Caregiver reconstruction of children's errors: The preservation of complexity in patterned systems

Madeline Meyers[1] & Daniel Yurovsky[1,2]

[1] University of Chicago/Stanford University???

[2] Carnegie Mellon University

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

Correspondence concerning this article should be addressed to Madeline Meyers, Madeline Meyers. E-mail: mcmeyers@????.edu

Abstract

Why do languages change? One possibility is they evolve in response to two competing pressures: (1) to be easily learned, and (2) to be effective for communication. In a number of domains, variation in the world's natural languages appears to be accounted for by different but near–optimal tradeoffs between these two pressures. Models of these evolutionary processes have used transmission chain paradigms in which errors of learning by one agent become the language input for the subsequent generation. However, a critical feature of human language is that children do not learn in isolation. Rather, they learn in communicative interactions with caregivers who draw inferences from their errorful productions to their intended interests. In a set of iterated reproduction experiments with both children and adults, we show that this supportive context can have a powerful stabilizing role in the development of artificial patterned systems, allowing them to achieve higher levels of complexity than they would by vertical transmission alone. Yet, the systems retain equivalent transmission accuracies—they are equally easy to transmit to the new generation. Thus, the caregiver plays a dual role as both a teacher and a protector of the patterned system as whole, facilitating its evolution to an optimal balance of learnability and communicability

Keywords: keywords

Word count: X

Caregiver reconstruction of children's errors: The preservation of complexity in patterned systems

The languages we speak today are not the same as the ones we spoke 300 years ago. Nor are they the same as the ones we spoke 500, 1000, or 2000 years ago. Why do languages change, aside from acquiring new vocabulary? One working theory is that they evolve to adapt to two dynamic competing pressures: (1) to be easily learned and transmitted, and (2) to be effective for communication (Kirby, Tamariz, Cornish, & Smith, 2015).

While children are often the actors who drive language evolution (Senghas, 2003), they differ from adults in their cognitive capabilities (Kempe, Gauvrit, & Forsyth, 2015), interests and early vocabularies, and conversation partners. As early language producers who are inundated with new information each day, children may be particularly biased towards simplification (Senghas, 2003). Indeed, when children are learning language, they often make simplification errors (Bowerman, 1982). This reflects the influence of the transmissibility pressure – children may latch onto word-forms which are simpler and thus easier to acquire. For example, if a child is asking for her bottle, she may be unable to produce the canonical label "bottle", and may produce the simplified form, "baba", instead. If this child grew up without competent speakers of the language and, unlike an average child, failed to have multiple opportunities to acquire the correct label, it is possible that she would retain and reproduce "baba", even to her children in the future. In this way, her error is retained in the language and sustained over generations. With too many of these simplification errors, however, a language can lose its ability to be effective for communication (Kempe et al., 2015). What enables languages to retain their communicative utility and expressivity in the face of these learnability pressures?

Children do not learn language in isolation, but they communicate with fluent speakers of the language–their parents and caregivers. Caregivers are able to combine their children's productions with both their own knowledge of the language as well as their knowledge of their children. This enables caregivers to be excellent interpreters of child utterances (Chouinard & Clark, 2003). Their interpretation skills may be a form of scaffolding for their children's language learning (Lustigman & Clark, 2019). Parents are able to successfully interpret a more simplified utterance, thus shouldering some of the complexity of the language – for a short time. Our hypothetical child may continue to call a bottle "baba" until she can handle the cognitive load of "bottle", and her caregivers will support this learning in multiple ways.

Caregivers, through their explicit interventions as well as their implicit modeling of correct language, may be scaffolding their children's language-learning, by providing a space for their children to simplify, as well as by re-introducing complexity into their communications. Adults can explicitly correct their children's language errors in various ways (e.g., by interruptions or repeating the correct word/grammatical form; Penner, 1987). Yet, children primarily learn language through listening to others talk, rather than explicit instruction (Romberg & Saffran, 2010). Thus, parents' modeling of accurate language constructions can have a powerful effect on reducing children's language errors: over time, children fix their own mistakes because they have had multiple opportunities to learn correct constructions from their caregivers (Hudson Kam & Newport, 2005). By way of this feedback, both implicit and explicit, children's simplification errors are corrected, and children are able to acquire adult-like speech. Eventually, when a child becomes an adult, they will not transmit the errors they previously had, but the correct forms they learned from their caregivers – as long as learning the correct forms is useful and necessary. Thus, over the course of a lifetime, the child language learner grows to become a parent language teacher, correcting their own children's errors. These error reconstructions may be a mechanism by which more complexity is retained in language over many lifetimes than children could sustain alone.

Using iterated reproduction to study language change

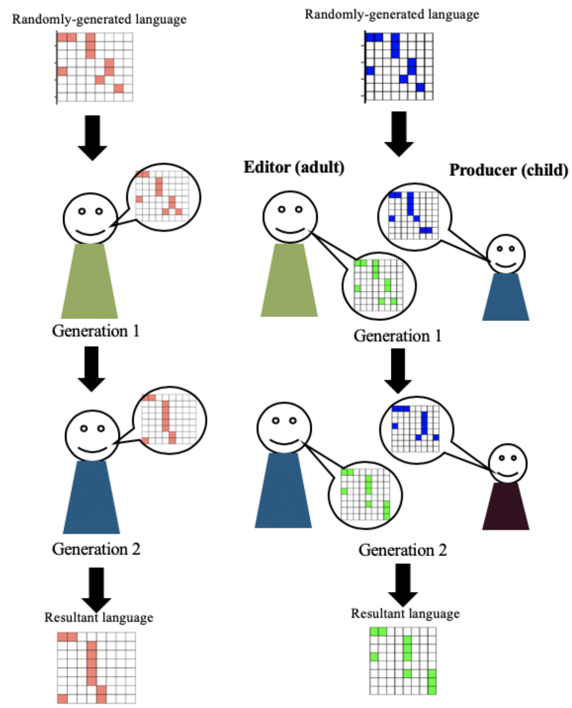

Figure 1. Experiments 1a, 2a, and 3a follow the conventional diffusion chain (iterated learning) paradigm, where a novel language is transmitted vertically through successive learning and recall. In Experiments 1b, 2b, and 3b, an element of horizontal transmission is added to the paradigm: novel language producers' reproductions are subject to alterations from a secondary participant, whose input is passed to the subsequent generation.

To model the impact of these competing pressures on language evolution in the laboratory, we use a diffusion chain paradigm developed by (Kirby, Dowman, & Griffiths, 2007). In this paradigm, one participant is trained on a randomly-generated language – e.g, a set of words created by arbitrarily pairing syllables together. The participant is later asked to recall the language, but inevitably makes some errors. Their errorful output becomes the training input for the subsequent participant, forming a transmission chain. This iterated learning process models the transmission of language over generations, with each participant unintentionally changing the language through their memory biases. The errors produced by participants reflect their memory or inductive biases–essentially, when the participant makes a mistake or mis-remembers, they rely on what they expect to see (Kalish, Griffiths, & Lewandowsky, 2007).

This paradigm has been used productively across a number of studies of cross-generational transmission in adults (Christiansen & Kirby, 2003; Kirby et al., 2007; Kirby, Griffiths, & Smith, 2014; Smith & Wonnacott, 2010), and children (Kempe et al., 2015; Raviv & Arnon, 2018). Various recent studies have also compared languages evolved over multiple generations (vertical transmission) to languages evolved by iterated use in the same conversational partners (horizontal transmission; Kirby et al., 2015). Indeed, research has shown that horizontal interaction between participants, specifically repair, does affect the language evolution process, resulting in increased communicative efficiency. However, repair's effects on communicative success (accuracy) are unclear (Micklos, Macuch Silva, & Fay, 2018). Typically, participants in horizontal transmission scenarios had similar levels of knowledge and similar cognitive constraints. This is different from children, who learn language in asymmetric knowledge situations, where their parent both knows more language and has an adult cognitive and executive-functioning (working-memory) system (1). We predict that this asymmetry may have a unique role in the evolution of language, allowing it to resist some of the simplifying pressure of transmissibility through adults' ability to maintain complexity while their children develop.

We adapted Kempe et al. (2015)'s non-linguistic iterated reproduction paradigm, as it has been used successfully with children (a similar task was used in non-human primates by Calidière, Smith, Kirby, & Fagot, 2014). This paradigm uses a stimulus set of novel grid patterns, which are akin to language in that they are patterned, structured systems–just as early language learning is attuning to and recognizing patterns in sound Yurovsky, Yu, and Smith (2012). We therefore use this paradigm, adapted to an iPad and Amazon Mechanical Turk format, to model the effect of introducing a secondary, error-correcting participant on the evolution of language. We hypothesize that these error-correctors (analogous to caregivers and teachers) are important not only to an individual's successful language acquisition, but also to the evolution of the language as a whole. This is because those who correct mistakes and provide feedback are able to protect against the strong transmissibility bias in early language producers by re-introducing and preserving complexity.

