

# Caregiver reconstruction of children's errors: the preservation of complexity in language

Madeline Meyers and Daniel Yurovsky  
{mcmeyers, yurovsky}@uchicago.edu  
Department of Psychology  
University of Chicago

## Abstract

Why do languages change? One possibility is they evolve due to two competing pressures: one, for the language to be easily transmitted to new generations—and hence simple—and another, for the language to be a useful, descriptive form of communication—and hence more complex. However, few studies have explored these pressures in the most skilled language learners: children. Conventional iterated learning studies focus on the transmission of a novel language from one adult to the next. However, this ignores one of the important features of language learning, namely, receiving feedback. This study compares adult performance on a conventional iterated learning task with their performance on a task which allows for error correction by a secondary participant. Results show that adding this error-correcting participant allows a greater level of complexity to be retained in the language compared with the baseline task. Data collection is ongoing with children, but current results suggest that editors (e.g. parents) may be playing a dual role in both child language acquisition and language evolution by re-introducing complexity into a given language.

**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 “Donde van ustedes?” of a group of work colleagues, but to address your friends, you use the informal “Donde váis 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? Why do languages change at all, aside from acquiring new vocabulary? One working theory is that languages evolve, like biological organisms, to adapt to two dynamic competing pressures: one, to be easily transmitted and learned (and hence simple), and another, to be a descriptive and effective system for communication (and hence complex) (Lupyan & Dale, 2010).

Children are often the actors who drive language evolution (Senghas, 2003), yet they differ from adults in their 1) cognitive capabilities, namely, memory systems (Kempe, Gauvrit, & Forsyth, 2015), 2) interests and early vocabularies, and 3) 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?

The following study tests a novel hypothesis for the maintenance of structure in language: Communicative inference by caregivers. Children's language learning is greatly influenced by those around them—especially the adults they talk to most. These caregivers determine the majority of their child's language 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). They may reconstruct their child's language in numerous ways—through explicit or implicit correction, or simply through modeling correct use of the language over time (Hudson Kam & Newport, 2005). By way of this feedback, 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. 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 generations than children could sustain alone.

## Using iterated learning to study language change

One way to study language acquisition in the lab is to use the iterated learning paradigm. This paradigm was created to study the effects of simplicity and informativity on intergenerational language evolution (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. Few iterated learning studies, however, have used children as research subjects. A study by Kempe et al. (2015) compared child (age 5-8) and adult performances on an iter-

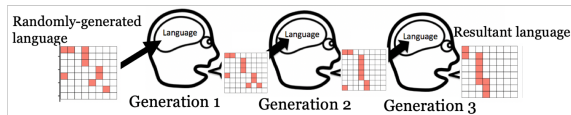


Figure 1: In the standard iterated learning paradigm, each participant is trained on input from the previous participant, creating a chain of generational transmission replete with implicit learning errors. This study uses grid pattern stimuli adapted from Kempe et al. (2015)

ated learning task using a novel dot-pattern paradigm. Their results found that structure emerged faster in children than adults—that is, the children’s patterns simplified much faster than the adult’s, allowing them to be easily reproducible earlier in the transmission chain. This study provides evidence of the importance of looking at both children and adults in an iterated learning paradigm, as they have different cognitive skills which affect their performance in language-learning tasks (Kempe et al., 2015). 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.

The following study uses an iterated learning (diffusion chain) paradigm with an adaptation of Kempe et al. (2015)’s original non-linguistic task to investigate how the pressures of descriptiveness and transmissibility operate in adults when their responses are subject to error correction by a secondary participant. 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 re-introducing and preserving complexity in language. All experiments were pre-registered on Open Science Framework and can be accessed at the following link: <https://osf.io/guzyf/>. [Right place to put?]

## Experiment 1: Baseline

### Methods

In a baseline experiment, adults participated in a standard 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.

The experimental task was designed online using JavaScript, HTML, and CSS, and was hosted as a web page

accessed through a server. All adult data was collected through Amazon Mechanical Turk, an online crowdsourcing site commonly used in psych studies (Buhrmester, Kwang, & Gosling, 2011). Participants were compensated with \$0.50 for their participation. Each participant’s patterns were stored using Google Sheets, so they could be transmitted to the next participant online in real-time.

