Caregiver reconstruction of children's errors preserves complexity in patterned systems

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

[1] Stanford University/University of Chicago

[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@uchicago.edu

¹¹   Abstract

¹²   Why do languages change? They may evolve in response to two competing pressures: (1) to

¹³ be easily learned, and (2) to be effective for expressive communication. In a number of

¹⁴ domains, variation in the world's natural languages appears to be accounted for by

¹⁵ near–optimal tradeoffs between these two pressures. Models of these evolutionary processes

¹⁶ have used transmission chain paradigms in which errors of learning by one agent become the

¹⁷ language input for the subsequent generation. However, a critical feature of human language

¹⁸ is that subsequent generations do not learn in isolation. Rather, they are children who learn

¹⁹ in communicative interactions with caregivers. These caregivers consistently draw inferences

²⁰ from children's errorful productions to their intended meanings. In a set of iterated

²¹ reproduction experiments with both children and adults, we show that this supportive

²² context can have a powerful stabilizing role in the development of patterned systems. While

²³ the systems achieve higher levels of complexity than they would by isolated transmission

²⁴ alone, they are equally easy to transmit to the new generation. Thus, caregivers play dual

²⁵ roles as both teachers and protectors of patterned systems as a whole, facilitating their

²⁶ evolution to an optimal balance of learnability and expressiveness.

²⁷   *Keywords:* keywords

²⁸   Word count: X

Caregiver reconstruction of children's errors preserves complexity in patterned systems

The ways we communicate today are not the same as the ways we communicated 2000, 500, or even 200 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 expressive communication (Kirby, Tamariz, Cornish, & Smith, 2015).

While children are often the drivers of language evolution (Senghas, 2003), they differ from adults in their cognitive capabilities (Kempe, Gauvrit, & Forsyth, 2015), interests, early vocabularies, and conversation partners [CITE]. Children may be particularly biased towards simplification because they are early language producers inundated with new information each day (Senghas, 2003). Indeed, when learning language, children often make simplification errors (Bowerman, 1982). This reflects the power of the learnability pressure – children may latch onto word-forms which are 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. In an isolated setting, her error could be sustained and retained in the language over generations. With too many of these simplification errors, 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?

Evidently, children do not learn language in isolation, but they communicate with fluent speakers of the language–their parents and caregivers. Caregivers do a lot of interpretive work, combining their children's productions with their own knowledges of the language as well as their children. As you may have experienced, it is much easier to understand a familiar versus an unfamiliar 2-year-old (Chouinard & Clark, 2003). Caregivers' interpretative skills may be a form of scaffolding for their children's language learning,

54 shouldering some of the complexity of the language until the supports can be removed

55 (Lustigman & Clark, 2018). Our hypothetical child may continue to call a bottle "baba"

56 until she can handle the cognitive load of "bottle", and her caregivers will support this

57 learning in multiple ways.

58   Caregivers support their children's language-learning by providing a space for their

59 children to simplify, as well as by re-introducing complexity into their communications.

60 Adults can explicitly correct their children's language errors in various ways (e.g., by

61 interruptions or repeating the correct word/grammatical form; Penner, 1987). Yet, children

62 primarily learn language through listening to others talk, rather than explicit instruction

63 (Romberg & Saffran, 2010). Parents' modeling of accurate language constructions can have a

64 powerful effect on reducing children's language errors: over time, children self-correct to

65 align to their language models (Hudson Kam & Newport, 2005). When a child grows, they

66 will not transmit the errors they previously had, but rather the correct forms they learned

67 from their caregivers. These caregiver reconstructions may be a mechanism by which more

68 expressiveness is retained in language than children could sustain alone.

69   Paragraph about why you would NEED TO LOSE complexity; i.e., the influence of the

70 expressivity pressure (here, just talking about how parents support not losing any

71 complexity).

Using iterated reproduction to study language change

73   To model the impact of these competing pressures on language evolution in the

74 laboratory, we use a diffusion chain paradigm developed by Kirby, Dowman, and Griffiths

75 (2007). In this paradigm, one participant is trained on a randomly-generated language – e.g,

76 a set of words created by arbitrarily pairing syllables together. The participant is later asked

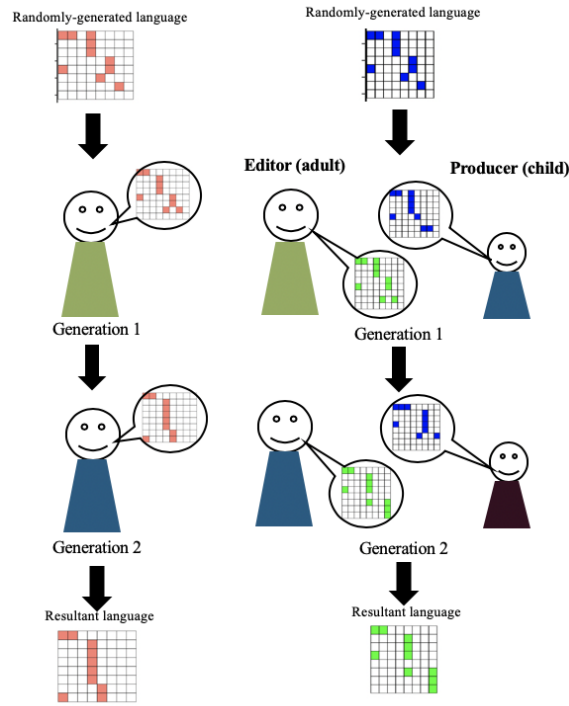77 to recall the language. Their inevitably errorful output becomes the training input for the