All experiments included in this paper were IRB-approved and were preregistered on Open Science Framework, and all tasks, data, and analysis code are available on Github (FOOTNOTE; see page XX for links).

Experiment 1a: Replicating Kempe et al. (2015)

We began by replication Kempe et al. (2015)'s experiment using a nonlinguistic stimulus to study the evolution of structure in an artificial symbolic system. Our motivations for using this paradigm were twofold. First, the stimuli lent themselves to algorithmic quantification of complexity. Second, Kempe et al. (2015) used this paradigm successfully with children, and our goal was to test our hypotheses not just in adult-adult chains, but also in child-child and child-adult chains (Experiments 3a and 3b). We chose to use the diffusion chain paradigm for six transmission generations. Although Kempe et al. (2015) original study used ten transmission generations, the results appeared to approach stable levels of complexity at six generations, so this number was chosen as our starting point.

Participants. Participants were 125 adults recruited on Amazon Mechanical Turk. Because five users failed to meet inclusion criteria, a larger number of participants was required to obtain the planned sample of 120. These participants were members of one of twenty diffusion chains, each of which had six generations. Each participant gave informed consent. The task was approximately eight minutes long, and subjects were compensated $0.50 for their participation.

Design and Procedure. Participants were asked to re-create patterns on a grid. Subjects were informed that they would see a target grid appear on their computer screen for ten seconds, followed by a picture (visual mask) displayed for three seconds. After the visual mask, participants viewed a blank 8x8 grid where they were given one minute to re-create the target grid (see **??**). A visual mask was used to ensure that the participants were storing the target patterns in working memory, rather than sensory memory (i.e., they were not re-producing the patterns from a transitory image; Phillips, 1974). Participants could click on any cell in the grid to change its color and could also remove any color placed. A counter on the screen showed how many cells had been colored, and it varied dynamically with the participant's clicks. After placing 10 colored blocks (called "stickers" in the experiment), participants could click a button to advance to the next trial (See Appendix Figures XXXXX for example grids). A timer was displayed on the screen, and participants were given an audio cue when they had fifteen seconds left.

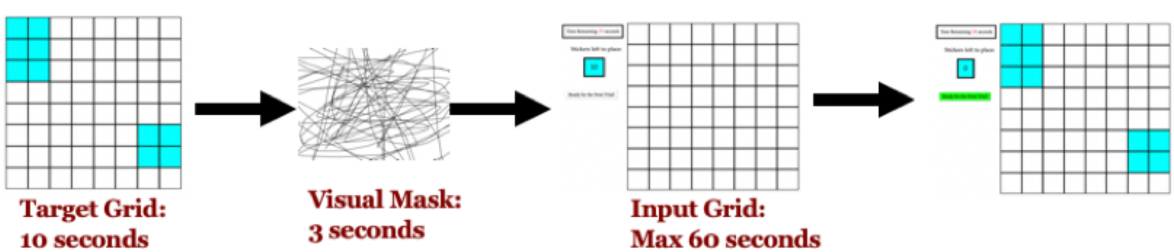

Figure 2. shows the experimental task (training trial shown) for experiments 1a, 2a, and 3a, as well as for producers in 1b, 2b, and 3b.

After completing one training and three practice trials, each participant completed 6 experiment trials. During the experimental trials, there was an additional display on the screen which informed the participant of how many trials they had left. Throughout the experimental trials, participants heard various engaging audio cues, including "You're doing great, keep it up!", "You're halfway there!", and "Just one more to go!". These were added to the task to add an additional level of engagement for data collection with children

(Experiments 3a and 3b). Participants in the first generation of each chain received the same initial grid patterns. These initial 8x8 grids were generated by randomly selecting 10 of the 64 possible cells to be filled using a random-number generator. Participants in subsequent generations received as their targets the outputs produced by the previous participant in their chain.

Prior to the experimental trials, all participants received the same training and practice trials. In the preliminary training trial, subjects viewed two 8x8 grids side-by-side and were instructed to make the blank grid on the right match the target grid on the left. Participants were unable to progress to the practice and experimental trials without reaching perfect accuracy on this first trial. The three practice trials followed the format in **??**; however, the target patterns were simpler to reproduce. Participants were required to meet a set of attention criteria for their data to be included in the transmission chain. If the participant scored less than 75% accuracy on the last two practice trials, or if they failed to select 10 cells before time ran out, their outputs were not transmitted to the next generation.

Analysis. Our primary measures of interest were reproduction accuracy and pattern complexity. Reproduction accuracy served as a proxy for transmissibility – higher reproduction accuracies indicated that the language was easier to learn. Reproduction accuracy was computed as the proportion of targets out of 10 placed in the exact same location on the target and input grids. This measure of accuracy did not count for the degree of error made by a participant–if they only misplaced a block by one unit, it was counted as incorrect, just as if they had misplaced the block by more than one unit.

Complexity served as a proxy for expressiveness. The ideal mechanism for measuring complexity is still contested, therefore, we followed Kempe et al. (2015) in using several metrics: algorithmic complexity, chunking, and edge length. Algorithmic complexity was calculated using the Block Decomposition Method, a measure of Kolmogorov-Chaitin Complexity applied to 2-dimensional patterns (Feldman, 2006; Zenil, Soler-Toscano, Dingle,

& Louis, 2014). This measure computes the length of the shortest Turing machine program required to produce the observed pattern. The shorter the program, the simpler the pattern. Chunking is the number of groups of colored blocks which share an edge. The more groups of blocks, the easier the pattern is to transmit, and the lower its complexity. Edge length is the total perimeter of the colored blocks and is similar to chunking. Implementation of these metrics was adapted from code provided by Gauvrit, Soler-Toscano, and Guida (2017). While algorithmic complexity was our primary dependent variable of interest, chunking and edge length served as additional measures to check the reliability of the Block Decomposition Method.

Results

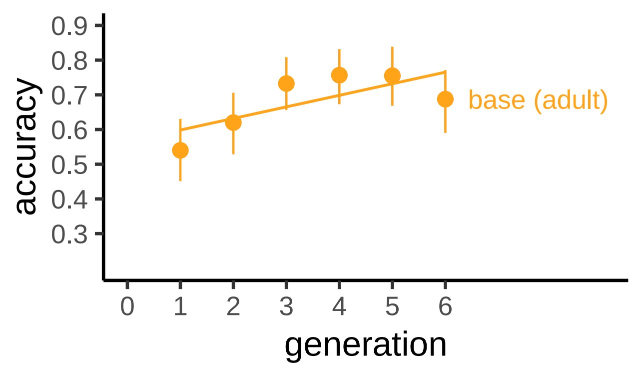

Figure 3. displays the transmission accuracy results from Experiment 1a. Error bars represent nintey-percent confidence intervals.

If iterated learning captures the hypothesized pressures of learnability and expressiveness, we predict that reproduction accuracy should increase, and complexity should decrease over generations. We tested these predictions with mixed–effects logistic regressions, first beginning with the most maximal model, and then reducing the model by first removing slopes, and then removing intercepts until the model converged. Our final model predicted accuracy and all three measures of complexity separately from fixed effects of generation and trial number, and random intercepts for participant and initial grid (e.g. `accuracy ~`

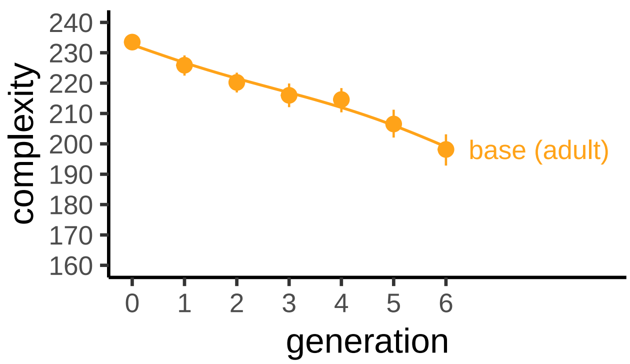

Figure 4. displays the algorithmic complexity results for Experiment 1a. Error bars represent nintey-percent confidence intervals.

