point (and so look just like the points on the lefthand side). But this isn't what we see – instead, the acceptability of (3d) is much lower than this. This is a superadditive effect for the embedded+island stimuli. So, the additional unacceptability of an island-crossing-dependency like (3d) – i.e., implicit knowledge of syntactic islands – appears as a superadditive interaction in these types of acceptability judgement plots. This superadditive acceptability judgment pattern appears for all four island types tested by Sprouse et al. (2012) from (2): Complex NP, Subject, Whether, and Adjunct islands. A modeled learner who can successfully acquire knowledge of these syntactic islands from its input should be able to reproduce this superadditive judgment pattern.

Figure 1: High-SES adult judgments demonstrating implicit knowledge of four syntactic islands via a superadditive interaction.



5 Linking children's input to syntactic island development

Pearl and Sprouse (2013) constructed a developmental computational model for learning these syntactic island constraints, linking children's input to the knowledge about *wh*-dependencies that children develop over time. In particular, this model implements a specific learning theory about how children use the *wh*-dependency information in their input to update their internal representations for *wh*-dependencies; these internal representations allow children to judge a particular *wh*-dependency as grammatical (or not). The model's learning theory assumes children can characterize a long-distance dependency as a syntactic path from the head of the dependency (e.g., *What* in (7)) through a set of structures that contain the tail (e.g., $\_\__{what}$) of the dependency, as shown in (7a)-(7b). These structures correspond to phrase types that make up *wh*-questions such as Verb Phrases (VP), Inflectional Phrases (IP), and Complementizer Phrases (CP), among others. Importantly, these are the structures that *wh*-dependencies could cross when forming *wh*-questions. Under this view, children simply need to learn which *wh*-dependencies have licit syntactic paths and which don't.

The learning process itself is implemented as a probabilistic learning algorithm that tracks local pieces of these syntactic paths. It breaks the syntactic path into a collection of syntactic trigrams that can be combined to reproduce the original syntactic path, as shown in (7c).[1] The modeled learner then tracks the frequencies of these syntactic trigrams in the input, based on the input it encounters, one data point at a time. The modeled learner later uses these learned frequencies to calculate probabilities for all syntactic trigrams comprising a *wh*-dependency[2] and so generate the probability of any *wh*-dependency (as shown in (8)- (9)). More specifically, any *wh*-dependency's probability is the product of the individual trigram probabilities that comprise its syntactic path, as shown in (10). Importantly, relying on the frequencies of syntactic trigrams (rather than the frequencies of entire *wh*-dependencies) allows the modeled learner to generate probabilities for any *wh*-dependency, including *wh*-dependencies that it's never seen before in its input. So, grammatical dependencies that are unattested can still have a higher probability than ungrammatical ones that are unattested, depending on the syntactic trigrams that comprise the *wh*-dependency.

The probability generated by the modeled learner corresponds to whether that dependency is allowed, with higher probabilities indicating grammatical dependencies and lower probabilities indicating ungrammatical dependencies. So, the modeled learner can generate judgments of *wh*-dependencies (e.g., grammatical vs. ungrammatical); if this learner can generate the same pattern of judgments that adults do, we can interpret this as the learner internalizing some version of the knowledge adults use to make those judgments. In this case, that means the modeled learner has internalized knowledge (via the syntactic trigrams) that allow it to capture syntactic island constraints. In this way, we can say that it's learned those syntactic island constraints.

(7) What did Falkor claim that Atreyu fought $\_\__{what}$?

a. Syntactic structures containing the *wh*-dependency:
What did [$_{IP}$ Falkor [$_{VP}$ claim [$_{CP}$ that [$_{IP}$ Atreyu [$_{VP}$ fought $\_\__{what}$]]]]]?

b. Syntactic path of *wh*-dependency:
start-IP-VP-CP$_{that}$-IP-VP-*end*

[1] For discussion of the empirical motivation for the modeling choices, including using trigrams and the aggregation of trigrams into a dependency, see Pearl and Sprouse (2013).

[2] It smooths these probabilities by adding 0.5 to all trigram counts. This allows the modeled learner to accept dependencies composed of trigrams it's never seen before, though it gives them a much lower probability than dependencies composed of trigrams it has in fact seen before. See Pearl and Sprouse (2013, 2015) for further discussion of this point.

c. Syntactic trigrams $T \in$ syntactic path:
= *start*-IP-VP
IP-VP-CP$_{that}$
VP-CP$_{that}$-IP
CP$_{that}$-IP-VP
IP-VP-*end*

(8) Smoothed probabilities of trigrams:
p(*start*-IP-VP) $\approx \frac{count(start-IP-VP)}{total\ count\ of\ all\ trigrams}$
...
p(IP-VP-*end*) $\approx \frac{count(IP-VP-end)}{total\ count\ of\ all\ trigrams}$

(9) Probability of new *wh*-dependency: What did Engywook tell Atreyu $\_\_{what}$?
Syntactic structures = What did [$_{IP}$ Engywook [$_{VP}$ tell Atreyu $\_\_{what}$?]]
Syntactic path = *start*-IP-VP-*end*
trigrams = *start*-IP-VP, IP-VP-*end*
Probability = p(*start*-IP-VP-*end*) = p(*start*-IP-VP)*p(IP-VP-*end*)

(10) General formula for generating a *wh*-dependency's probability:
$\prod_{trigrams \in T} p(trigram)$

We note that this developmental learning model requires children to have certain (potentially sophisticated) knowledge and abilities before they can utilize the learning strategy implemented by this model. Core assumptions of the model require that the child be able to (i) parse sentences into phrase structure trees, (ii) extract the syntactic paths for the dependencies, (iii) track the frequency of the syntactic trigams, and (iv) calculate the probability for the complete syntactic path of the dependency, based on its trigrams. It remains for future work to determine when these core pieces are available in children – once they are, children would be able to harness the input the way this learning model does.[3]

Still, with these prerequisities in place, Pearl and Sprouse (2013) found that a modeled learner using high-SES child-directed speech as input could generate the superadditivity patterns for all four islands in (2). So, the learning theory encoded by the modeled learner predicts development of implicit syntactic island knowledge for these islands, given high-SES child input. In terms of developmental outcomes, high-SES children do indeed seem to be sensitive to Complex NP islands by age four (de Villiers et al., 2008)[4]. So, the predicted developmental outcome for Complex NP islands aligns with the assessed outcome for high-SES children by age four, and therefore supports the plausibility of the learning theory encoded by the model.

[3]We also note that we're using this learning model as a novel way to assess input quality, on the basis of prior work (Pearl and Sprouse, 2013) that applies it to high-SES children's input. Because of this focus, we won't discuss the theoretical implications of this learning strategy for questions of innateness with respect to the knowledge needed and assumed by the model; we instead refer interested readers to the discussion in Pearl and Sprouse (2013).

[4]See Section 6.4 for more detailed discussion of this interpretation of children's *wh*-dependency knowledge, based on the stimuli used in de Villiers et al. (2008).

6 Input analysis across SES

Here we assess input quality across SES, focusing on the information necessary for the development of the syntactic island knowledge for the four islands in (2). We first want to identify any quantitative differences between the high-SES and low-SES input samples we have in terms of the *wh*-dependencies and resulting syntactic trigrams available; recall that these dependencies and trigrams are the foundation of the development of syntactic island constraints, based on the learning theory in the model of Pearl and Sprouse (2013). We'll identify any quantitative input differences via quantitative analysis of the distribution of *wh*-dependencies and syntactic trigrams available, as well as an estimation of the quantity of *wh*-dependency input available across SES by age four.

We then want to identify any differences between the high-SES and low-SES input in terms of how well the available *wh*-dependencies and syntactic trigrams are predicted to scaffold the development of syntactic island knowledge. That is, whether any significant quantitative differences exist or not, does low-SES input differ from high-SES input in how syntactic development is predicted to occur by age four, based on that input? We'll answer this question by applying the same computational learning model from Pearl and Sprouse (2013) to realistic estimates of the child-directed speech that high-SES and low-SES children encounter by age four, and comparing the predicted developmental outcomes.

6.1 Input samples

High-SES. Our high-SES input samples are the data used by Pearl and Sprouse (2013), and come from the structurally-annotated Brown-Adam (Brown, 1973), Brown-Eve (Brown, 1973), Valian (Valian, 1991), and Suppes (Suppes, 1974) corpora from the CHILDES Treebank (Pearl and Sprouse, 2013). These data are child interactions involving 24 children between the ages of one and a half and four, containing 101,838 utterances with 20,923 *wh*-dependencies.

