

I Forgot but It's Okay: Learning about Island Constraints under Child-Like Memory Constraints

Niels Dickson, Richard Futrell, and Lisa Pearl

1 Introduction

A remarkable feat of children's developing language abilities is their success in acquiring complex syntactic patterns in their native language(s) at a young age. Specifically, English-speaking children as young as four (De Villiers et al., 2008) (and perhaps younger (Hirzel, 2022)) seem to interpret *wh*-dependencies in adult-like ways. To investigate how children are able to acquire this knowledge from their input, we start with a *wh*-dependency learning theory that has been formally articulated and evaluated via computational cognitive modeling (Dickson et al., 2022), and found to succeed at matching observed behavior that signals knowledge of *wh*-dependencies. Here, we evaluate this theory's ability to handle more realistic learning scenarios that incorporate children's memory limitations. We find that modeled learners who implement the learning theory of Dickson et al. (2022), while also contending with memory constraints, can still capture most of the previously observed behavior.

We first briefly review the specific knowledge about *wh*-dependencies, known as "syntactic islands", that serves as the target of acquisition, along with the behavioral data that signals knowledge of syntactic islands. We then discuss the learning theory of Dickson et al. (2022), which assumes children are trying to identify linguistic representations that allow them to efficiently parse their surrounding language data. We describe how we implement this theory in a computational cognitive model (drawing on O'Donnell (2015)'s Fragment Grammar approach), which articulates the modeled learner's hypothesis space of possible linguistic representations, the modeled learner's data intake, and the modeled learner's inference computation. We then turn to how memory limitations are implemented and incorporated into the modeled learners, specifically as a recency effect impacting the modeled learner's data intake. We review the realistic input sample the memory-impacted learners learn from before presenting our results, which highlight the learners' success at generating target behavior patterns. We then discuss what differences we observe in the memory-impacted learners (as compared to idealized learners with no memory limitations), the implications of our results for the acquisition of *wh*-dependencies, and future work that can help us further understand this acquisition process in children.

*Niels Dickson, University of California, Irvine, nielsd@uci.edu. Richard Futrell, University of California, Irvine, rfutrell@uci.edu. Lisa Pearl, University of California, Irvine, lpearl@uci.edu.

2 The target of acquisition

2.1 Target knowledge: Syntactic islands

Critical *wh*-dependency knowledge is which ones are and aren't acceptable in the language. Consider the *wh*-dependencies contained in the utterances in (1).

- (1) a. Who does Jack think the necklace is for *__who*?
b. * Who does Jack think the necklace for *__who* is expensive?

Sentence (1a) includes a dependency between the *wh*-word (“Who”) and a position later in the utterance (the object of the preposition “for”). This dependency is generally acceptable to English speakers, in contrast to the unacceptable dependency contained in (1b) between the utterance-initial “What” and the object of “for” in the subject of the embedded clause “is expensive”.

One way these unacceptable dependencies have been described is that they cross structures called “syntactic islands” (Ross, 1967): the metaphor is that certain structures (e.g., the complex subject of the embedded clause “the necklace for *__*”) act as a barrier that a dependency can't cross (i.e., an island that the dependent element is stuck on). Thus, knowledge of the syntactic islands of the language is a sophisticated type of syntactic knowledge about *wh*-dependencies that children must acquire, and English-learning children seem to demonstrate some of this knowledge by age four (De Villiers et al., 2008).

2.2 Target behavior

Past research has interpreted certain patterns of behavior in controlled experiments as a signal that the participants have knowledge about syntactic islands (e.g. De Villiers et al., 2008; Sprouse et al., 2012; Liu et al., 2022). More specifically, participants generate these observable behavior patterns, using their internalized knowledge of syntactic islands. So, these behavior patterns serve as an observable target we can compare a modeled learner's output against. In particular, if a modeled learner can generate the target behavior patterns that humans generate, we assume that the modeled learner must have an appropriate internal representation of syntactic islands. Here, we focus on three behavioral patterns that reflect adult-like knowledge of syntactic islands (shown in Figure 1).

2.2.1 Superadditive judgments: The “island difference”

The first behavior pattern derives from acceptability judgments of selected *wh*-dependency stimuli sets (e.g., those in (2)) that control for two factors: dependency length (main vs. embedded) and presence of an island structure (non-island vs. island) in the utterance (Sprouse et al., 2012).