`generation + trial + (1|subject) + (1|initialGrid)`. Reproduction accuracy increased significantly over generations (3; $\beta = 0.15$, t = 3.64, p = < .001). Complexity on all three measures (algorithmic complexity, chunking, and edge length) decreased significantly over generations, as shown in **??** (β_{BDM} = -26.52, t = -7.10, p = < .001; $\beta_{chunking}$ = -1.14, t = -5.81, p = < .001; β_{edge} = -4.38, t = -7.38, p = < .001). Trial number, or how far along the subject was in the task, was not a significant predictor in any model.

Experiment 1b: Introducing an Editor

This project aims to model the effects of the simplicity pressure present in language learning–retaining too much complexity in language is infeasible and unproductive. Without any scaffolds, adults are unable to retain the original complexity of the symbol system, as shown in Experiment 1a–instead, the patterns simplify to a more easily-transmissible level of complexity. What happens if scaffolding supports are introduced into the transmission process, just as they are in real-life with knowledgeable speakers, caregivers, and teachers?

In order to add an element of feedback from a more experienced interlocutor to the iterated reproduction process, we adapted the task from Experiment 1a to include a secondary, “editing” participant. This participant was analogous to a caregiver who protects their child from acquiring and perpetuating errors in symbol systems (i.e., language).

Participants. Participants in Experiment 1b were 289 adults recruited on Amazon Mechanical Turk. Approximately 17% (n=49) of participants in Experiment 1b were excluded from analysis due to failure to meet accuracy requirements on the practice trials (n=38) or failure to select the necessary number of targets on one or more experimental trials (n=11). More participants who were designated as "editors" failed to meet accuracy requirements (n=23) compared to those who were designated as "producers" (n=15). This resulted in a total of 240 participants included in the analysis. These participants occupied one of twenty diffusion chains and one of six generations. Each participant gave informed consent and was compensated with $0.50 for their participation in this 8-minute task.

Design and Procedure. A primary participant was designated as a "producer" and completed the same task as in Experiments 1a and 1b (see **??**). As before, the "producer" completed an iterated reproduction task, where they were told to re-create patterns on a grid. After completing the experiment, a secondary, "editing" participant was given an adapted task. Throughout the study, including in training and practice trials, editors were not told to re-create patterns, but to edit patterns to resemble a target grid exactly. Editors in this experiment viewed the same target grid as producers, but instead of seeing an empty input grid, they were given a grid prepopulated with 10 elements. These were the elements the previous producer had generated. The editing participant could then change the 10 items' positions. There was no "reset" button during this task, so data reflected participants' initial instincts. In Experiment 2a, a generation consisted of a producer, who re-created the target grid, and an editor, who altered the producer's re-creation to match the same target grid. The editor's changed pattern was used as the target grid for the subsequent generation.

Analysis. As in Experiments 1a and 1b, our primary measures of analysis were accuracy and complexity. Transmission accuracy was calculated as the proportion of 10 targets placed correctly, while complexity was measured using the Block Decomposition

Method of algorithmic complexity (Zenil et al., 2014), as well as measures of chunking and edge length, as in Experiment 1a (Gauvrit et al., 2017).

Results

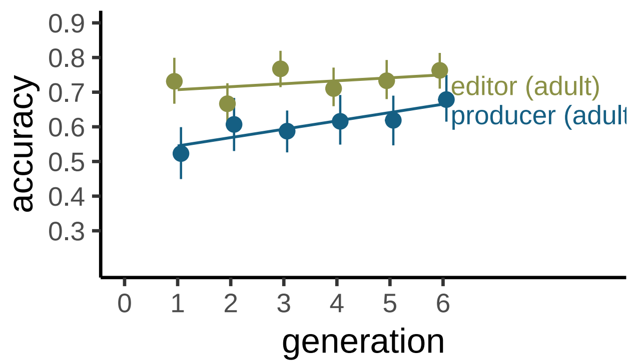

Figure 5. displays the results for transmission accuracies from Experiment 1b. Error bars represent nintey-percent confidence intervals.

5 shows the transmission accuracy results by editors and producers in Experiment 1b. We fit a linear mixed-effects model of the form `accuracy ~ condition * log(generation) + trial + (1|initialGrid) + (trial|subject)`. There was no main effect of generation, meaning that the patterns were not produced more accurately over transmissions ($\beta_{log(generation)}$ = 0.02, t = 0.78, p = .434). There was, however, a significant effect of condition, with producers having lower transmission accuracies compared to editors ($\beta_{producer}$ = -0.20, t = -3.50, p = .001). There was an additional significant effect of trial, with later trials having higher accuracies than earlier trials in the task (β_{trial} = 0.01, t = 2.23, p = .023).

6 shows the relationship between the complexity of editors' and producers' patterns. In each generation, the producer decreased the complexity of the pattern, and the editor was able to compensate for some of this loss by re-introducing complexity. We fit the same model as above, this time predicting algorithmic complexity rather than transmission accuracy.

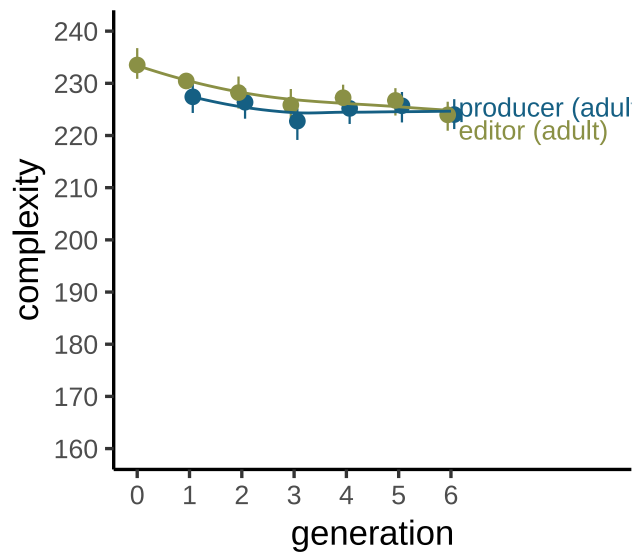

Figure 6. shows the results of Experiment 1b for algorithmic complexity. Example patterns produced by participants at generations 0 (initial, randomly-generated pattern), 6, and 12 (resultant pattern) are shown.

Trends were consistent across the additional measure of edge length and chunking. There was a main effect of generation, with both producers and editors producing simpler patterns over transmissions ($\beta_{log(generation)}$ = -4.36, t = -3.32, p = .001). Producers had lower complexity values than editors ($\beta_{producer}$ = -5.54, t = -2.19, p =.030). There was an additional significant effect of trial, with participants creating more complex patterns as they completed more trials in the task (β_{trial} = 0.77, t = 3.08, p = .002).

Experiment 2: Replication Study

In line with Kempe et al. (2015), we found that accuracy increased across generations and complexity decreased on all three measures in Experiment 1a. In Experiment 1b, we introduced an element of horizontal transmission into the diffusion chain paradigm by adding a secondary "editing" participant who adjusted the producer's patterns to match a target. However, in both studies, we were interested in the shape of the trends observed – namely, whether there were differences in the rates of change with successive generations. We thus replicated both experiments 1a and 1b while increasing the number of generations from six to twelve. This replication would not only increase the strength of our findings, but it would also help to estimate the shape of the functions modeled by the algorithmic complexity findings.

Experiment 2a: Adult Baseline Replication

This experiment replicated the task from Experiment 1a with the addition of twice as many chains and generations.

Participants. Participants were 519 adults recruited on Amazon Mechanical Turk. Approximately 8% (n=39) of participants in Experiment 2a were excluded due to failure to meet accuracy requirements on the practice trials or failure to select the complete number of cells on one or more experimental trials. This resulted in a total of 480 participants included in the analysis. These participants were members of one of forty diffusion chains, each of which had twelve generations. Each participant gave informed consent and was compensated with $0.50 for their participation in this 8-minute task.

Design and Procedure. The task in Experiment 2a was identical to Experiment 1a. Participants were told to reproduce patterns on a grid, and their responses were passed to the next subject in the transmission chain. They completed one training, three practice, and six experimental trials.

Results

For sample patterns produced by participants during the task, see Appendix Figure XXX. The results of this experiment replicated those found in Experiment 1a. Reproduction accuracies increased significantly over generations, as shown in 7 ($\beta = 0.09$, t = 0.09, p = < .089).