Figure 1. Experiments XX follow the conventional diffusion chain paradigm, where a novel language is vertically transmitted through successive recall. In Experiments XX, an element of horizontal transmission is added to the paradigm: novel language producers' reproductions are subject to alterations from a secondary participant.

subsequent participant, forming a transmission chain. This iterated learning process models the transmission of language over generations, with each participant unintentionally changing the language. The errors produced by participants reflect their memory or inductive biases as their errors reflect their expectations (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

88 has shown that horizontal interaction between participants, especially repair, increases

89 communicative efficiency. [ADD/EDIT/ELABORATE HERE] However, repair's effects on

90 communicative success (accuracy) are unclear (Micklos, Silva, & Fay, 2018). Participants in

91 horizontal transmission scenarios most commonly have similar levels of knowledge and

92 similar cognitive constraints. This differs from human language learners (children), who

93 learn language in asymmetric knowledge situations, where their parent both knows more

94 language and has an adult cognitive and executive-functioning system (1). We predict that

95 this asymmetry may have a unique role in the evolution of language, allowing it to resist

96 some of the simplifying pressure of learnability through adults' ability to maintain

97 expressiveness while their children develop. [AND USEFULNESS]

98   In this study, we adapted Kempe et al. (2015)'s non-linguistic iterated reproduction

99 paradigm, as it has been used successfully with children (a similar task was used in

100 non-human primates by Calidière, Smith, Kirby, & Fagot, 2014). This paradigm uses a

101 stimulus set of novel grid patterns, which are akin to language in that they are patterned,

102 structured systems. We adapted this task to model the effect of introducing a secondary,

103 error-correcting participant. We hypothesize that these error-correctors (analogous to

104 caregivers and teachers) are facilitators of both individuals' successful language acquisition

105 and also the evolution of language as a whole. Those who correct mistakes and provide

106 feedback may protect against the strong learnability bias in early language producers by

107 re-introducing and preserving expressiveness. [YET, USEFULNESS]

108   All experiments included in this paper were IRB-approved and were preregistered on

109 Open Science Framework. All tasks, pilot and experimental data, and analyses are available

110 on Github (FOOTNOTE; see page XX for links).

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

We began by replicating 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 with children and adults. We first replicated the study with 6 generations (see Github code), and then with 12 generations to ensure reliability and observe trends over transmissions. The following data is from trials spanning 12 generations.

Participants.   Participants were 519 adults recruited on Amazon Mechanical Turk. Approximately 8% (n=39) of participants in this experiment either failed to meet accuracy requirements on practice trials or failed to select the complete number of items on one or more experimental trials and were therefore exculded from analyses. This resulted in a total of 480 total participants. Participants were members of one of forty diffusion chains, each of which had twelve transmission generations. Each participant provided informed consent and was compensated with $0.50 for their participation in this 8-minute task.

Design and Procedure.   After providing consent, participants 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. Next, participants viewed a blank 8x8 grid where they were given one minute to re-create the target grid (see 2). 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. Only after placing 10

¹³⁶ colored blocks (called "stickers" in the experiment) could participants advance to the next

¹³⁷ trial (See Appendix Figures XXXXX for example grids). A timer was displayed on the

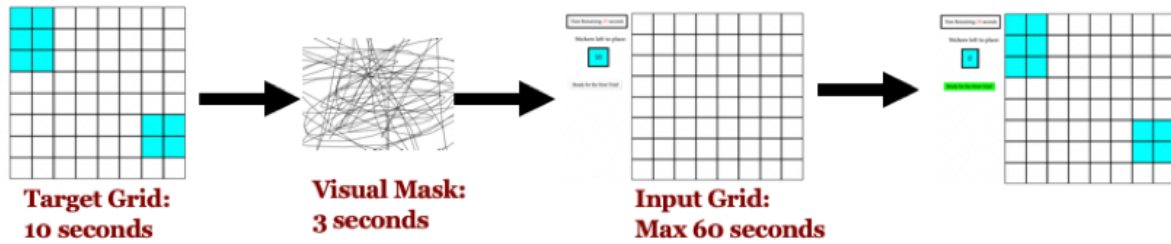¹³⁸ screen, and participants were given an audio cue when they had fifteen seconds left.

