A new way to identify if variation in children's input could be developmentally meaningful: Using computational cognitive modeling to assess input across socio-economic status for syntactic islands

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

Abstract

While there are always differences in children's input, it is unclear how often these differences impact language development – that is, are *developmentally meaningful* – and why they do (or do not) do so. We describe a new approach using computational cognitive modeling that links children's input to predicted language development outcomes, and can identify if input differences are potentially developmentally meaningful. We use this approach to investigate if there is developmentally-meaningful input variation across socio-economic status (**SES**) with respect to the complex syntactic knowledge called syntactic islands. We focus on four island types with available data about the target linguistic behavior. Despite several measurable input differences for syntactic island input across SES, our model predicts this variation not to be developmentally meaningful: it predicts no differences in the syntactic island knowledge that can be learned from that input. We discuss implications for language development variability across SES.

Key Words: input variation, child-directed speech, socioeconomic status, computational cognitive modeling, syntactic islands, *wh*-dependencies, quantitative approaches

1 Introduction

1.1 Identifying if input differences are developmentally meaningful

There is 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 causes different trajectories (e.g., measurable delays in knowledge development) or different knowledge to develop (e.g., dialectal variation). So, while input differences may appear, the input is not different when it comes to supporting language development. However, some input variation does indeed impact language development – this variation is then developmentally meaningful.

For instance, developmentally-meaningful input deficits would lead to language development delays. As a concrete example, we have evidence that language development delays appear across socio-economic status (**SES**), with lower-SES children behind their higher-SES peers for different components of language development (e.g., vocabulary development: Hart and Risley 1995; Hoff 2003, language processing: Fernald et al. 2013). Importantly, variation in children's input can often predict later language development (Hart and Risley, 1995; Huttenlocher et al., 2002, 2010; Rowe, 2012; Weisleder and Fernald, 2013; Hirsh-Pasek et al., 2015; Schwab and Lew-Williams, 2016), suggesting a causal link between observed input variation and language development variation, including the observed language development delays across SES.

Still, when we identify developmental delays that may be linked to variation in children's input, it is often unclear which of the known delays may be caused (at least in part) by which specific input differences, and why. Certainly, there are observed differences in total input quantity as well as the composition of the input across SES (though input differences also exist within SES: Blum 2015; Sperry et al. 2018). For instance, when it comes to total input quantity at the word level, some studies have found that lower-SES children may encounter significantly fewer words of caregiver speech than their higher-SES peers (Hart and Risley, 1995; Schwab and Lew-Williams, 2016). For input composition, 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 is often unclear is whether a specific measurable input difference matters for developing a specific component of language. For instance, there are components that do not appear to be delayed across SES, despite the input differences (e.g., some types of complex syntactic knowledge: de Villiers et al. 2008; Vasilyeva et al. 2008). That is, some aspects of language development remain constant despite contextual variability that surfaces as measurable input differences (Hoff, 2006). Moreover, there are many components where we simply do not know if there are developmental delays across SES, despite known input variation.

From an intervention perspective, if we believe an input-based language delay is occurring, it is important to understand what aspect of the input has the disparity so that interventions can target that aspect. That is, not only is it useful to know that a developmentally-meaningful input difference exists, but it is useful to know exactly what part of the input is in fact impacting the development of specific language knowledge and why. So, being able to causally link children's input to their developing language knowledge is valuable, because this link allows us to predict if a measurable input difference will potentially cause a difference in language development.

One way to make this causal link between children's input and their developing knowledge, often measured via some observable behavior, is to use computational cognitive modeling (e.g., Pearl and Sprouse, 2013, 2015, 2019, 2021; Scontras and Pearl, 2021; Pearl, 2021; Dickson et al., 2022). A computational cognitive model aimed at explaining some component of language development can concretely implement a specific learning theory that describes how the input is used by children to update their hypotheses about language over time; children's language knowledge is then reflected in their observable language behavior. In this way, computational cognitive modeling connects theories of language development, empirical data on children's input, and child behavioral experiments. Thus, a computational cognitive 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). In other words, a computational cognitive model can test hypotheses about what particular aspects of the input may matter and why. More specifically, we can use a computational cognitive model to predict if a measurable input difference will matter for the development of a specific component of language knowledge – that is, when a difference is predicted to be developmentally meaningful, and why it is predicted to be developmentally meaningful.

This computational cognitive modeling approach complements a standard way that relies on correlation to determine if a measurable input difference is developmentally meaningful: 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 lower-SES students to improve their language comprehension: Huttenlocher et al. 2002). If input-based intervention is indeed effective, this is support that the language input difference caused the observed language outcome difference, and was therefore developmentally meaningful. However, *why* that input disparity caused the language development outcome difference is still unknown. Moreover, carefully designing, implementing, and evaluating such interventions can often be costly in terms of both time and resources. Computational cognitive modeling can offer a way to predict beforehand if an input difference is likely to cause a language development difference, and so help inform the design of intervention-based approaches that assess if an input difference is developmentally meaningful.

Importantly, because a computational cognitive model describes exactly how the input can cause the predicted knowledge to develop, the model can also determine if an observed input difference is predicted *not* to be developmentally meaningful. That is, the model can identify contextual variation surfacing in children's input that is predicted not to impact language development (Hoff, 2006). In this case, we would expect an input-based intervention targeting that aspect of the input to be ineffective at improving children's development of the language knowledge that depends on that input aspect.

1.2 Input differences for syntactic island knowledge

Here, we harness this computational cognitive modeling approach to 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 called *syntactic islands* that concerns *wh*-dependencies, such as *wh*-questions (e.g., the acceptable *Who did Lily think the pretty kitty was for?* vs. the far less acceptable *Who did Lily think the kitty for was pretty?*). In syntactic theory (Chomsky, 1965; Ross, 1967; Chomsky, 1973), syntactic islands are structures that interfere with *wh*-dependencies, so that *wh*-dependencies crossing them are far less acceptable (sometimes called "ungrammatical"). Knowledge of syntactic islands thus allows speakers to judge which *wh*-dependencies in their language are more vs. (far) less acceptable; that is, even if speakers have never heard a particular *wh*-dependency before, they can use their knowledge of syntactic islands to judge how acceptable it is. This ability to judge dependency acceptability means that speakers with knowledge of syntactic islands have internalized something quite sophisticated about the syntax of *wh*-dependencies: not simply how to understand the *wh*-dependencies that occur in their language, but also (i) how acceptable different *wh*-dependencies are, and (ii) which ones are far less acceptable (and therefore unlikely to occur) because those *wh*-dependencies cross syntactic islands. From a developmental perspective, we can then investigate how children come to have this knowledge about syntactic islands, and more specifically, how children's input influences that language development.

We first briefly review what is currently known about the development of *wh*-dependency knowledge, particularly with respect to syntactic islands. We then discuss syntactic island knowledge in more detail, and describe the particular syntactic islands we focus on; we selected these islands due to the available empirical data on the behavior that signals successful knowledge development (specifically, judgment data from adults and children). We then review a computational cognitive model for learning syntactic islands that specifies how the input causes the relevant knowledge to develop (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 implemented in the model specifies that the relevant aspect of children's input involves *wh*-dependencies, which rely on "*wh*-words" like *what* and *who* in English (among others). We additionally summarize prior modeling results by Pearl and Sprouse (2013) where the model learned from higher-SES child input and successfully demonstrated knowledge of four syntactic islands, as evidenced by the acceptability judgment patterns it predicted. We hypothesize that children across SES would use the same learning process to learn about syntactic islands from their input, as specified by the learning theory implemented in the computational cognitive model. With this hypothesis in hand, we then use the same computational cognitive model to investigate the impact of input variation across SES for learning about syntactic islands.

We begin by looking at the distributions of *wh*-dependencies in American English child-directed speech (**CDS**) between higher-SES and lower-SES populations. We first provide a descriptive corpus analysis comparing higher-SES to lower-SES input. We then assess total input quantity differences by deriving realistic estimates of the total quantity of *wh*-dependencies that higher-SES vs. lower-SES children would hear by age four; age four is when children across SES seem to demonstrate some knowledge about one of the syntactic island types we investigate (de Villiers et al., 2008). This input quantity assessment highlights what can potentially be a significant difference in total quantity of *wh*-dependencies that children hear across SES by age four.

With realistic estimates of the input data to higher-SES and lower-SES children, we then provide a computational cognitive modeling analysis of the input composition, using the model of Pearl and Sprouse (2013). The model predicts the syntactic island knowledge that higher-SES and lower-SES children would be able to acquire on the basis of their *wh*-dependency input by age four, as evidenced by the acceptability judgment patterns they would generate for a variety of *wh*-dependencies.

Our computational cognitive modeling analysis predicts that the lower-SES input supports the development of knowledge about the four syntactic islands we investigate by age four just as well as the higher-SES input does. This is true despite the differences in both total quantity and the distributions of *wh*-dependencies. Our results thus suggest that the input variation across SES is not developmentally meaningful by age four; that is, the input for learning about these four syntactic islands does not fundamentally differ across SES. This result accords with known developmental evidence for one type of syntactic island, and predicts additional developmental similarities for the other three types we investigate here.

Interestingly, our modeling analysis predicts that a syntactic building block involving complementizer *that* (e.g., *that* in *Who do you think that Lily likes?*) is crucial for successfully developing knowledge of two syntactic island types. This building block comes from a different *wh*-dependency type in higher-SES CDS vs. lower-SES CDS, which highlights that surface input composition differences may mask deeper input composition similarities. We discuss limitations of our current findings, model predictions that are testable with future work, and implications for variability in language development across SES.

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 and a little about the *wh*-dependency input.

For *wh*-dependency knowledge, higher-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). Higher-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 higher-SES children). This knowledge includes sensitivity to preferred interpretations of certain *wh*-dependencies – that is, which interpretations are more or less preferred because those interpretations depend on which *wh*-dependencies are more or less preferred.

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 preferred interpretation has *how* modifying the main clause verb *learn* (i.e., a possible answer is "from a recipe book"); the strongly dispreferred interpretation has *how* modifying the embedded clause verb *bake* (i.e., 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 preferred 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 strongly dispreferred 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, developmental outcomes by age four across SES are similar with respect to preferred and dispreferred interpretations for certain *wh*-dependencies; these interpretations rest on children being sensitive to how preferred (or dispreferred) the different *wh*-dependencies themselves are. These developmental outcome similarities suggest that input differences across SES for these types of *wh*-dependency knowledge should not be developmentally meaningful.

Still, 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 these types of *wh*-dependency knowledge despite any input variation that might be present. More generally, much remains unknown, including (i) the input variation present across SES for learning about *wh*-dependencies, (ii) how the input scaffolds the development of this complex syntactic knowledge, (iii) why any input variation present does not lead to different developmental outcomes for certain *wh*-dependency knowledge across SES by certain ages, and (iv) whether any input variation present is developmentally meaningful for other types of *wh*-dependency preferences that have yet to be assessed in children across SES.