- (2) a. Who *__who* thinks the necklace is expensive? MAIN | NON-ISLAND
b. What does Jack think *__what* is expensive? EMBEDDED | NON-ISLAND

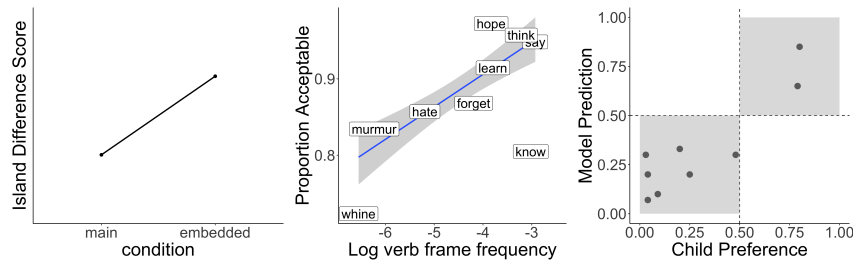


Figure 1: Three behavioral patterns representing target knowledge of syntactic islands. From left to right, the “Island Difference” pattern illustrating the unacceptability of island-crossing *wh*-dependencies (Sprouse et al., 2012), the impact of lexical information (main-verb) on *wh*-dependency acceptability (Liu et al., 2022), and child interpretation preferences for potentially ambiguous *wh*-dependencies (De Villiers et al., 2008).

- c. Who *who* thinks the necklace for Lily is expensive? MAIN | ISLAND
d. *Who does Jack think the necklace for *who* is expensive? EMBEDDED | ISLAND

The main feature of this pattern is a particularly low acceptability score for the embedded island questions like (2d) (beyond what we might expect from the two factors of length and presence of an island). The pattern can be summarized by plotting the difference in acceptability between the island vs. non-island sentences (acceptability of non-island embedded sentence minus the acceptability of the island embedded; with this calculation repeated for the two main clause dependencies). This is done in the left panel of Figure 1) where a positive slope in the line from the main to the embedded difference score indicates a particularly low embedded island acceptability score. We take this pattern to be a behavioral target for acquisition, and so our modeled learners will aim to generate this behavior when given *wh*-dependency stimuli sets like (2).

2.2.2 The impact of verb frequency

The second behavior pattern captures a sensitivity to lexical frequency in acceptability ratings of *wh*-dependencies. Liu et al. (2022) observed that the frequency of the main verb impacts the judged acceptability of the *wh*-dependency, as shown in the center panel of Figure 1.

In particular, participants judged the acceptability of utterances of the form “What did Lily VERB that Jack bought?”, with different verbs appearing in the VERB position, such as “say” and “whine.” For each of the stimuli, the authors estimated the frequency that the verb in the VERB position uses the syntactic frame in the utterance (i.e., a sentential complement like “that Jack bought”). Liu et al. (2022) found that this verb-frame frequency correlated with acceptability of the *wh*-dependencies. The center panel of Figure 1 shows this positive correlation,

plotting the log-transformed frequency of the verb frame¹ against the judged acceptability. As with the island difference pattern, we take this correlation as a behavioral target for acquisition, and so our modeled learners will aim to generate this pattern when tested on these stimuli.

2.2.3 Interpretation preferences

The third behavior is interpretation preferences for potentially ambiguous *wh*-questions (De Villiers et al., 2008). Children were presented with a story and asked a related *wh*-question that had two possible interpretations: one where the dependency is resolved in the main clause and one where the dependency is resolved in the embedded clause. For example, after hearing a story about a boy who got hurt and was talking about it later, children were asked a question like “How did the boy say *__how*₁ he hurt *__how*₂ himself?” If children’s answer was about how the saying occurred, they would seem to have resolved the dependency in the main clause (*how*₁); if instead their answer was about how the hurting occurred, they would seem to have resolved the dependency in the embedded clause (*how*₂).