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

¹³⁹　　　Participants first completed one training and three practice trials. Each participant

¹⁴⁰ subsequently completed 6 experiment trials, which were presented randomly. During these

¹⁴¹ trials, there was an additional display on the screen which informed the participant of how

¹⁴² many trials they had left to complete. Throughout the trials, participants heard various

¹⁴³ engaging audio cues, including "You're doing great, keep it up!", "You're halfway there!",

¹⁴⁴ and "Just one more to go!". These were added to the task to add an additional level of

¹⁴⁵ engagement for future experiments with children (Experiments XX). Participants in the first

¹⁴⁶ generation of each chain received the same initial grid patterns. These initial 8x8 grids were

¹⁴⁷ generated by randomly selecting 10 of 64 possible cells to be filled using a random-number

¹⁴⁸ generator. Participants in subsequent generations received as their targets the outputs

¹⁴⁹ produced by the previous participant in their chain.

¹⁵⁰　　　Prior to the experimental trials, all participants received the same training and

¹⁵¹ practice trials. In the preliminary training trial, subjects viewed two 8x8 grids side-by-side

¹⁵² and were instructed to make the blank grid on the right match the target grid on the left.

¹⁵³ 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 2;

however, the target patterns were simpler to reproduce. 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 data were excluded from analyses and 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 percentage of targets out of 10 placed in the exact same

location on the target and input grids. This measure of accuracy did not account 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, or how much information was

conveyed. 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 block groups.

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 ensure reliability of the
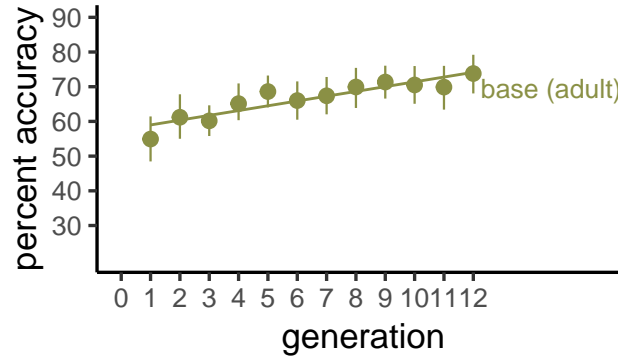
Block Decomposition Method.

¹⁷⁹ **Results**

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

¹⁸⁰   If iterated learning captures the effects of the hypothesized learnability and

¹⁸¹ expressiveness pressures, we predict that reproduction accuracy should increase and

¹⁸² complexity should decrease over generations. BECAUSE WHY? ADD. We tested these

¹⁸³ predictions with mixed–effects logistic regressions, first beginning with the most maximal

¹⁸⁴ model and reducing it by first removing slopes and then removing intercepts until the model

¹⁸⁵ converged. Our final model predicted accuracy and all three measures of complexity

¹⁸⁶ separately from fixed effects of generation and trial number, with random intercepts for

¹⁸⁷ participant and initial grid (e.g. `accuracy` $\sim$ `log(generation) + trial + (1|subject)`

¹⁸⁸ `+ (1|initialGrid)`. Reproduction accuracies increased significantly over generations

¹⁸⁹ (figure 3; $\beta = 0.09$, t = 6.13, p = < .001). Complexity on all three measures (algorithmic

¹⁹⁰ complexity, chunking, and edge length) decreased significantly over generations, as shown in

¹⁹¹ figure 4 (β_{BDM} = -0.04, t = -5.11, p = < .001; $\beta_{chunking}$ = -0.82, t = -13.03, p = < .001;

¹⁹² β_{edge} = -1.57, t = -9.85, p = < .001). Trial number, or how far along the subject was in the

¹⁹³ task, was not a significant predictor in any model. For sample patterns produced by

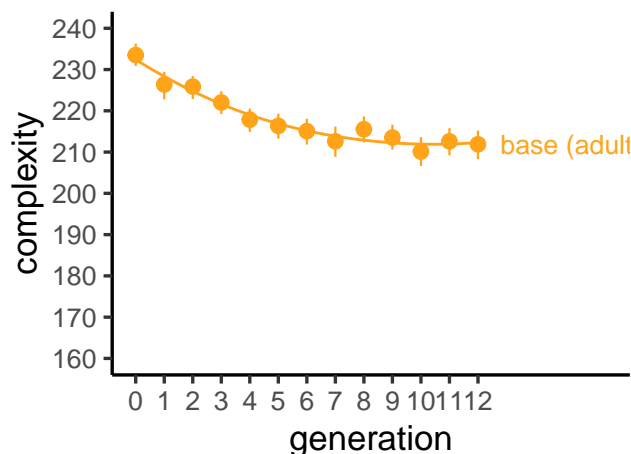¹⁹⁴ participants during the task, see Appendix Figure XXX.

Figure 4. shows the results of Experiment 1a for algorithmic complexity. Example patterns produced by participants at generations 0 (initial, randomly-generated pattern), 6, and 12 (resultant pattern) are shown.