3 Syntactic islands

A key component of syntactic knowledge is the ability to have long-distance dependencies, where there is a relationship between two words that are not next to each other. Long-distance dependencies, such as the *wh*-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 *wh*-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 next 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, adult speakers find different *wh*-dependencies to be more or less acceptable (sometimes referred to as "allowed" or "grammatical" vs. "disallowed" or "ungrammatical"), with some *wh*-dependencies being far less acceptable than others. As mentioned previously, this marked decrease in acceptability has been attributed to specific syntactic structures, called syntactic islands, that interfere with long-distance dependencies (Chomsky, 1965; Ross, 1967; Chomsky, 1973). Four example syntactic islands are in (2), with * indicating very low acceptability and [...] highlighting the proposed island structure that interferes with a *wh*-dependency 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 knowledge that allows the appropriate preferences for long-distance *wh*-dependencies. This knowledge allows them to recognize that the questions in (2) are far less acceptable, while the questions in (1) are much moreso. We note that this recognition is a measurable behavior of children's internalized knowledge. That is, distinguishing more acceptable questions like (1) from far less acceptable questions like (2) is one way to indicate knowledge of the relevant syntactic islands (whatever form that knowledge may take).

4 Assessing knowledge of syntactic islands

Previous work assessing children's knowledge of syntactic islands has focused on which interpretations of *wh*-dependencies are preferred, rather than the relative acceptability 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 is easier to ask children if they prefer a particular interpretation that relies on a certain *wh*-dependency (something more similar to naturalistic communication) rather than asking children directly how acceptable they find that *wh*-dependency (something more meta-linguistic that requires reasoning about language forms). Suppose children disprefer a certain interpretation (e.g., "*What is Jane drawing a monkey that is drinking milk with?*" with *what* interpreted as "the straw"); this (dis)preference can be interpreted as children finding the *wh*-dependency that the interpretation relies on (e.g., "*What is Jane drawing [a monkey that is drinking milk with* $\_\_{what}$*]?*") less acceptable. So, this behavior can then be interpreted as children knowing about the syntactic island that interferes with that *wh*-dependency (e.g., a Complex NP island). That is, when children disprefer a particular interpretation, this indirectly indicates their knowledge of a particular syntactic island: the syntactic island interfering with the *wh*-dependency that the dispreferred 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., in the previous work of Sprouse et al. 2012). When the stimuli are carefully designed (as discussed below), relative differences in judged acceptability can be used to compare the acceptability of island-crossing *wh*-dependencies against the acceptability of *wh*-dependencies that do not cross islands, yet are similar in other important ways to the island-crossing ones. The key idea is that knowledge of the relevant syntactic island is signaled when the island-crossing *wh*-dependency is still judged as far less acceptable (Sprouse et al., 2012). We therefore follow Sprouse et al. (2012), and use acceptability judgment data to indicate knowledge of syntactic islands, and follow Pearl and Sprouse (2013, 2015) in using these acceptability judgment patterns as a measurable target state for development. In particular, following Pearl and Sprouse (2013), the computational cognitive model we implement will attempt to predict the appropriate acceptability judgment patterns found by Sprouse et al. (2012) that indicate knowledge of different syntactic islands.

Sprouse et al. (2012) investigated the four islands from (2). A sample stimuli 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 absence/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 supposed to be far less acceptable than the others.

(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 higher-SES adults tested by Sprouse et al.

(2012). We briefly review the logic behind this interpretation, as described in Sprouse et al. (2012).

For example, consider the Complex NP plot in the top row, where 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 is 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). However, this is not what we see. Instead, the acceptability of (3d) is much lower than this. This much-lower acceptability is a superadditive effect for the embedded+island stimuli. So, the additional unacceptability of an island-crossing-dependency like (3d) – interpreted by Sprouse and colleagues (Sprouse et al., 2012; Pearl and Sprouse, 2013, 2015) as 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.

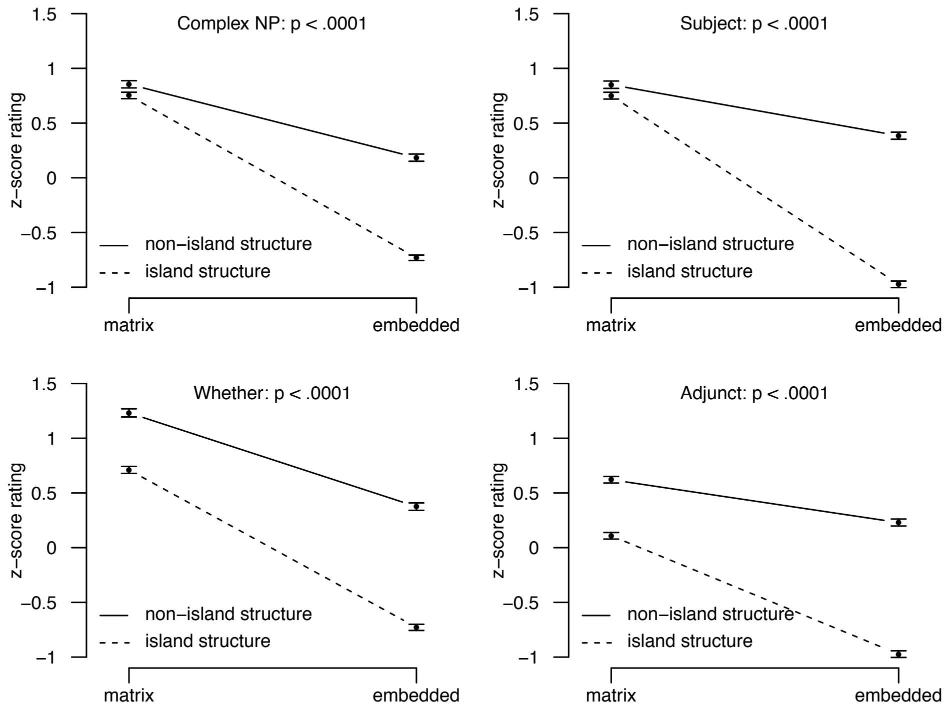
5 Linking children's input to syntactic island development

From a computational cognitive modeling standpoint, a modeled learner who can successfully acquire knowledge from its input of any of the four syntactic islands, as measured via acceptability judgments like those of Sprouse et al. (2012), should be able to reproduce the superadditive judgment pattern described above. So, the target behavior for successful development is generating the superadditive judgment pattern for a set of *wh*-dependency stimuli associated with a particular syntactic island. Pearl and Sprouse (2013) proposed a concrete learning theory – the first of its kind – to specify a precise quantitative link between children's input and this measurable output behavior, and then implemented this learning theory in a computational cognitive model.[1]

This learning theory is based on the intuition that children will learn what they can from all the *wh*-dependencies available in the input, rather than ones that are identical to the *wh*-dependencies they need to judge the acceptability of. To do this, the learning theory proposes that children

[1]We note that there are several more recent computational modeling approaches using non-symbolic frameworks such as LSTMs (see Linzen and Baroni 2021 for a review) that have also been used to learn about syntactic knowledge, including syntactic islands. However, these models do not, to our knowledge, implement a concrete learning theory – or at least not one that is easy to interpret from the model (see Pearl 2019 and Linzen and Baroni 2021 for more discussion on this point). Thus, these models contrast with the Pearl & Sprouse model used here, which implements an easy-to-interpret learning theory for syntactic islands. Another more recent computational cognitive model by Dickson et al. (2022) encodes an easy-to-interpret learning theory that learns about syntactic islands as a by-product of learning how to efficiently represent the structure of *wh*-dependencies. We discuss alternative modeling approaches further in the general discussion.

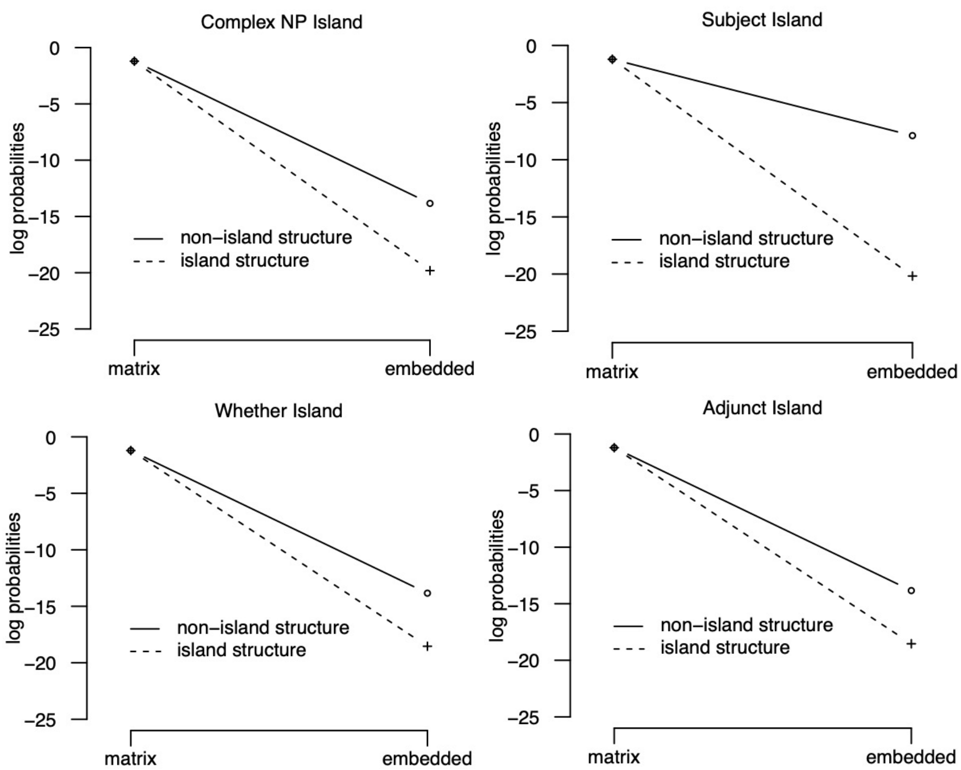
Figure 1: Higher-SES adult acceptability judgments from Sprouse et al. (2012), showing means and standard deviations of adult judgments. These judgments are interpreted as demonstrating implicit knowledge of four syntactic islands via a superadditive interaction of acceptability judgments for the selected *wh*-dependencies that cross dependency length (matrix vs. embeddeded) with the absence/presence of an island structure (non-island structure vs. island structure) in a 2 x 2 factorial design.



break *wh*-dependencies they encounter into smaller building blocks that can be used to construct any *wh*-dependency, and not necessarily just the *wh*-dependencies they have encountered before. So, these smaller building blocks are the internalized knowledge that corresponds to syntactic island knowledge. That is, by drawing on these learned building blocks, children can generate acceptability judgements, just as they would presumably draw on their syntactic island knowledge to generate acceptability judgments.

Pearl and Sprouse (2013) evaluated their computational cognitive model by allowing it to learn from a realistic sample of higher-SES CDS, and then seeing if it could generate the superadditive acceptability judgment patterns from Sprouse et al. (2012). They found that the modeled learner could indeed generate the appropriate patterns (see Figure 2). This finding supported the learning theory implemented in the model for explaining the development of syntactic island knowledge in higher-SES children. Additionally, the specific finding that *wh*-dependencies crossing Complex NP islands are far less acceptable (Figure 2, upper left) aligns with higher-SES child *wh*-dependency (dis)preferences at age four for *wh*-dependencies crossing Complex NP islands (de Villiers et al., 2008); this alignment also supports the learning theory implemented in the model. Because the model could match available data on output behavior when it learned from children's input, we use it here as a tool for evaluating variation in children's input.