The right panel of Figure 1, on the x-axis, plots the proportion of time children preferred an embedded-clause *wh*-dependency, given nine different *wh*-questions. The y-axis shows possible model predictions. We take the target pattern to be qualitative, assessing whether the modeled learner’s predictions can align with a threshold of above or below 50%, as shown by the grey boxes in the right panel of Figure 1. That is, child and model preference for an embedded dependency occurring less than 50% of the time are in the bottom lefthand grey quadrant, while child and model preferences for an embedded dependency occurring more than 50% of the time are in the top righthand grey quadrant.

3 Learning theory: Efficient linguistic representation

The modeled learner of Dickson et al. (2022) implements a learning theory adapted from O’Donnell (2015), which encodes the following intuition: children can acquire relevant knowledge about *wh*-dependencies by learning to efficiently represent *wh*-dependencies. More specifically, if children have an efficient representation of *wh*-dependencies (i.e., the right chunks that can be put together to generate any *wh*-dependency), then island-crossing *wh*-dependencies will emerge as being far less acceptable (i.e., low probability) because these *wh*-dependencies are represented with some low-probability chunks. Dickson et al. (2022)’s learning theory implementation focuses on an efficient representation of a subpart of the utterance that contains the *wh*-dependency itself (Pearl and Sprouse, 2013;

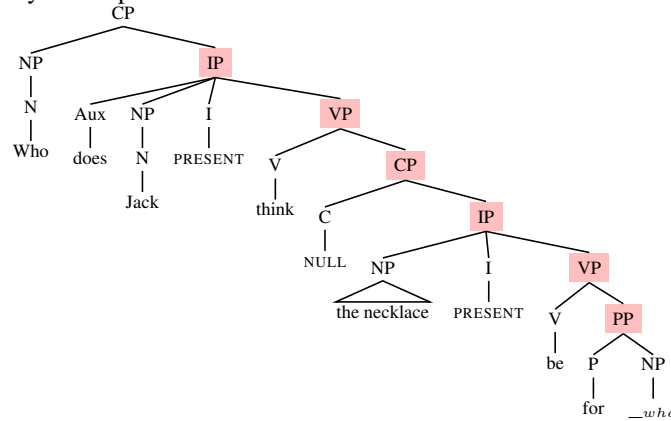
¹Note that log-transformed probabilities range from $-\infty$ to 0, with higher probabilities having a log probability closer to 0.

Pearl and Bates, 2022), termed the “syntactic path” (see (3) below). Below we discuss the modeled learner’s intake, which relies on the syntactic path, the learner’s hypothesis space of possible *wh*-dependency chunks, and the learner’s inference process.

3.1 Learner intake

As mentioned above, the modeled learner aims to identify an efficient representation of the syntactic path of a *wh*-dependency, which is a subpart of the utterance containing the *wh*-dependency. In (3), we see the syntactic path for the utterance “Who does Jack think the necklace is for *__who?*”.²

(3) Syntactic path structural nodes



More specifically, the syntactic path can be thought of as the syntactic nodes that “contain” the *wh*-dependency (Pearl and Sprouse, 2013), focusing on the path of child to parent nodes that contain the “gap” (e.g., *__who*) and eventually contain the *wh*-word (e.g., *who*, as shown in (3)). In (3), the structural nodes of the syntactic path can be represented with the highlighted sequence IP-VP-CP-IP-VP-PP. Syntactic paths may also contain lexical information (e.g., the complementizer is NULL for the CP: Pearl and Sprouse 2013; Pearl and Bates 2022). Here, we assume all lexical information connected to the head of the syntactic path nodes (i.e., NULL for the complementizer of the CP, “think” for the main VP’s verb, etc.) is included in the modeled learner’s intake: $IP_{PRESENT}-VP_{think}-CP_{NULL}-IP_{PRESENT}-VP_{be}-PP_{for}$.

3.2 Possible representations: Fragment Grammar (FG) model

To identify an efficient representation of *wh*-dependencies, our learning theory assumes children must identify efficient chunks that can be combined to gen-

²We note that this tree assumes a particular syntactic phrase structure representation, but a syntactic path can be defined for any syntactic tree structure.

erate any *wh*-dependency. Here, we define the modeled learner’s hypothesis space of potential *wh*-dependency chunks using Fragment Grammars (**FGs**) (O’Donnell, 2015) – see Dickson et al. (2022) for mathematical details of the hypothesis space specification. The basic idea is that the modeled learner considers the space of all possible *wh*-dependency chunks (“fragments”) and identifies a set of chunks that can be used to generate any *wh*-dependency (the “grammar”). An efficient FG will strike a balance between the simplicity of the representation and the ability to capture the data (see Figure 2).