Low-SES. Our low-SES CDS input samples come from a subpart of the HSLLD corpus (Dickinson and Tabors, 2001) in CHILDES (MacWhinney, 2000), and SES was defined according to maternal education and annual income. Maternal education ranged from 6 years of schooling to some post-high school education. Annual income didn't have hard lower and upper bounds; instead, 70% of the families reported an annual income of $20,000 or less, while 21% of the families reported an income of over $25,000. The annual income of the remaining 9% was unreported. In this dataset, we focused on the Elicited Report, Mealtime, and Toy Play sections, which represent more naturalistic interactions. We also drew our samples from Home Visit 1, which recorded child language interactions involving children between the ages of three and five. Our sample contained 31,875 utterances and 3,904 *wh*-dependencies directed at 78 children. We extracted and syntactically annotated all *wh*-dependencies following the format of the CHILDES Treebank (Pearl and Sprouse, 2013).

***Wh*-dependency coding.** The structural annotations of the *wh*-dependencies in each sample indicate the syntactic structure necessary to characterize the syntactic paths of *wh*-dependencies. We coded the syntactic paths of the dependencies (as in (7b) and shown below with a different example in (11)). Following Pearl and Sprouse (2013), the *CP* phrase structure nodes were further subcategorized by the lexical item serving as complementizer, such as CP_{that}, $CP_{whether}$, CP_{if}, and CP_{null}. This allows the modeled learner of Pearl and Sprouse (2013) to distinguish dependencies judged by high-SES adults to be grammatical, like (11a), from those judged to be ungrammatical, like (11b). With these syntactic paths characterizing *wh*-dependencies, we can then assess the distribution of the *wh*-dependencies in each input sample.

(11) a. Who do you think $\_\_{who}$ read the book?
syntactic path: *start-IP-VP-*CP_{null}*-IP-end*
b. *Who do you think that $\_\_{who}$ read the book?
syntactic path: **start-IP-VP-*CP_{that}*-IP-end*

6.2 Descriptive corpus analyses

***Wh*-dependencies.** Our corpus analyses found 12 *wh*-dependencies in common between the high-SES and low-SES child input samples (out of 26 total in the high-SES and 16 total in the low-SES).[5] So, the high-SES input contained 14 *wh*-dependency types not in the low-SES input, and the low-SES input contained 4 *wh*-dependency types not in the high-SES input, shown in Table 1.

Interestingly, the last dependency type in Table 1 found only in the low-SES child input (e.g., *What do you think that happens?*) is an example of a "*that*-trace" violation judged ungrammatical as by high-SES adults (Cowart, 1997). Because adults are producing these child-directed speech samples, the presence of this dependency in low-SES child input represents a difference across SES with respect to adult knowledge of specific *wh*-dependencies; in particular, low-SES adults potentially believe this *wh*-dependency is grammatical, unless each instance was a speech error.

When we compare the rate of *wh*-dependencies across SES, we find that the *wh*-dependency rate is considerably higher in high-SES child-directed speech (high-SES: 20.5%, low-SES: 12.2%). Over time (as detailed in section 6.4), this rate difference can lead to a considerable difference in the quantity of *wh*-dependencies encountered.

Still, there's a striking similarity when we look at the most frequent *wh*-dependencies types across SES: the two dependency types that account for the vast majority of the low-SES *wh*-dependency input (85.8%) are the same two that account for the vast majority of the high-SES input (89.5%), and they occur in about the same proportions (shown in (12)). This suggests a high-level qualitative similarity in the *wh*-dependency input across SES, despite the individual *wh*-dependency differences.

(12) Proportions of the two most frequent *wh*-dependency types across SES
a. 1^{st} most frequent: *start-IP-VP-end* (e.g., *What did Lily read* $\_\_{what}$*?*)
75.5% low-SES, 76.7% high-SES
b. 2^{nd} most frequent: *start-IP-end* (e.g., *What* $\_\_{what}$ *happened?*)
10.3% low-SES, 12.8% high-SES

[5] A more detailed description of the *wh*-dependency distribution across SES is available in Appendix A.1.

Table 1: *Wh*-dependencies and syntactic trigrams unique to speech samples directed at high-SES and low-SES children, respectively. Unique syntactic trigrams are on the same row as the unique *wh*-dependencies they come from.

***wh*-dependencies**	**syntactic trigrams**
only high-SES	
start-IP-VP-CP$_{for}$-IP-VP-PP-end (e.g., *What did she put on for you to dance to $\_\_{what}$?*)	*IP-VP-CP$_{for}$*, *VP-CP$_{for}$-IP*, *CP$_{for}$-IP-VP*
start-IP-VP-CP$_{null}$-IP-VP-IP-VP-IP-VP-end (e.g., *What did he think she wanted to pretend to steal $\_\_{what}$?*)	
start-IP-VP-CP$_{null}$-IP-VP-IP-VP-PP-end (e.g., *Who did he think she wanted to steal from $\_\_{who}$?*)	
start-IP-VP-CP$_{null}$-IP-VP-NP-end (e.g., *What did he think she said $\_\_{what}$ about it?*)	
start-IP-VP-CP$_{null}$-IP-VP-PP-PP-end (e.g., *What did he think she wanted out of $\_\_{what}$?*)	
start-IP-VP-CP$_{that}$-IP-VP-end (e.g., *What did he think that she stole $\_\_{what}$?*)	*CP$_{that}$-IP-VP*
start-IP-VP-IP-end (e.g., *What did he want $\_\_{who}$ to steal the necklace?*)	*VP-IP-end*
start-IP-VP-IP-VP-IP-VP-PP-end (e.g., *Who did he want her to pretend to steal from $\_\_{who}$?*)	
start-IP-VP-IP-VP-NP-end (e.g., *What did he want to say $\_\_{what}$ about it?*)	
start-IP-VP-IP-VP-NP-PP-end (e.g., *What did she want to steal more of $\_\_{what}$?*)	
start-IP-VP-NP-end (e.g., *What did she say $\_\_{what}$ about the necklace?*)	*VP-NP-end*
start-IP-VP-PP-CP$_{null}$-IP-VP-end (e.g., *What did she feel like he saw $\_\_{what}$?*)	*VP-PP-CP$_{null}$*, *PP-CP$_{null}$-IP*
start-IP-VP-PP-NP-PP-end (e.g., *What do you put it on top of $\_\_{what}$?*)	*VP-PP-NP*, *PP-NP-PP*
start-IP-VP-PP-IP-VP-end (e.g., *What did he think about stealing $\_\_{what}$?*)	
only low-SES	
start-IP-VP-CP$_{null}$-IP-VP-NP-PP-end (e.g., *What did he think it was a movie of $\_\_{what}$?*)	
start-IP-VP-IP-VP-IP-VP-PP-IP-VP-end (e.g., *What did you want to try to plan on doing $\_\_{what}$?*)	
start-IP-VP-PP-IP-VP-end (e.g., *What did she think about buying $\_\_{what}$?*)	
start-IP-VP-CP$_{that}$-IP-end (e.g., *What do you think that $\_\_{what}$ happens?*)	*CP$_{that}$-IP-end*

Syntactic trigrams. For syntactic trigrams, which serve as the building blocks of *wh*-dependencies under the Pearl & Sprouse learning strategy, our corpus analysis found 19 syntactic trigrams in common between the high-SES and low-SES child input samples (out of 29 total for the high-SES and 20 total in the low-SES). So, the high-SES input contained 10 syntactic trigrams not in the low-SES input, and the low-SES input contained 1 syntactic trigram not in the high-SES input, shown in Table 1.[6]

Notably, just as with the *wh*-dependency analysis, the most frequent syntactic trigrams are very similar across SES. The three trigram types that account for the majority of the trigrams (85.0%) in the low-SES *wh*-dependency input are the same three that account for the majority of the trigrams (87.9%) in the high-SES *wh*-dependency input, and they occur in about the same proportions (shown in (13)). So, as with the *wh*-dependencies, this suggests a high-level qualitative similarity in the syntactic trigram input across SES, despite the individual syntactic trigram differences.

(13) Proportions of the three most frequent trigram types across SES

a. 1^{st} most frequent: *start-IP-VP*
41.4% low-SES, 41.8% high-SES

b. 2^{nd} most frequent: *IP-VP-end*
38.9% low-SES, 40.0% high-SES

c. 3^{rd} most frequent: *start-IP-end*
4.7% low-SES, 6.1% high-SES

6.3 Input distribution comparisons

To more precisely quantify how similar the input distributions are for both *wh*-dependencies and syntactic trigrams across SES, we use the Jensen-Shannon divergence (**JSDiv**) (Endres and Schindelin, 2003). JSDiv values range from 0 to 1, with 0 indicating identical distributions. That is, higher JSDiv values indicate greater divergence in the distributions, while values closer to 0 indicate distributions that are more similar. In this way, JSDiv analysis provides a way to quantify similarity between distributions; this makes it useful as a comparative measure, where different distributions are assessed for their relative similarity to each other.

