

Caregiver reconstruction of children's errors: the preservation of complexity in language (improve title)

Anonymous CogSci submission

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 (e.g. kinship categories, color terms), variation in the world's natural languages appears to be accounted for by different but near-optimal tradeoffs between these two pressures (Regier, Kemp, & Kay, 2015). Models of this evolutionary process have focused on transmission chains in which errors of learning by one agent become the language input for the next generation. However, a critical feature of human language is that children do not learn in isolation, but rather in communicative interactions with caregivers who can draw inferences from a child's errorful productions to their intended interests. In a set of iterated learning experiments, we show that this supportive context can have a powerful stabilizing role in the development of artificial languages, allowing them achieve higher levels of asymptotic complexity than they would by transmission alone.

Keywords: communication; language acquisition; language evolution; iterated learning

Introduction

How do you ask a group of people where they are going in Spanish? In Spain, the answer depends on the group: you might ask "¿Dónde van ustedes?" of a group of work colleagues, but to address your friends, you use the informal "¿Dónde vais vosotros?" instead. In Mexican Spanish, this distinction has disappeared, and the "ustedes" form is used exclusively. Why did Spanish change in this way, simplifying and shedding the formal second person plural? One working theory is that languages evolve to adapt to two dynamic competing pressures: (1) ease of learning and transmission, and (2) effective communication (Lupyan & Dale, 2010).

When children are learning language, they often make simplification errors (Bowerman, 1982). If a child is asking for some milk, they may ask for a "baba" (bottle). Thus, the child has coined a new word, which her parents understand. This simplification works well for the child, until she starts to learn her animal words, and calls a sheep "baba". The child language-learner has shown the effects of the simplicity pressure in language: if she calls both bottles and sheep "baba", she has to learn one less word. Does she go through her life, believing that "baba" is the label for these disparate objects? It's possible, if she never has a need to discriminate between them, she might even pass this label on to her children. In this way, errors in language can be passed on to the next generation, through transmission from one speaker to the next.

It reflects the needs of the language speakers: if our hypothetical language learner grew up in a world where there was no need to distinguish between bottles and sheep, there is no need to retain this extra complexity (and extra memory load).

Children are often the actors who drive language evolution (Senghas, 2003), yet they differ from adults in their cognitive capabilities, namely, memory systems (Kempe, Gauvrit, & Forsyth, 2015), interests and early vocabularies, and conversation partners. Therefore, though children are skilled language learners, their developing cognitive systems prevent aspects of language that are difficult to learn and remember from being passed on—pushing languages towards simplicity (Hudson Kam & Newport, 2005; Senghas, 2003). But, languages that become too simple can lose the ability to be effective for communication (Kirby, Griffiths, & Smith, 2014). What enables languages to retain their communicative utility in the face of these learnability pressures?

Indeed, most Americans grow up in a world where it is useful to discriminate sheep and bottles – and they must learn the different labels for these objects. It is often caregivers, through their explicit interventions as well as their implicit modeling of correct language, who may be reintroducing descriptiveness into a language where it would otherwise be lost. Children's language learning is greatly influenced by those around them – especially the adults they talk to most. These caregivers control much of their child's linguistic input, and are responsible for seeing that their children develop effective and useful systems of communication. Even the youngest children are not passive learners of language — they are active participants, engaging in conversations with their parents. These adults are experts both in the language and in the children themselves, as they understand the child's intuitions, personality, and context. Caregivers play an important interpretive role in these interactions through their ability to understand the intended target of children's errorful productions (Chouinard & Clark, 2003). 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, parent's 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 learned the 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 grows to be an adult, they will not transmit the errors they had as a child, but the correct forms of speech 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 structure is retained in language over many lifetimes than children could sustain alone.

Using iterated learning to study language change

In order to study the effects of the descriptiveness and transmissibility biases on inter-generational language evolution, we will be using the iterated learning paradigm (Kirby et al., 2014). In an iterated learning paradigm, one participant is trained on a randomly-generated language—for example, a set of words created by arbitrarily pairing syllables together. The participant is later asked to recall the language, and their responses are given as training input for the next subject, thus creating a transmission chain. This iterated process mimics the transmission of language across generations, with each participant unintentionally changing the language through their memory biases. The vast majority of iterated learning studies have used adult participants (Christiansen & Kirby, 2003; Kirby, Dowman, & Griffiths, 2007; Kirby et al., 2014; e.g., Smith & Wonnacott, 2010; structure & signals, 2014), while few studies have used children as research subjects (Kempe et al., 2015; Raviv & Arnon, 2018). However, language evolution cannot be fully grasped using this paradigm with only separate adult or child learning chains, because language learning does not occur only within the same age group (horizontal transmission), or only across age groups (vertical transmission), but it occurs dynamically, in both directions. In a true language-acquisition environment, a child receives both language input and feedback from their caregiver and uses it to interact with their peers throughout life, eventually growing into a new teacher-caregiver.