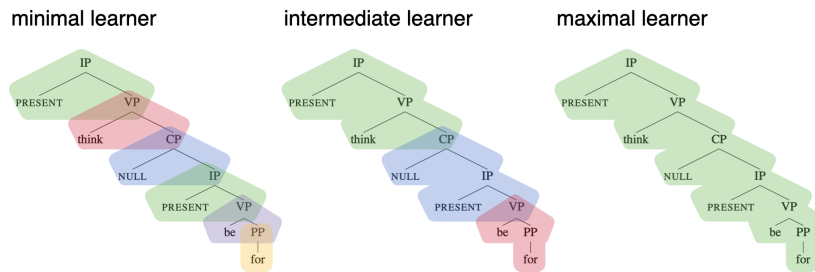


Figure 2: Some possible FG representations that capture the syntactic path of the *wh*-dependency in (3) with different-sized chunks, demonstrating the trade-off in learned representations based on the size of the stored chunk.

A learner who chooses minimal-size chunks (on the left of Figure 2) would end up with fewer chunks overall (i.e., a simpler representation). These chunks would have a higher probability because they’re often used, and can be flexibly combined to generate *wh*-dependencies. However, to generate a specific *wh*-dependency, many chunks have to be used, and so the generated *wh*-dependency includes that higher “construction” cost.

In contrast, a learner who chooses maximal-sized chunks (on the right of Figure 2) would end up with many chunks overall (one chunk for each unique *wh*-dependency in the input, which is a more complex representation). These chunks would have a lower probability because they’re far less often used, and would not be able to easily capture new *wh*-dependencies that haven’t been encountered in the input. However, to generate a specific *wh*-dependency, only one chunk has to be used, and so the generated *wh*-dependency has a low construction cost.

A learner who strikes a balance between these two extremes and chooses chunks of different sizes is an “intermediate learner” (center of Figure 2). This learner forms generalizations about the data to chunk frequently co-occurring structures (maintaining a low construction cost like the “maximal” learner) but keeps the learned pieces small enough to maintain a low-complexity, flexible representation (leveraging the advantage of the “minimal” learner). Thus, intermediate chunks can be a more efficient representation than either minimal or maximal chunks – and the learner’s inference process involves how to find just the right

intermediate chunks.

3.3 Learner inference

To search the hypothesis space of possible *wh*-dependency chunks, the modeled learners follow Dickson et al. (2022), adapted from O’Donnell (2015), and use Bayesian inference, a plausible inference computation for young children (see Pearl (2021) for more detailed discussion). The modeled learners here use a computational-level implementation of Bayesian inference (Marr, 2010; Pearl, 2023b), and search the hypothesis space using Metropolis-Hastings sampling. In particular, the modeled learner samples a potential FG (i.e., a set of chunks) from the set of possible chunks, parses the entire available input with this sampled FG, and tracks how probable the data are with this FG representation. The learner then makes an informed adjustment to the learned representation and adopts this adjustment if the probability of the data under the adjusted representation increases (O’Donnell, 2015). Over many iterations, this process leads to the modeled learner adopting a high-probability FG.

To be clear, we don’t assume that children are capable of accomplishing this mental computation of Bayesian inference in this way – it seems unlikely they can hold a detailed representation of all their input data over many years in mind, for one thing. However, we *are* committed to children performing Bayesian inference, likely approximating this mental computation as best they can with the cognitive resources they have available.

4 A more realistic modeled learner with memory limitations

4.1 The role of memory in *wh*-dependency processing

An assumption of the previous modeled learners of Dickson et al. (2022) is that they can perfectly extract and represent the data they learn from: the syntactic paths of the *wh*-dependencies in their input. In other words, when the modeled learner processes the input, none of the information is lost, and all of the information contributes to the learned representation. However, children have limited cognitive resources, including memory limitations, that likely would impact their ability to accurately extract and represent *wh*-dependency information. In particular, we know that memory plays an important part in processing dependencies (McElree et al., 2003), and children’s short term memory, along with related abilities like encoding information with context and maintaining attention, develops over time (Paris, 1978; Gathercole et al., 2004; Fandakova et al., 2014). Given this, we investigate the impact of memory limitations on modeled learners searching for an efficient FG representation of *wh*-dependencies.