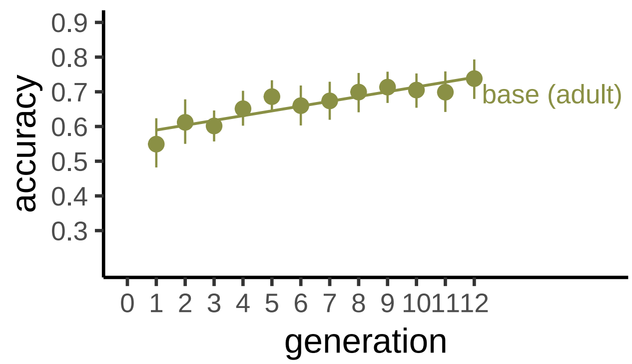

Figure 7. shows the results from Experiment 2a for transmission accuracy. Error bars represent nintey-percent confidence intervals.

**FIX THIS WHOLE SECTION BECAUSE ITS NOT EXPONENTIAL U DID AN OOPS 8 shows the results for algorithmic complexity. Algorithmic complexity appeared to follow an exponential function of the form y = e-ax + b. We therefore fit an exponential mixed-effects regression model predicting complexity from fixed effects of generation and trial number, and random intercepts for participant and initial grid (e.g. `log(complexity)` ~ `log(generation+1) + trial + (1|subject) + (1|initial)`. Algorithmic complexity decreased over generations, and the rate of change decreased as well (β_{BDM} = -0.04, t = -5.11, p = < < .001). Similar trends were found with chunking and edge length, the alternate measures of complexity ($\beta_{chunking}$ = -0.82, t = -13.03, p = < < .001; β_{edge} = -1.57, t = -9.85, p = < < .001). As in Experiment 1a, trial number was not a significant predictor in any model.

Experiment 2b: Adult–Adult Dyad Replication

This experiment replicated the task from Experiment 1b with the addition of twice as many chains and generations. As in Experiment 1b, a secondary participant was assigned to be an “editor”, completing a variation of the standard iterated reproduction task.

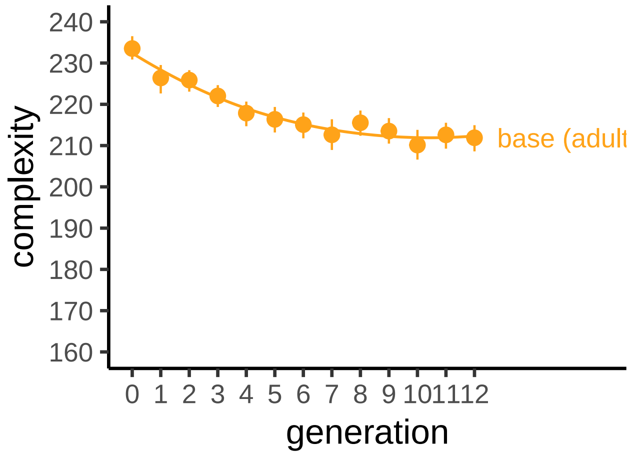

Figure 8. shows the results of Experiment 2a for algorithmic complexity. Example patterns produced by participants at generations 0 (initial, randomly-generated pattern), 6, and 12 (resultant pattern) are shown.

Participants. Participants in Experiment 2b were 1031 adults recruited on Amazon Mechanical Turk. Approximately 8% (n=71) of participants in Experiment 2b were excluded from analysis due to failure to meet accuracy requirements on the practice trials or failure to select the necessary number of targets on one or more experimental trials. This resulted in a total of 960 participants included in the analysis. These participants occupied one of forty transmission chains and one of twelve generations. Each participant gave informed consent and was compensated with $0.50 for their participation in this 8-minute task.

Design and Procedure. The procedure in this task replicated that of Experiment 1b. As before, one participant was designated as a "producer", who completed the standard iterated reproduction task. This participant was told to re-create patterns on a blank grid to match a target they had seen displayed for ten seconds. After the completion of one training, three practice, and six experimental trials, the producer's data was passed to a second, "editing" participant. This participant was told to "fix" the producer's patterns to match the same target, which they had also seen displayed for ten seconds. The editor's edits were passed as the target for the subsequent generation. Together, the producer and editor occupied one “generation”.

Analysis and Results

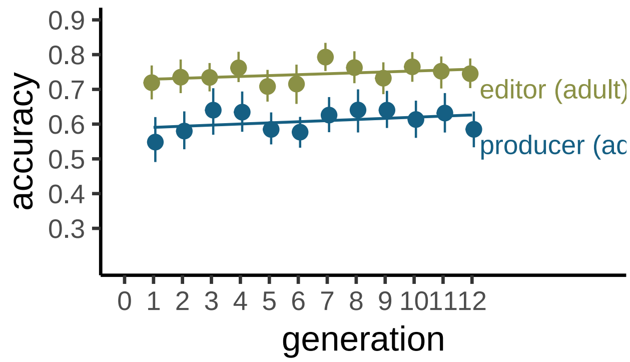

Figure 9. displays the results for transmission accuracies from Experiment 1b. Error bars represent nintey-percent confidence intervals.

As in previous experiments, the data were analyzed for transmission accuracy and three measures of complexity. 9 displays the results for transmission accuracy. According to a linear mixed-effects model predicting group from generation and trial number and controlling for random effects of subject and initial grid, reproduction accuracies between groups were significantly different ($\beta_{condition-producer}$ = -0.14, t = -4.11, p = < < .001). Neither the editors’ nor producers’ transmission accuracies increased significantly over generations ($\beta_{log(generation+1)}$ = 0.02, t = 1.34, p = .181).

**FIX EXPONENTIAL FUNCTION?? 10 shows the relationship between the complexity of editors’ and producers’ patterns. In each generation, the producer decreased the complexity of the pattern, and the editor was able to compensate for some of this loss by re-introducing complexity into the grid patterns. As in Experiment 2a, we fit an exponential function to this data. There was no significant effect of trial number in this model. There was., however, a main effect of generation, with both the editors’ and producers’ patterns decreasing in complexity over generations ($\beta_{log(generation)}$ = -0.02, t = -6.45, p < .181).

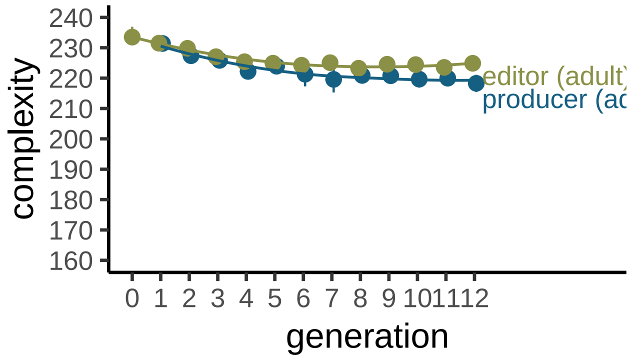

Figure 10. displays the algorithmic complexity results for Experiment 1a. Error bars represent nintey-percent confidence intervals.

Additionally, the producers had significantly lower levels of complexity compared to the editors ($\beta_{producer}$ = -0.02, t = -4.11, p < < .001).

Experiment 2 Results

```
## Warning: Removed 1 rows containing missing values (geom_pointrange).
```

11 shows the combined accuracy results of Experiments 2a and 2b. We fit a linear mixed–effects model predicting condition (baseline adult, producer, or editor) from accuracy, log(generation), and trial number, including random effects of subject and initial grid. There were significant main effects of generation and condition, where accuracy increased over generations ($\beta_{log(generation)}$ = 0.05, t = 5.48, p < < .001). Producers had significantly lower transmission accuracies compared to the other groups ($\beta_{producer}$ = -0.03, t = -1.10, p < .270). Producers did not show significantly smaller increases in accuracy over generations ($\beta_{producer*log(generation)}$ = -0.03, t = -1.89, p < .059). Trial number was not a significant predictor in this model (β_{trial} =0, t = -0.28, p .777).

12 shows the combined complexity results for Experiments 2a and 2b. As in previous experiments, we fit an exponential function to this data. Algorithmic complexity of the patterns decreased over generations ($\beta_{log(generation)}$ = -3.88, t = -4.88, p < < .001).

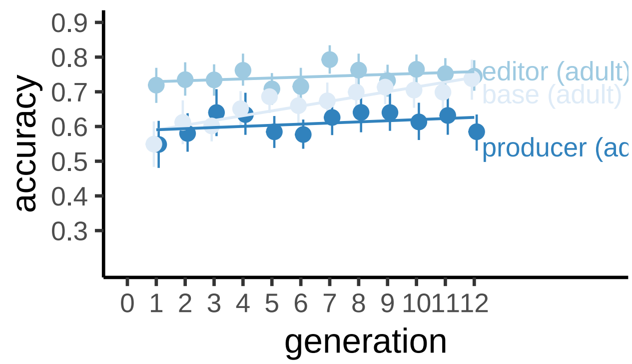

Figure 11. shows the results from Experiment 2a for transmission accuracy. Error bars represent nintey-percent confidence intervals.