With this in mind, we additionally use JSDiv to assess child-directed speech in comparison to adult-directed speech and text, in order to provide a comparison baseline for the similarity across input samples of both *wh*-dependencies and syntactic trigrams. In particular, we assess how similar the low-SES and high-SES CDS *wh*-dependency and trigram distributions are to those in high-SES adult-directed speech (**ADS**) and adult-directed text (**ADT**) samples from Pearl and Sprouse (2013), based on the *wh*-dependencies and syntactic trigrams in common across these corpus samples. The adult-directed corpora are described in Table 2. This JSDiv analysis will reveal which factors impact *wh*-dependency and syntactic trigram distributions more: SES, whether the speech is directed at children or adults, or whether the input is speech-based vs. text-based.

***Wh*-dependencies.** Figure 2 shows the results of the JSDiv analysis for *wh*-dependencies, calculated over the distribution of the 9 *wh*-dependencies (shown in Table 3) that these four corpora had in common. We see that low-SES CDS and high-SES CDS are the most similar in *wh*-dependency distribution (JS: 0.00445), and appear to be twice as similar as the next closest comparison, which

[6] A more detailed description of the syntactic trigram distribution across SES is available in Appendix A.2.

Table 2: Corpora statistics for low-SES CDS (L-CDS), high-SES CDS (H-CDS), high-SES adult-directed speech (H-ADS), and high-SES adult-directed text (H-ADT) samples used for JSDiv analysis.

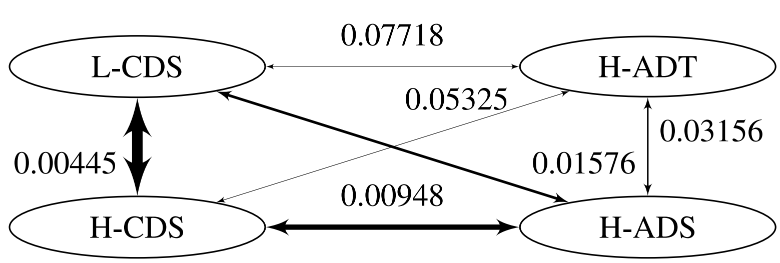
corpora	# utterances	# *wh*-dependencies	# children	ages
L-CDS	31,875	3,904	78	3 - 5
H-CDS	101,838	20,923	25	1 - 5
H-ADS	74,576	8,508	N/A	N/A
H-ADT	24,243	4,230	N/A	N/A

Table 3: The nine *wh*-dependencies shared across all four corpora that are used in the JSDiv analysis.

Shared dependencies	Example utterance	Corpora percentage
start-IP-*end*	*Who saw it?*	10.3% - 33.0%
start-IP-VP-*end*	*Who did she see?*	63.3% - 76.7%
start-IP-VP-CP$_{null}$-IP-*end*	*Who did he think stole it?*	0.1% - 0.6%
start-IP-VP-CP$_{null}$-IP-VP-*end*	*What did he think she stole?*	0.2% - 1.1%
start-IP-VP-CP$_{null}$-IP-VP-PP-*end*	*What did he think she wanted it for?*	<0.1% - 0.1%
start-IP-VP-IP-VP-*end*	*What did he want her to steal?*	1.3% - 7.5%
start-IP-VP-IP-VP-IP-VP-*end*	*What did he want her to pretend to steal?*	<0.1%
start-IP-VP-IP-VP-PP-*end*	*What did she want to get out from under?*	<0.1% - 0.8%
start-IP-VP-PP-*end*	*Who did she steal from?*	1.3% - 4.3%

is high-SES CDS vs. high-SES ADS (JS: 0.00948). This affirms a quantitative similarity across SES in child *wh*-dependency input, in terms of *wh*-dependency distribution. Moreover, these results highlight that CDS *across* SES is more similar than CDS vs. ADS *within* SES. That is, whether the speech is directed at children or adults matters more than whether speech is coming from a high-SES or low-SES population. We also note that these JSDiv results accord with intuitions that speech of any kind is more similar to other speech than it is to text: high-SES ADS diverges more from high-SES ADT (JS: 0.03156) than it does from either high-SES CDS (JS: 0.00948) or low-SES CDS (JS: 0.01576).

Figure 2: JSDiv analyses for low-SES CDS (L-CDS), high-SES CDS (H-CDS), high-SES adult-directed speech (H-ADS), and high-SES adult-directed text (H-ADT). Line thickness corresponds to similarity, with thicker lines indicating more similar distributions.

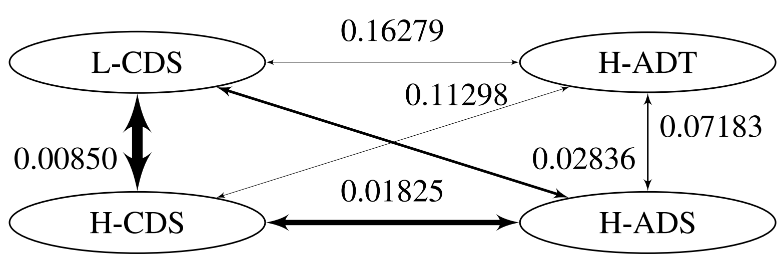


Syntactic trigrams. Figure 3 shows the results of the JSDiv analysis for syntactic trigrams, calculated over the distribution of the 14 trigrams shown in Table 4 (see Table A2 in Appendix A.2 for the full list of trigrams) that these four corpora had in common across all *wh*-dependencies. These trigrams accounted for 99.5-99.8% of the total trigrams in these corpora. As with the analysis of the *wh*-dependencies, we see the same pattern emerge: (i) low-SES CDS is more similar to high-SES CDS (JSDiv: 0.00850) than any other input type, and (ii) all speech is more similar to other types of speech than to text (speech vs. speech: JSDiv=0.00850-0.02836; speech vs. text: JSDiv=0.07183-0.16279).

Table 4: Distribution of the 14 syntactic trigrams across child-directed Low-SES (L-CDS) and High-SES (H-CDS), as well as High-SES adult-directed speech (H-ADS) and text (H-ADT).

Syntactic trigrams	Syntactic trigram percentage
CP_{null}-*IP-VP*	0.1% - 0.7%
CP_{null}-*IP-end*	<0.1% - 0.3%
IP-VP-CP_{null}	0.3% - 0.7%
IP-VP-CP_{that}	<0.1%
IP-VP-IP	0.9% - 4.0%
IP-VP-NP	<0.1% - 0.1%
IP-VP-PP	0.8% - 2.5%
IP-VP-end	38.5% - 39.9%
VP-CP_{null}*-IP*	0.3 - 0.7%
VP-CP_{that}*-IP*	<0.1%
VP-IP-VP	0.9% - 4.0%
VP-PP-end	0.8% - 2.3%
start-IP-VP	38.6% - 41.7%
start-IP-end	4.7% - 19.0%

Figure 3: JSDiv analyses for low-SES CDS (L-CDS) trigrams, high-SES CDS (H-CDS) trigrams, high-SES adult-directed speech (H-ADS) trigrams, and high-SES adult-directed text (H-ADT) trigrams. Line thickness corresponds to similarity, with thicker lines indicating more similar distributions.



Distributional analysis summary. Our distributional analyses suggest that the input children encounter for learning about syntactic islands is very similar across SES. In particular, both the *wh*-distributions and the syntactic trigram distributions appear quite similar, despite some individual *wh*-dependency and trigram differences. However, it's unclear if even these comparatively small differences could lead to different predicted developmental outcomes (e.g., a predicted delay at age four for certain syntactic islands). This is because even small input differences could be developmentally meaningful. So, using realistic estimates of high-SES and low-SES child-directed speech input, does the learning model predict similar developmental outcomes, or not? To assess this, we need to first determine what realistic estimates of high-SES and low-SES input are for children learning about syntactic islands.

6.4 Realistic estimates derived from empirical data

Child behavioral evidence (Otsu, 1981; de Villiers and Roeper, 1995; de Villiers et al., 2008) suggests that four-year-old children across SES know about a syntactic island constraint similar to the Complex NP island constraint we investigate here. In particular, four-year-olds don't allow an interpretation corresponding to island-crossing *wh*-dependencies like the one in (14a). This *wh*-dependency is similar to an island-crossing dependency we investigate here, shown in (14b); importantly, both *wh*-dependencies cross a complex NP (i.e., "the cat that was..." and "the claim that Atreyu..."), and so might be viewed as accessing similar knowledge about syntactic islands.

(14) *Wh*-dependencies crossing Complex NP islands

a. from Otsu (1981); de Villiers and Roeper (1995) and (de Villiers et al., 2008)
"What did the boy fix [the cat that was lying on the table with $\_\_{what}$]?"

b. from Sprouse et al. (2012)
"Who did Falkor make [the claim that Atreyu fought $\_\_{who}$]?