We adapted Kempe et al. (2015)'s child-friendly language paradigm to model the effect of a cooperative caregiver on the evolution of language. We hypothesize that these error-correctors (analogous to caregivers and teachers) are pivotal not only to an individual's successful language acquisition, but also to the evolution of a language as a whole. This is because those who correct mistakes and provide feedback are able to protect against the transmissibility (simplicity) bias, which is likely stronger in early language learners, by reintroducing and preserving complexity in language.

All experiments were pre-registered on Open Science Framework, and all data and code can be accessed at the following link: <https://osf.io/guzyf/>.

Experiment 1: Replicating Kempe et al. (2015)

In a baseline experiment, adults participated in a standard baseline iterated learning study, using stimuli adapted from Kempe et al. (2015). Participants were told to reproduce patterns on grids, and each user's responses were used as training input for the subsequent participant.

Method

Participants Participants in Experiment 1 were 120 adults recruited on Amazon Mechanical Turk. These participants were divided into 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 **I THINK YOU MIGHT WANT TO DESCRIBE GENERAL TRIALS INSTEAD, ACTUALLY. AND THEN SAY WE STARTED WITH 3 PRACTICE TRIALS** Participants in Experiment 1 were asked to re-create patterns on a grid. Participants provided consent. After a consent screen, participants first viewed a training trial with two 8x8 grids on the screen – a target grid, with 10 cells colored in, and a blank grid. They were told to make the blank grid match the target grid exactly, and were unable to progress until the grids were identical. Following this trial, participants were informed that they would see a target grid appear on the screen for 10 seconds, followed by a picture (a visual mask) displayed for 3 seconds. After the visual mask, participants viewed a blank 8x8 grid where they were given 60 seconds to re-create the target grid. Participants could click on any cell in the grid to change its color, and could also undo any color placed. A counter on the screen showed how many targets had been colored, and it varied dynamically with the participant's clicks. After placing 10 targets, participants could click a button to advance to the next trial. After completing 3 Practice trials, which were identical for all participants in all chains, participants were informed that the study would begin.

Each participant then completed 6 Experiment trials. Participants in the first generation of each chain received the same initial grids for Experiment trials. These initial 8x8 grids were generated by randomly selecting 10 of the 64 possible cells to be filled. Participants in subsequent chains received as their targets the outputs produced by their parent in the chain.

Participants' performance on the practice trials were used as an attention check to determine whether their data would be passed to the next participant. If the participant scored less than 75% accuracy on the last 2 practice trials, or if they failed to select 10 cells before time ran out, their outputs were not transmitted to the next generation. In Experiment 1, five participants failed to meet these criteria and were excluded from analysis.

Results and Analysis

Our primary measures of interest were reproduction accuracy and pattern complexity. Reproduction accuracy served as a proxy for transmissibility – higher reproduction accuracies indicate that the “language” is easier to learn. Reproduction accuracy was computed as the proportion of targets (out of 10) which were placed in the same location on the target and input grids.

Complexity served as a proxy for descriptiveness. We followed Kempe et al. (2015) in using several measures of complexity: algorithmic complexity, chunking, and edge length. Algorithmic complexity is calculated using the Block Decomposition Method, a measure of Kolmogorov-Chaitin Complexity applied to 2-dimensional patterns (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 the complexity is. Edge length is the total perimeter of the colored blocks. If all blocks were in one chunk, the edge length would be low, and the complexity of the pattern would likely be lower compared to if none of the chosen targets shared an edge. Implementation of these metrics was adapted from code provided by Gauvrit, Soler-Toscano, & Guida (2017).

Results

If iterated learning captures the hypothesized pressures of expressiveness and transmissibility, we predict that over generations reproduction accuracy should increase and complexity should decrease. We tested these predictions with mixed-effects logistic regressions, predicting 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. $\text{accuracy} \sim \text{generation} + \text{trial} + (1|\text{subject}) + (1|\text{initialGrid})$).