More specifically, we implement memory-impacted modeled learners who exhibit a recency effect (Anderson and Milson, 1989), where more recent information is more likely to be remembered. In humans, this effect follows a power-

law distribution: recent items have a high probability of being remembered, but this probability quickly decreases for the items farther away from the most recent item (see Figure 3). Here, the items to be remembered are the units in the syntactic path. So, a recency effect causes our memory-impacted learners to forget some of the syntactic path units (typically, the ones further from the end of the path) and be forced to learn an efficient FG representation for the pieces of the syntactic paths that remain.

4.2 Recency effect implementation

To implement this recency effect, we altered the syntactic paths of the *wh*-dependencies in the modeled learner’s data intake, removing (i.e., “forgetting”) lexical items in these syntactic paths with a probability proportional to their position in the path (see (4)), which is based on a power-law implementation of a recency effect (Anderson and Milson, 1989).

$$(4) \quad \text{remember probability} = p_{rem} = \frac{1}{(\text{position} + 1)^\alpha}$$

In particular, a lexical item’s “remember probability” p_{rem} depends on its position relative to the end of the syntactic path. Lower positions are closer to the end (more recent), and so result in higher p_{rem} values. The α parameter represents the “forgetting rate” and controls the shape of the remember probability curve, with higher values resulting in more-recent items being remembered relatively more often (Figure 3). When $\alpha=0$, all positions are remembered perfectly (remember probability = 1.00). For $\alpha>0$, the more recent lexical items (lower positions relative to the end of the syntactic path) are remembered (somewhat or much) more often relative to less-recent positions.

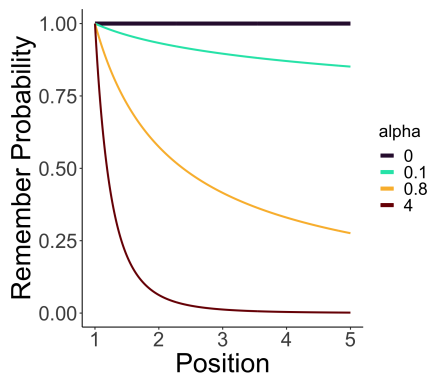


Figure 3: Probability of remembering a lexical item at different positions relative to the end of the syntactic path, based on forgetting rate α values.

Table 1 illustrates the impact of different forgetting rates (implemented with

α) for the example syntactic path from (3). In particular, the probability of the item being remembered is shown, relative to its position.

	PRESENT	<i>think</i>	NULL	PRESENT	<i>be</i>	<i>for</i>
Position	6	5	4	3	2	1
$\alpha = 0.1$	0.82	0.84	0.85	0.87	0.90	0.93
$\alpha = 0.8$	0.21	0.24	0.28	0.33	0.42	0.57

Table 1: Impact of different forgetting rates α on syntactic path lexical items, showing the probability of an item being remembered, given its position (Position) relative to the end of the syntactic path.

When a lexical item is forgotten, the modeled learner replaces it with a generic “unknown” (UNK) symbol. For instance, with $\alpha=0.8$, the probability of remembering the item in position 6 (PRESENT) is 0.21, and so with probability $1-0.21=0.79$, this item is forgotten when the syntactic path instance is encountered in the input, and replaced with UNK. That is, the syntactic path used by the modeled learner for learning becomes $IP_{UNK}-VP_{think}-CP_{NULL}-IP_{PRESENT}-VP_{be}-PP_{for}$.

5 Modeled learner input

Modeled learner input was drawn from a realistic sample of child-directed speech interactions, directed at children between the ages of one and a half and five, from the CHILDES Treebank (Pearl and Sprouse, 2013). We identified 12,704 *wh*-dependencies from this sample. We then estimated the total *wh*-dependencies encountered by children during a plausible learning period (from 18 months to 4 years old: Perkins and Lidz 2021; Pearl and Bates 2022), considering the average waking hours (Davis et al., 2004), utterances per hour (Rowe, 2012), and *wh*-dependency frequency in children’s input. This estimate yielded 2,146,324 *wh*-dependencies total for our modeled learners to learn from, distributed according to the sample of 12,704 *wh*-dependencies.