Experiment 1b: Introducing an Editor

As Experiment 1a has shown, without scaffolds, adults are unable to retain the original complexity of a patterned system. Instead, patterns simplify to more easily-transmissible levels of complexity. What happens if scaffolding supports are provided to adults during the transmission process, just as they are in real-life language learning? We focus here on introducing indirect feedback as a scaffold. Just as adults re-introduce complexity by interpreting and silently editing childrens' simplified utterances, we allow a secondary group of participants to edit previous participants' productions.

Participants. Similarly to Experiment 1a, we first ran this experiment with half as many chains and generations, but present a larger sample here (see Github for previous results). Participants in Experiment 1b were 1031 adults recruited on Amazon Mechanical Turk. Approximately 8% (n=71) of participants in Experiment 1b were excluded from analysis due to failure to meet accuracy requirements in 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 analyses. These participants occupied one of forty diffusion

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. A primary participant was designated as a "producer"

and completed the same task as in Experiment 1a (see 2). As before, the "producer"

completed an iterated reproduction task, where they re-created 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 told to edit,

rather than re-create patterns to match a target grid. Editors in this experiment viewed the

same target grid as producers, but were given a prepopulated grid to edit. The grid was

prepopulated with 10 elements placed by the previous producer. 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 this experiment, a full 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 Experiment 1a, our primary measures of analysis were accuracy and

complexity. Transmission accuracy was calculated as the percentage 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 shows the transmission accuracy results from 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
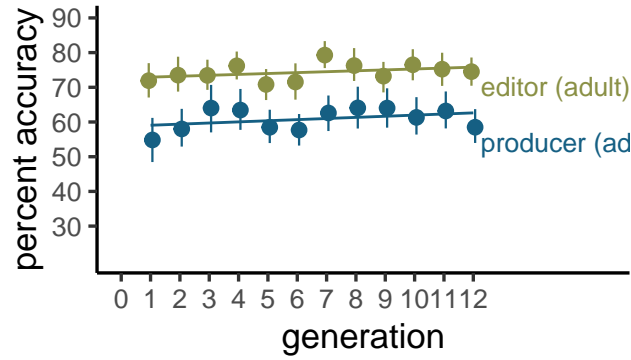
Figure 5. displays the results for transmission accuracies from Experiment 1b. Error bars represent nintey-percent confidence intervals.

that the patterns were not produced more accurately over transmissions ($\beta_{log(generation)} =$ 0.02, t = 1.34, p = .181). There was, however, a significant effect of condition, with producers having lower transmission accuracies compared to editors ($\beta_{producer}$ = -0.14, t = -4.11, p = < .001). There was no effect of trial number.

Figure 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. 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 ($\beta_{log(generation)}$ = -0.02, t = -6.45, p = < .001). Producers had lower complexity values than editors ($\beta_{producer}$ = -0.02, t = -4.59, p =< .001). There was no significant effect of trial number in this model.

Experiment 1 Results

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

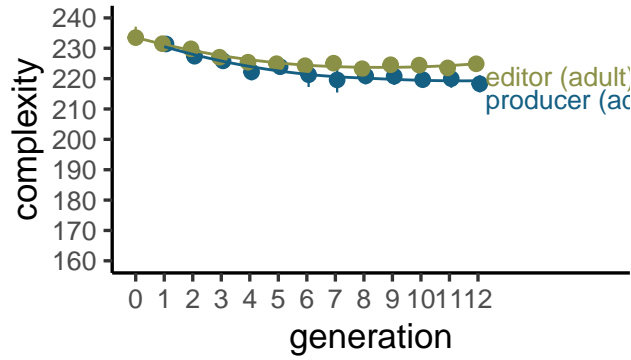Figure 7 shows the combined accuracy results of Experiments 1a and 1b. We fit a

Figure 6. displays the algorithmic complexity results for Experiment 1b. Error bars represent nintey-percent confidence intervals.

250 linear mixed–effects model predicting condition (baseline adult, producer, or editor) from

251 accuracy, log(generation), and trial number, including random effects of subject and initial

252 grid. There were significant main effects of generation and condition, where accuracy

253 increased over generations ($\beta_{log(generation)}$ = 0.09, t = 6.59, p < .001).

254   Both producers and editors had significantly smaller increases in transmission

255 accuracies over generations compared to the baseline adults ($\beta_{producer*log(generation)}$ = -0.06, t

256 = -3.51, p < .001; $\beta_{editor*log(generation)}$ = -0.07, t = -3.64, p < .001). Trial number was not a

257 significant predictor in this model (β_{trial} =0, t = -0.28, p = .777).

