Caregiver reconstruction of children’s errors: the preservation of structure in language

Madeline Meyers & Dan Yurovsky

The University of Chicago

**Themes**:

* Systematicity (kempe really pushed this, but im not sure about the links that they talk about with reducing complexity = more systematicity, I guess it makes sense that increased within chain similarity = more systematic, maybe?)
* Complexity/simplicity
* Correction of errors / language learning as a collaborative process
* Language evolution lacking kid research [really iterated learning should be longitudinal]
  + Do kids have the same cognitive systems for learning novel languages as adults? We know kids are a lot BETTER at learning languages than adults—does this mean they also change languages in different ways? Ex. Spanish
  + Learnability vs Descriptiveness
* Non-linguistic task to study language in people with different levels of experience
* Why iterated learning?
  + Cool way to see language evolution in the lab b.c its hard to see in real-world
    - NSL
      * Gained descriptiveness when had both horizontal & vertical input! (is this substantiated>>> unclear)
    - Other people use it to see simplicity vs descriptiveness

**Background** (from EFF Research Proposal):

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 an effective system for communication (and hence informative; Lupyan & Dale, 2010).

Transition Children are inundated with thousands(?) of novel words each day, and they have different strategies to parse this information. Children are known for being experts at language learning (CITE), but does this mean they have balance the pressures that drive language evolution differently? \*children are really good language learners, and this skill might relate to them having a different balance of factors that drive language evolution\* Children are often the actors who drive language evolution\*\*indepth\* (Senghas, 2003), yet they differ from adults in their 1) cognitive capabilities, namely, memory systems \*\*indepth (Kempe et al., 2015), 2) interests and early vocabularies\*\*indepth compared to L2 or just compared to adults? Should you be making the comparsion between kids and adults or between kid lang acq and adult lang acq? Probably the later!!\*\*, and 3) conversation partners. Therefore, children’s 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 \*\*indepth(Kirby et al., 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 their caregivers. 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—they understand the child’s intuitions, personality, and context. Caregivers play an important interpretive role in these interactions by their ability to understand the intended target of children’s errorful productions \*\*indepth(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). These reconstructions may be a mechanism by which more structure is retained in language than children could sustain alone.

\*talk about how this might be different from adult language learners / adult fixing errors\*

\*introduce the idea of systematicity into these arguments\*

**Iterated Learning & Diffusion Chain Paradigms**

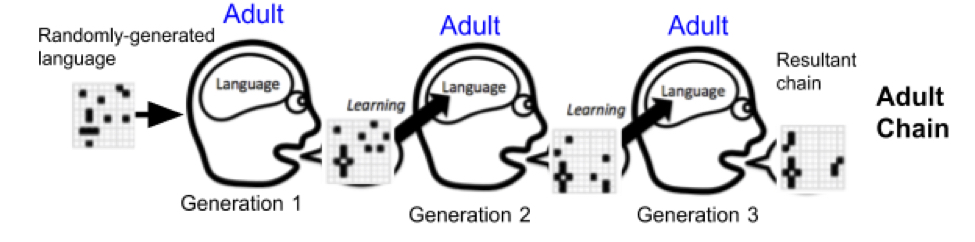
One way to study language learning in the lab is to use the iterated learning paradigm. Created to study the effects of simplicity and informativity on inter-generational language evolution, this method is useful for testing the proposed hypotheses (Kirby et al., 2014). In an iterated learning paradigm, 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 (see Figure 1). 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 (5-8) and adult performances on an iterated learning task using a novel dot-pattern paradigm (showed in Figure 1). 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 easier to reproduce 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 can never be fully grasped using this paradigm with only separate adult or child earning 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 in both directions. In a true language-acquisition situation, a child receives a great deal of language input from their caregiver and uses it to interact with their peers throughout life, eventually growing into a new teacher-caregiver. \*add NSL info\*