6 Results

Given this realistic input estimate of *wh*-dependencies to learn from, the modeled learners then use Bayesian inference to identify an efficient FG representation comprised of different-sized chunks that can be used to generate any *wh*-dependency. Below we report the modeled learners’ ability to generate the target behavior patterns that signal knowledge of syntactic islands. To demonstrate the impact of memory limitations, we show the performance both of learners with no memory limitations as a baseline as well as memory-impacted learners. The selected results come from modeled learners that have different rates of forgetting, yielding different average probabilities of forgetting any particular lexical item in a syntactic path: $\alpha=0$ (0% forgotten), $\alpha=0.1$ (9% forgotten), $\alpha=0.8$ (52%

forgotten), and $\alpha=4$ (96% forgotten). We ran 20 modeled learners for each of the forgetting rates (generating a new set of *wh*-dependency paths for each learner with some lexical items removed).³

6.1 Target pattern: The island difference

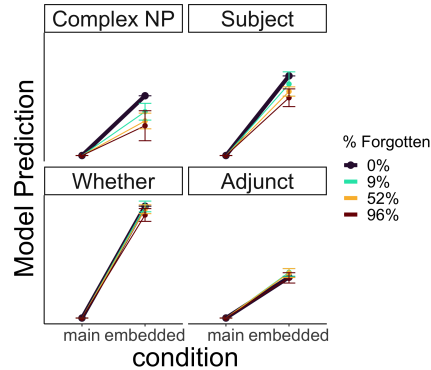


Figure 4: Generated acceptability judgment patterns, with 95% confidence intervals, from modeled learners with different levels of forgetting. All modeled learners are able to generate the target “island difference” pattern, indicated with a positive slope.

In Figure 4, we see that all modeled learners are able to generate the target behavior pattern of the “island difference”, which encodes the superadditive acceptability judgment pattern of Sprouse et al. (2012) and reflects internalized knowledge of syntactic islands. That is, both idealized learners with perfect memory (0% forgotten) and memory-impacted learners (9-96% forgotten) can generate the positive slope that indicates the appropriate acceptability judgment pattern. So, even under severe memory constraints (e.g., lexical items forgotten 96% of the time), learners identifying efficient *wh*-dependency chunks can succeed at acquiring *wh*-dependency knowledge that allows them to generate this behavior signaling knowledge of syntactic islands.

6.2 Target pattern: The impact of verb frequency

In Figure 5, we see that modeled learners without heavy memory loss (lexical items forgotten 0-52% of the time) are able to generate the target behavior pattern, where the frequency of the verb frame positively correlates with the judged acceptability of the selected *wh*-dependency. That is, both idealized learners (0% forgotten) and some memory-impacted learners (9-52% forgotten) can generate

³The training and testing sets, the output data, and relevant code generating the results can be found at <https://github.com/nielswd23/noisy-deps>.

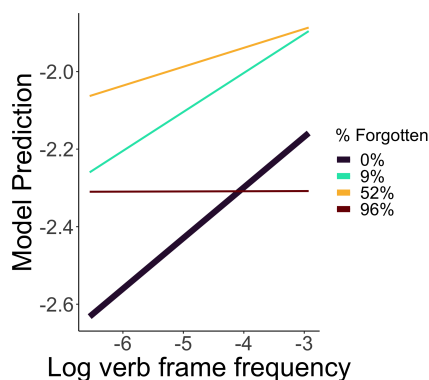


Figure 5: Generated acceptability judgment patterns from modeled learners with different levels of forgetting. Modeled learners without heavy memory loss ($\leq 52\%$ forgotten) are able to generate the target pattern that shows the impact of verb frequency, indicated with a positive slope.

the positive slope that indicates the appropriate acceptability judgment pattern. However, a learner who forgets lexical items 96% of the time can't. Intuitively, this failure by a severely-impacted learner is less surprising because this pattern depends more on specific lexical items (i.e., the main verb). So, it seems likely that this modeled learner didn't see enough lexical item data to notice that the specific lexical item of the main verb mattered.