258   Figure 8 shows the combined complexity results for Experiments 1a and 1b.

259 Algorithmic complexity of the patterns decreased over generations ($\beta_{log(generation)}$ = -9.08, t

260 = -10.23, p < .001). Notably, editors and producers had significantly higher levels of

261 algorithmic complexity across generations compared to the adult baseline condition from

262 Experiment 1a ($\beta_{producer*log(generation)}$ = 2.79, t = 2.79, p < .005; $\beta_{editor*log(generation)}$ = 4.05, t

263 = 4.05, p < .001). There was an overall effect of trial number, with later trials having higher

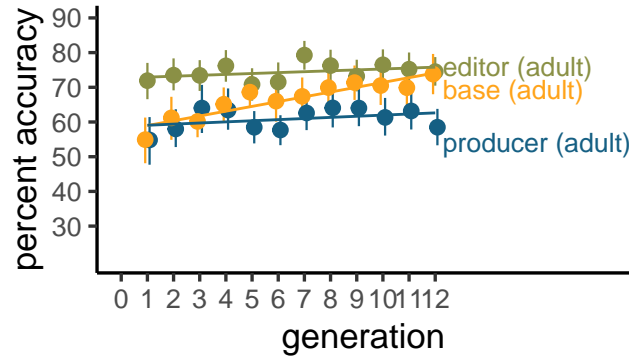264 levels of complexity (β_{trial} = 0.28, t = 2.07, p = .039).

Figure 7. shows the results from experiments 1a and 1b for transmission accuracy. Error bars represent nintey-percent confidence intervals.

Discussion

The results of Experiment 1a 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, while transmission accuracies increases over generations. This shows that generational transmission creates a bottleneck in the evolutionary process, where patterned systems quickly lose expressivity (complexity), but approach a level of learnability which makes them easier to reproduce. Over generations, the rate of simplification decreased [SHOWN WHERE?]. Thus, this system was not losing all of its descriptiveness, instead reaching a balance point between learnability and expressivity.

The results of Experiment 1b attempt to more closely model pattern-system transmissions analogous to language-learning. These experiments introduce a secondary, more knowledgeable participant into the iterated reproduction process. These editing participants had higher transmission accuracies, reflecting their increased working-memory capacities which contributed to greater "knowledge" of the task. In Experiment 1b, editors were not simply completely re-creating the patterns made by "producers", but they were fixing their errors. [WHERE SHOW THIS? ADD TO RESULTS SECTION]

The algorithmic complexity results of Experiment 1b show that the loss in complexity
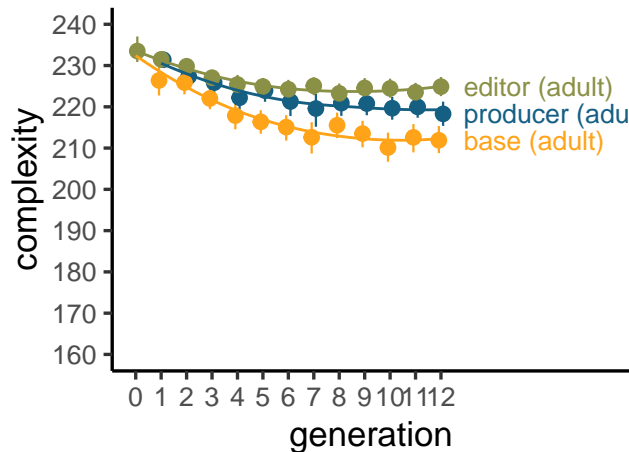
Figure 8. shows the combined results of experiments 1a and 1b for algorithmic complexity.

²⁸² from Experiments 1a is not permanent cultural regression (Henrich, 2004), as complexity can

²⁸³ be reintroduced in the patterned system by way of a secondary participant. Do these results

²⁸⁴ hold, however, when there is a true working memory imbalance between participants,

²⁸⁵ notably when the task is completed by adults and *real* children?

Experiment 3a: Child Baseline

²⁸⁷ Although past iterated learning studies are meant to mirror language-learners, few use

²⁸⁸ children as participants (Kempe et al., 2015; Raviv & Arnon, 2018). It is likely that there

²⁸⁹ are significant differences between a child-like adult (i.e., the producers in Experiment 1b)

²⁹⁰ and true children. The results of this experiment will inform us of whether iterated learning

²⁹¹ studies can equate adult learners with child learners.

²⁹² This experiment is identical to Experiment 1a, in that it is meant to gauge a baseline

²⁹³ for how children perform during an iterated transmission task. Unfortunately, due to data

²⁹⁴ collection restraints, we collected a smaller sample of participants than in Experiment 1a.

²⁹⁵ **Data Collection.** Children in this experiment completed the task on iPads, which

²⁹⁶ have been shown to be effective media for conducting experiments, as they are engaging and

297 intuitive to use (Frank, Sugarman, Horowitz, Lewis, & Yurovsky, 2016). Conducting the

298 experiment on iPads also allows us to retain reliability across conditions, as both adults and

299 children completed the task online using technology. Participants completed the task at one

300 of three locations: a local science museum; a University campus, or a local private school.