Figure 2: Higher-SES child judgments generated from the computational cognitive model in Pearl and Sprouse (2013). These generated judgements can be interpreted as demonstrating implicit knowledge of four syntactic islands via a superadditive interaction of acceptability judgments for the selected *wh*-dependencies that cross dependency length (matrix vs. embeddeded) with the absence/presence of an island structure (non-island structure vs. island structure) in a 2 x 2 factorial design. Log probabilities correspond to acceptability judgments, with log probabilities closer to 0 indicating higher acceptability.



The model's learning theory assumes children can characterize a *wh*-dependency as a syntactic path from the head of the dependency (e.g., *What* in (7)) through a set of phrase structures that contain the tail (e.g., $\_\_{what}$) of the *wh*-dependency, as shown in (7a)-(7b). These structures correspond to phrase types that make up *wh*-dependencies, such as Verb Phrases (VP), Inflectional Phrases (IP), and Complementizer Phrases (CP), among others. Importantly, these are the structures that *wh*-dependencies would cross to create the link between the head of the dependency and the tail of the dependency. Under this view, children simply need to learn how acceptable the syntactic paths are for different *wh*-dependencies, which cross different phrase structures.

The learning process itself is implemented as a probabilistic learning algorithm that tracks local pieces (i.e., the building blocks) of these syntactic paths. The learning algorithm assumes the learner breaks the syntactic path into a collection of "syntactic trigrams" (groups of three units derived from the syntactic path) that can be combined to reproduce the original syntactic path, as shown in (7c).[2] The modeled learner then tracks the frequencies of these syntactic trigrams in the

[2]For discussion of the motivation for the model's implementation choices, including using information only from *wh*-dependencies, using trigrams as opposed to n-grams of other sizes, the specification of the

input, encountering one data point at a time. After the learning period is complete, the modeled learner uses these learned frequencies to calculate probabilities for all syntactic trigrams potentially comprising a *wh*-dependency[3] 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 has never seen before in its input. So, an unseen acceptable *wh*-dependency can still have a higher probability than an unseen one that is less acceptable, depending on the syntactic trigrams comprising each *wh*-dependency.

(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*

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:
$\prod_{trigrams \in T} p(trigram)$

trigrams as comprised of these particular phrase structures, when special *start* and *end* symbols are added, calculating trigram probabilities, and the method of aggregating trigrams into a *wh*-dependency, see Pearl and Sprouse (2013).

[3]The modeled learner smooths these probabilities by adding 0.5 to all trigram counts. This smoothing allows the modeled learner to generate a non-zero probability for *wh*-dependencies composed of trigrams it has never seen before. However, it gives these *wh*-dependencies a much lower probability than *wh*-dependencies composed of trigrams it has in fact seen before. See Pearl and Sprouse (2013, 2015) for further discussion of this point.

The probability generated by the modeled learner corresponds to how acceptable the *wh*-dependency is predicted to be. In this way, the modeled learner can generate judgments of *wh*-dependencies. If the learner can generate the same pattern of judgments that adults do, we can interpret this predicted judgment behavior 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 replicate the knowledge contained in syntactic islands. So, we can interpret this as the modeled learner having learned about those syntactic islands.

For the stimuli sets used by Sprouse and colleagues (Sprouse et al., 2012; Pearl and Sprouse, 2013, 2015), each *wh*-dependency stimulus can be transformed into its respective syntactic path (see Table 1). Then, the syntactic trigram probabilities learned from children's input can be used by the modeled learner to generate predicted acceptability judgments. This is the process that allowed Pearl and Sprouse (2013) to generate the judgment patterns in Figure 2, which matched higher-SES adult judgment patterns and so were interpreted as the modeled learner successfully developing knowledge of those four syntactic islands, when given higher-SES children's input.

Table 1: Syntactic paths for experimental stimuli that the modeled learner can generate acceptability judgments for, in a 2x2 factorial design varying dependency length (*matrix* vs. *embedded*) and absence/presence of an island structure (*non*-island vs. *island*). Island-spanning dependencies are indicated with a *.

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

We note that the learning theory implemented in this computational cognitive model requires children to have certain (potentially sophisticated) knowledge and abilities in place. More specifically, children are assumed to be able to reliably (i) parse utterances in their input into phrase structure trees, (ii) extract the syntactic paths for the *wh*-dependencies, (iii) track the frequency of the syntactic trigams, and (iv) calculate the probability for the complete syntactic path of a *wh*-dependency, based on its syntactic trigrams. It remains for future work to determine when children are able to accomplish these prerequisite tasks, especially if there is variation with respect to when they can. However, once children can indeed do these things, children would be able to harness the input the way this computational cognitive model does.

6 Input analysis across SES through age four

Here we assess input variation across SES, focusing on the information necessary for developing knowledge of the four syntactic islands in (2). The learning theory reviewed above assumes that the relevant input aspect is the *wh*-dependencies and the syntactic trigrams that comprise those *wh*-dependencies. So, we consider information available to children across SES in both the *wh*-dependencies and the syntactic trigrams. Because prior child behavioral work indicates that both higher-SES and lower-SES four-year-olds disprefer *wh*-dependencies crossing Complex NP islands (Otsu, 1981; de Villiers and Roeper, 1995; de Villiers et al., 2008)[4], we consider variation present in children's input across SES through age four.

We begin characterizing children's input for learning about syntactic islands by providing a descriptive analysis of the *wh*-dependencies and syntactic trigrams available in samples of higher-SES and lower-SES CDS.[5]. We then estimate the total quantity of *wh*-dependency input available across SES through age four, finding a potentially large difference in the total quantity of *wh*-dependencies.

We then use the computational cognitive model from Pearl and Sprouse (2013) to predict the syntactic island knowledge children would learn by age four from their input. More specifically, the modeled learner learns from the estimated *wh*-dependency input that higher-SES and lower-SES children encounter by age four, in terms of both the total quantity of *wh*-dependencies encountered and the distributions of those *wh*-dependencies. The modeled learner then predicts the acceptability judgments that would be generated by higher-SES and lower-SES children for the four sets of stimuli from Sprouse et al. (2012). We see if these predicted acceptability judgments suggest any input-based differences across SES by age four, which would signal that differences in the *wh*-dependency input were indeed developmentally meaningful. Conversely, similarity in the predicted acceptability judgment patterns would signal that *wh*-dependency input differences are predicted not to be developmentally meaningful.

6.1 Input samples

Higher-SES. Our higher-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.

Lower-SES. Our lower-SES input samples come from a subpart of the HSLLD corpus (Dickinson and Tabors, 2001) in CHILDES (MacWhinney, 2000), where SES was defined according to maternal education and annual income. Maternal education ranged from 6 years of schooling to

[4]We note that the *wh*-dependencies we refer to as crossing Complex NP islands are referred to in those prior studies as dependencies crossing argument barriers with a relative clause.

[5]Appendix B additionally provides an information-theoretic analysis quantifying how similar the *wh*-dependency and syntactic trigram distributions are in CDS across SES, compared to these distributions within SES but across child-directed vs. adult-directed speech.

some post-high school education. Annual income did not 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 manually annotated all *wh*-dependencies with syntactic structure, following the format of the CHILDES Treebank, as described in the accompanying documentation for the CHILDES Treebank[6] (Pearl and Sprouse, 2013).

Limitations of corpus samples. Because we draw our samples from already existing corpora freely available through CHILDES, they do differ on other factors besides SES. Such factors include age range of the children sampled, number of children sampled, gender ratios of the children sampled, size of the samples, and myriad factors related to the child language interactions themselves, including specific topics of conversation and contexts in which the interactions occurred. Though there are overlaps for some of these factors, such as age range (three- and four-year-olds) as well as some topics and contexts of interactions (meal times and toy-playing sessions), it is certainly possible that the non-SES-based differences between these samples impact the *wh*-dependency distributions.

With respect to the age range differences in these samples, analyses from Pearl and Sprouse (2013) suggest that there is little difference in *wh*-dependency distribution when comparing higher-SES CDS between one and four years old with adult-directed speech. Because the differences between CDS and adult-directed speech are generally more pronounced than CDS at different ages, this prior analysis suggests that the age range differences in the samples here may not impact the *wh*-dependency distributions so much. However, a valuable avenue for future work is to collect data across SES that more explicitly controls for many other factors in order to know more clearly which factors do and do not impact the *wh*-dependency distribution in the input.

***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), 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 subcategorization allows the modeled learner to distinguish dependencies judged by higher-SES adults to be more acceptable, like (11a), from those judged to be far less acceptable, like (11b) (Cowart, 1997). 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*

[6]This documentation is available with the downloaded corpus at https://www.socsci.uci.edu/~lpearl/CoLaLab/CHILDESTreebank/childestreebank.html and at https://childes.talkbank.org/derived/ (called the Pearl_Sprouse_Corpus at that URL).

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*-dependency types in common between the higher-SES and lower-SES child input samples (out of 26 total in the higher-SES and 16 total in the lower-SES).[7] So, the higher-SES input sample contained 14 *wh*-dependency types not in the lower-SES input sample, and the lower-SES input sample contained 4 *wh*-dependency types not in the higher-SES input sample, as shown in the lefthand column of Table 2.

We see first that there is a striking similarity in the two most frequent *wh*-dependency types across SES: the same two account for the vast majority of *wh*-dependency types in children's input across SES (higher-SES: 89.5%, lower-SES: 85.8%), and these two types seem to occur in similar proportions (shown in (12)).[8] This suggests a high-level distributional 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 $\_{what}$?*)
76.7% higher-SES, 75.5% lower-SES

b. 2^{nd} most frequent: *start-IP-end* (e.g., *What $\_{what}$ happened?*)
12.8% higher-SES, 10.3% lower-SES

When we compare the rate of *wh*-dependencies across SES (i.e., how often an utterance has a *wh*-dependency), we find another difference, with *wh*-dependencies occurring more frequently in higher-SES CDS (higher-SES: 20,932/101,383 = 20.5%, lower-SES: 3,904/31,875 = 12.2%; two-proportion z-test: z=33.3, p<.01). Over time (as detailed in section 6.3), this rate difference can lead to a considerable difference in the total quantity of *wh*-dependencies encountered.

Syntactic trigrams. For syntactic trigrams, which serve as the building blocks of *wh*-dependencies under the Pearl & Sprouse learning theory, our corpus analysis found 19 syntactic trigrams in common between the higher-SES and lower-SES child input samples (out of 29 total for the higher-SES and 20 total in the lower-SES). So, the higher-SES input sample contained 10 syntactic trigrams not in the lower-SES input sample, and the lower-SES input sample contained 1 syntactic trigram not in the higher-SES input sample, shown in the righthand column of Table 2.[9]

As might be expected from the *wh*-dependency descriptive analysis, the most frequent syntactic trigrams are also very similar across SES; this is because these trigrams come from the most

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

[8] In fact, despite the sample size differences (20,923 vs. 3,904), the most frequent *wh*-dependency proportion (76.7% higher-SES vs. 75.5% lower-SES) is indeed not significantly different across these samples (two-proportion z-test: z=1.62, p=.10). However, the second most frequent *wh*-dependency proportion (12.8% higher-SES vs. 10.3% lower-SES) does seem to be different, despite the surface similarity in proportions (two proportion z-test: z=4.34, p<.01).