Reproduction accuracy increased significantly over generations ($\beta = 0.033$, $t = 3.146$, $p = .002$). Figure ?? shows the results for accuracy. Complexity on all three measure decreased significantly over generations ($\beta_{BDM} = 0.033$, $t = 3.146$, $p = .002$; $\beta_{chunking} = 0.033$, $t = 3.146$, $p = .002$; $\beta_{edge} = 0.033$, $t = 3.146$, $p = .002$). Trial Number was not a significant predictor in any model. Figure ?? shows the results for algorithmic complexity.

Experiment 2: Replication and extension of Experiment 1

Experiment 2, replicated the task from Experiment 1 with the addition of twice as many chains and generations. We replicated the task with a larger sample in order to approximate the shape of the algorithmic complexity curve. Particularly, we were interested in whether complexity asymptoted over generations.

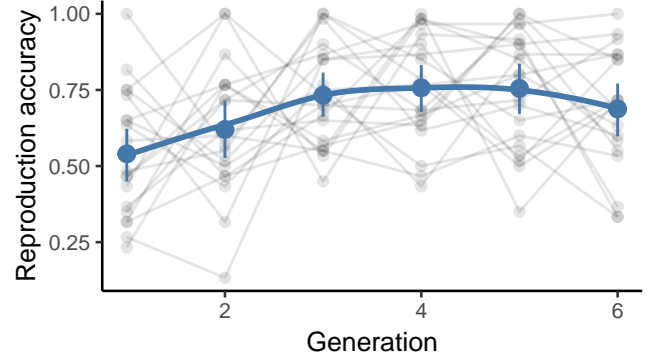


Figure 1: Experiments 1 and 2 show increases in accuracy over transmission generations. CHANGE POINT SIZE

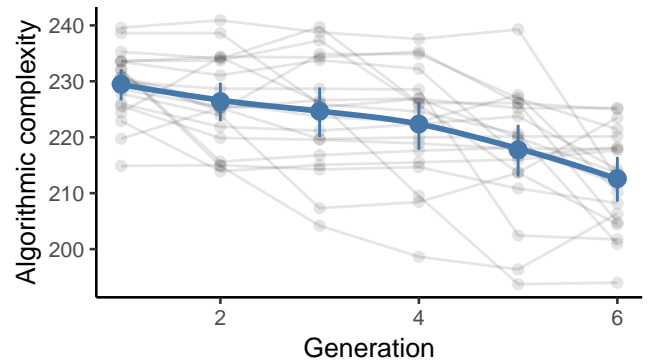


Figure 2: Experiments 1 and 2 show increases in accuracy over transmission generations. CHANGE POINT SIZE

Method

Participants

Participants in Experiment 2 were 519 adults recruited on Amazon Mechanical Turk. These participants were divided into forty diffusion chains, each of which had twelve generations. Each participant gave informed consent, and was compensated with \$0.50 for their participation.

Design and Procedure

The task in Experiment 2 was identical to Experiment 1. Participants were told to reproduce patterns on a grid, and their responses were passed to the next subject in the transmission chain.

Approximately 8% ($n=39$) of participants in Experiment 2 were excluded from analysis due to failure to meet accuracy requirements on the practice trials or failure to select the complete number of targets on one or more experimental trials. This resulted in a total of 480 participants included in the analysis.

Results

The results of this experiment replicated those found in Experiment 1. Reproduction accuracy increased significantly over generations ($\beta = 0.061$, $t = 6.07$, $p = < .001$). Figure ?? shows the results for accuracy.

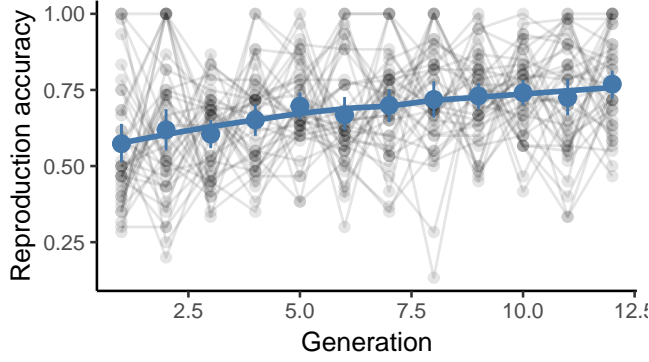


Figure 3: Experiments 1 and 2 show increases in accuracy over transmission generations. CHANGE POINT SIZE