Given this, here we assume that age four, across SES, is the age by which children learn this one type of syntactic island constraint (i.e., one relating to complex NPs) on the basis of their input. With this age of acquisition, we can look to the amount of *wh*-dependency data children hear by that age as the amount of data they encounter before acquiring this knowledge. That is, by age four and before age five, children across SES have learned one of the four types of syntactic island constraints we examine here.[7] We can quantify that amount of data for both high-SES and low-SES children, on the basis of the data samples we have here, coupled with knowledge about the frequency of syntactic input more generally to children of different ages and from different SES backgrounds. This provides us a detailed comparison of the relative quantities of relevant input (i.e., the *wh*-dependencies) that children across SES would encounter before learning about Complex NP islands. We can use this same age, and therefore the estimates of relevant data quantities across SES, as a milestone at which to assess input quantity across SES for learning syntactic island constraints.

When children's learning period plausibly starts. To begin learning about constraints on *wh*-dependencies, children must be able to process *wh*-dependency structure. Current research suggests that children begin to represent the full structure of *wh*-dependencies (e.g., *wh*-questions and relative clauses) at 20 months (Seidl et al., 2003; Gagliardi et al., 2016; Perkins and Lidz, 2020). So, we take 20 months as the starting point of the learning period for syntactic island constraints.

[7] This means that a baseline check for the learning model is whether it predicts that Complex NP islands can be learned from the input that both high-SES and low-SES children encounter by age four.

How many hours awake during the learning period. Taking four years old as the end point of the learning period for syntactic islands, the estimated learning period is then from 20 months through the end of age four (59 months). We can estimate the number of hours awake by drawing on Davis et al. (2004), who summarize the hours asleep for young children at different ages, as shown in Table 5. In particular, one-year-olds sleep about 14 hours a day (awake for 10), two-year-olds sleep about 13 hours a day (awake for 11), three-year-olds sleep about 12 hours a day (awake for 12), and four-year-olds sleep about 11.5 hours a day (awake for 12.5). Based on this, we can estimate the hours awake between age 20 months and age 59 months, and sum them to estimate the total hours awake during this learning period. Our calculations in Table 5 yield about 14174 hours awake.

Table 5: Calculating the total hours awake for children between the ages of 20 and 59 months, the estimated learning period for syntactic islands. These calculations are based on waking hours per day (waking) and total waking hours. Cumulative hours awake are shown at age one (20-23 months), two (24-35 months), three (36-47 months), and four (48-59 months).

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

How many *wh*-dependencies during the learning period. Based on the number of hours awake during the learning period, we can then estimate the input quantity with respect to *wh*-dependencies by drawing on estimates of the number of utterances children from different SES backgrounds hear per hour and our own samples of the rate of *wh*-dependencies in children's input. Hart and Risley (1995) estimate that children whose parents are professional class (and so are high-SES) hear about 487 utterances per hour; children whose parents are working class (and so are low-SES) hear about 301 utterances per hour. Our corpus estimates of *wh*-dependency rate suggest that high-SES children's input consists of about 20.5% *wh*-dependencies (20,923 *wh*-dependencies of 101,838 utterances), while low-SES children's input consists of about 12.2% *wh*-dependencies (3,904 *wh*-dependencies of 31,857 utterances). Table 6 shows the resulting quantities of *wh*-dependencies heard during the learning period across SES: 1,418,193 *wh*-dependencies for high-SES children and 522,539 *wh*-dependencies for low-SES children.

Summary and implications. Our estimated quantities of *wh*-dependencies heard during the estimated learning period for syntactic islands are quite different across SES: high-SES children are estimated to encounter nearly three times as many *wh*-dependencies as low-SES children. Yet, developmental trajectory data (Otsu, 1981; de Villiers and Roeper, 1995; de Villiers et al., 2008) suggest that for Complex NP island knowledge, the lower quantity that low-SES children encounter is sufficient; that is, despite the significant difference in the quantity of relevant input, low-SES children aren't delayed relative to high-SES children in their acquisition of Complex NP island knowledge. So, we know that the input quantity difference that we found, even when coupled with

Table 6: Calculating the total *wh*-dependencies (total *wh*-dep) that high-SES and low-SES children encounter between the ages of 20 and 59 months, the estimated learning period for syntactic islands. These calculations are based on about 14,174 waking hours between these ages, and estimated rates of utterances per hour (utt/hr) and *wh*-dependencies in the input (*wh*-dep/utt) across SES.

SES	**hours**	*	**utt/hour**	*	***wh*-dep/utt**	=	**total *wh*-dep**
high-SES	14174	*	487	*	20932/101838	=	1,418,193
low-SES	14174	*	301	*	3904/31875	=	522,539

the small differences in *wh*-dependency and syntactic trigram distributions, isn't developmentally meaningful for learning about Complex NP island knowledge.

But, what about the other three island types (Subject, Whether, and Adjunct)? Is this difference in input quantity predicted to delay their development, when coupled with the small differences we found in *wh*-dependency and syntactic trigram distributions? We can answer this question by using the developmental computational model from Pearl and Sprouse (2013) that implements a specific learning theory linking children's input to their predicted knowledge development. In particular, when given a realistic quantity of children's *wh*-dependency input (the amount encountered by age four) comprised of *wh*-dependencies that are distributed the same as in our corpus samples, we can see whether there are predicted differences in the knowledge acquired by age four across SES. If the learning theory implemented by the model is plausible, it should find that the input data for low-SES children support the acquisition of Complex NP island knowledge as well as the input data for high-SES children. We can then see if the model predicts that the input data for low-SES children support the acquisition of the other three island types we investigate as well as the input data do for high-SES children – or not.

6.5 Developmental computational modeling analysis

To evaluate input quality for learning about these syntactic islands, we implemented a modeled learner using the learning strategy described by Pearl and Sprouse (2013), which links input distributions of *wh*-dependencies and the resulting syntactic trigrams to development of syntactic islands knowledge. We then compared what this modeled learner would learn about these syntactic islands by age four, when given the realistic estimates of high-SES and low-SES input described in the previous section.

More specifically, for each input set, the modeled learner estimated syntactic trigram probabilities and could then generate probabilities for any desired *wh*-dependency, whether seen or unseen in its input. The *wh*-dependencies that the model needed to generate probabilities for were those that American English adults had given acceptability judgments for in Sprouse et al. (2012), corresponding to the four islands from (2). Recall that each island stimuli set therefore had four dependency types: matrix+non-island, embedded+non-island, matrix+island, and embedded+island; the embedded+island stimulus in each case involved a *wh*-dependency that crossed a syntactic island, and so was ungrammatical. These experimental stimuli can be characterized by the syntactic paths shown in Table 7.[8]

[8]Note that many of the grammatical dependencies for each island type (e.g., matrix+non-island and ma-

Table 7: Syntactic paths for experimental stimuli that acceptability judgments were generated for, in a 2x2 factorial design varying dependency length (*mat*rix vs. *emb*edded) and presence of an island structure (*non*-island vs. *island*). Ungrammatical island-spanning dependencies are indicated with *.

		Complex NP islands	***Subject islands***
mat	non	*start*-IP-*end*	*start*-IP-*end*
emb	non	*start*-IP-VP-CP$_{that}$-IP-VP-*end*	*start*-IP-VP-CP$_{null}$-IP-*end*
mat	island	*start*-IP-*end*	*start*-IP-*end*
emb	island	**start*-IP-VP-NP-CP$_{that}$-IP-VP-*end*	**start*-IP-VP-CP$_{null}$-IP-NP-PP-*end*
		Whether islands	***Adjunct islands***
mat	non	*start*-IP-*end*	*start*-IP-*end*
emb	non	*start*-IP-VP-CP$_{that}$-IP-VP-*end*	*start*-IP-VP-CP$_{that}$-IP-VP-*end*
mat	island	*start*-IP-*end*	*start*-IP-*end*
emb	island	**start*-IP-VP-CP$_{whether}$-IP-VP-*end*	**start*-IP-VP-CP$_{if}$-IP-VP-*end*

Predicted knowledge by age four. The modeled learner from Pearl and Sprouse (2013) learns from the realistic input samples estimated in the previous section for high-SES and low-SES children. This modeled learner can then generate probabilities for the four sets of experimental stimuli of Sprouse et al. (2012), corresponding to Complex NP, Subject, Whether, and Adjunct islands. To aid comparison of predicted learning outcomes across SES, the resulting log probabilities for each *wh*-dependency type from the stimuli are shown in Table 8. Log probabilities are reported for each dependency because the probabilities are very small numbers (due to the multiplication of syntactic trigram probabilities).[9] Figure 4 shows these log probabilities plotted on interaction plots for each of the four island types.