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

Table 2: *Wh*-dependencies and syntactic trigrams unique to speech samples directed at higher-SES and lower-SES children, respectively. Unique syntactic trigrams are on the same row as the unique *wh*-dependencies they come from.

***wh*-dependencies**	**syntactic trigrams**
only higher-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., *Who 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 lower-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*

frequent *wh*-dependency type, *start-IP-VP-end*. More specifically, the two trigram types that collectively account for the majority of the trigrams in the *wh*-dependency input (*start-IP-VP, IP-VP-end*) are the same across SES and account for the vast majority of the input (higher-SES: 81.7%, lower-SES: 80.3%). Moreover, these two syntactic trigrams occur in similar proportions [10] (shown in (13)). So, as with the *wh*-dependency types, this descriptive analysis suggests a high-level distributional similarity in the syntactic trigram input across SES, despite the individual syntactic trigram differences.

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

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

b. 2^{nd} most frequent: *IP-VP-end*
39.9% higher-SES, 38.9% lower-SES

6.3 Realistic estimates of total input quantity across SES through age four

To estimate the total quantity of *wh*-dependency data that children from different SES backgrounds encounter through age four, we can draw on available empirical data sources to estimate both how long children have to learn (i.e., the learning period) and how much data they encounter during that learning period. More specifically, we can estimate when children would begin harnessing the *wh*-dependency information in their input (i.e., when the learning period for syntactic islands could plausibly start), how much time passes between that starting point and age four (i.e., the length of the learning period), and how many *wh*-dependencies children across SES would encounter during that learning period.

When children's learning period plausibly starts. To begin learning about the relative acceptability of different *wh*-dependencies, children must be able to process the structure of *wh*-dependencies. 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 estimate 20 months as the starting point of the learning period for syntactic islands, which depend on *wh*-dependencies.

How much time 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 estimate the number of hours awake by drawing on Davis et al. (2004), who summarize the hours asleep for young children at different ages (one through four), as shown in Table 3. Based on these estimates, we can then estimate the hours awake between 20 months and 59 months, and sum those hours to estimate the total hours awake during this learning period. Our calculations in Table 3 yield about 14,174 hours awake (≈850,450 minutes awake).

[10] As with the *wh*-dependency analysis, despite the sample size differences (43,786 vs. 8,464), the first and second most frequent syntactic trigram proportions (1^{st} most frequent: 41.8% higher-SES vs. 41.4% lower-SES; 2^{nd} most frequent: 39.9% higher-SES vs. 38.9% lower-SES) are not significantly different across these samples (two-proportion z-test for the 1^{st} most frequent: z=0.68, p=.49; for the 2^{nd} most frequent: z=1.72, p=.085).

Table 3: Calculating the total hours (cumulative waking hrs) and minutes (cumulative waking mins) 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 hrs**
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
				cumulative waking mins
				14174.17 * 60 min/hour
				850450.2

How many *wh*-dependencies during the learning period. Based on the estimated minutes awake during the learning period, we can then estimate the total quantity of *wh*-dependencies children encounter. More specifically, we estimate this quantity by drawing on estimates of the number of utterances children from different SES backgrounds hear per minute and our own corpus samples of the rate of *wh*-dependencies in children's input.

To estimate utterances per minute across SES, we draw on work by Rowe (2012) and Hoff-Ginsberg (1998). Rowe (2012) examined word tokens per minute at ages 18 months, 30 months, and 42 months across SES, finding that quantity of word tokens per minute appears to remain steady (rather than increasing). So, we assume here that the rate of utterances per minute across SES also remains the same during the learning period from 20 months to 59 months. Hoff-Ginsberg (1998) identified average rates of utterances per minute for children age 21 to 24 months from families with different SES backgrounds: (i) parents who were college-educated and worked in professional positions (which we will associate with higher-SES), and (ii) parents who were high-school educated and worked in semi-skilled, unskilled, or service positions (which we will associate with lower-SES). The higher-SES children heard 15.8 utterances per minute (standard deviation 4.2), while the lower-SES children heard 13.0 utterances per minute (standard deviation 4.2). To capture 95% of each population, we consider the range of utterance rates within two standard deviations from the average, as shown in Table 4 (higher-SES: 7.4-24.2 utterances/minute; lower-SES: 4.6-21.4 utterances/minute).

Our corpus estimates of *wh*-dependency rate suggest that higher-SES children's input consists of about 20.5% *wh*-dependencies (20,923 *wh*-dependencies of 101,838 utterances), while lower-SES children's input consists of about 12.2% *wh*-dependencies (3,904 *wh*-dependencies of 31,857 utterances). Table 4 shows the resulting range of total *wh*-dependency quantity heard during the learning period across SES: 1,293,545-4,230,241 for higher-SES children, and 479,144-2,229,063 for lower-SES children. While there are some points where there appear to be similar total quantities of *wh*-dependencies in children's input across SES (e.g., 2 standard deviations below the higher-SES average = 1,293,545 while the lower-SES average = 1,354,103), there can be a marked disparity in total quantity. On average, higher-SES children will hear about twice as many *wh*-dependencies as lower-SES children ($\frac{2{,}761{,}893}{1{,}354{,}103}$=2.04). In the most extreme case, higher-SES children at the top of the higher-SES range (2 standard deviations above the average: 4,230,241) hear nearly 9 times as many *wh*-dependencies as lower-SES children at the bottom of the lower-SES range (2 standard deviations below the average: 479,144): $\frac{4,230,241}{479,144}$=8.8.

Table 4: Calculating the range of total *wh*-dependencies (total *wh*-dep) that higher-SES and lower-SES children encounter between the ages of 20 and 59 months, the estimated learning period for syntactic islands. These calculations are based on 850,450.2 waking minutes between these ages, estimated ranges of utterance rates per min (utt/min), based on average rates (average) and standard deviations (s.d.) across SES, and *wh*-dependencies in the input (*wh*-dep/utt) across SES.

	utt/min	*****	**min**	*****	***wh*-dep/utt**	**=**	**total *wh*-dep**
higher-SES		*	850,450.2	*	20,932/101,838		
- 2 s.d.	7.4					=	1,293,545
- 1 s.d.	11.6					=	2,027,719
average	15.8					=	2,761,893
+ 1 s.d.	20.0					=	3,496,067
+ 2 s.d.	24.2					=	4,230,241
lower-SES		*	850,450.2	*	3,904/31,875		
- 2 s.d.	4.6					=	479,144
- 1 s.d.	8.8					=	916,624
average	13.0					=	1,354,103
+ 1 s.d.	17.2					=	1,791,583
+ 2 s.d.	21.4					=	2,229,063

6.4 Summary and implications of corpus analyses

Our descriptive corpus analyses highlight both high-level similarities and differences in the distributions of *wh*-dependency information in children's input across SES. Children's input is similar with respect to the most frequent *wh*-dependency types and syntactic trigrams, as well as how frequent they are; children's input is different with respect to specific *wh*-dependency types and syntactic trigrams unique to each sample, as well as the rate of *wh*-dependencies in the input. Moreover, our estimate of the total quantity of *wh*-dependencies heard during the estimated learning period for syntactic islands (through age four) highlights how the total quantity can be quite different across SES, with higher-SES children potentially hearing nearly nine times the quantity of *wh*-dependencies as lower-SES children.

However, recall that for at least one syntactic island type we investigate (Complex NP islands), children across SES seem to have developed a similar (dis)preference by age four for *wh*-dependencies crossing that island (Otsu, 1981; de Villiers and Roeper, 1995; de Villiers et al., 2008). So, we might expect that the input differences across SES that we have found so far are not developmentally meaningful by age four for learning a dispreference for *wh*-dependencies crossing Complex NP islands. This is a prediction we can evaluate using the computational cognitive model from Pearl and Sprouse (2013). Note that each island type involves different syntactic structures – therefore, even if knowledge of one syntactic island type can develop from children's input (e.g., Complex NP islands), there is no guarantee that knowledge of all these island types can develop from that same input.

Of course, as noted previously, there is suggestive evidence from prior modeling work by Pearl and Sprouse (2013) that higher-SES input can support development of all four syntactic island types. However, the input sample used in those prior analyses is not as realistic as the range we explore in our own modeling analyses here, summarized in Table 4. Thus, our analysis with a more realistic range of higher-SES input will serve as a more comprehensive comparison to our analysis with lower-SES input, and thus of input variability across SES for learning about syntactic islands.

6.5 Computational cognitive modeling analysis

We conducted the computational cognitive modeling analysis by implementing a modeled learner that uses the learning theory of Pearl and Sprouse (2013), and then allowing that modeled learner to learn from the estimated input samples described above. In particular, the modeled learner learned from the range of quantities of *wh*-dependencies estimated for higher-SES children by age four, with the *wh*-dependencies distributed as in our higher-SES corpus sample; similarly, the modeled learner learned from the range of quantities of *wh*-dependencies estimated for lower-SES children by age four, distributed as in our lower-SES corpus sample. 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.

We then demonstrate what this modeled learner would learn about the syntactic island types we investigate from its input, as measured by its predicted judgments on the *wh*-dependency stimuli from Sprouse et al. (2012), reviewed in (3)-(6) and characterized by the syntactic paths in Table 1. The target state for development is adult-like acceptability judgment patterns (which are superadditive, as in Figure 1). As mentioned above, previous computational cognitive modeling results from Pearl and Sprouse (2013) using higher-SES input were able to generate this superadditive judgment pattern for all four syntactic island types, as shown in Figure 2. Our current analysis will see if the higher-SES predicted judgment patterns replicate when using more realistic estimates of higher-SES input encountered by age four. We will additionally be able to predict the lower-SES judgment patterns resulting by age four, and see how those compare to the predicted higher-SES judgment patterns. In this way, we will be able to compare the input across SES by age four for learning about these four syntactic island types.