301 All experimenters were trained by the first author to follow a script, and were IRB-approved

302 to work with children.

303   At the science museum, experimenters arranged a table next to a popular children's

304 exhibit. Interested families approached the experimenters at the table, who informed them of

305 the general study procedures and obtained written and verbal consent. An experimenter

306 introduced the child to the task, explaining that they would be playing a memory game.

307 Children were given headphones in order to hear audio cues. The experimenter aided the

308 child in the first training trial, demonstrating how "stickers" (colored blocks) could be placed

309 or removed by tapping on the screen. Additional guidance was given during the first three

310 practice trials only if necessary. If the child asked the experimenter a question during the

311 task, the experimenter replied, "Just do your best." After completion of the study, children

312 received their choice of stickers as compensation.

313   The procedure was similar at the University campus. Participants were recruited

314 through an online database and scheduled appointments to come into the laboratory.

315 Children completed the study in a quiet, kid-friendly room. Participants were compensated

316 with $10 and children received a book or toy for their participation.

317   At the local private school, consent forms were distributed to a single class prior to

318 data collection. Those children who returned signed consent forms participated in the study.

319 Children completed the study in a quiet room, and received stickers for their participation.

320 **Participants UPDATEEEEEE MEEEEEE START EDITING HERE.**

321 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 the 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.

Design and Procedure.   The task for this experiment was identical to experiments 1a. Participants were in one of 20 chains, each of which contained 6 generations.

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 percentage 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. Chunking and edge length were used as additional measures of complexity.

Results UPDATEEE NOT UPDATED DATA

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 9. Transmission accuracies increased significantly over generations ($\beta_{log(generation)} = 0.28$, t = 10.96, p = $<< .001$).

10 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
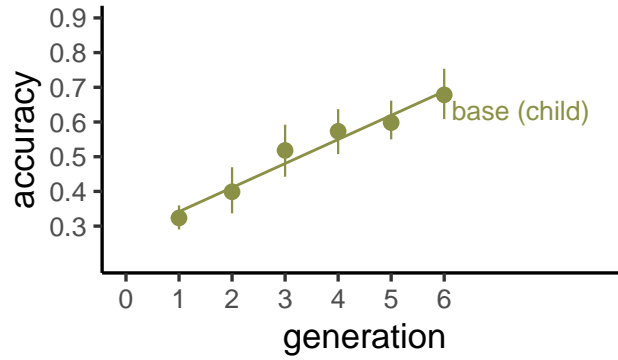
Figure 9. shows the results from Experiment 3a for transmission accuracy. Error bars represent nintey-percent confidence intervals.

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.

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
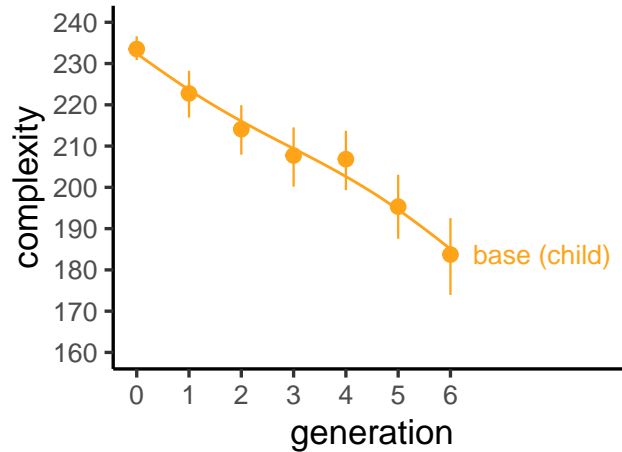
Figure 10. shows the results of Experiment 3a for algorithmic complexity.

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.

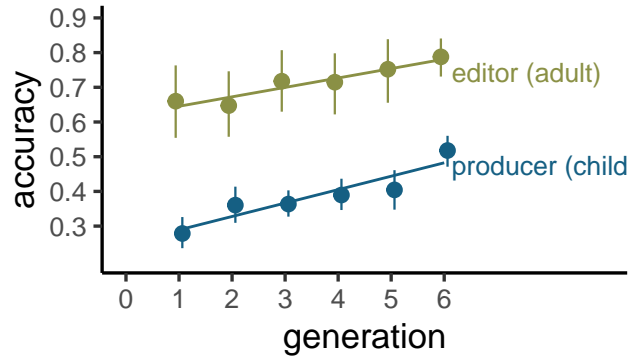(2015), and additionally due to time constraints on collecting data with children.

Figure 11. shows the results from Experiment 3b for transmission accuracy. Error bars represent nintey-percent confidence intervals.