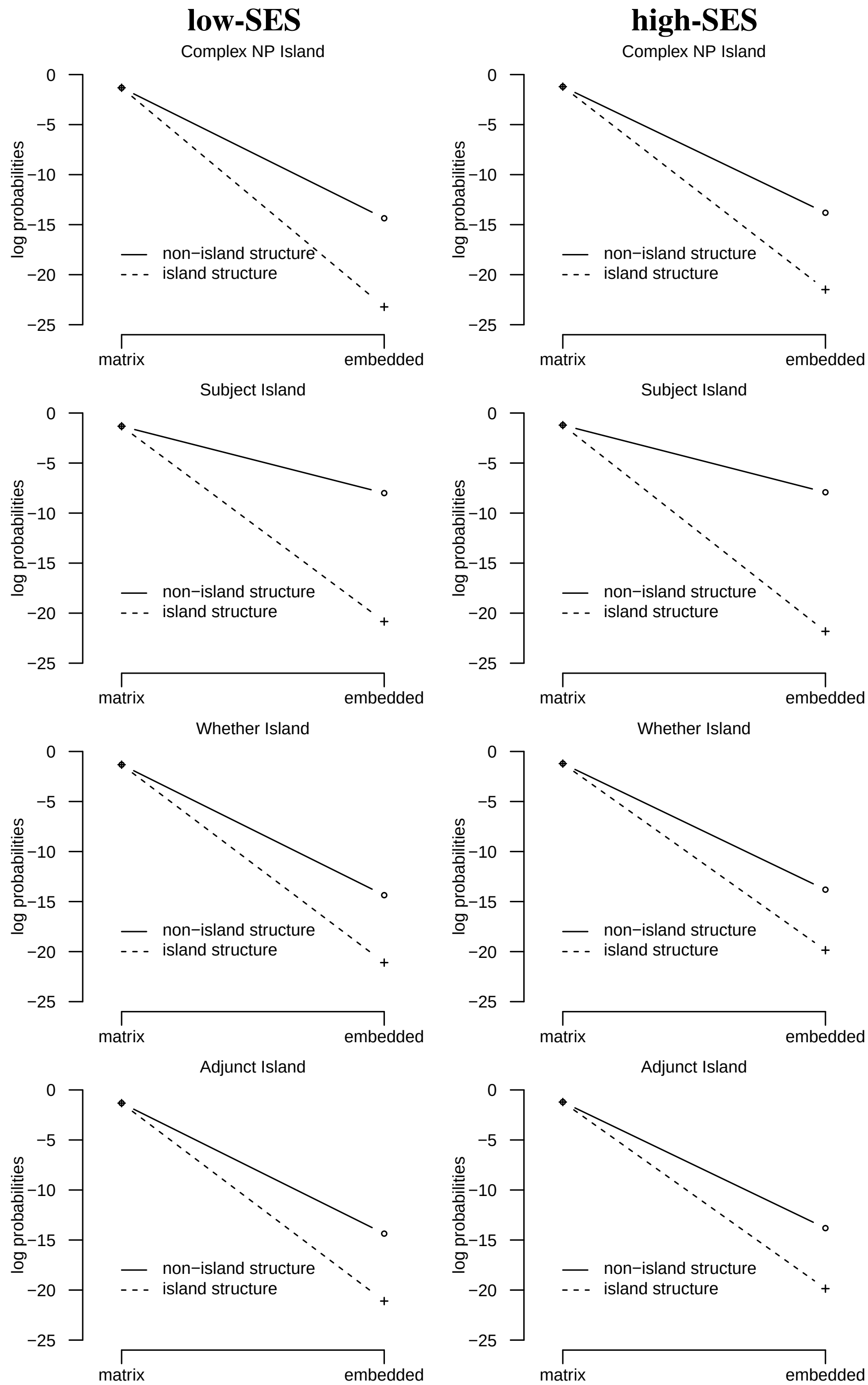
Table 8: Log probabilities of different *wh*-dependencies, representing predicted judgments, for modeled learners learning from estimates of low-SES child-directed speech (L-CDS) and high-SES child-directed speech that children hear by age four.

	L-CDS	H-CDS
Grammatical dependencies		
start-IP-*end*	-1.32	-1.21
start-IP-VP-CP$_{null}$-IP-*end*	-8.00	-7.92
start-IP-VP-CP$_{that}$-IP-VP-*end*	-14.36	-13.81
Island-spanning dependencies		
start-IP-VP-NP-CP$_{that}$-IP-VP-*end*	-23.22	-21.49
start-IP-VP-CP$_{null}$-IP-NP-PP-*end*	-20.84	-21.82
start-IP-VP-CP$_{whether}$-IP-VP-*end*	-19.86	-21.10
start-IP-VP-CP$_{if}$-IP-VP-*end*	-19.86	-21.10

trix+island) are characterized by the same syntactic path (e.g., *start*-IP-*end*). This means that the generated judgments from the modeled learner will be the same for those stimuli.

[9]For log probabilities, less negative numbers are equivalent to higher probabilities. For example, log(.001) = log(10^{-3}) = -3, while log(.000001) = log(10^{-6}) = -6.

Figure 4: Four-year-old child judgments predicted from a modeled learner learning from low-SES (left column) and high-SES (right column) input data, demonstrating the same implicit knowledge of four syntactic islands that appears as a superadditive interaction.



We can see that a core pattern emerges when learning from either low-SES or high-SES CDS: all grammatical dependencies have higher probabilities (equivalent to less negative log probabilities) than the island-spanning dependencies. In particular, grammatical dependencies have log probabilities ranging from -1.21 to -14.36, while island-spanning dependencies range from -19.86 to -23.22. So, even the least acceptable grammatical dependency (with log probability -14.36) is predicted to be over 300,000 times more acceptable than the most acceptable ungrammatical dependency (with log probability -19.86), because $\frac{10^{-14.36}}{10^{-19.86}} \approx 316,228$. The predicted judgments for these stimuli in Figure 4, represented by these log probabilities, show the superadditive interactions that indicate implicit knowledge of the four islands.

The predicted judgments for Complex NP islands (Figure 4, first row) align with the empirical data we have across SES that children are sensitive to this island type by age four (de Villiers et al., 2008). So, the learning model correctly predicts observed developmental outcomes (in this case, no difference) by age four for Complex NP islands. This correct prediction then supports the plausibility of the learning theory implemented by the modeled learner, which assumes children pay attention to a certain aspect of the input (all *wh*-dependencies), and use this information a particular way that involves specific prior knowledge and abilities being in place (recall the discussion from section 5).

The learning model then predicts, on the basis of this same view of the relevant input and how children use that input to learn, that there should be no developmental outcome difference by age four across SES for the other three syntactic islands: Subject, Whether, and Adjunct. These predictions can be tested experimentally in future child behavioral work. If they are indeed true, and there's no difference in syntactic island knowledge for all four of these islands by age four across SES, then this would additionally support our basic finding: low-SES input is qualitatively similar to high-SES input, when it comes to the development of this syntactic island knowledge. Importantly, because of the learning theory implemented concretely by the modeled learner, we understand why this is: any observable differences in the input don't affect input quality for the part of that input that scaffolds knowledge of these syntactic islands.

7 Discussion

Our results suggest that the *wh*-dependency input, and in turn the syntactic trigram input, that low-SES children receive is similar in many respects to the input of high-SES children, despite the large differences in quantity of *wh*-dependency input by age four. In particular, any input differences across SES aren't predicted to be developmentally meaningful with respect to learning this syntactic island knowledge. So, while there may be some surface-level quantitative differences in input across SES, there don't appear to be qualitative differences. That is, surface input differences mask deeper input similarities across SES for the development of this syntactic island knowledge.

More specifically, our developmental computational modeling results serve as predictions of children's learning behavior for these four syntactic islands, and predict no learning outcome differences due to input differences across SES. Prior child behavioral work (de Villiers et al., 2008) in fact found no learning outcome differences by age four for one island type (Complex NP islands). Importantly, this correct prediction provides empirical support to the learning theory implemented by the learning model; so, the model's predictions for other complex syntactic knowledge (such as the other three syntactic islands) gain credibility. Moreover, because the learning model allows us to understand how the input leads to the development of this syntactic knowledge, we can understand why the observable input differences aren't predicted to be developmentally meaningful. Below, we discuss some interesting input differences, other model predictions for complex syntactic knowledge, and the plausibility of the prior knowledge and abilities assumed by the learning theory implemented in the learning model.

7.1 Interesting input differences

There's a striking difference in the exact *wh*-dependency distribution across SES that's predicted to be crucial for acquisition success for two of the syntactic island types. This difference involves a particular structural building block, which comes from dependencies that are characterized with CP_{that}.

As noted before in (11), the only distinction between certain dependencies judged grammatical and certain dependencies judged ungrammatical by high-SES adults is the complementizer. In particular, high-SES adults judge the dependency as ungrammatical when it has complementizer *that* (e.g., CP_{that}) but grammatical when it has the null complementizer (e.g., CP_{null}): *Who do you think (*that) $\_\__{who}$ read the book?* Another key example is the difference between grammatical dependencies with complementizer *that* (15a) and ungrammatical dependencies with complementizers like *whether* (whether islands) or *if* (adjunct islands) (15b). Again, the only difference in the syntactic path of these dependencies is the CP building block, which is CP_{that} for the dependency judged grammatical and $CP_{whether}$ or CP_{if} for the dependencies judged ungrammatical.