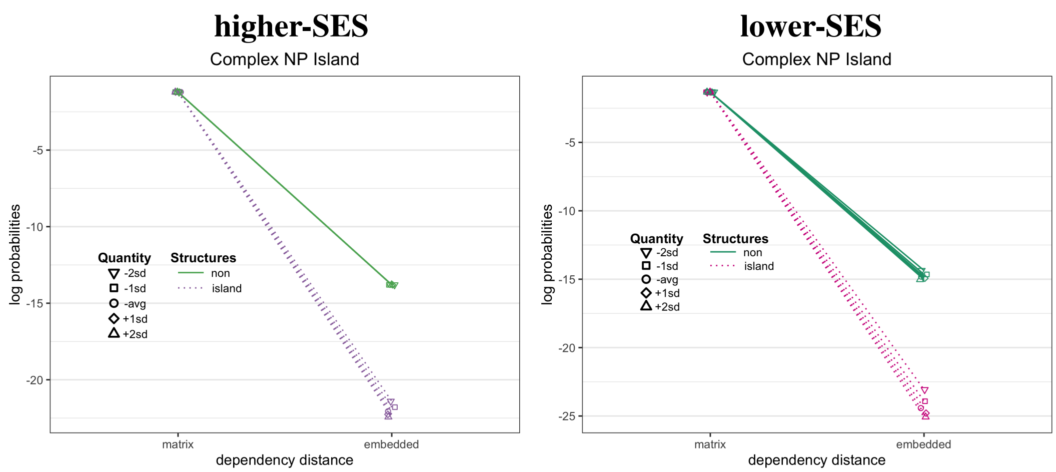
6.5.1 Analysis implementation and visualization

For each SES type (higher vs. lower), a modeled learner was run on 1000 representative input sets sampled according to the relative frequencies of the *wh*-dependencies in our corpus samples; each input set matched the estimated input quantity being modeled (2 standard deviations below average, 1 standard deviation below average, average, 1 standard deviation above average, 2 standard deviations above average). Averages of these 1000 runs for each SES type and estimated input quantity are plotted in Figures 3 and 4, with the log probability averages and standard deviations for each *wh*-dependency stimuli type available in Appendix C. Standard deviations were not plotted as they were too small to appear on the graphs.

6.5.2 Complex NP islands

The computational cognitive modeling analysis for Complex NP islands predicts acceptability judgment patterns for the *wh*-dependency stimuli from Sprouse et al. (2012), as shown in Figure 3. For higher-SES child-directed input (left side of Figure 3), we see the same superadditive judgment pattern that higher-SES adults had in Sprouse et al. (2012), and which the prior computational cognitive modeling analysis of Pearl and Sprouse (2013) found. This judgment pattern can be interpreted as demonstrating implicit knowledge of the Complex NP island. In particular, the island-crossing dependency (an embedded dependency with an island structure in it) is far less acceptable than expected if its acceptability were solely based on it being an embedded dependency with an island structure present in the utterance. Thus, these results support prior computational cognitive modeling work suggesting that higher-SES input can lead to implicit knowledge of the Complex NP island, as assessed by the superadditive judgment pattern.

Figure 3: Predicted four-year-old child judgments for Complex NP stimuli by a modeled learner learning from higher-SES (left) and lower-SES (right) input data ranges (2 standard deviations below average (-2sd), 1 standard deviation below average (-1sd), average (avg), 1 standard deviation above average (+1sd), 2 standard deviations above average (+2sd)). Averages are shown from 1000 modeled learner runs per input range. Both interaction plots show the superadditive pattern that appears in adult judgments of these *wh*-dependencies, given the factorial design crossing dependency distance (matrix vs. embedded) with the absence/presence of an island structure in the utterance (non vs. island).



We see this same judgment pattern in the predicted judgments derived from lower-SES child input (right side of Figure 3). So, these results additionally suggest that there is no predicted difference in Complex NP island knowledge by age four across SES. In particular, both higher-SES and lower-SES children should find *wh*-dependencies that cross Complex NP islands to be far less acceptable. These results align with prior child behavioral data from de Villiers et al. (2008) suggesting that children across SES disprefer *wh*-dependencies crossing Complex NP islands. That is, our computational cognitive modeling results predict that four-year-olds across SES should judge such *wh*-dependencies as much less acceptable, which seems to be true.

So, the computational cognitive model correctly predicts that (i) higher-SES children should disprefer *wh*-dependencies that cross Complex NP islands, and that (ii) lower-SES children should also disprefer these *wh*-dependencies. Moreover, a more precise prediction is that both higher-SES and lower-SES children should show the same, adult-like superadditive acceptability judgment pattern on this *wh*-dependency stimuli set by age four. Taken together, these results suggest there is no predicted developmentally-meaningful difference by age four in children's input across SES for learning about the Complex NP island, and this prediction aligns with currently available empirical evidence. With this in mind, we now turn to the predictions for the other three island types.

6.5.3 Subject, Whether, and Adjunct islands

The computational cognitive modeling analysis for Subject, Whether, and Adjunct islands predicts acceptability judgment patterns for the *wh*-dependency stimuli from Sprouse et al. (2012), as shown in Figure 4. For higher-SES child-directed input (left side of Figure 4), we see the same superadditive judgment pattern that higher-SES adults had in Sprouse et al. (2012), and which the prior computational cognitive modeling analysis of Pearl and Sprouse (2013) found. This judgment pattern can be interpreted as demonstrating implicit knowledge of Subject, Whether, and Adjunct islands. In particular, the island-spanning dependencies (embedded dependencies with an island structure in them) are far less acceptable than expected if their acceptability were solely based on them being embedded dependencies with an island structure present in the utterance. Thus, these results support prior computational cognitive modeling work suggesting that higher-SES input can lead to implicit knowledge of Subject, Whether, and Adjunct islands, as assessed by the superadditive judgment pattern.

We see this same judgment pattern in the predicted judgments derived from lower-SES child input (right side of Figure 4). So, these results additionally suggest that there is no predicted difference in Subject, Whether, or Adjunct island knowledge by age four across SES. In particular, both higher-SES and lower-SES children by age four should find *wh*-dependencies that cross Subject, Whether, and Adjunct islands to be far less acceptable.

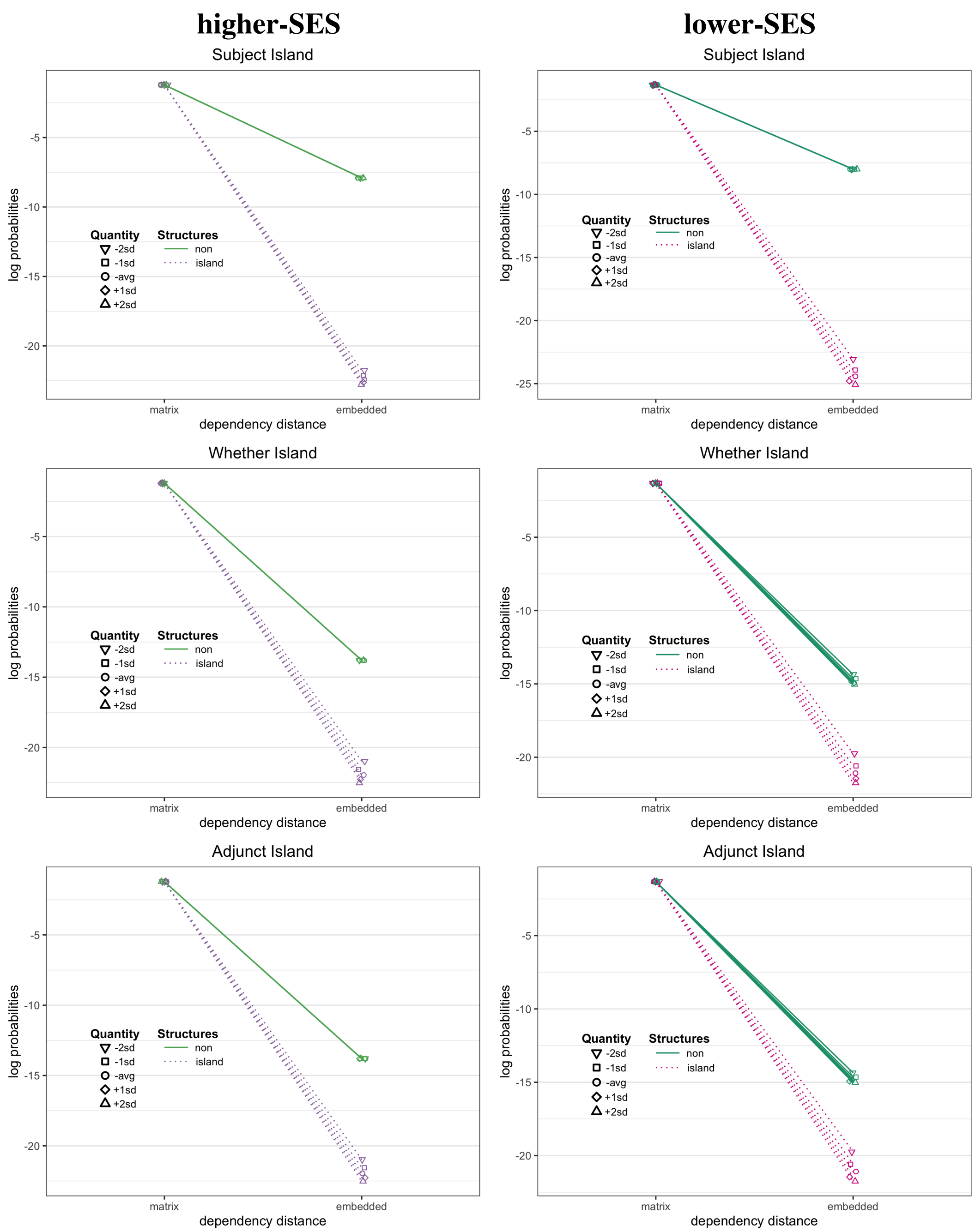
So, as with the Complex NP island, the computational cognitive model predicts that (i) higher-SES children should disprefer *wh*-dependencies that cross Subject, Adjunct, and Whether islands, and (ii) lower-SES children should also disprefer these *wh*-dependencies. As with the Complex NP island type, a more precise prediction is that both higher-SES and lower-SES children should show the same, adult-like superadditive acceptability judgment pattern on these *wh*-dependency stimuli sets by age four. Taken together, these results suggest there is also no predicted developmentally-meaningful difference in children's input by age four across SES for learning about Subject, Whether, or Adjunct islands.

6.5.4 Summary of modeling results

As mentioned above, our computational cognitive modeling analysis predicts no difference in children's knowledge across SES by age four about these four island types, as assessed by acceptability judgment patterns for specific sets of *wh*-dependencies. These predictions can be tested experimentally in future child behavioral work that gathers acceptability judgments.

If these predictions are indeed true, and there is no difference in acceptability judgments for all four of these island types by age four across SES, then those future behavioral results would additionally support our basic finding: lower-SES input is equivalent to higher-SES input when it comes to the development of this syntactic island knowledge. That is, the measurable input differences across SES are not developmentally meaningful. Importantly, because of the learning theory implemented concretely by the modeled learner, we understand why this result occurs, both in general and more specifically. In general, the observable differences in the *wh*-dependency distributions in children's input across SES do not matter for the part of that input that scaffolds

Figure 4: Predicted four-year-old child judgments for Subject, Whether, and Adjunct stimuli by a modeled learner learning from higher-SES (left column) and lower-SES (right column) input data ranges (2 standard deviations below average (-2sd), 1 standard deviation below average (-1sd), average (avg), 1 standard deviation above average (+1sd), 2 standard deviations above average (+2sd)). Averages are shown from 1000 modeled learner runs per input range. All interaction plots show the superadditive pattern that appears in adult judgments of these *wh*-dependencies, given the factorial design crossing dependency distance (matrix vs. embedded) with the absence/presence of an island structure in the utterance (non vs. island).



knowledge of these syntactic islands. More specifically, the necessary building blocks (i.e., the specific syntactic trigrams associated with each *wh*-dependency) appear in the appropriate relative frequencies in children's input across SES.