More specifically, the verbs become too rare in the input for the severely-impacted modeled learner to include them in the learned chunks of the grammar. For example, in lower levels of forgetting, the learner can generate higher probability for *wh*-dependencies that include "say" (which appears frequently in the input) by using a learned chunk that includes "say" within a larger structure. This contrasts with *wh*-dependencies that include other verbs (e.g. "whine"), which don't appear frequently in the input; therefore, the learner doesn't learn larger, higher-probability chunks that include "whine". However, with 96% forgetting, "say" also doesn't appear frequently enough to include in a larger, higher-probability chunk. In fact, no verbs do. So, all *wh*-dependencies with this structure have the same probability, because they all use the same chunks that don't include the verb lexical items. More generally, learners who rely on the efficient chunking strategy can only succeed at acquiring the appropriate *wh*-dependency knowledge if their memory limitations are less severe.

6.3 Target pattern: Interpretation preferences

In Figure 6, we see that all modeled learners behave qualitatively the same, and are able to generate most of the target behavior preferences. That is, both idealized learners with perfect memory (0% forgotten) and memory-impacted

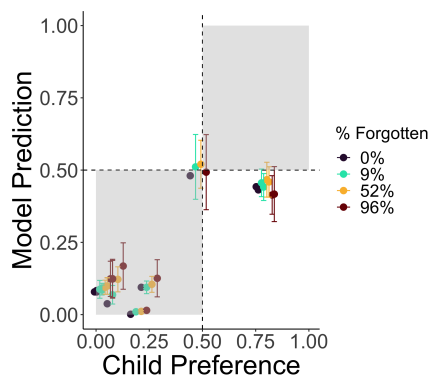


Figure 6: Generated modeled learner interpreted preferences compared against observed child interpretation preferences at varying forgetting rates. Each point plots the mean model score (preference for the long-distance dependency interpretation) on the y-axis for an individual test item. The error bars indicate a 95% confidence interval over the 20 modeled learner runs.

learners (9-96% forgotten) can generate nearly all the preference patterns (7 out of 9) that children have. So, even with memory constraints, modeled learners identifying efficient *wh*-dependency chunks can succeed at acquiring most of the *wh*-dependency knowledge that allows them to generate this behavior signaling knowledge of syntactic islands. For the remaining patterns that weren't captured, future work can investigate if different implementations of this learning theory (perhaps with different memory or data intake assumptions – see discussion in section 7) can capture more of the observed target behavior.

6.4 Results summary

We find that even memory-impacted modeled learners are still able to capture the majority of the target behavior patterns, just as idealized learners with perfect memory can. In many cases, even very severe memory limitations (lexical items forgotten 96% of the time) don't hinder the modeled learners from learning representations that allow them to generate the target behavior patterns. We interpret these results as supporting the plausibility of the learning theory implemented in the modeled learners, which involves learners trying to identify efficient chunks for their representation of *wh*-dependencies. That is, while we may not know for certain exactly how much information is lost due to children's memory limitations, it seems likely that this learning theory is viable as a way for children to learn about syntactic islands.

7 General discussion and future work

Our findings support the proposed learning theory from Dickson et al. (2022), which focuses on a learning an efficient representation of the input, by demonstrating that modeled learners implementing this theory can acquire relevant *wh*-dependency knowledge about syntactic islands (Sprouse et al., 2012; Liu et al., 2022; De Villiers et al., 2008) even when they operate with memory constraints. More broadly, our findings align with the goal of developing cognitively-plausible theories of acquisition that evaluate how well learning theories would actually work for children who are developing both their linguistic (and non-linguistic) knowledge along with their linguistic (and non-linguistic) processing abilities Pearl (2023a,b).

However, there are limitations to the current learner implementations that can be investigated in the future. For instance, the learners here implemented a memory constraint based on the established psychological phenomenon of the recency effect. However, other (more sophisticated) memory constraints are possible, such as those anchored in word predictability (more predictable words are better remembered: Hahn et al. 2022). We can additionally consider memory constraints that cause learners not only to forget lexical items, but also the syntactic structure associated with those items.