Subjects in Experiment 1 were told that in this task, they would be re-creating patterns on a grid. After a consent screen, subjects 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. Subjects could click on any cell in the grid to change its color, and could also undo any color placed. A counter was displayed on the screen which showed how many targets had been colored, and it varied dynamically with the participant’s clicks. This was implemented in order to help the subject place exactly 10 targets. Participants could only place up to 10 targets at a time, and could not advance to the next trial unless they had selected 10 targets. Subjects had 60 seconds to complete each trial, which was displayed as a timer on the screen, and an audio cue reminded them when they had only 20 seconds left to complete their pattern. There were additional audio cues such as sparkling sounds and words of encouragement throughout the task. After 3 practice trials, participants were told that the study would begin. The subject’s performance on the practice trials was used as an attention check to determine whether their data would be passed to the next participant. If the subject scored less than 75% accuracy on the last 2/3 scored practice trials, or if the subject failed to select 10 targets before time ran out, their data would be marked as “unavailable” to the next user in the chain. There were 6 experimental trials, where participants viewed either the randomly-generated grids or the grids passed from a previous participant in the chain, rather than the simpler practice target grids.

Transmission chains consisted of 6 generations each, and 20 separate chains were run in Experiment 1, for a total of 120 participants. All chains in the following experiments began with the same initial target grids. The participant in the first generation received the initial grid as their target, and each subsequent generation received the previous generation’s inputs as their targets.

The initial 8x8 grids were generated randomly. Each colored target corresponded to a number which was generated using Excel’s random number generator. A random number 1-64 was assigned to each cell in the grid, and the 10 random numbers generated for each of the 6 practice trials determined the pattern viewed by participants in the first generation. The

6 grid patterns which began the iteration all had initial similar transmission accuracies of approximately 55%. All initial patterns remained constant across all chains and conditions of the study; this allows a direct comparison of aggregate data. Aside from the first practice trial, where participants were required to reproduce a pattern perfectly, participants never received feedback on their responses. ##Analysis In order to measure the effects of transmissibility and descriptiveness pressures on this dot-pattern-language, reproduction accuracy and complexity were computed. Reproduction accuracy represented the transmissibility pressure: if the participant was able to pass the identical language to the next generation. Complexity was computed to measure the competing effects of both pressures: a higher-complexity pattern contained a greater degree of descriptiveness, while a lower-complexity pattern contained a greater degree of simplicity and transmissibility.

## Reproduction Accuracy

Reproduction accuracy was calculated as the proportion of targets (out of 10) which were placed in the same location on the target and input grids. This measure does not account for the degree of error for targets placed in incorrect locations.

## Complexity

To calculate the complexity of the grid patterns produced, we used three measures: algorithmic complexity, chunking, and edge length. All analysis code was adapted from Gauvrit, Soler-Toscano, & Guida (2017). 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). Essentially, BDM uses the coding theorem method to break each two-color 8x8 grid into 4x4 patterns. If each 4x4 pattern is  $p$ , the formula for computing BDM is [needs to be formatted]  $\sum p (\log_2(np) + K(p))$  (Kempe et al., 2015). Here,  $K(p)$ , or the complexity of a 4x4 pattern, is defined as the length of the program taken as input by a Turing machine, and  $nP$  is the frequency of each pattern  $p$  (Zenil, Soler-Toscano, Delahaye, & Gauvrit, 2015). This measure of complexity accounts for the shortcomings of other complexity measures such as entropy, as it accounts for the condition where a checkerboard pattern is not probable, yet is not typically defined as complex. In this case, BDM would see this pattern as having a lower complexity than a different randomly-generated 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 (Gauvrit et al., 2017).

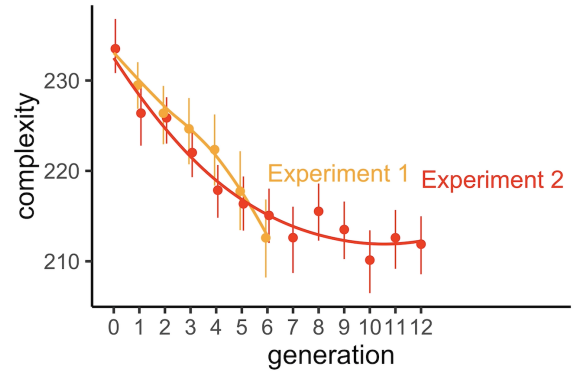


Figure 2: Experiments 1 and 2 show decreases in algorithmic complexity over time. Experiment 2 begins to asymptote.