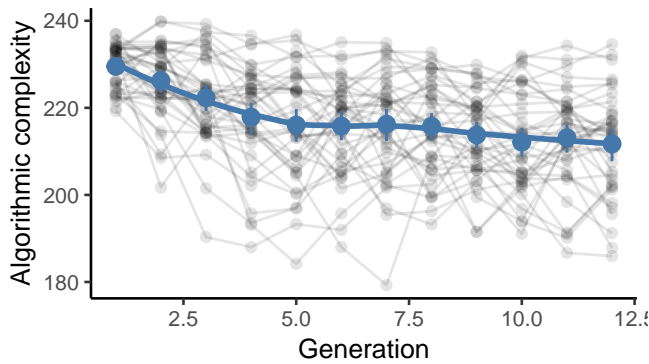


Figure 4: Experiments 1 and 2 show increases in accuracy over transmission generations. CHANGE POINT SIZE

Figure ?? shows the results for algorithmic complexity. Algorithmic complexity appeared to follow (W.C.??) an exponential function of the form $y = e^{-x} + b$. We therefore fit an exponential mixed-effects regression model predicting complexity from fixed effects of generation and trial number, and random intercepts for participant, chain, and initial grid (e.g. $\log(\text{complexity}) \sim \text{generation} + \text{trial} + (1|\text{subject}) + (1|\text{initial}) + (1|\text{chain})$). Algorithmic complexity decreased and asymptoted over generations ($\beta_{BDM} = -0.046$, $t = -12.368$, $p < .001$). Similar trends were also found with chunking and edge length, the alternate measures of complexity (CHECK THESE MODELS B/C THE LOG EXPONENTIAL DOESN'T WORK $\beta_{\text{chunking}} = -0.881$, $t = -17.286$, $p < .001$; $\beta_{\text{edge}} = -1.76$, $t = -15.766$, $p < .001$).

Experiment 3: Introducing an interlocutor

In order to add an element of feedback from a more experienced interlocutor to the iterated-learning process, we adapted the task from Experiments 1 and 2 to include a secondary, “editing” participant. This participant was analogous to a caregiver who protects their child from learning incorrect forms of language.

Method

Participants

Participants in Experiment 3 were 1031 adults recruited on Amazon Mechanical Turk. These participants were divided into forty diffusion chains, each of which had twelve generations. Each participant gave informed consent, and was compensated with \$0.50 for their participation.

Design and Procedure

In the third, dyad experiment, a primary participant was designated to be a “learner”, and completed the same task as in Experiment 1 and Experiment 2. They were told to reproduce patterns on a grid. A secondary participant – the “fixer” – was given an adapted task. Throughout the study, fixers were not told to re-create patterns, but to fix patterns to resemble a target grid exactly. Fixers in this experiment viewed the same target grid as learners, but instead of seeing an empty input grid, they saw a grid with 10 elements filled in – the elements that the previous learner had submitted. The participant could then edit the 10 items’ positions. There was no “reset” button during this task, so produced patterns reflect participants’ initial instincts.

In Experiment 3, a generation consisted of a learner, who re-created the target grid, and a fixer, who then received the same target grid as well as the learner’s input grid to edit. The fixer’s edited pattern was used as the target grid for the subsequent generation.

Approximately 8% ($n=71$) of participants in Experiment 3 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.

Analysis and Results

As in Experiments 1 and 2, our primary measures of analysis were accuracy and complexity. These measures were computed using the same methods as in the previous experiments.

Fixers and learners had significantly different pattern reproduction accuracies. According to a linear mixed-effects model (need to put in formula? Did an lmer with gen and condition (learner/fixer) as fixed effects, with all the same random effects), the accuracies between groups were significantly different ($\beta_{\text{condition-child}} = -0.087$, $t = -11.38$, $p < .001$). The fixers’ transmission accuracies did not increase significantly over generations ($\beta_{\text{fixers}} = 0.01$, $t = 1.135$, $p = .257$), while the accuracy of the learners showed a marginally significant increase ($\beta_{\text{learners}} = 0.014$, $t = 1.511$, $p = .132$).