(15) a. What do you think that Jack read $\_\__{what}$?
syntactic path: *start-IP-VP-CP_{that}-IP-VP-end*
b. *What do you wonder whether/if Jack read $\_\__{what}$?
syntactic path: **start-IP-VP-$CP_{whether/if}$-IP-VP-end*

So, it's important that the child encounter *wh*-dependencies in her input that involve complementizer *that* (and not ones that involve complementizers *whether* or *if*). When this happens, the probabilistic learning strategy we used here can leverage the CP_{that} building block to predict that (15a) should be judged as better than (15b).

However, dependencies involving CP_{that} are actually fairly rare in naturalistic usage. Pearl and Sprouse (2013) only found 2 of 20,923 (0.0096%) in high-SES CDS (along with 7 of 8,508 (0.082%) in high-SES ADS and 2 of 4,230 (0.048%) in high-SES ADT). For high-SES children, this would correspond to approximately three to four *wh*-dependencies with CP_{that} every month.[10] In our low-SES CDS sample, there are 2 of 3,094 (0.051%) dependencies involving CP_{that}, which corresponds to approximately eight to nine *wh*-dependencies with CP_{that} every month.[11] This calculation highlights that low-SES children actually hear a crucial building block for certain syntactic islands more often in their input than high-SES children do, despite low-SES children hearing fewer *wh*-dependencies overall before age four.

Interestingly, the type of *wh*-dependency in children's input that contains the crucial CP_{that} building block also differs across SES. In the high-SES CDS sample, both dependencies involving CP_{that} are of the same type: *start-IP-VP-CP_{that}-IP-VP-end* instances like (15a). However, in our low-SES CDS sample, the CP_{that} building block comes from a different *wh*-dependency type,

[10] With an estimated high-SES quantity of 1,418,211 *wh*-dependencies between 1;8 and 4;11 (40 months), this is $\frac{1{,}418{,}211}{40} \approx$35,455/month. $\frac{2\ CP_{that}}{20{,}923}$*35,455=3.39 *wh*-dependencies with CP_{that} per month.

[11] With an estimated low-SES quantity of 522,914 *wh*-dependencies between 1;8 and 4;11 (40 months), this is $\frac{522{,}914}{40} \approx$13,073/month. $\frac{2\ CP_{that}}{3{,}094}$*13,073=8.45 *wh*-dependencies with CP_{that} per month.

which happens to be a "*that*-trace violation" judged ungrammatical by high-SES adults: *start-IP-VP-CP*$_{that}$*-IP-end* instances like (16).

(16) What do you think that $\_\__{what}$ happens?
What do $[_{IP}$ you $[_{VP}$ think $[_{CP_{that}}$ that $[_{IP}$ $\_\__{what}$ $[_{VP}$ happens]]]]]?
syntactic path: *start-IP-VP-CP*$_{that}$*-IP*

So, the presence of this *wh*-dependency type, which is ungrammatical in the high-SES dialect, is predicted to provide the crucial CP$_{that}$ building block necessary for the acquisition of Whether and Adjunct islands. That is, the key linguistic experience that would allow a child learning from low-SES CDS to acquire the same syntactic knowledge as a high-SES child does actually comes from data that's ungrammatical for a high-SES child. This underscores the power of learning strategies that generate the linguistic knowledge of larger structures from smaller building blocks, like the learning theory implemented in the modeled learner here. In particular, children with different input experiences who rely on smaller building blocks may be able to find evidence for the same building blocks in different places.

7.2 Other model predictions

The developmental model has offered predictions for the knowledge that four-year-olds across SES should have about Subject, Whether, and Adjunct islands. As noted above, the model allows us to see that the predictions about knowledge of Whether and Adjunct islands rely on the input *wh*-dependencies containing CP$_{that}$. For low-SES children, these *wh*-dependencies are ones judged ungrammatical by high-SES speakers (i.e., *that*-trace violations); yet, if these are the *wh*-dependencies providing crucial data for low-SES children, then low-SES adults should produce these *wh*-dependencies. To produce these dependencies as something other than a speech error, low-SES adults should therefore find these *wh*-dependencies grammatical, unlike high-SES adults. This is a prediction that can be tested with future adult behavioral work.

More generally, because the learning model uses syntactic trigrams to generate a probability for any *wh*-dependency, we can therefore generate predictions (in the form of relative probabilities) for a child's judgment about any *wh*-dependencies of interest across SES. This includes the other *wh*-dependencies investigated by previous child behavioral work (de Villiers et al., 2008), as well as ones yet to be investigated, such as *that*-trace violations. Any predictions can then be evaluated against existing or future child behavioral work.

7.3 Learning prerequisites

Leveraging the *wh*-dependency and syntactic trigram information that the developmental model relies on isn't trivial. More concretely, several foundational knowledge components and processing abilities must be "good enough" to scaffold acquisition of syntactic islands the way the developmental model assumes. First, the child must know about syntactic phrase structure; she must be able to use that phrase structure knowledge to extract the syntactic path of a *wh*-dependency in real time (including accurately identifying where the *wh*-word is understood). As noted in section 6.4, current research suggests children begin to represent the full structure of *wh*-dependencies at 20 months (Seidl et al., 2003; Gagliardi et al., 2016; Perkins and Lidz, 2020), which is why we took that age as the starting point for our modeled learner. Yet, it's possible that that there's variation across SES on when this ability is good enough for the learning strategy implemented in the developmental model, as there are known delays in language processing in low-SES children compared to their high-SES counterparts (Fernald et al., 2013; Weisleder and Fernald, 2013).

The child must also know to break syntactic paths into smaller trigram building blocks that can be used to generate a probability for any *wh*-dependency; she must be able to identify these syntactic trigrams in real time. As with extracting the syntactic path, it's possible that a "good enough" version of this ability could be delayed in low-SES children relative to their high-SES counterparts because it involves language processing.

In addition, the child must know to track the relative frequency of the syntactic trigrams and know to combine these syntactic trigrams to generate the probability for a new *wh*-dependency; she must be able to do both of these in real time. These components rely on statistical learning abilities, as they involve sensitivity to input frequencies and the ability to aggregate probabilistic information. Recent work on statistical learning abilities across SES (Eghbalzad et al., 2016, 2021) found no differences by age 8. It's therefore possible that younger children across SES also wouldn't differ in statistical learning abilities.

More generally, it's possible that the components reviewed above that are related to language processing are delayed in low-SES children, while the domain-general components related to statistical learning aren't. Any delays could lead to low-SES children being less able to harness the complex syntactic information available in their input, even if the necessary information is in fact there (as our developmental computational model predicts). However, prior child behavioral work suggests that any delays present are surmounted by the time children are four years old when it comes to learning about Complex NP islands (de Villiers et al., 2008), as there are no delays across SES. So, these behavioral results suggest that the necessary prerequisites for learning about syntactic islands are good enough across SES for the developmental model predictions here to be plausible.

8 Conclusion

We have aimed to provide a new way for identifying developmentally-meaningful input differences, harnessing developmental computational modeling. Developmental computational modeling can be used to assess input quality by predicting what children should be able to learn from their input. If input variation is developmentally meaningful, then the model predicts learning outcome differences; in contrast, the model predicts similar learning outcomes when the input variation isn't developmentally meaningful. To demonstrate this technique, we applied it to input variation across SES related to the development of syntactic island knowledge; our model predicted that there were no developmentally-meaningful input differences. So, input quality for syntactic islands is predicted to be the same across SES. One predicted developmental similarity is confirmed by prior child behavioral work, and so lends plausibility to the remaining predictions of developmental similarity. Perhaps more importantly, because the developmental learning model provides an explicit link between the input and linguistic knowledge development, we know (i) why observable input differences aren't predicted to be developmentally meaningful, (ii) what parts of the input are predicted to be especially important, and (iii) where those important parts appear in different input samples that reflect different linguistic input experiences.

This result broadens the body of research on linguistic input variation across SES to include the nature of the input for more complex syntactic knowledge. To our knowledge, this is the first comparison across SES concerning the input for learning about these syntactic islands. Our results suggest that if we do see developmental differences in syntactic island knowledge across SES, it's not because of the information available in the input. Instead, children's ability to harness that information may differ. In short, the syntactic islands information is predicted to be there for children to use, no matter their SES – a key developmental step may instead be for them to figure out how to use it.

References

Blum, S. (2015). Wordism": Is there a teacher in the house. *Journal of Linguistic Anthropology*, 25(1):74–75.

Brown, R. (1973). *A first language: The early stages*. Harvard University Press, Cambridge, MA.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. The MIT Press, Cambridge.

Chomsky, N. (1973). Conditions on transformations. In Anderson, S. and Kiparsky, P., editors, *A Festschrift for Morris Halle*, pages 237–286. Holt, Rinehart, and Winston, New York.