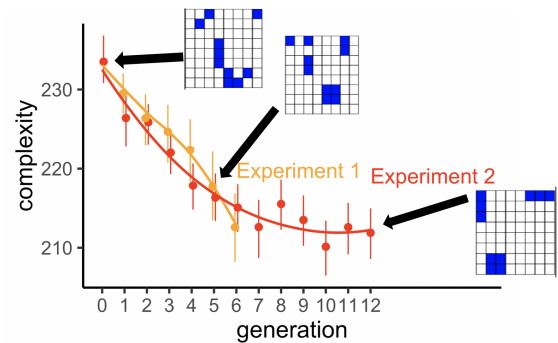


Figure 3: Experiments 1 and 2 show decreases in algorithmic complexity over time. Experiment 2 begins to asymptote.

## Experiment 1 Results

We set out to test whether the accuracy and complexity of participants' pattern reproductions changed over successive generations. Specifically, we expected to see an increase in participant's transmission accuracies, relating to a decrease in the computational complexity of the patterns and an increase in reproduction accuracy.

Reproduction accuracy increased linearly over 6 generations. We fit a linear model to this data, which was significant ( $\log(\text{generation})$  Estimate = 0.11,  $p < 0.001$ ). The trend in complexity was consistent across the measures of algorithmic complexity, chunking, and edge length. Algorithmic complexity decreased over generations, although the shape of the trend was difficult to determine. A linear model also showed a significant negative effect of generation ( $\log(\text{generation})$  Estimate = -12.0,  $p < 0.001$ ). These findings replicated those found by Kempe et al. (2015).

## Experiment 2: Baseline Replication

In Experiment 2, we replicated our task from Experiment 1, but with the addition of twice as many chains and generations. We replicated the task with a larger sample in order to determine the shape of the complexity curve—if it was linear, exponential, or another shape, and if it reached a stable

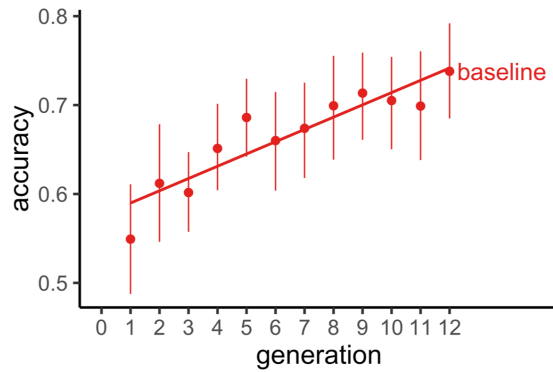


Figure 4: Reproduction accuracy increases over generations.

asymptote of complexity over generations.

## Methods

Transmission chains consisted of 12 generations each, and there were 40 separate chains. A total of 518 adults participated in this study. Approximately 7% ( $n=38$ ) of participants in the 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, four times the number of participants in Experiment 1.

## Experiment 2 Results

The results of this experiment replicated those found in Experiment 1. Percent accuracy increased linearly over generations ( $\log(\text{generation})$  Estimate: 0.05,  $p < 0.001$ ), reflecting an increase in transmissibility of the patterns. Algorithmic complexity, shown in figure X, appeared to follow an exponential function of the form  $y = e^{-x} + b$ . This result was also found in the alternate measures of complexity (chunking and edge length). The algorithmic complexity of the patterns decreased, and began to asymptote. We fit an exponential model (include model formula?), which was significant ( $\log(\text{generation})$  Estimate = -0.04,  $p < 0.001$ ).

## Experiment 3: Dyad

In order to add an element of error-correction to the iterated-learning process, we adapted the task from Experiments 1 and 2 to include an editor participant, analogous to a caregiver who protects their child from learning incorrect forms of language.

## Methods

In the third, dyad experiment, A primary participant was designated as a “learner”, and completed the same task as in the baseline experiment. A secondary participant—the “fixer”—was given an adjusted task, where instead of reproducing a pattern on a grid, they were told to “fix the errors” on a grid pattern in order to match the same target pattern. The fixer’s responses were passed as target input for the subsequent learner.