D ON T UNDERSTAND Notably, editors and producers had significantly higher levels of algorithmic complexity across generations compared to the adult baseline condition from Experiment 2a (β_{editor} = 0.049, t = 9.618, p < 0.001; $\beta_{producer}$ = 0.028, t = 5.509, p < .001). There was no significant effect of trial number in this model.

Experiments 1-2 Discussion

The combined results of Experiments 1a and 2a replicate the adult results found in a similar task by Kempe et al. (2015). In a standard iterated reproduction experiment, the complexity of adults' produced patterns decreased over transmission generations to reach an asymptote. This shows that generational transmission creates a bottleneck in the evolutionary process, where patterned systems quickly lose complexity, but appear to reach a level of simplicity which is easier to reproduce. These results are supported by transmission accuracy findings, which show that participants become better at reproducing the target patterns over generations. These effects are robust, as they were replicated in two experiments. The pressure of passing a patterned system to a new participant modeled the simplicity pressure on language learning. In Experiment 2a, we see that the system did not simplify to nothingness but reached a more stable level of simplicity over transmissions. Thus, this system was not losing all of its descriptiveness, instead reaching a balance point

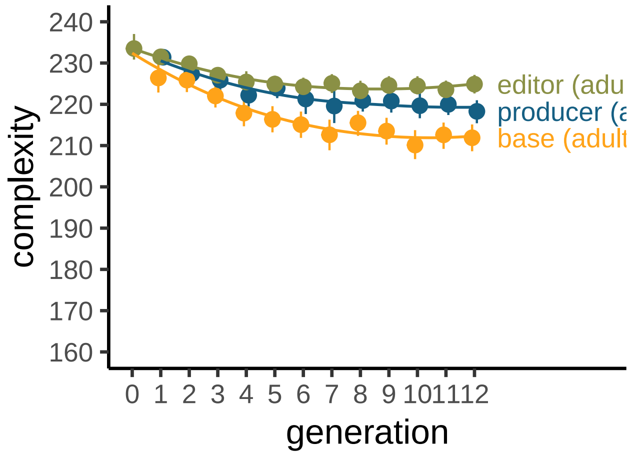

Figure 12. shows the results of Experiment 2a for algorithmic complexity. Example patterns produced by participants at generations 0 (initial, randomly-generated pattern), 6, and 12 (resultant pattern) are shown.

between transmissibility and expressivity.

In a few instances, we also see significant effects of trial number, where the more trials a participant completes, the more complex their patterns become. Perhaps this is a fatigue issue, where participants are placing blocks randomly when they forget their exact location. However, this is contrary to the practice effect, which predicts that participants are actually becoming better at the task and are simply retaining more complexity because they remember the patterns better over time (Donovan & Radosevich, 1999).

The results of Experiments 1b and 2b attempt to more closely model pattern-system transmissions (loosely analogous to language-learning) using an iterated reproduction paradigm. These experiments introduce a secondary participant into the iterated reproduction process–a caregiver-like participant who has more knowledge about the novel system. These editing participants had higher transmission accuracies, reflecting their greater "knowledge" about the task and ability to reproduce more-accurate patterns compared to producers, who had to re-create grid patterns from scratch. This relationship is meant to mirror parents who have a relatively easier time recalling, editing, and producing language compared to early child language learners. Indeed, in Experiments 1b and 2b, "editors" were not simply completely re-creating the patterns made by "producers", but they were fixing their errors.

The algorithmic complexity results of Experiment 1b and 2b show that the loss in complexity from Experiments 1a and 1b is not permanent cultural regression (Henrich, 2004), as complexity can be reintroduced in the patterned system by way of a secondary participant. When the iterated reproduction process begins to resemble the true process of language-learning, where there is an imbalance in knowledge during horizontal transmission, a lesser amount of complexity was lost during transmission. In Experiment 2b, the editor's corrected "language" was passed to the next producer in the chain, representing a child who, after many years of being corrected by their own parent, becomes a parent, and, in turn, passes their language to the next generation. Due to the higher transmission accuracies, or knowledge, of editors, they were able to compensate for some (though not all) of the producers' losses in complexity. Do these results hold, however, when there is a true working memory imbalance between participants, when the task is completed by adults and *real* children?

Experiment 3a: Child Baseline

While experiments 1 and 2 were meant to mimic child language-learners, it is likely that there are significant differences between a child-like adult (i.e., the producers in Experiments 1b and 2b) and true children. As the goal of this project is to understand the role that children and adults–mirroring caregivers and parents–play in language evolution, we make a comparison between adults and children on the same iterated reproduction task. Although past iterated learning studies are meant to mirror language-learners, few use children as participants (Kempe et al., 2015; Raviv & Arnon, 2018).

This experiment is identical to Experiment 1a, in that it is meant to gauge a baseline for how children perform during an iterated transmission task. Children in this experiment complete the task on iPads, primarily at a science museum in Chicago. iPads have been shown to be effective media for conducting experiments, especially with children, as they are engaging and intuitive to use (Frank, Sugarman, Horowitz, Lewis, & Yurovsky, 2016). Conducting the experiment on iPads also allows us to retain a high degree of comparability between experiments 1a-b and 2a-b, as although adults participated remotely online, and children participated in-person with an experimenter nearby, both groups of participants are completing the same task using technology.

The results of this experiment will inform us of whether iterated learning studies can equate child learners with adult learners–do children respond the same way to stimuli as adults? Do their iterated reproductions show similar trends as adults?

Data Collection (can definitely be cut a lot) UPDATE!! Data collection is in progress for this experiment and will likely be completed by the end of Summer 2019. Therefore, reported results are incomplete and not finalized. Data for experiments 3a and 3b were collected simultaneously from August 2018-June 2019. Participants completed the task at one of three locations: The Museum of Science and Industry (MSI), Chicago; the University of Chicago campus, and a private school in Chicago. The majority of data was collected at the MSI. Data collection at all three locations followed similar procedures. All experimenters were trained by the first author to follow a script, and all experimenters were IRB-approved and trained to collect data at all locations.

At the MSI, experimenters from the Communication and Learning Lab arranged a table next to a popular exhibit for children under ten years old. A sign advertised available studies for the target age range, 6-8 years old. Interested families approached the experimenters at the table, who informed them of the general study procedures (a short, 8-minute iPad task), and obtained written consent. Many families also completed an optional demographics sheet, which included questions about caregiver education levels and the languages the child hears at home. After written consent was provided by a parent or legal guardian, and verbal consent was obtained from the child, children went with one of the experimenters to a nearby bench. The experimenter introduced the child to the task, explaining that they would be playing a memory game. Children were given headphones in order to hear the various audio cues throughout the task. The experimenter aided the child in the first training trial, demonstrating how "stickers" (colored blocks) could be placed or removed by tapping on the screen. Additional guidance was given during the first three practice trials if necessary, for example, if the child did not understand that they needed to place exactly 10 stickers down on each trial, or if they did not understand the reproduction element of the task. After the training and practice trials were complete, the experimenter ceased verbal contact with the child, except if the child wanted to end the study. If the child asked the experimenter a question, or expressed frustration about the difficulty of the task, the experimenter replied, "Just do your best." After completion of the study, children received their choice of one or two stickers as compensation.

For participants who completed the study at the University of Chicago, the procedure was similar. The only differences were that participants were recruited through the UChicago Center for Early Childhood Research database, and scheduled appointments to come into the laboratory. After providing written consent (from parents/legal guardians) and verbal consent (from the child), children were taken into a separate, quiet room. Experimenters were trained to follow the same procedure described above. Participants were compensated with $10 and children received a book or toy for their participation.

At the local private school, consent forms were distributed to a kindergarten class prior to data collection. Those children who returned signed consent forms participated in the study. At the beginning of the school day, experimenters brought small groups of children to a quiet room, where they completed the task following the same procedure above. Children were compensated with stickers.

Participants. Participants consisted of 89 children ages 6-8 ($\mu = 6.92$ years; 54% male). 83 children completed the task at the Museum of Science and Industry, Chicago, and 2 participants completed the task on the University of Chicago campus. A small number of participants' data (n=4) were collected at a private school in Chicago. 29 children were removed from the data set due to failure to complete the task, failure to select ten blocks on all experimental trials, or failure to meet accuracy requirements on two out of three practice trials. These participants were removed from the transmission chains, and their re-creations were not passed to the subsequent participant. This results in a total sample size of 60 participants ($\mu = 6.95$ years; 53% male) which makes up 50% of our goal of 120 participants. The included data is relatively evenly-distributed across generations.