7 Discussion

Our computational cognitive modeling analysis suggests that higher-SES child input is equivalent to lower-SES child input with respect to how the *wh*-dependency input can support the development of certain syntactic island knowledge by age four. This is true despite the small differences in *wh*-dependency distribution and the potentially large differences in total quantity of *wh*-dependency input encountered by age four. Notably, small distributional differences could have mattered, as children's learning is often impacted by relative frequency differences of different items in their input (e.g., see Ramscar et al. 2013a) and Ramscar et al. 2013b). Yet, we did not find this – instead, any measurable *wh*-dependency input differences across SES are not predicted to be developmentally meaningful with respect to learning this syntactic island knowledge. That is, surface input differences mask deeper input similarities across SES.

One benefit of our computational cognitive modeling approach is that it implements a learning theory specifying a causal link between children's input and their observable language behavior. In particular, it makes predictions about children's observable behavior (here: acceptability judgments for *wh*-dependencies at age four) that can be evaluated against existing and future child behavioral data. Current data from de Villiers et al. (2008) align with the predictions for Complex NP islands, supporting the learning theory implemented in the computational cognitive model. We note again that, to our knowledge, this is the first learning theory of this kind for syntactic islands that is specified enough to generate precise, testable predictions from children's input. Thus, we believe it is valuable to continue evaluating the learning theory's predictions against empirical data, though of course future work may explore other learning theories for syntactic islands and evaluate their predictions against available empirical data.

In particular, future child behavioral work can investigate the specific predicted acceptability judgements for Complex NP islands, to further evaluate both the learning theory and the prediction that there should be no difference in this Complex NP island knowledge across SES by age four. Future child behavioral studies can also investigate the predictions for the other three island types (Subject, Whether, and Adjunct), where the computational cognitive modeling analysis also predicts no differences across SES by age four.

Below, we first discuss some interesting input differences across SES involving the complementizer *that*, which the learning theory implemented by the computational cognitive model identifies as important for the development of certain syntactic island knowledge. We then turn to other testable model predictions for related syntactic knowledge concerning *wh*-dependencies. We then consider the plausibility of the prior knowledge and abilities assumed by the learning theory implemented in the model; these prerequisites are also potential points of variation across SES that could therefore impact when children across SES could harness the information in their input in the way the learning theory proposes. We additionally discuss limitations of this computational cognitive model, and consider alternative computational modeling approaches that can be used to evaluate developmentally-meaningful input variation.

7.1 Interesting input differences involving complementizer *that*

There is a striking difference in the exact *wh*-dependency distribution across SES that is predicted by the learning theory to be crucial for learning about two of the syntactic island types, Whether and Adjunct islands. This input difference involves particular structural building blocks, which come from *wh*-dependencies that have the complementizer *that* and so are characterized by syntactic trigrams with CP_{that} in them.

As noted before in (11), the only distinction between certain *wh*-dependencies judged more acceptable and other *wh*-dependencies judged less acceptable by higher-SES adults is the complementizer. With respect to the *wh*-dependencies we have investigated here, *wh*-dependencies like (14a) with complementizer *that* are judged as more acceptable, while equivalent *wh*-dependencies like (14b) with complementizers like *whether* (Whether islands) or *if* (Adjunct islands) are judged as far less acceptable. Again, the only difference in the syntactic path of these *wh*-dependencies is CP_{that} for the *wh*-dependency in (14a) and $\text{CP}_{whether}$ or CP_{if} for the *wh*-dependencies in (14b).

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

This instance highlights that it is important for children to encounter *wh*-dependencies in their input that involve complementizer *that* (and not *whether* or *if*), if children are to learn about Whether and Adjunct islands the way the learning theory here proposes. When children do in fact encounter *wh*-dependencies with complementizer *that* (CP_{that}), the learning theory here can leverage the CP_{that} piece to predict that (14a) should be judged as more acceptable than (14b).

However, *wh*-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.[11] Based on our estimated input ranges by age four for higher-SES children, this would correspond to about three to ten *wh*-dependencies with CP_{that} every month.[12] In our lower-SES CDS sample, there are 2 of 3,094 (0.051%) *wh*-dependencies involving CP_{that}. Based on our estimated input ranges by age four for lower-SES children, this would correspond to about six to 29 *wh*-dependencies with CP_{that} every month.[13] If these corpus samples are accurate, this calculation highlights that lower-SES children could actually hear a crucial building block far more often in their input than higher-SES children do (i.e., lower-SES: 29 times vs. higher-SES: ten times per month even at the highest input estimates); this is true despite higher-SES children likely hearing more *wh*-dependencies overall before age four. That is, input quantity for this particular input aspect (i.e., *wh*-dependencies involving

[11]They additionally found that CP_{that} *wh*-dependencies are rare in both high-SES adult-directed speech (7 of 8,508 = 0.082%) and adult-directed text (2 of 4,230 = 0.048%).

[12]Two standard deviations below the average: CP_{that} rate $\frac{2}{20932}$ * 1,293,545 *wh*-dependencies in the learning period = 124; 124 / 40 months in the learning period = 3.1 CP_{that} *wh*-dependencies per month. Two standard deviations above the average: CP_{that} rate $\frac{2}{20932}$ * 4,230,241 *wh*-dependencies in the learning period = 404; 404/40 months in the learning period = 10.1 CP_{that} *wh*-dependencies per month.

[13]Two standard deviations below the average: CP_{that} rate $\frac{2}{3094}$ * 479,144 *wh*-dependencies in the learning period = 245; 245 / 40 months in the learning period = 6.1 CP_{that} *wh*-dependencies per month. Two standard deviations above the average: CP_{that} rate $\frac{2}{3094}$ * 2,229,063 *wh*-dependencies in the learning period = 1142; 1142/40 months in the learning period = 28.6 CP_{that} *wh*-dependencies per month.

CP_{that}) is estimated to be more for lower-SES children, rather than for higher-SES children, in contrast to total *wh*-dependency quantity.

Interestingly, the type of *wh*-dependency in children's input that contains the crucial CP_{that} building block also appears to differ across SES, based on our corpus samples. In the higher-SES sample, both CP_{that} dependencies are of the same type: *start-IP-VP-*$\textit{CP}_{that}$*-IP-VP-end* instances like (14a). However, in our lower-SES CDS sample, the CP_{that} building block comes from a different *wh*-dependency type, which happens to be a "*that*-trace violation" judged as much less acceptable by higher-SES adults (Cowart, 1997): *start-IP-VP-*$\textit{CP}_{that}$*-IP-end* instances like (15).

(15) 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-*$\textit{CP}_{that}$*-IP*

That is, the key linguistic experience allowing a lower-SES child to acquire the same syntactic knowledge about Whether and Adjunct islands as a higher-SES child actually comes from data that would be unlikely to occur in a higher-SES child's input. It is unlikely to occur because that data type is judged less acceptable by higher-SES adults, who produce the CDS. This finding underscores the power of learning theories that generate the linguistic knowledge of larger structures (such as *wh*-dependencies) from smaller building blocks (such as syntactic trigrams), like the learning theory 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 (e.g., syntactic trigrams involving CP_{that}) in different places (e.g., different *wh*-dependencies involving CP_{that}).

However, we note again that these findings and implications rest on the accuracy of our corpus samples. In particular, for the lower-SES CP_{that} *wh*-dependencies, it is possible that these *wh*-dependency instances were speech errors from the adult speakers. We feel this possibility is less likely, as the two *wh*-dependency instances came from two different speakers, and so are more likely to reflect naturalistic lower-SES usage. Still, future work can evaluate this prediction that these *wh*-dependencies would in fact be judged as acceptable by lower-SES adults.

However, suppose these *wh*-dependency instances in the lower-SES corpus samples were in fact speech errors and so are unlikely to occur in lower-SES children's input in general (this would be because lower-SES adults would find them as unacceptable as higher-SES adults do). In that case, we would not expect lower-SES children in general to encounter these CP_{that} *wh*-dependencies. Because these were the only *wh*-dependencies in our lower-SES sample containing CP_{that}, we might then expect that lower-SES children do *not* in fact encounter any CP_{that} *wh*-dependencies. Without the crucial CP_{that} building block in lower-SES children's input, the learning theory would predict that lower-SES children would not in fact judge *wh*-dependencies crossing Whether and Adjunct islands as any less acceptable than *wh*-dependencies crossing embedded clauses with complementizer *that*. That is, the learning theory would predict no difference in judged acceptability of the *wh*-dependencies in (14a) and (14b). So, lower-SES children would not learn the same syntactic knowledge as higher-SES children with respect to Whether and Adjunct islands, as reflected in judged acceptability of the relevant *wh*-dependencies.

In this situation, the computational cognitive modeling analysis would predict a developmentally-meaningful input difference across SES for Whether and Adjunct islands. In particular, higher-SES children's input would be predicted to support the development of this knowledge, while lower-SES children's input would not. More specifically, lower-SES children would be predicted to *not* have the adult-like superadditive judgment pattern by age four for the Whether and Adjunct *wh*-dependency stimuli, in contrast with higher-SES children.

To explore whether this input situation is in fact occurring, there are at least two specific things we can investigate in future work, using both corpus and behavioral techniques. First, we can analyze larger samples of lower-SES input to see if and how *wh*-dependencies with CP_{that} occur. The CHILDES database (MacWhinney, 2000) has additional data from the HSLLD corpus (Dickinson and Tabors, 2001) that we drew from for our lower-SES corpus sample here, as well as other lower-SES CDS samples in the Hall (Hall and Tirre, 1979) and the Brown-Sarah (Brown, 1973) corpora.

Second, we can use behavioral techniques to evaluate whether lower-SES adults judge as acceptable the specific *wh*-dependency with CP_{that} that we found in our lower-SES sample (i.e., the "*that*-trace violation"). If so, this would support the plausibility of lower-SES adults using this *wh*-dependency type in lower-SES children's input, rather than it being a speech error. Lower-SES children would then be likely to encounter this *wh*-dependency type, and importantly, the CP_{that} building block it contains. If instead lower-SES adults find that CP_{that} *wh*-dependency type less acceptable (as higher-SES adults do), this would suggest that the instances in our lower-SES corpus sample were speech errors. In that case, lower-SES children would not be likely to encounter this *wh*-dependency type in their input in general. Information about the CP_{that} building block, used to learn about Whether and Adjunct islands, would need to come from some other type(s) of *wh*-dependency involving CP_{that}, if lower-SES children are to learn about these islands the way higher-SES children are proposed to do.

7.2 Other predictions

While our investigation here focused on four island types and the specific *wh*-dependency stimuli related to them, where empirical data were already available about their judged acceptability, the learning theory is capable of generating predictions for any *wh*-dependency. Recall that this is because the learning theory proposed that all *wh*-dependencies are comprised of the same building blocks (i.e., the syntactic trigrams). So, the learning theory proposes that children are learning about those building blocks from their input, and then can use those building blocks to judge the acceptability of any *wh*-dependency.