Coles-White, D., de Villiers, J. G., Roeper, T., et al. (2004). The emergence of barriers to wh-movement, negative concord, and quantification. In Brugos, A., Micciulla, L., and Smith, C., editors, *The proceedings of the 28th annual Boston University Conference on Language Development*, pages 98–107, Somerville, MA. Cascadilla Press.

Cowart, W. (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgements*. Thousand Oaks, CA: Sage.

Davis, K. F., Parker, K. P., and Montgomery, G. L. (2004). Sleep in infants and young children: Part one: normal sleep. *Journal of Pediatric Health Care*, 18(2):65–71.

De Villiers, J., Roeper, T., and Vainikka, A. (1990). The acquisition of long-distance rules. In Frazier, L. and de Villiers, J., editors, *Language processing and language acquisition*, pages 257–297. Kluwer Academic, Boston.

De Villiers, J. G. and Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development*, 17(1):1037–1060.

de Villiers, J. and Roeper, T. (1995). Relative clauses are barriers to wh-movement for young children. *Journal of Child Language*, 22(2):389–404.

de Villiers, J., Roeper, T., Bland-Stewart, L., and Pearson, B. (2008). Answering hard questions: Wh-movement across dialects and disorder. *Applied Psycholinguistics*, 29(1):67–103.

Dickinson, D. K. and Tabors, P. O. (2001). *Beginning literacy with language: Young children learning at home and school.* Paul H Brookes Publishing.

Eghbalzad, L., Deocampo, J., and Conway, C. (2016). Statistical Learning Ability Can Overcome the Negative Impact of Low Socioeconomic Status on Language Development. In *Proceedings of the 38th annual meeting of the Cognitive Science Society*, pages 2129–2134, Austin, TX.

Eghbalzad, L., Deocampo, J. A., and Conway, C. M. (2021). How statistical learning interacts with the socioeconomic environment to shape children's language development. *PloS One*, 16(1):e0244954.

Endres, D. M. and Schindelin, J. E. (2003). A new metric for probability distributions. *IEEE Transactions on Information theory*.

Fernald, A., Marchman, V. A., and Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, 16(2):234–248.

Gagliardi, A., Mease, T. M., and Lidz, J. (2016). Discontinuous development in the acquisition of filler-gap dependencies: Evidence from 15-and 20-month-olds. *Language Acquisition*, 23(3):234–260.

Hart, B. and Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. P.H. Brookes, Baltimore, MD.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, 74(5):1368–1378.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, 26(1):55–88.

Huttenlocher, J., Vasilyeva, M., Cymerman, E., and Levine, S. (2002). Language input and child syntax. *Cognitive Psychology*, 45(3):337–374.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., and Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, 61(4):343–365.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Lawrence Erlbaum Associates, Mahwah, NJ.

McDaniel, D., Chiu, B., and Maxfield, T. L. (1995). Parameters for wh-movement types: Evidence from child English. *Natural Language & Linguistic Theory*, 13(4):709–753.

Otsu, Y. (1981). *Universal Grammar and syntactic development in children: Toward a theory of syntactic development*. PhD thesis, Massachusetts Institute of Technology.

Pearl, L. (in press). Modeling syntactic acquisition. In Sprouse, J., editor, *Oxford Handbook of Experimental Syntax*. Oxford University Press.

Pearl, L. and Sprouse, J. (2013). Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition*, 20:19–64.

Pearl, L. and Sprouse, J. (2015). Computational modeling for language acquisition: A tutorial with syntactic islands. *Journal of Speech, Language, and Hearing Research*, 58:740–753.

Perkins, L. and Lidz, J. (2020). Filler-gap dependency comprehension at 15 months: The role of vocabulary. *Language Acquisition*, 27(1):98–115.

Roeper, T. and Seymour, H. N. (1994). The place of linguistic theory in the theory of language acquisition and language impairment. In Levy, Y., editor, *Other children, other languages: Issues in the theory of language acquisition*, pages 305–330. Erlbaum, Hillsdale, NJ.

Ross, J. (1967). *Constraints on variables in syntax*. PhD thesis, MIT, Cambridge, MA.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, 83(5):1762–1774.

Rowe, M. L., Leech, K. A., and Cabrera, N. (2017). Going beyond input quantity: Wh-questions matter for toddlers' language and cognitive development. *Cognitive Science*, 41:162–179.

Schwab, J. F. and Lew-Williams, C. (2016). Language learning, socioeconomic status, and child-directed speech. *Wiley Interdisciplinary Reviews: Cognitive Science*, 7:264—275.

Seidl, A., Hollich, G., and Jusczyk, P. W. (2003). Early understanding of subject and object wh-questions. *Infancy*, 4(3):423–436.

Sperry, D. E., Sperry, L. L., and Miller, P. J. (2018). Reexamining the verbal environments of children from different socioeconomic backgrounds. *Child development*.

Sprouse, J., Wagers, M., and Phillips, C. (2012). A test of the relation between working memory capacity and syntactic island effects. *Language*, 88(1):82–124.

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

Vainikka, A. and Roeper, T. (1995). Abstract operators in early acquisition. *Linguistic Review*, 12:275–312.

Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition*, 40(1):21–81.

Valian, V. and Casey, L. (2003). Young children's acquisition of wh-questions: The role of structured input. *Journal of child language*, 30(1):117–143.

Weisleder, A. and Fernald, A. (2013). Talking to children matters early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11):2143–2152.

A Appendices

A.1 *Wh*-dependency distribution across SES

Table A1 shows the distribution of *wh*-dependencies across the different corpora, including the low-SES and high-SES child-directed speech, as well as high-SES adult-directed speech and adult-directed text.

Table A1: Distribution of *wh*-dependencies in child-directed Low-SES (L-CDS) and High-SES (H-CDS), as well as High-SES adult-directed speech (H-ADS) and text (H-ADT). Percentages are shown for syntactic paths, based on the total *wh*-dependencies in each corpus, with the quantity observed in the corpus on the line below. An example of each syntactic path is given below the path. Dependencies used in the Jensen-Shannon divergence (JSDiv) analysis are in **bold**. The dependency in the Low-SES dialect that's judged to be ungrammatical in the High-SES dialect is in ***bold italics***.

Distribution of *wh*-dependencies in the input				
Syntactic path and example utterance	L-CDS	H-CDS	H-ADS	H-ADT
IP	10.3%	12.8%	17.2%	33.0%
Who saw it?	402	2680	1464	1396
IP-VP	75.5%	76.7%	73.0%	63.3%
What did she see?	2949	16039	6215	2677
IP-VP-AdjP-IP-VP	0.0%	0.0%	<0.1%	0.1%
What are you willing to see?	0	0	1	5
IP-VP-AdjP-IP-VP-PP	0.0%	0.0%	<0.1%	0.0%
What are you willing to go to?	0	0	1	0
IP-VP-AdjP-PP	0.0%	0.0%	<0.1%	<0.1%
What are they good for?	0	0	1	1
IP-VP-CP$_{for}$-IP-VP-PP	0.0%	<0.1%	0.0%	0.0%
What did she put on for you to dance to?	0	1	0	0
IP-VP-CP$_{null}$-IP	0.1%	0.1%	0.6%	0.3%
Who did he think stole it?	5	24	52	12
IP-VP-CP$_{null}$-IP-VP	0.9%	1.1%	0.4%	0.2%
What did he think she stole?	39	236	30	8
IP-VP-CP$_{null}$-IP-VP-IP-VP	<0.1%	0.1%	<0.1%	0.0%
What did he think she wanted to steal?	3	28	3	0
IP-VP-CP$_{null}$-IP-VP-IP-VP-IP-VP	0.0%	<0.1%	0.0%	0.0%
What did he think she wanted to pretend to steal?	0	2	0	0
IP-VP-CP$_{null}$-IP-VP-IP-VP-IP-VP-PP	0.0%	0.0%	<0.1%	0.0%
Who did he think she wanted to pretend to steal from?	0	0	1	0
IP-VP-CP$_{null}$-IP-VP-IP-VP-PP	0.0%	<0.1%	0.0%	0.0%
Who did he think she wanted to steal from?	0	1	0	0
IP-VP-CP$_{null}$-IP-VP-NP	0.0%	<0.1%	<0.1%	<0.1%
What did he think she said about it?	0	1	5	1
IP-VP-CP$_{null}$-IP-VP-NP-PP	<0.1%	0.0%	0.0%	0.0%
What did he think it was a movie of?	3	0	0	0
IP-VP-CP$_{null}$-IP-VP-PP	0.1%	0.1%	<0.1%	<0.1%
What did he think she wanted it for?	4	28	5	1
IP-VP-CP$_{null}$-IP-VP-PP-PP	0.0%	<0.1%	0.0%	0.0%
What did he think she wanted out of?	0	1	0	0