Design and Procedure. The task for this experiment was identical to experiments 1a and 2a. Participants viewed a target grid pattern for ten seconds, followed by a visual mask, followed by a blank grid. Children had to tap on the grid to place colored blocks, while a counter dynamically marked how many (out of ten) were left to be placed. A timer counted down from 60 seconds to let the participant know how much time they had left on a particular trial. Auditory cues were presented throughout the task, to encourage the child ("You're doing great!") and to let them know when they were running out of time. After completing one training and three practice trials, participants completed six experimental trials. Participants were in one of 20 chains, each of which contained 6 generations. Children were compensated with stickers for their participation in this 6-10-minute task.

Analysis. As in previous experiments, the patterns produced by participants were analyzed using a measure of accuracy as well as multiple measures of complexity. Accuracy was measured by the number of blocks (out of ten) placed in exactly the same position as in the target pattern. Our primary measure of complexity was calculated using the Block Decomposition Method (see page X for a detailed explanation of this method). Chunking and edge length were used as additional measures of complexity.

Results

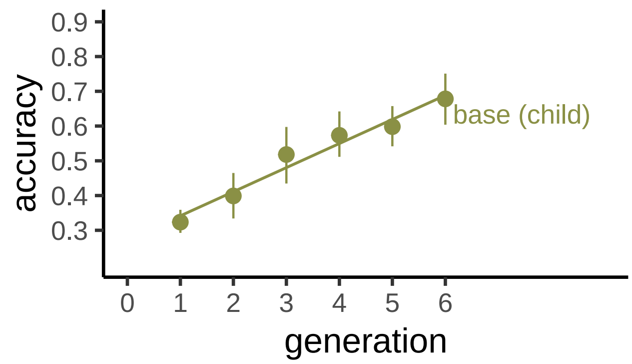

Figure 13. shows the results from Experiment 3a for transmission accuracy. Error bars represent nintey-percent confidence intervals.

For sample patterns produced by children in the study, see Appendix Figure X. As in previous conditions, we fit a linear mixed-effects model to the data, predicting transmission accuracy from log(generation), including random effects from subject and initial grid. Results for accuracy are shown in 13. Transmission accuracies increased significantly over generations ($\beta_{log(generation)}$ = 0.28, t = 10.96, p = $<<$ .001).

14 shows the results for algorithmic complexity. Again, we fit a linear mixed-effects model, this time predicting algorithmic complexity from log(generation). As in previous experiments, the results from algorithmic complexity were in line with the additional measures of complexity. Algorithmic complexity decreased significantly over generations ($\beta_{log(generation)}$ = -26.52, t = -7.10, p = $<<$ .001). Trial number was not a significant predictor in any model.

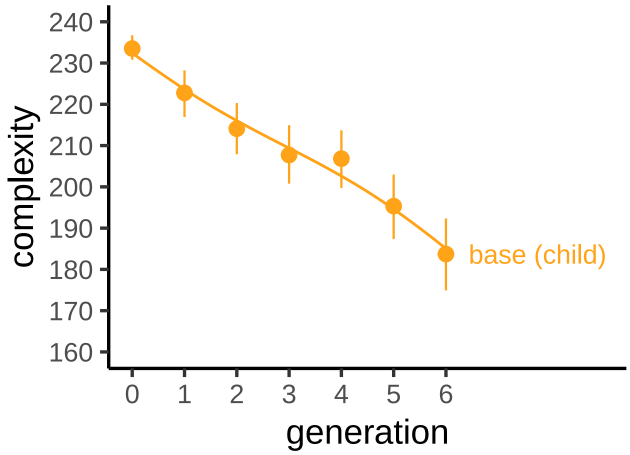

Figure 14. shows the results of Experiment 3a for algorithmic complexity.

Experiment 3b: Adult–Child Dyad

This experiment investigates whether introducing a secondary participant alters the patterned system's evolution. This experiment uses the same paradigm as Experiments 1b and 2b, with one participant designated as a "producer" and another as an "editor". Notably, in this experiment, children are designated as "producers" and adults are designated as "editors". Thus, this condition is the closest step towards to the goal of understanding the impact of caregiver-child language correction in the language evolution process.

Data Collection. Data collection with children was identical to Experiment 3a. Participants were recruited, and data were collected at one of three locations: The Museum of Science and Industry, the University of Chicago, or a local private school.

Participants UPDATE. Participants consisted of 103 children ages 6-8 ($\mu = 6.84$ years; 45% male) and 123 adults. 76 participants completed the task at the Museum of Science and Industry, Chicago, and 18 children completed the task on the University of Chicago campus. A small number of participants' data (n=9) were collected at a private school in Chicago. All adults completed the task online on Amazon Mechanical Turk. 6 children were removed from the dataset due to failure to select ten blocks on all experimental trials, or failure to meet accuracy requirements on two out of three practice trials. 27 adults were removed from the dataset due to failure to select ten blocks on all experimental trials or failure to meet accuracy requirements.

Design and Procedure. Because children were designated as producers in this dyad task, the task procedure was identical to that of Experiment 3a. Adults, who were designated to be "editors", completed the task online on Amazon Mechanical Turk. Therefore, their procedure was identical to that of the editors in Experiments 1b and 2b. As in Experiment 1b, we chose to complete the task with twenty diffusion chains, transmitted over six generations. This choice was made after observing the results of Kempe et al. (2015), and additionally due to time constraints on collecting data with children.

Results

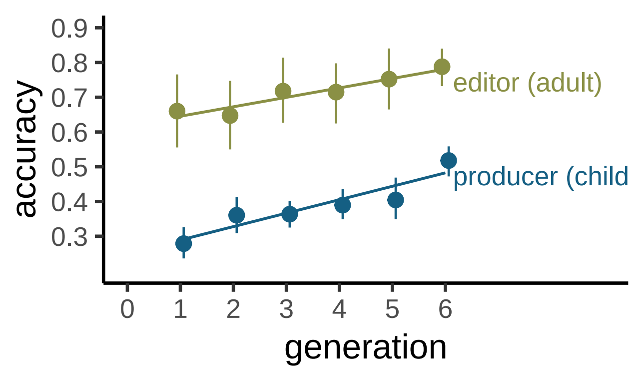

Figure 15. shows the results from Experiment 3b for transmission accuracy. Error bars represent nintey-percent confidence intervals.

15 shows the transmission accuracy results by editors and producers in Experiment 3b. We fit a linear mixed-effects model predicting accuracy from group and log(generation) and trial number and controlling for random effects of subject and initial grid. There was no main effect of generation, meaning that the patterns did not become significantly easier to produce over transmission generations (in line with findings from Exp. 2b) ($\beta_{log(generation)} = 0.10$, t =

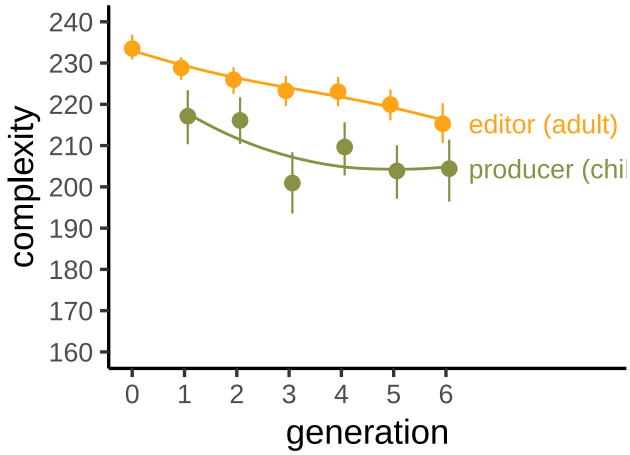

Figure 16. shows the results of Experiment 3b for algorithmic complexity.

3.02, p = .003). There was a significant main effect of condition, with child producers having lower transmission accuracies than adult editors ($\beta_{producers}$ = -0.32, t = -0.32, p < < .001).

16 shows the relationship between the complexity of adult editors' and child producers' patterns. In each generation, the producer decreased the complexity of the pattern, and the editor was able to compensate for some of this loss by re-introducing complexity. As in previous experiments, we fit a linear mixed-effects model to this data to predict complexity from group, generation and trial number, and random intercepts for participant and initial grid (e.g. complexity - condition + generation + trial + (trial|subject) + (1|initialGrid). Only results from the measure of algorithmic complexity are reported, however, trends were consistent across edge length and chunking. There was a marginally-significant main effect of log(generation), with patterns decreasing over transmissions ($\beta_{log(generation)}$ = -9.37, t = -2.67, p = .008). There was also a marginally-significant difference between the algorithmic complexities of producers and editors ($\beta_{producer}$ = -16.36, t = -2.56, p =.011). There were no significant effects of trial number in either complexity or accuracy models.