6 shows the relationship between the complexity of fixers’ and learners’ patterns. In each generation, the learner decreases the complexity of the pattern, and the fixer is able to compensate for some of this loss. AS in Experiment 2, we fit an exponential model to the data. Both conditions show decreases in pattern complexity over generations ($\beta_{\text{learners}} = -0.029$, $t = -4.498$, $p < .001$; $\beta_{\text{fixers}} = -0.018$, $t = -6.831$,

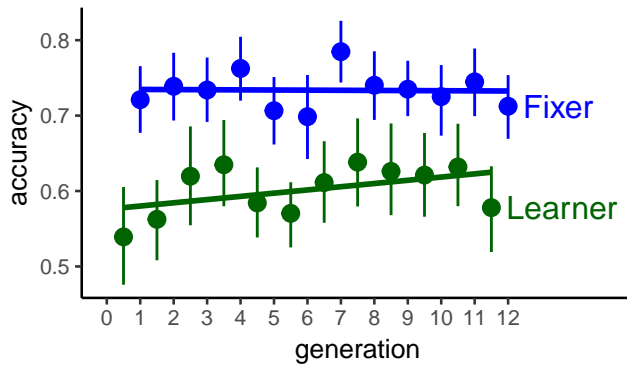


Figure 5: In the dyad task, reproduction accuracy stays relatively constant across generations. Fixers have significantly higher accuracies than learners. CHANGE POINT SIZE

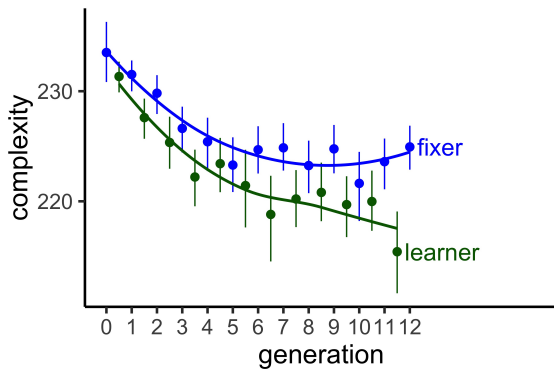


Figure 6: Fixers reintroduce algorithmic complexity which is lost by learners in the dyad condition.

$p = < .001$), although the effect of generation is stronger for learners compared to fixers ($\beta_{\text{generation}} = -4.843$, $t = -9.869$, $p = < .001$). These results hold true for all three measures of complexity (Do i need to report all of these stats??).

7 shows that the presence of an editor does help retain complexity in the grid patterns. The addition of a fixer into the task allowed a higher degree of complexity to be retained in the language over time ($\beta_{\text{condition-child}} = -3.584$, $t = -6.155$, $p = < .001$). Additionally, it appeared that the patterns in the dyad condition asymptoted sooner than in the baseline condition (Stats for this?).

General Discussion

Although Experiments 1-3 used a non-linguistic task, we were able to measure change in a culturally-transmitted, learned symbol system. In Experiments 1 and 2, language simplified rapidly and dramatically, reflecting the strong pressure towards simplification in language learning. These findings replicated those of Kempe et al. (2015): when transmitting an artificial language of grid patterns, complexity in the language was lost.

However, the results of Experiment 3 show that this loss is not permanent, but can be reintroduced in the language

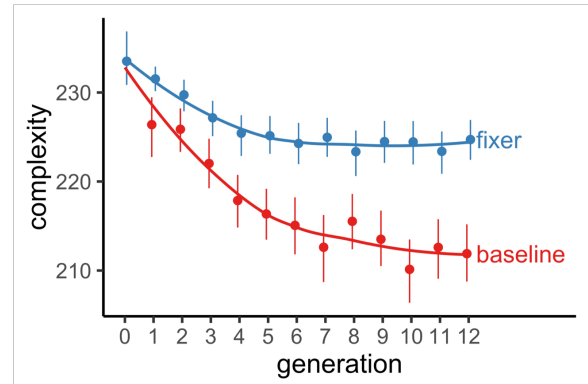


Figure 7: The presence of a fixer in the dyad condition causes a much greater level of algorithmic complexity to be retained across the evolution of a novel language. CHANGE POSITION IN PAPER