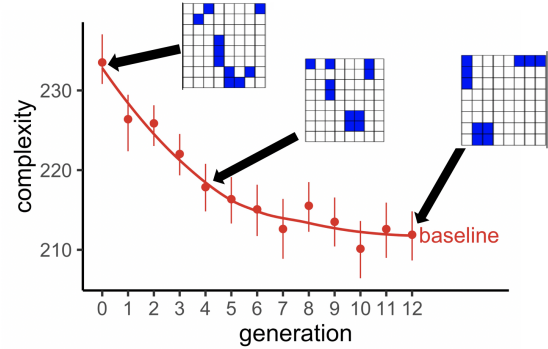


Figure 5: Algorithmic complexity decreases over time, and reflects simplifications in the patterns produced by participants.

Those in the “fixer” condition in the dyad experiment were given an adapted task. The only difference was that throughout the study, they were not told to re-create the target grid, but to fix a grid they saw to make it resemble the target grid exactly. Essentially, fixers in the dyad condition viewed the same target grid as the learners, but instead of seeing a blank input grid, they saw a grid that already had 10 elements filled in – the elements that the previous learner had submitted. The participant could then click and unclick the elements to edit their positions. There was no “reset” button on these patterns, so they reflect participants first memory instincts.

In the dyad condition, a generation consisted of a learner, who re-created the target grid, and a fixer, who received the same target grid as well as the learner’s input grid as their grid to edit. The fixer’s final input was used as the target grid for the subsequent generation.

Transmission chains consisted of 12 generations each, and 40 separate chains were run during each condition of the study. A total of 1,038 adults participated in this study. Approximately 7% ( $n=78$ ) 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.

## Experiment 3 Results

We expected to find significant accuracy differences between learners and fixers: because learners had a more difficult task, the strain on their working memories was larger, analogous to a child language learner who is inundated with new words to remember every day. As the fixers had a less difficult task—they viewed the target as well as what the learner had produced, we expected their reproduction accuracies to be greater than the learner’s. Additionally, a higher reproduction accuracy in the learners compared to the fixers would convey that the fixers were, in fact, correcting errors made by the learners.

Indeed, we see that fixers and learners had significantly dif-

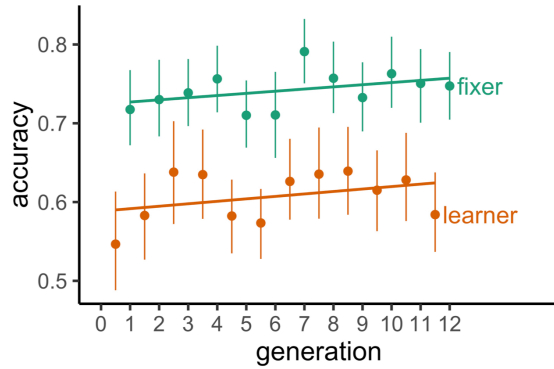


Figure 6: In the dyad task, reproduction accuracy stays relatively constant across generations. Fixers have significantly higher accuracies than learners.

ferent pattern reproduction accuracies (Fig X). According to a one-way AOV, these accuracies were significantly different ( $F = 177.3$ ,  $p < 0.001$ ). The accuracy of the fixers did not increase significantly over generations ( $\log(\text{generation})$  Estimate =  $-0.01$ ,  $p = 0.3$ ), while the accuracy of the learners showed a marginally significant increase ( $\log(\text{generation})$  Estimate =  $0.02$ ,  $p = 0.06$ ).

We expected to see a trend in complexity where learners simplified the language and fixers reintroduced or compensated for some of this loss in complexity. Indeed, as shown in Figure X, we see this trend.

Overall, we expected to see a smaller decrease in algorithmic complexity in the fixers over generations compared to in Experiment 2. When fitting an exponential model to the fixer's complexities, we see a decrease over generations ( $\log(\text{generation})$  Estimate =  $-0.02$ ,  $p < 0.001$ ). Again, the results of algorithmic complexity were in line with those seen in other measures of complexity (chunking and edge length). This is significantly different from the results seen in Experiment 2, namely, the addition of a fixer into the task allowed a higher degree of complexity to be retained in the language over time (AOV,  $F = 33.66$ ,  $p < 0.001$ ). Additionally, it appeared that the patterns in the dyad condition asymptoted sooner than in the baseline condition—the fixers allowed the language to reach a stable level of complexity earlier on in the transmission chain.

## General Discussion

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.