There are in fact additional data available about children's preferences and dispreferences for certain *wh*-dependencies across SES (e.g., from de Villiers et al. 2008). So, the learning theory itself can be evaluated by seeing how well it can capture those known preferences. For instance, de Villiers et al. (2008) found that four-year-olds across SES prefer a *wh*-dependency like *What did he fix the table with* $\_\__{what}$*?* (with syntactic path *start*-IP-VP-PP-*end*) over a *wh*-dependency crossing a Complex NP syntactic island. This preference is easily captured by comparing the probabilities generated by the model learning from either higher-SES or lower-SES input data: the probability for the preferred *wh*-dependency is much higher[14], yielding a prediction that children across SES prefer that *wh*-dependency, just as children across SES actually do.

Of course, there are many *wh*-dependencies for which we do not know children's preferences

[14]Higher-SES: the preferred dependency has a predicted log probability about 10^{18} times more probable than the dispreferred one. Lower-SES: the preferred dependency has a predicted log probability about 10^{21} times more probable than the dispreferred one.

(e.g., the *that*-trace violations discussed above). In these cases, the model's predictions can be used to design future child behavioral studies that can evaluate those predictions. In addition, because the model generates more precise predictions about judged acceptability patterns (for which we do not currently have child behavioral data) rather than simple preference, future child behavioral studies can be designed to test predicted acceptability judgment patterns in children across SES.

7.3 Learning prerequisites and possible variation

It is not trivial to leverage the information from *wh*-dependencies that the learning theory relies on. More concretely, several foundational knowledge components and processing abilities must be "good enough" to learn the specific syntactic island knowledge investigated here the way the learning theory assumes. First, children must know about syntactic phrase structure; they 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 6.3, 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 learners. Yet, it is possible that that there is variation across SES on when this ability is good enough, as there are known delays in language processing in lower-SES children compared to their higher-SES counterparts (Fernald et al., 2013).

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

In addition, children 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; they 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 is therefore possible that younger children across SES also would not differ in statistical learning abilities, though of course they might.

More generally, it is possible that the components reviewed above that are related to language processing are delayed in lower-SES children, while the domain-general components related to statistical learning are not. Any delays could lead to lower-SES children being less able to harness the complex syntactic information available in their input as early as higher-SES children do. This inability to harness information would occur even if the necessary information is in fact there (as our modeling analysis predicts it to be). However, prior child behavioral work by de Villiers et al. (2008) suggests that any delays present are surmounted by the time children are four years old when it comes to learning certain preferences about Complex NP islands, as there are no delays across SES. So, those prior behavioral results suggest that the necessary prerequisites for learning about syntactic islands are good enough across SES for some amount of time before age four. This then means the computational cognitive model predictions here are likely plausible by age four.

7.4 Using computational models to evaluate input variation

The computational cognitive model we used here to evaluate input variation seemed reasonable because prior work demonstrated its ability to learn from children's input and match available empirical data on observable behavior. Yet, this model has limitations. For instance, this model currently only learns about *wh*-dependencies, rather than implementing a more general-purpose syntactic learning theory. That is, it is unclear if the model can be used to learn about other syntactic phenomena involving dependencies (e.g., binding relations between pronouns like *him* and their antecedents like *Atreyu* in *Jareth banished Atreyu*$_a$ *after meeting him*$_a$).[15] If we believe children do not use a learning strategy tuned to *wh*-dependencies specifically, then the computational cognitive modeling analysis here may not accurately represent what children would learn from their input.

Another limitation is that the model here operates over the abstract representations of phrase structure. While it is generally uncontroversial that children have abstract representations they rely on when learning from their input, the exact form of those representations is often not agreed upon. In contrast, models that learn from less-abstract representations that are easier to agree upon, such as words, may serve as alternative input evaluation tools. Several recent computational models learn by trying to predict the next word in a sequence, and along the way, these models internalize a variety of syntactic knowledge, including knowledge about syntactic islands (e.g., Wilcox et al. 2018; Futrell et al. 2019; Chaves 2020; Warstadt et al. 2020; Wilcox et al. 2021). To the extent we believe the computations that these models perform are equivalent to the mental computations that children perform, future work can use these models to evaluate input variation as we have done here.

More generally, future work can aim to use the modeling approach demonstrated here to evaluate input variation, relying on whatever computational cognitive model seems reasonable. However, it is indeed important that the chosen model be a plausible implementation for what children could be doing to extract information from their input and learn from that extracted information. When the particular computational cognitive model is plausible in this way, we can be more confident in using that model to evaluate whether input variation is potentially developmentally meaningful, as we have done here.

8 Conclusion

We have provided a new approach for identifying if and when variation in children's input could be developmentally meaningful. This approach harnesses computational cognitive modeling and complements existing behavioral approaches. In particular, a computational cognitive model can be used to assess if a particular measurable difference is likely to be developmentally meaningful; the model does so by predicting what children should be able to learn from their input, because the model concretely implements a theory of learning from that input. If input variation is potentially developmentally meaningful, then the model predicts different learning outcomes; in contrast, if input variation is not developmentally meaningful, the model predicts similar learning outcomes.

One practical benefit of this approach is that it is typically less costly to implement in terms of time and resources, compared to behavioral approaches that assess developmental outcomes and then look for correlations with children's input. However, this approach does require that

[15] See Pearl and Sprouse (2013) for more discussion.

reasonable samples of children's input are available, as well as a learning theory that specifies how the input causes linguistic knowledge to develop over time. Still, with the input samples and learning theory in hand, the computational cognitive modeling approach can provide a "first pass" input variation assessment, which can predict if input differences are likely to matter. These predictions can be followed up by targeted behavioral work evaluating the predictions, and thus offer a way to guide future research relying on behavioral approaches.

To demonstrate the computational cognitive modeling approach, 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 by age four, as equivalent outcomes were predicted to occur for all the island types we investigated, despite measurable input differences. One predicted developmental similarity about a specific island type aligns with prior child behavioral work, though more targeted behavioral work can investigate the precise outcome predictions for that island type as well as the predictions for the other island types. More generally, because the learning theory implemented in the model provides an explicit link between the input and language knowledge development, this approach can help us better understand (i) when and why observable input differences are not 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 language input experiences.

This result broadens the body of research on language input variation across SES to include the nature of the input for more complex syntactic knowledge, such as syntactic islands. This is the first comparison across SES that uses a computational cognitive modeling approach to investigate the impact of input variation with respect to learning about syntactic island knowledge. Our results suggest that if we do see developmental differences in syntactic island knowledge across SES, it is not because of meaningful differences in the information available in the input. Instead, children's ability to harness that information may differ. In short, the information for learning about these syntactic islands 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.

Chaves, R. P. (2020). What Don't RNN Language Models Learn About Filler-Gap Dependencies? *Proceedings of the Society for Computation in Linguistics*, 3(1):20–30.

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. 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.

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.

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

Dickson, N., Pearl, L., and Futrell, R. (2022). Learning constraints on *wh*-dependencies by learning how to efficiently represent *wh*-dependencies: A developmental modeling investigation with Fragment Grammars. *Proceedings of the Society for Computation in Linguistics*, 5(1):220–224.

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.

Futrell, R., Wilcox, E., Morita, T., Qian, P., Ballesteros, M., and Levy, R. (2019). Neural language models as psycholinguistic subjects: Representations of syntactic state. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1.

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.

Hall, W. S. and Tirre, W. C. (1979). The Communicative Environment of Young Children: Social Class, Ethnic, and Situational Differences. Technical Report No. 125.

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

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., Yust, P. K., and Suma, K. (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science*, 26:1071–1083.

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.

Hoff-Ginsberg, E. (1998). The relation of birth order and socioeconomic status to children's language experience and language development. *Applied Psycholinguistics*, 19(4):603–629.

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.
Linzen, T. and Baroni, M. (2021). Syntactic Structure from Deep Learning. *Annual Review of Linguistics*, pages 195–212.
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. (2019). Fusion is great, and interpretable fusion could be exciting for theory generation: Response to Pater. *Language*, 95(1):e109–e114.
Pearl, L. (2021). 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.
Pearl, L. and Sprouse, J. (2019). Comparing solutions to the linking problem using an integrated quantitative framework of language acquisition. *Language*.
Pearl, L. and Sprouse, J. (2021). The acquisition of linking theories: A Tolerance and Sufficiency Principle approach to deriving UTAH and rUTAH. *Language Acquisition*, pages 1–32.
Perkins, L. and Lidz, J. (2020). Filler-gap dependency comprehension at 15 months: The role of vocabulary. *Language Acquisition*, 27(1):98–115.
Ramscar, M., Dye, M., and Klein, J. (2013a). Children value informativity over logic in word learning. *Psychological Science*, 24(6):1017–1023.
Ramscar, M., Dye, M., and McCauley, S. (2013b). Error and expectation in language learning: The curious absence of mouses in adult speech. *Language*, 89(4):760–793.
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.
Scontras, G. and Pearl, L. S. (2021). When pragmatics matters more for truth-value judgments: An investigation of quantifier scope ambiguity. *Glossa: A journal of general linguistics*, 6(1).
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.

Vasilyeva, M., Waterfall, H., and Huttenlocher, J. (2008). Emergence of syntax: Commonalities and differences across children. *Developmental Science*, 11(1):84–97.

Warstadt, A., Parrish, A., Liu, H., Mohananey, A., Peng, W., Wang, S.-F., and Bowman, S. R. (2020). BLiMP: The benchmark of linguistic minimal pairs for English. *Transactions of the Association for Computational Linguistics*, 8:377–392.

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

Wilcox, E., Futrell, R., and Levy, R. (2021). Using computational models to test syntactic learnability. *https://ling.auf.net/lingbuzz/006327*.

Wilcox, E., Levy, R., Morita, T., and Futrell, R. (2018). What do RNN language models learn about filler-gap dependencies? In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*. Association for Computational Linguistics.

A Appendices

A.1 *Wh*-dependency distribution across SES

Table A1 shows the distribution of *wh*-dependencies across the different corpora, including the lower-SES and higher-SES CDS corpora, as well as higher-SES adult-directed speech and adult-directed text corpora. The *wh*-dependencies in common across all four corpora are used when calculating the Jensen-Shannon divergence analyses in Appendix B.

Table A1: Distribution of *wh*-dependencies in lower-SES CDS (L-CDS) and higher-SES CDS (H-CDS), as well as higher-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 in common across all four corpora are in **bold**. The dependency in the lower-SES CDS sample that is judged to be far less acceptable by higher-SES adults is in ***pink 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

Distribution of *wh*-dependencies in the input				
Syntactic path and example utterance	L-CDS	H-CDS	H-ADS	H-ADT
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
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%

Distribution of *wh*-dependencies in the input				
Syntactic path and example utterance	L-CDS	H-CDS	H-ADS	H-ADT
Where was she at in the building?	0	0	2	0
IP-VP-PP-NP-PP What do you put it on top of?	0.0% 0	<0.1% 2	0.0% 0	0.0% 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 lower-SES CDS (L-CDS) and higher-SES CDS (H-CDS), as well as higher-SES adult-directed speech (H-ADS) and adult-directed text (H-ADT). The syntactic trigrams in common across all four corpora are used when calculating the Jensen-Shannon divergence analyses in Appendix B.