Distribution of *wh*-dependencies in the input				
Syntactic path and example utterance	L-CDS	H-CDS	H-ADS	H-ADT
IP-VP-CP$_{that}$-IP	<0.1%	0.0%	0.0%	0.0%
What do you think that happens?	2	0	0	0
IP-VP-CP$_{that}$-IP-VP	0.0%	<0.1%	<0.1%	<0.1%
What did he think that she stole?	0	2	5	2
IP-VP-CP$_{that}$-IP-VP-IP-VP	0.0%	0.0%	<0.1%	0.0%
What did he think that she wanted to steal?	0	0	1	0
IP-VP-CP$_{that}$-IP-VP-PP	0.0%	0.0%	<0.1%	0.0%
Who did he think that she wanted to steal from?	0	0	1	0
IP-VP-IP	0.0%	<0.1%	<0.1%	0.0%
Who did he want to steal the necklace?	0	9	2	0
IP-VP-IP-VP	7.5%	5.6%	3.4%	1.3%
What did he want her to steal?	296	1167	287	57
IP-VP-IP-VP-IP-VP	<0.1%	<0.1%	<0.1%	<0.1%
What did he want her to pretend to steal?	2	11	6	1
IP-VP-IP-VP-IP-VP-PP	0.0%	0.2%	<0.1%	0.0%
Who did he want her to pretend to steal from?	0	43	6	0
IP-VP-IP-VP-IP-VP-PP-IP-VP	<0.1%	0.0%	0.0%	0.0%
What did you want to try to plan on doing?	1	0	0	0
IP-VP-IP-VP-NP	0.0%	<0.1%	0.0%	0.0%
What did he want to say about it?	0	6	0	0
IP-VP-IP-VP-NP-IP-VP	0.0%	0.0%	0.0%	<0.1%
What did he have to give her the opportunity to steal?	0	0	0	1
IP-VP-IP-VP-NP-PP	0.0%	<0.1%	<0.1%	0.0%
What did she want to steal more of?	0	1	1	0
IP-VP-IP-VP-PP	0.8%	0.4%	0.4%	<0.1%
What did she want to steal from?	35	74	33	4
IP-VP-IP-VP-PP-PP	0.0%	0.0%	0.0%	<0.1%
What did she want to get out from under?	0	0	0	1
IP-VP-NP	0.0%	0.2%	0.1%	0.1%
What did she say about the necklace?	0	52	10	5
IP-VP-NP-IP-VP	0.0%	0.0%	<0.1%	<0.1%
What did he give her the opportunity to steal?	0	0	1	2
IP-VP-NP-PP	<0.1%	<0.1%	<0.1%	0.0%
What was she a member of?	1	7	6	0
IP-VP-PP	4.0%	2.5%	4.3%	1.3%
Who did she steal from?	159	524	369	57
IP-VP-PP-CP$_{null}$-IP	0.0%	0.0%	<0.1%	0.0%
What did she feel like was a very good place?	0	0	1	0
IP-VP-PP-CP$_{null}$-IP-VP	0.0%	<0.1%	0.0%	0.0%
What did she feel like he saw?	0	1	0	0
IP-VP-PP-IP-VP	<0.1%	0.0%	<0.1%	0.0%
What did she think about buying?	2	0	3	0
IP-VP-PP-NP	0.0%	0.0%	<0.1%	0.0%
Where was she at in the building?	0	0	2	0
IP-VP-PP-NP-PP	0.0%	<0.1%	0.0%	0.0%

Distribution of *wh*-dependencies in the input				
Syntactic path and example utterance	L-CDS	H-CDS	H-ADS	H-ADT
What do you put it on top of?	0	2	0	0
IP-VP-PP-NP-PP-IP-VP What is she in the habit of doing?	0.0% 0	0.0% 0	<0.1% 1	0.0% 0
IP-VP-PP-PP What does he eat out of?	0.5% 1	0.1% 22	0.0% 0	0.0% 0
IP-VP-PP-IP-VP What did he think about stealing?	0.0% 0	<0.1% 1	0.0% 0	0.0% 0

A.2 Syntactic trigram distribution across SES

Table A2 shows the distribution of the syntactic trigrams across the different corpora, including the low-SES and high-SES child-directed speech, as well as high-SES adult-directed speech and adult-directed text. The shared syntactic trigrams were used when calculating the Jensen-Shannon divergence (JSDiv) analyses.

Table A2: Distribution of the syntactic trigrams across child-directed Low-SES (L-CDS) and High-SES (H-CDS), as well as High-SES adult-directed speech (H-ADS) and text (H-ADT). The 14 shared trigrams used in the JSDiv analysis are in **bold**.

Distribution of trigrams in the input				
Trigrams	L-CDS	H-CDS	H-ADS	H-ADT
AdjP-IP-VP	0.0% 0	0.0% 0	<0.1% 2	<0.1% 5
AdjP-PP-end	0.0% 0	0.0% 0	<0.1% 1	<0.1% 1
CP$_{for}$-IP-VP	0.0% 0	<0.1% 1	0.0% 0	0.0% 0
CP$_{null}$-IP-VP	0.6% 49	0.7% 298	0.2% 44	0.1% 10
CP$_{null}$-IP-end	<0.1% 5	<0.1% 24	0.3% 53	0.2% 12
CP$_{that}$-IP-VP	0.0% 0	<0.1% 2	<0.1% 7	<0.1% 2
CP$_{that}$-IP-end	<0.1% 2	0.0% 0	0.0% 0	0.0% 0
IP-VP-AdjP	0.0% 0	0.0% 0	<0.1% 3	<0.1% 6
IP-VP-CP$_{for}$	0.0% 0	<0.1% 1	0.0% 0	0.0% 0
IP-VP-CP$_{null}$	0.6% 54	0.7% 321	0.6% 96	0.3% 22
IP-VP-CP$_{that}$	<0.1% 2	<0.1% 2	<0.1% 7	<0.1% 2
IP-VP-IP	4.0% 340	3.2% 1398	2.1% 353	0.9% 65
IP-VP-NP	<0.1%	0.1%	0.1%	0.1%

Distribution of trigrams in the input				
Trigrams	L-CDS	H-CDS	H-ADS	H-ADT
	4	67	23	9
IP-VP-PP	2.4% 202	1.6% 698	2.5% 423	0.8% 63
IP-VP-end	38.9% 3292	39.9% 17487	38.5% 6553	37.4% 2753
NP-IP-VP	0.0% 0	0.0% 0	<0.1% 1	<0.1% 3
NP-PP-IP	0.0% 0	0.0% 0	<0.1% 1	0.0% 0
NP-PP-end	<0.1% 4	<0.1% 10	<0.1% 7	0.0% 0
PP-CP$_{null}$-IP	0.0% 0	<0.1% 1	<0.1% 1	0.0% 0
PP-IP-VP	<0.1% 3	<0.1% 1	<0.1% 4	0.0% 0
PP-NP-PP	0.0% 0	<0.1% 2	<0.1% 1	0.0% 0
PP-NP-end	0.0% 0	0.0% 0	<0.1% 2	0.0% 0
PP-PP-end	<0.1% 1	<0.1% 23	0.0% 0	<0.1% 1
VP-AdjP-IP	0.0% 0	0.0% 0	<0.1% 2	<0.1% 5
VP-AdjP-PP	0.0% 0	0.0% 0	<0.1% 1	<0.1% 1
VP-CP$_{for}$-IP	0.0% 0	<0.1% 1	0.0% 0	0.0% 0
VP-CP$_{null}$-IP	0.6% 54	0.7% 321	0.6% 96	0.3% 22
VP-CP$_{that}$-IP	<0.1% 2	<0.1 2	<0.1% 7	<0.1 2
VP-IP-VP	4.0% 340	3.2% 1389	2.1% 351	0.9% 65
VP-IP-end	0.0% 0	<0.1% 9	<0.1% 2	0.0% 0
VP-NP-IP	0.0% 0	0.0% 0	<0.1% 1	<0.1% 3
VP-NP-PP	<0.1 4	<0.1% 8	<0.1 7	0.0% 0
VP-NP-end	0.0% 0	0.1% 59	<0.1% 15	<0.1% 6
VP-PP-CP$_{null}$	0.0% 0	<0.1% 1	<0.1% 1	0.0% 0

Distribution of trigrams in the input				
Trigrams	L-CDS	H-CDS	H-ADS	H-ADT
VP-PP-IP	<0.1 3	<0.1% 1	<0.1 3	0.0% 0
VP-PP-NP	0.0% 0	<0.1% 2	<0.1% 3	0.0% 0
VP-PP-PP	<0.1 1	<0.1% 23	0.0% 0	<0.1 1
VP-PP-end	2.3% 198	1.5% 671	2.4% 416	0.8% 62
start-IP-VP	41.4% 3502	41.7% 18283	41.5% 7049	38.6% 2835
start-IP-end	4.7% 402	6.1% 2680	8.6% 1464	19.0% 1396