Although Experiments 1-3 used a non-linguistic task, we

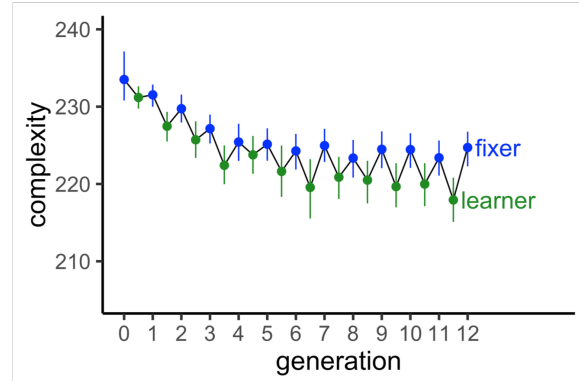


Figure 7: Fixers reintroduce algorithmic complexity which is lost by learners in the dyad condition.

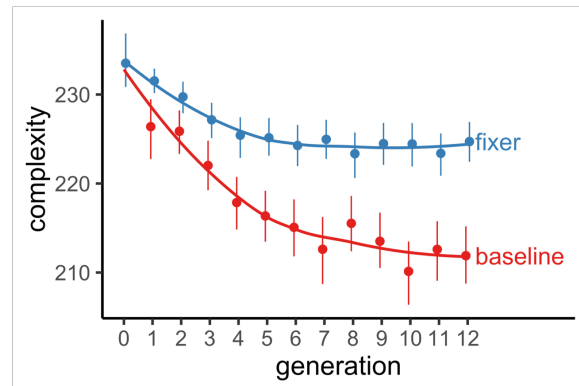


Figure 8: The presence of a fixer in the dyad condition causes a much greater level of algorithmic complexity to be retained across the evolution of a novel language.

were able to measure change in a culturally-transmitted, learned symbol system. In the baseline experiment, language simplified rapidly and dramatically, reflecting the strong pressure towards simplification in language learning. However, 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. In the dyad task, the fixers represented parents or teachers, as they had an easier memory task, shown by their higher transmission accuracies. This reflects the greater ease that adults have when retrieving the correct form of a word compared to children—their memory for most words is stronger compared to a child language-learner. The 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. Not only did editors re-introduce complexity into the language, allowing for a greater amount of complexity to be retained over time, but they also helped the language reach an asymptote—a stable level of complexity—sooner than in Experiments 1 and 2.



This stability in complexity did not mean that the language stopped changing, but that the descriptiveness and transmissibility pressures were in balance. In fact, the learner’s (analogous to children’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 optimally 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.

### Future Directions

These experiments aim to describe how parents and children uniquely affect the language evolution process. However, the results described above were obtained with adult participants. For this reason, we are currently collecting data with children in both the baseline and dyad conditions, to see if these pressures operated differently with real children. Data collection is ongoing with children ages 6-8 at the Museum of Science and Industry in Hyde Park, Chicago. Children are participants in the baseline task, and are learners in the dyad task, with adult mTurkers as the fixers. Children complete the task on an iPad, and receive their choice of stickers as compensation. iPad tasks have many advantages over other research methods, including the paper-and-sticker task used by Kempe et al. (2015) because the use of an iPad reduces the completion time of the study and is engaging for young children (Frank, Sugarman, Horowitz, Lewis, & Yurovsky, 2016). Parents of the children in the study completed an additional child information sheet about the child’s language experiences and home environment. We expect to see similar trends in complexity over time with children as were seen with adults, namely, that in the dyad task, adults reintroduce complexity which is lost by child language learners. However, we expect sharper initial decreases in complexity with children, in line with the findings of Kempe et al. (2015). We plan to conduct a set of qualitative analyses on the patterns produced by adults and children, in order to see whether children are simply making more errors than adults, or if they are making fundamentally different errors, perhaps reflecting their differential cognitive language-learning systems.

### Acknowledgements

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

### References

- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon’s mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1), 3–5.
- Chouinard, M. M., & Clark, E. V. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child Language*, 30(3), 637–669.
- 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.
- 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., 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.
- Senghas, A. (2003). Intergenerational influence and ontogenetic development in the emergence of spatial grammar in nicaraguan sign language. *Cognitive Development*, 18, 511–531.
- Zenil, H., Soler-Toscano, F., Delahaye, J.-P., & Gauvrit, N. (2015). Two-dimensional kolmogorov complexity and an empirical validation of the coding theorem method by compressibility. *PeerJ Computer Science*, e23.
- 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.

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