Experiment 3 Results

```
## Warning: Removed 1 rows containing missing values (geom_pointrange).
```

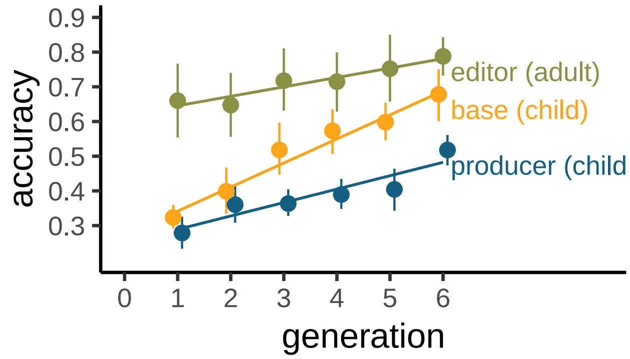

Figure 17. shows the combined results from Experiment 3 for transmission accuracy. Error bars represent nintey-percent confidence intervals.

17 shows the combined accuracy results of Experiments 3a and 3b. We fit a linear mixed-effects model predicting condition (child baseline, child producer, or adult editor) from accuracy and generation, including random effects of subject and initial grid. There was a main effect of generation, with all conditions showing increases in transmission accuracies over generations ($\beta_{log(generation)}$ = 0.22, t = 10.60, p < .001). Baseline children and producers also had lower transmission accuracies than editors (β_{child} = -0.02, t = -0.53, p .598). [is there another comparison to put in here??] 18 shows the combined complexity results for Experiments 3a and 3b. As in previous experiments, there was a main effect of generation, where complexity decreased significantly across generations ($\beta_{log(generation)}$ = -18.59, t = -7.40, p < .001). However, the complexity of producers' and baseline children's patterns decreased more over generations compared to editors ($\beta_{child*log(generation)}$ = 8.38, t = 2.19, p .029). There were no significant effects of trial order in either the accuracy or complexity models. [need to print stats?]

Experiment 3 Discussion

The results of Experiments 3a and 3b continue to push the diffusion chain paradigm further towards modeling true processes of language-learning, this time modeling language-learning by children–those who are the best language-learners. In Experiment 3a,

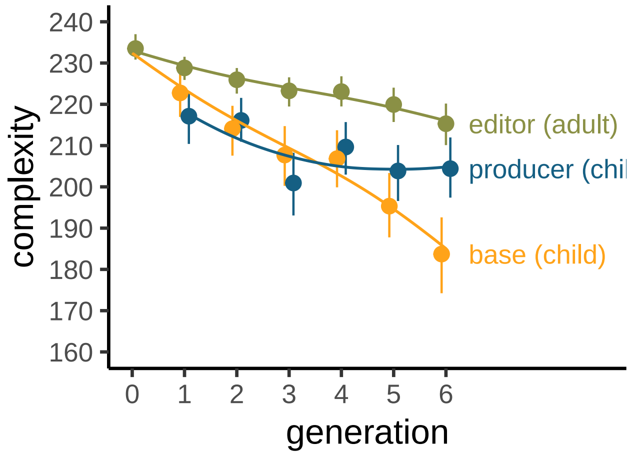

Figure 18. shows the combined results of Experiment 3 for algorithmic complexity.

as in Experiment 1a, we see a dramatic linear decrease in algorithmic complexity over generations, coupled with a linear increase in transmission accuracy. This reflects the effects of the learnability pressure, which, as hypothesized, is especially strong in children. It also replicates the findings of Kempe et al. (2015).

In Experiment 3b, we see similar trends as were expected. Editors (adults) and producers (children) had significantly different reproduction accuracies, with editors being better at re-creating the target, just as they were in Experiment 2b. However, unlike in Experiment 2b or 1b, the reproduction accuracies of producers increased over generations. Thus, children were becoming better at reproducing the grid patterns over time, perhaps pointing to features of the grids which were facilitating easier transmission. Baseline children (Experiment 3a) had significantly higher transmission accuracies than child producers (Exp. 3b). However, the trends shown are somewhat different from those seen in Exps. 1b and 2b, with the child producer and baseline conditions being more similar than the adult baseline and producer conditions. Adult editors had significantly higher levels of complexity compared to baseline children. Thus, the addition of an adult editor–analogous to a parent or caregiver–allowed a significantly higher level of complexity to be retained in the language.

Experiment 2–3 Results [WHAT KIND OF COMPARISONS DO WE WANT TO MAKE HERE? EXP 2 VS 3, OR MORE FINE-GRAINED ADULTS V KIDS? THESE PLOTS ARE A LOT, WHAT TO DISPLAY AND WHAT TO NOT (eg maybe just editors and baseline?)] [LOOK OVER MODEL RESULTS AND WRITING TO MAKE SURE INCLUDE THINGS THAT ARE ACCURATE & IMPORTANT, RN PLUG & CHUG]

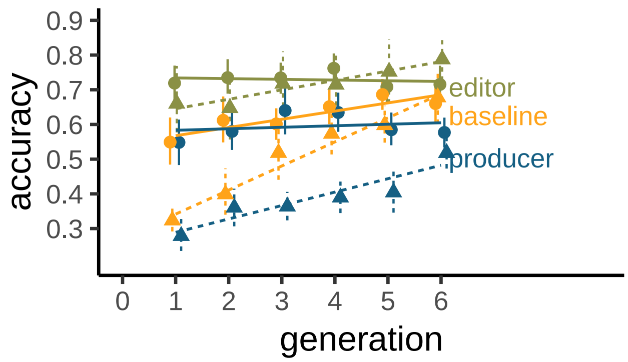

Figure 19. shows the combined results from Experiments 2 and 3 for transmission accuracy. Error bars represent nintey-percent confidence intervals.

In order to compare between Experiments 2 and 3, we subset the data from Experiment 2 to only the first six generations. 19 displays the comparison between transmission accuracies across Experiments 2 and 3. To compare across experiments, we fit a linear mixed effects model of the form *FORMAT ME* lmer(accuracy – person x condition x log(generation+1) + trial + (1|initialGrid) + (trialCount|subject). We found main effects for Experiments 3a & 3b (child–involved experiments), editors, and generation. Baseline children and producers in Experiment 3b had lower percent accuracies (β_{child} = -0.37, t =-8.19 , $p < .001$). Editors (Experiments 2b and 3b) had significantly higher accuracies ($\beta_{editor} = 0.21$, t = 5.62, $p < .001$). Overall, percent accuracies increased across generations ($\beta_{log(generation)} = 0.09$, t = 6.68, $p < .001$).

REWRITE, THIS IS WEIRD AND THERES A LOT OF INTERACTIONS TO LOOK OVER, MIGHT BE DIFF FROM WHAT WRITTEN HERE, DEPENDS WHAT

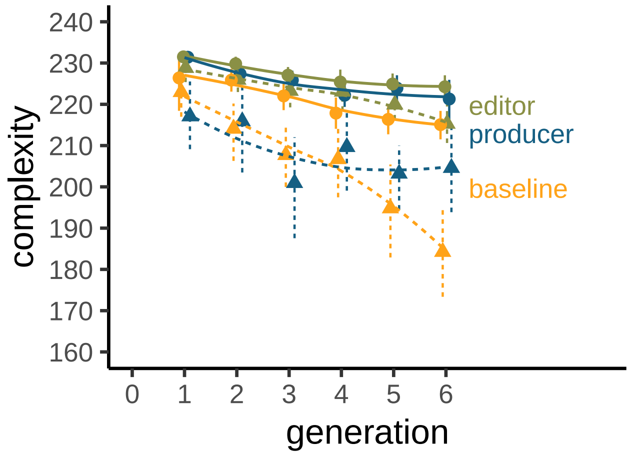

Figure 20. shows the combined results of Experiment 3 for algorithmic complexity.

YOU WANT TO REPORT Additionally, adult editors in Experiment 3b and children in the baseline condition had accuracies which increased more over generations, although the editors' accuracies in both experiments increased less than the baseline conditions over generations ($\beta_{child*log(generation)}$ = 0.18, t = 6.41, p < .001; $\beta_{editor*log(generation)}$ = -0.37, t = -2.978, p = .00298).