Results

³⁷⁷ 11 shows the transmission accuracy results by editors and producers in Experiment 3b. ³⁷⁸ We fit a linear mixed-effects model predicting accuracy from group and log(generation) and ³⁷⁹ trial number and controlling for random effects of subject and initial grid. There was no main ³⁸⁰ effect of generation, meaning that the patterns did not become significantly easier to produce ³⁸¹ over transmission generations (in line with findings from Exp. 2b) ($\beta_{log(generation)} = 0.10$, t = ³⁸² 3.02, p = .003). There was a significant main effect of condition, with child producers having ³⁸³ lower transmission accuracies than adult editors ($\beta_{producers}$ = -0.32, t = -0.32, p $<<$ .001).

³⁸⁴ 12 shows the relationship between the complexity of adult editors' and child producers' ³⁸⁵ patterns. In each generation, the producer decreased the complexity of the pattern, and the ³⁸⁶ editor was able to compensate for some of this loss by re-introducing complexity. As in ³⁸⁷ previous experiments, we fit a linear mixed-effects model to this data to predict complexity ³⁸⁸ from group, generation and trial number, and random intercepts for participant and initial ³⁸⁹ grid (e.g. complexity - condition + generation + trial + (trial|subject) + (1|initialGrid). ³⁹⁰ Only results from the measure of algorithmic complexity are reported, however, trends were ³⁹¹ consistent across edge length and chunking. There was a marginally-significant main effect of ³⁹² log(generation), with patterns decreasing over transmissions ($\beta_{log(generation)}$ = -9.37, t =
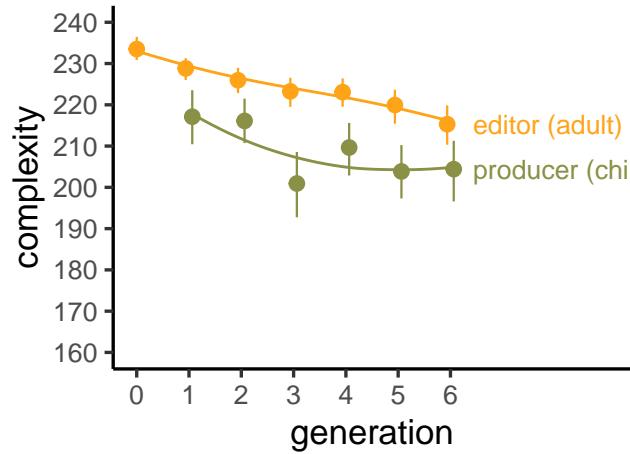
Figure 12. shows the results of Experiment 3b for algorithmic complexity.

-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).

13 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??] 14 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 =
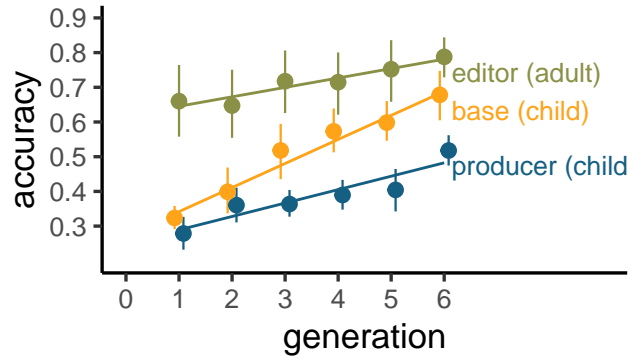
Figure 13. shows the combined results from Experiment 3 for transmission accuracy. Error bars represent nintey-percent confidence intervals.

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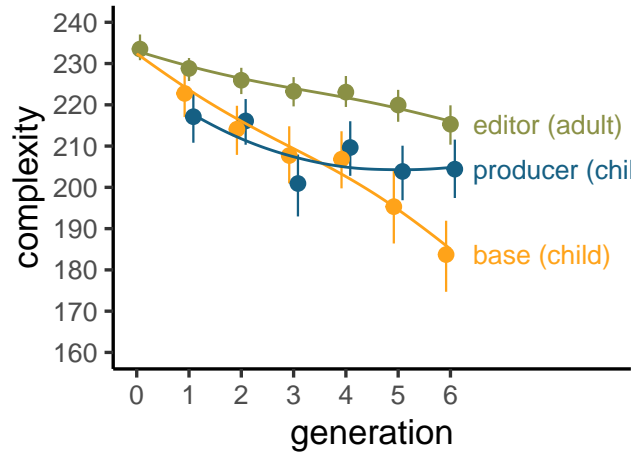
Figure 14. shows the combined results of Experiment 3 for algorithmic complexity.

(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]

In order to compare between Experiments 2 and 3, we subset the data from Experiment 2 to only the first six generations. 15 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
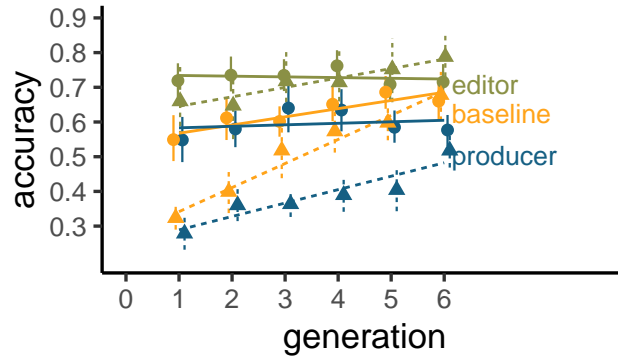
Figure 15. shows the combined results from Experiments 2 and 3 for transmission accuracy. Error bars represent nintey-percent confidence intervals.