Table A2: Distribution of the syntactic trigrams across lower-SES CDS (L-CDS) and higher-SES CDS (H-CDS), as well as higher-SES adult-directed speech (H-ADS) and text (H-ADT). The 14 trigrams in common across all four corpora 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

Distribution of trigrams in the input				
Trigrams	L-CDS	H-CDS	H-ADS	H-ADT
IP-VP-IP	4.0% 340	3.2% 1398	2.1% 353	0.9% 65
IP-VP-NP	<0.1% 4	0.1% 67	0.1% 23	0.1% 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

Distribution of trigrams in the input				
Trigrams	L-CDS	H-CDS	H-ADS	H-ADT
VP-PP-CP$_{null}$	0.0% 0	<0.1% 1	<0.1% 1	0.0% 0
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	41.8% 18283	41.5% 7049	38.6% 2835
start-IP-end	4.7% 402	6.1% 2680	8.6% 1464	19.0% 1396

B Input distribution comparisons

One way to quantify how similar (or not) the input distributions are for both *wh*-dependencies and syntactic trigrams across SES is to 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 JSDiv analysis useful as a comparative measure, where different distributions are assessed for their relative similarity to each other.

With this in mind, we use JSDiv to assess CDS 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 lower-SES and higher-SES CDS *wh*-dependency and trigram distributions are to those in higher-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 A3. This JSDiv analysis can thus suggest 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. Of course, this analysis is limited by the corpus samples available. In particular, including samples of lower-SES adult-directed speech and lower-SES adult-directed text would provide a more complete testbed for the JSDiv analysis with respect to the factors above. However, the analysis based on the currently-available samples seems a useful preliminary assessment.

***Wh*-dependencies.** Figure 5 shows the results of the JSDiv analysis for *wh*-dependencies, calculated over the distribution of the 9 *wh*-dependencies (shown in Table A4) that these four corpora had in common. We see that lower-SES CDS and higher-SES CDS are the most similar in *wh*-dependency distribution (JS: 0.00445), and are more similar than the next closest comparison,

Table A3: Corpora statistics for lower-SES CDS (L-CDS), higher-SES CDS (H-CDS), higher-SES adult-directed speech (H-ADS), and higher-SES adult-directed text (H-ADT) samples used for the JSDiv analysis.

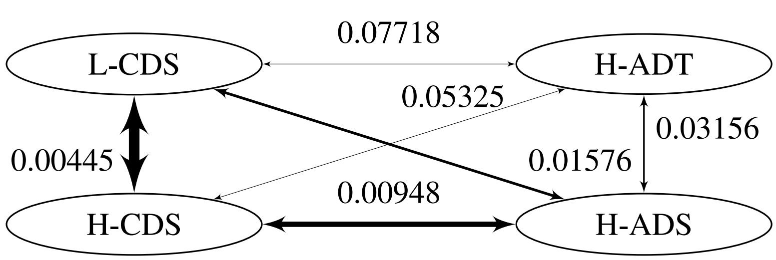
corpora	# utterances	# *wh*-dependencies	# children	ages
L-CDS	31,875	3,904	78	3 - 5
H-CDS	101,838	20,923	25	1 - 5
H-ADS	74,576	8,508	N/A	N/A
H-ADT	24,243	4,230	N/A	N/A

Table A4: The nine *wh*-dependencies shared across all four corpora that are used in the JSDiv analysis, and the percentage of the corpora the *wh*-dependency comprises.

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%

which is higher-SES CDS vs. higher-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. We can tentatively intepret this result as follows: whether the speech is directed at children or adults matters more than whether speech is coming from a higher-SES or lower-SES population. However, as mentioned above, this interpretation would be strengthened by having samples of lower-SES adult-directed speech and lower-SES adult-directed text for a fuller comparison. Still, we note that these JSDiv results accord with intuitions that speech of any kind is more similar to other speech than it is to text: higher-SES ADS diverges more from higher-SES ADT (JS: 0.03156) than it does from either higher-SES CDS (JS: 0.00948) or lower-SES CDS (JS: 0.01576).

Figure 5: JSDiv analyses for lower-SES CDS (L-CDS), higher-SES CDS (H-CDS), higher-SES adult-directed speech (H-ADS), and higher-SES adult-directed text (H-ADT). Line thickness corresponds to similarity, with thicker lines indicating more similar distributions.

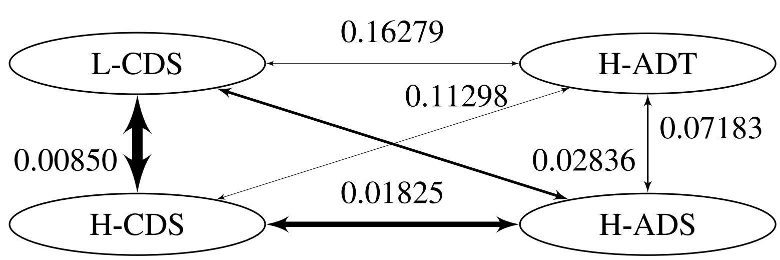


Syntactic trigrams. Figure 6 shows the results of the JSDiv analysis for syntactic trigrams, calculated over the distribution of the 14 trigrams shown in Table A5 (see Table A2 in Appendix A.2 for the full list of trigrams) that these four corpora had in common across all *wh*-dependencies. These trigrams 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) lower-SES CDS is more similar to higher-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 A5: The 14 syntactic trigrams shared across all four corpora that are used in the JSDiv analysis, and the percentage of the corpora the syntactic trigram comprises.

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



Distributional analysis summary. Our JSDiv 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.

C Predicted acceptability judgments

To aid comparison of predicted acceptability judgments across SES, Table A6 shows the resulting log probability averages and standard deviations from 1000 model runs for each *wh*-dependency type from the stimuli. Log probability averages and standard deviations (rather than plain probability averages and standard deviations) are reported for each *wh*-dependency type because the probabilities are very small numbers, due to the multiplication of syntactic trigram probabilities.[16]

Table A6: Log probability averages (with standard deviations in parentheses) from 1000 model runs, representing predicted judgments, for different syntactic paths characterizing *wh*-dependencies in the stimuli from Sprouse et al. (2012). Log probabilities are generated by modeled learners learning from estimates of higher-SES CDS (Higher-SES) and lower-SES CDS (Lower-SES) heard through age four. Results are shown for quantity ranges of estimated input: 2 standard deviations below the average (-2 sd), 1 standard deviation below the average (-1 sd), the average (avg), 1 standard deviation above the average (+1 sd), and 2 standard deviations above the average (+2 sd). *Wh*-dependencies that are judged as more acceptable by higher-SES adults are on the top, while island-spanning dependencies (indicated with *) that are judged as far less acceptable are on the bottom.

Higher-SES log probability avg (sd)					Lower-SES log probability avg (sd)				
-2 sd	-1 sd	avg	+1 sd	+2 sd	-2 sd	-1 sd	avg	+1 sd	+2 sd
start-IP-_end_									
-1.21	-1.21	-1.21	-1.21	-1.21	-1.32	-1.32	-1.32	-1.32	-1.32
(.0011)	(.00083)	(.00072)	(.00066)	(.00060)	(.0020)	(.0014)	(.0012)	(.0010)	(.00090)
start-IP-VP-CP$_{null}$-IP-_end_									
-7.91	-7.91	-7.91	-7.91	-7.91	-8.00	-8.00	-8.00	-8.00	-8.00
(.014)	(.011)	(.0095)	(.0087)	(.0079)	(.023)	(.016)	(.014)	(.012)	(.011)
start-IP-VP-CP$_{that}$-IP-VP-_end_									
-13.80	-13.80	-13.80	-13.80	-13.80	-14.36	-14.65	-14.81	-14.93	-15.03
(.12)	(.096)	(.084)	(.070)	(.064)	(.057)	(.039)	(.033)	(.029)	(.026)
Complex NP: *_start_-IP-VP-NP-CP$_{that}$-IP-VP-_end_									
-21.40	-21.79	-22.06	-22.26	-22.43	-23.07	-23.92	-24.42	-24.79	-25.07
(.040)	(.033)	(.028)	(.024)	(.022)	(.020)	(.014)	(.012)	(.010)	(.0089)
Subject: *_start_-IP-VP-CP$_{null}$-IP-NP-PP-_end_									
-21.76	-22.15	-22.41	-22.62	-22.78	-20.73	-21.30	-21.63	-21.88	-22.07
(.018)	(.015)	(.012)	(.011)	(.0096)	(.023)	(.017)	(.014)	(.013)	(.011)
Whether/Adjunct: *_start_-IP-VP-CP$_{whether/if}$-IP-VP-_end_									
-20.98	-21.56	-21.96	-22.27	-22.52	-19.75	-20.59	-21.09	-21.46	-21.75
(.00067)	(.00054)	(.00046)	(.00042)	(.00037)	(.0011)	(.00082)	(.00071)	(.00057)	(.00054)

We first observe that the standard deviations are always quite low, which reflects the consistency with which the modeled learners converge on these predicted probabilities, despite the different input sets that were learned from. We can also see that the total input quantity differences within SES seem to matter less than the input *wh*-dependency distribution across SES. For instance, higher-

[16]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.

SES results for the more acceptable *wh*-dependencies (top of Table A6) are identical across the entire range of total input quantities (from 2 standard deviations below the average to 2 standard deviations above). Notably, these final log probabilities differ from the lower-SES results for the same *wh*-dependencies, though the lower-SES results differ little from themselves across the range of total input quantities.

We note also that the results for Whether and Adjunct island-spanning *wh*-dependencies are identical for a given input quantity and SES class (e.g., the log probability for 2 standard deviations below the input average for higher-SES = -20.98). This is because the syntactic paths for these *wh*-dependency types are identical except for the complementizer used ($CP_{whether}$ vs. CP_{if}). Because both these complementizers never appear in *wh*-dependencies in children's input (either higher-SES or lower-SES), the syntactic trigrams using those complementizer building blocks have the same very low probability that is assigned to trigrams never observed in the input.

More generally, we can also observe that a core pattern emerges when learning from either higher-SES or lower-SES CDS: all dependencies judged as more acceptable by higher-SES adults have higher probabilities (equivalent to less negative log probabilities) than the island-spanning dependencies. In particular, dependencies judged as more acceptable have log probabilities ranging from -1.21 to -15.03, while island-spanning dependencies range from -19.75 to -25.07. So, even the least acceptable dependency that does not span an island (with log probability -15.03: Lower-SES, +2 sd, *start*-IP-VP-CP_{that}-IP-VP-*end*) is predicted to be much more acceptable than the most acceptable dependency spanning an island (with log probability -19.75: Lower-SES, -2 sd, *start*-IP-VP-$CP_{whether/if}$-IP-VP-*end*). (For this particular comparison, the more acceptable dependency has a probability $\frac{10^{-15.03}}{10^{-19.75}} \approx 52,481$ times higher.) We note that because human acceptability judgments likely rely on additional factors beyond *wh*-dependency probability, the exact "amount" of relative acceptability may not map directly to human acceptability judgments. However, following Pearl and Sprouse (2013), we assume that the probability of a *wh*-dependency is a significant component of its judged acceptability, and so we expect the relative patterns of acceptability to hold (i.e., which *wh*-dependencies are judged more vs. less acceptable), as indicated by these predicted probabilities.