Another assumption of the current work concerns the learner’s intake. Here, the modeled learners were restricted to the syntactic path, which excluded other (potentially relevant) parts of the structure. Future work could include the rest of the utterance structure as part of the intake, forcing the modeled learners to learn an efficient representation for the entire utterance structure that contains the *wh*-dependency.

In addition, because the strategy of identifying efficient chunks is not limited to learning about syntactic islands, future work may also be able to evaluate whether this learning theory can capture other sophisticated *wh*-dependency knowledge that involves multiple interpreted positions (gaps), such as across-the-board extraction and parasitic gaps (Ross, 1967).

References

- John R Anderson and Robert Milson. 1989. Human memory: An adaptive perspective. *Psychological Review* 96(4):703.
- Katherine Finn Davis, Kathy P Parker, and Gary L Montgomery. 2004. Sleep in infants and young children: Part one: normal sleep. *Journal of Pediatric Health Care* 18(2):65–71.
- Jill De Villiers, Thomas Roeper, Linda Bland-Stewart, and Barbara Pearson. 2008. Answering hard questions: Wh-movement across dialects and disorder. *Applied Psycholinguistics* 29(1):67–103.
- Niels Dickson, Lisa Pearl, and Richard Futrell. 2022. Learning constraints on wh-dependencies by learning how to efficiently represent wh-dependencies: A develop-

- mental modeling investigation with fragment grammars. *Proceedings of the Society for Computation in Linguistics* 5(1):220–224.
- Yana Fandakova, Myriam C Sander, Markus Werkle-Bergner, and Yee Lee Shing. 2014. Age differences in short-term memory binding are related to working memory performance across the lifespan. *Psychology and Aging* 29(1):140.
- Susan E Gathercole, Susan J Pickering, Benjamin Ambridge, and Hannah Wearing. 2004. The structure of working memory from 4 to 15 years of age. *Developmental psychology* 40(2):177.
- Michael Hahn, Richard Futrell, Roger Levy, and Edward Gibson. 2022. A resource-rational model of human processing of recursive linguistic structure. *Proceedings of the National Academy of Sciences* 119(43):e2122602119.
- Mina Robinson Hirzel. 2022. *Island Constraints: What is there for children to learn?*. Ph.D. thesis, University of Maryland, College Park.
- Yingtong Liu, Rachel Ryskin, Richard Futrell, and Edward Gibson. 2022. A verb-frame frequency account of constraints on long-distance dependencies in english. *Cognition* 222:104902.
- David Marr. 2010. *Vision: A computational investigation into the human representation and processing of visual information*. MIT press.
- Brian McElree, Stephani Foraker, and Lisbeth Dyer. 2003. Memory structures that subserve sentence comprehension. *Journal of memory and language* 48(1):67–91.
- Timothy J O’Donnell. 2015. *Productivity and reuse in language: A theory of linguistic computation and storage*. MIT Press.
- Scott G Paris. 1978. The development of inference and transformation as memory operations. In *Memory Development in Children*, Psychology Press, pages 129–156.
- Lisa Pearl. 2023a. Computational cognitive modeling for syntactic acquisition: Approaches that integrate information from multiple places. *Journal of Child Language* 50(6):1353–1373.
- Lisa Pearl. 2023b. Modeling syntactic acquisition. In Jon Sprouse, editor, *The Oxford Handbook of Experimental Syntax*.
- Lisa Pearl and Alandi Bates. 2022. 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. *Journal of Child Language* pages 1–34. <https://doi.org/doi:10.1017/S0305000922000514>.
- Lisa Pearl and Jon Sprouse. 2013. Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition* 20:19–64.
- Lisa S Pearl. 2021. How statistical learning can play well with Universal Grammar.
- Laurel Perkins and Jeffrey Lidz. 2021. Eighteen-month-old infants represent nonlocal syntactic dependencies. *Proceedings of the National Academy of Sciences* 118(41):e2026469118.
- John Robert Ross. 1967. Constraints on variables in syntax. .
- Meredith L Rowe. 2012. A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child development* 83(5):1762–1774.
- Jon Sprouse, Matt Wagers, and Colin Phillips. 2012. A test of the relation between working-memory capacity and syntactic island effects. *Language* pages 82–123.