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 (β_{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 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).

16 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 ($\beta_{log(generation)}$ =
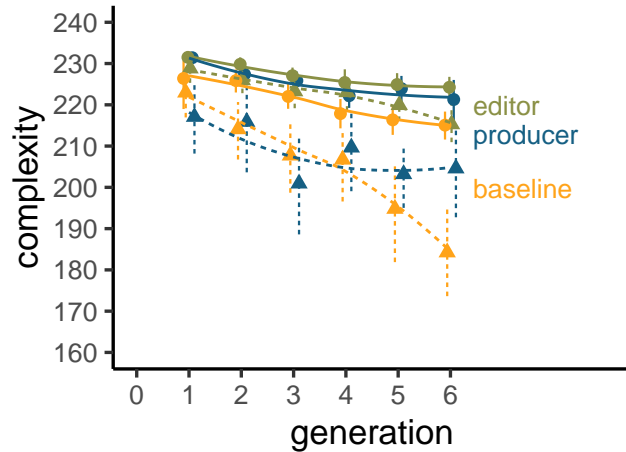
Figure 16. shows the combined results of Experiment 3 for algorithmic complexity.

458 -8.97, t = -9.29, p < .001). There was no main effect of whether the participant was an

459 editor (Experiments 2b, 3b) or whether they were in the baseline condition (NEED

460 REPORT?)(β_{editor} = 2.472, t = 0.724, p= 0.469). There were multiple interaction effects,

461 whereby the editors in Experiment 3b (child-adult dyad) had INSERT HERE. While

462 participants in Experiments 3a and 3b had complexity values which decreased more over

463 transmission generations compared with Experiments 2a and 2b ($\beta_{child*log(generation)}$ = -16.35,

464 t = -6.58, p < .001), the editors in Experiment 3b (child-adult dyad) decreased significantly

465 less over generations compared to the children in the baseline condition

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

General Discussion

468   Due to the higher transmission accuracies, or knowledge, of editors, they were able to

469 compensate for some (though not all) of the producers' losses in complexity

470   In Experiments 1a and 2a, patterns produced by adults over transmission generations

471 in an iterated reproduction task simplified rapidly and dramatically, reflecting the strong

472 transmissibility pressure in memory-based tasks related to early language learning. With

473 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)

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 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

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.

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

569   *Learning and Development, 1*(2), 151–195.

570   Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning:

571   Intergenerational knowledge transmission reveals inductive biases. *Psychonomic*

572   *Bulletin & Review, 14*(2), 288–294.

573   Kempe, V., Gauvrit, N., & Forsyth, D. (2015). Structure emerges faster during cultural

574   transmission in children than in adults. *Cognition, 136*, 247–254.

575   Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of

576   language. *Proceedings of the National Academy of Sciences, 104*(12), 5241–5245.

577   Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language.

578   *Current Opinion in Neurobiology, 28*, 108–114.

579   Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication

580   in the cultural evolution of linguistic structure. *Cognition, 141*, 87–102.

581   Lustigman, L., & Clark, E. V. (2018). Exposure and feedback in language acquisition: Adult

582   construals of children's early verb-form use in hebrew. *Journal of Child Language*,

583   1–24.

584   Micklos, A., Silva, M. V., & Fay, N. (2018). The prevalence of repair in studies of language

585   evolution. *12th International Conference on the Evolution of Language:(EVOLANG*

586   *XII)*, 316–318.

587   Penner, S. G. (1987). Parental responses to grammatical and ungrammatical child

588   utterances. *Child Development, 58*(2), 376–384.

589   Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual

590   memory. *Perception & Psychophysics*, *16*(2), 283–290.

591   Raviv, L., & Arnon, I. (2018). Systematicity, but not compositionality: Examining the

592   emergence of linguistic structure in children and adults using iterated learning.

593   *Cognition*, *181*, 160–173.

594   Romberg, A. R., & Saffran, J. (2010). Statistical learning and language acquisition. *WIREs*

595   *Cognitive Science*, *1*, 906–914.

596   Senghas, A. (2003). Intergenerational influence and ontogenetic development in the

597   emergence of spatial grammar in nicaraguan sign language. *Cognitive Development*,

598   *18*, 511–531.

599   Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated

600   learning. *Cognition*, *116*, 444–449.

601   Zenil, H., Soler-Toscano, F., Dingle, K., & Louis, A. A. (2014). Correlation of automorphism

602   group size and topical properties with program-size complexity evaluations of

603   graphs and complex networks. *Physica A*, *404*, 341–358.