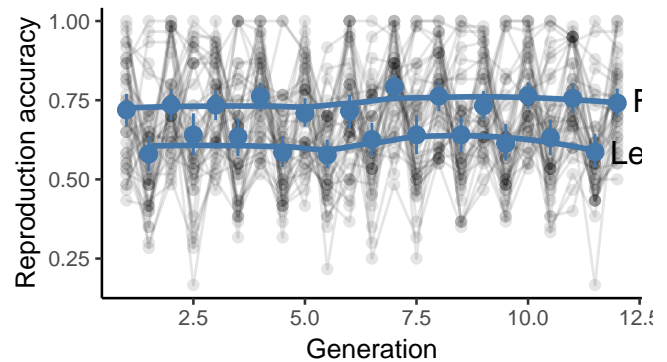


Figure 8: Experiments 1 and 2 show increases in accuracy over transmission generations. CHANGE POINT SIZE

by way of a secondary participant who helps bring the language towards a stable level of complexity. When the iterated-learning process begins to resemble the true process of language-learning, where children speak with and are subject to correction by those more competent in the language, a lesser amount of complexity was lost during transmission. Additionally, this stable level of complexity is much higher, and is reached earlier in the transmission chain with the help of a fixing participant. This stability in complexity did not mean that the language stopped changing, but that the descriptiveness and transmissibility pressures were in balance. Fixers in Experiment 3 represented caregivers – they were more accurate at reproducing the language, and could therefore be seen as more fluent speakers of the language, just as adults are of their native languages. The learners, on the other hand, had a more difficult task, which greater strained their working memories, similar to the strain on a child language learner who is inundated with new words each day. The fixer's corrected language was passed to the next learner 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 optimal language to the next generation.

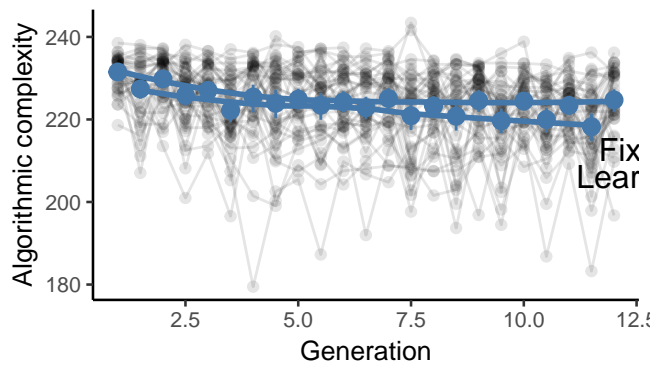


Figure 9: Experiments 1 and 2 show increases in accuracy over transmission generations. CHANGE POINT SIZE

Due to the higher accuracy by fixers, and therefore greater knowledge of the language, the fixers were able to compensate for some (not all) of the loss in complexity seen by the learners by editing their patterns.

In Experiment 3, the learner's reproduction accuracies were actually increasing over generations. Despite the stable level of complexity, learners found the language easier to reproduce over evolution. Although a high level of descriptiveness was retained in the language, transmissibility was increasing, without the simplicity pressure weighing in. Perhaps the language was becoming stable and complex, with the symbol-patterns changing to be both descriptive and useful, while being easily transmissible. This reflects the optimal evolutionary response to these two competing pressures.

When a caregiver or teacher prevents their child from growing up to believe that “baba” is the word for both “bottle” and “sheep”, they are not only helping their individual child become a competent speaker of the language, but they are also re-introducing complexity, and helping the language system as a whole from simplifying to disuse. Data collection is ongoing with children ages 6-8 at a local science museum in both the Experiment 2 and Experiment 3 tasks, in order to investigate whether the pressures of similarity and complexity affect children similarly to how they affect adults in early language-learning conditions.

We do not learn language as passive listeners, who absorb a proportion of the the linguistic input they hear. Therefore, we cannot measure language learning only through measuring input, nor through measuring only linguistic output. Languages are both learned and changed through conversations, with feedback and error correction, to evolve to the needs of the language's users. Therefore, we must study language learning in process, to see how it adapts and evolves with communicative interactions.

All code for these analyses are available at
<https://github.com/mcmeyers/iteratedlearning>

Acknowledgements

This research was funded by a James S. McDonnell Foundation Scholar Award to DY.

References

- Bowerman, M. (1982). U shaped behavioral growth. In (pp. 101–145). Academic Press.
- 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.
- 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.
- Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1(2), 151–195.
- 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.
- Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE*, 5(1), 1–10.
- Penner, S. G. (1987). Parental responses to grammatical and ungrammatical child utterances. *Child Development*, 58(2), 376–384.
- 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.
- Regier, T., Kemp, C., & Kay, P. (2015). 11 word meanings across languages support efficient communication. *The Handbook of Language Emergence*, 87, 237.
- 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.
- structure, E. of combinatorial, & signals. (2014). Verhoef, tessa and kirby, simon and de boer, bart. *Journal of Pho-*

netics, 43, 57–68.

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