20 shows the comparison between Experiments 2 and 3 for algorithmic complexity. As with accuracy, we fit a mixed-effects linear model predicting algorithmic complexity from experiment and condition, fitting effects for display order, initial grid and subject. There was a main effect of generation, with all patterns simplifying over transmissions ($\beta_{log(generation)}$ = -8.97, t = -9.29, p < .001). There was no main effect of whether the participant was an editor (Experiments 2b, 3b) or whether they were in the baseline condition (NEED REPORT?)(β_{editor} = 2.472, t = 0.724, p= 0.469). There were multiple interaction effects, whereby the editors in Experiment 3b (child-adult dyad) had INSERT HERE. While participants in Experiments 3a and 3b had complexity values which decreased more over transmission generations compared with Experiments 2a and 2b ($\beta_{child*log(generation)}$ = -16.35, t = -6.58, p < .001), the editors in Experiment 3b (child-adult dyad) decreased significantly less over generations compared to the children in the baseline condition

($\beta_{child*editor*log(generation)}$ = 10.49, t = 2.76, p .006).

General Discussion

In Experiments 1a and 2a, patterns produced by adults over transmission generations in an iterated reproduction task simplified rapidly and dramatically, reflecting the strong transmissibility pressure in memory-based tasks related to early language learning. With children ages 6-8, we see a similarly rapid, dramatic, and linear decrease in complexity of patterned systems over generations. These findings replicated those of Kempe et al. (2015): when transmitting an artificial patterned system, complexity was lost.

Editors in Experiment 1b, 2b, and 3b represented caregivers – they were more accurate at reproducing the grid patterns and could therefore be seen as more fluent speakers of the "language". The producers, on the other hand, had a more difficult task, which greater strained their working memories, similar to the strain on a child language producer who is exposed to many new words each day. Indeed, when there were real children introduced into the iterated reproduction process, we see that editors still preserve and re-introduce complexity into the patterned systems. In fact, adult editors are able to re-introduce complexity to the level that adults attained on their own. Therefore, it seems that adults are able to compensate for children's losses in complexity. Notably, producers in the child-adult dyad condition showed increases in transmission accuracies, unlike the editors in the same condition. Although the grid patterns were continuing to simplify, the patterns appeared to be evolving to be easier to transmit for children, but not for adults.

Despite the use of a non-linguistic task, we were able to measure change in a culturally-transmitted, reproduced symbol system. When an element of horizontal transmission more closely resembling the relationship between caregivers and children is introduced into the typical diffusion chain paradigm, a greater level of complexity is retained in an evolving "language", as adults are able to re-introduce and protect against oversimplification.

General Discussion (second one; integrate)

In a number of iterated reproduction studies with both children and adults, we show the impact of introducing an element of horizontal transmission into the iterated reproduction paradigm. Additionally, we show similar, yet slightly different findings for adults and children, pointing to the importance of involving those who are the most frequent–and best–language-learners in studies meant to model language evolution.

Both adults and children show similar trends in baseline tasks. Experiments 1a, 2a, and 3a, describe how both children and adults show increases in the learnability and decreases in the expressiveness of randomly-generated patterned systems. However, when a secondary participant is introduced, it is pivotal whether they are a child or a child-like adult. Results show that, in child-adult dyads (Exp. 3b), adults are able to reintroduce complexity to the point where adults were on their own (to the level of Exp. 2a). Editors in this task were able to prevent the language from oversimplification by children. Similarly, real caregivers do not resort to taking children's simplified utterances literally but infer complexity into their productions. Additionally, we see from Experiment 4 that children and adults may not be able to be equated in iterated reproduction tasks, as they are perhaps not simply more errorful-adults, but they make different types of errors during these reproduction tasks. Thus, children and adults could have different patterned-system priors, which may relate to the different strategies and skills they bring to the early language-learning process.

Additionally, the results of this project suggest that introducing an element of horizontal transmission, namely, error-correction by a more knowledgeable participant, changes the system-transmission process. Yet, error-correction, whether explicit or implicit, is a constant, common part of the language-learning process. Therefore, when we attempt to model language evolution in the laboratory, we should include the relationships and phenomena which are found commonly in transmission.

Although the system of grid patterns transmitted and reproduced in this study is quite different, and quite abstracted from the language-learning process, we were able to successfully collect data with young children. The task was engaging, and it allowed us to manipulate the task difficulty level for comparison between adults and children. Many attempts to study language evolution with children in the past have failed, as the tasks used are either too simple or are too difficult for kids (Raviv & Arnon, 2018). However, it is important and necessary, in order to study language evolution, to study those who are most often learning and changing language (Senghas, 2003).

This study is a first, reliable step towards future studies implementing both horizontal and vertical transmission between children and adults in the patterned-system reproduction process. Overall, the results suggest that when a caregiver prevents their child from growing up to believe that "baba" is the word for "bottle", they are not only helping their individual child become a competent speaker of the language, but they are also helping the language system as a whole from oversimplifying.

We do not learn language as passive listeners, who absorb a proportion of the linguistic input they hear. Therefore, we cannot measure language learning only through measuring input, nor through measuring only linguistic output. Language is both learned and changed through conversation to evolve to the needs of its users. Therefore, in order to understand how language adapts to and evolves with communicative interactions, we should study language learning in process.

All experiments were pre-registered on Open Science Framework, and all data and code will be made available through GitHub after de-anonymization.

Acknowledgements

This work was funded by a James McDonnell Foundation Scholar Award to DY as well as an Earl R. Franklin Fellowship Award and a PRISM Research Grant to MM at the University of Chicago.

References

Bowerman, M. (1982). U shaped behavioral growth. In (pp. 101–145). Academic Press.

Calidière, N., Smith, K., Kirby, S., & Fagot, J. (2014). Cultural evolution of systematically structured behaviour in a non-human primate. *Proceedings of the Royal Society B*, *281*.

Chouinard, M. M., & Clark, E. V. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child Language*, *30*(3), 637–669.

Christiansen, M. H., & Kirby, S. (2003). Language evolution. In M. H. Christiansen & S. Kirby (Eds.), (pp. 1–15). Oxford University Press.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, *84*(5), 795 **.

Feldman, J. (2006). An algebra of human concept learning. *Journal of Mathematical Psychology*, *50*(4), 339–368.

Frank, M. C., Sugarman, E., Horowitz, A. C., Lewis, M. L., & Yurovsky, D. (2016). Using tablets to collect data from young children. *Journal of Cognition and Development*, *17*(1), 1–17.

Gauvrit, N., Soler-Toscano, F., & Guida, A. (2017). A preference for some types of complexity comment on "perceived beauty of random texture patterns: A preference for complexity". *Acta Psychologica*, *174*, 48–53.

Henrich, J. (2004). Demography and cultural evolution: How adaptive cultural processes can produce maladaptive losses: The tasmanian case. *American Antiquity*, *69*(2), 197–214.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in languagae formation and change. *Language Learning and Development*, *1*(2), 151–195.

Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin & Review*, *14*(2), 288–294.

Kempe, V., Gauvrit, N., & Forsyth, D. (2015). Structure emerges faster during cultural transmission in children than in adults. *Cognition*, *136*, 247–254.

Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, *104*(12), 5241–5245.

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28*, 108–114.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Lustigman, L., & Clark, E. V. (2019). Exposure and feedback in language acquisition: Adult construals of children's early verb-form use in hebrew. *Journal of Child Language*, *46*(2), 241–264.

Micklos, A., Macuch Silva, V., & Fay, N. (2018). The prevalence of repair in studies of language evolution. *PsychXArv******, ******(***), ***.

Penner, S. G. (1987). Parental responses to grammatical and ungrammatical child utterances. *Child Development*, *58*(2), 376–384.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*(2), 283–290.

Raviv, L., & Arnon, I. (2018). Systematicity, but not compositionality: Examining the emergence of linguistic structure in children and adults using iterated learning. *Cognition*, *181*, 160–173.

Romberg, A. R., & Saffran, J. (2010). Statistical learning and language acquisition. *WIREs Cognitive Science*, *1*, 906–914.

Senghas, A. (2003). Intergenerational influence and ontogenetic development in the emergence of spatial grammar in nicaraguan sign language. *Cognitive Development*, *18*, 511–531.

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*, 444–449.

Yurovsky, D., Yu, C., & Smith, L. B. (2012). Statistical speech segmentation and word learning in parallel: Scaffolding from child-directed speech. *Frontiers in Psychology*, *3*, 374*****.

Zenil, H., Soler-Toscano, F., Dingle, K., & Louis, A. A. (2014). Correlation of automorphism group size and topolical properties with program-size complexity evaluations of graphs and complex networks. *Physica A*, *404*, 341–358.