\*importance of using a non-linguistic task\*

At the beginning stages, language learning is the learning of different audio patterns, re-creating them, and memorizing what they refer to/imply. \*\*talk more about beginning lang dvt, why this task works\*\*

\*introduce four conditions and why you have them -- adult adult to also provide a standard for correction because otherwise is confounded (also might be issue that there is more to fix when the errors are more significant, but I guess this is what the real-world problem is)\*

\*benefit of having an online task; able to collect more than one data point per participant\*



**Conditions**:

1. **Adult Baseline** (Complete): Ran 152 U.S. adults on Amazon Mechanical Turk
   1. ~15 were removed due to errors in the data saving method
   2. ~15 (12.5%) were excluded from chains due to failure to meet accuracy requirements:
      1. Subjects either timed out on a trial and failed to select 10 items or
      2. Subjects scored less than 75% combined accuracy on the first two practice trials

For a total of 120 usable data points; 20 full chains of 6 participants (generations) each

1. **Adult-Adult Dyad** (Complete): Ran 289 U.S. adults on Amazon Mechanical Turk. Subjects were alternately assigned to be a “child” or a “parent”; where the “child” performed the same task as in the baseline condition; and the “parent” performed a slightly different task, where instead of re-creating the grid; they were charged with fixing an already-filled-in grid
   1. A total of 49 people were excluded from the analysis due to failure to meet accuracy requirements:
      1. 11 subjects timed out & did not select exactly 10 items on one or more trials
      2. 38 subjects scored less than 75% combined accuracy on the second two out of three practice trials (23 adults and 15 children, or 20% of adults and 12.5% of children)

For a total of 240 usable data points; this represents 20 full chains of 6 generations, with one generation being made up of one child and one parent

1. **Child Baseline** (in progress): Plan to obtain 120 usable data points from children ages 6-8 at the Museum of Science and Industry in Hyde Park, Chicago. Currently have run 11 participants, with 8 usable data points. Chains were not divided based on the child’s age, all children were mixed around between 6-8y/o in the chains.
2. **Adult-Child Dyad** (in progress): Plan to obtain 240 usable data points by running 120 children ages 6-8 at the Museum of Science and Industry in Hyde Park, Chicago, and 120 adults on Amazon Mechanical Turk. This condition will have to be run generation by generation, as it will require batches of data from the museum alternating with batches of data gathered online.

**Initial Hypotheses:**

Caregivers help retain a higher level of complexity in language while still retaining a similar percent accuracy: translated, this means that parents are the reason children can communicate complicated concepts that are out of their memory range—concepts and things that they would normally simplify. Parents may reintroduce this complexity by “translating” for the child, or for providing the child with more descriptive vocabulary; when normally a child would just lose the descriptiveness (and therefore would be a less effective communicator).

For this study, it means that we expect to see higher levels of complexity retained in the dyad vs. baseline conditions, where there is still a similar level of percent accuracy from targets to inputs.

Isn’t this just a memory task? Of course a memory task would be easier when people have extra cues? \*ADDRESS THESE QUESTIONS\*

We expected to replicate the findings of Kempe et al. (2015) in both adult and baseline conditions. Their findings were XXXX.

We therefore expected our see our measures of complexity decrease and asymptote, and percent accuracy increase over time (linearly? Asymptote b/c people will always make errors?)

Adult Baseline:

* Replicate Kempe et al. (2015)
* Asymptotes

Child Baseline:

* Replicate Kempe et al. (2015)
* Asymptotes
* No age effects between 6 and 8
* Much sharper initial simplification, but then still has emergence of structure

Do the dyad conditions reach a different, stable state, or do they just slow the process of the baseline state?

# of changes made

Adult-Adult Dyad:

* Asymptotes
* No initial hypothesis about where it asymptotes to—could be adult level (slower) or higher than adult level (retains more complexity)

Adult-Child Dyad:

* Asymptotes
* No age effects within children
* No hypotheses about where it asymptotes to –maybe not until gets to child level (just slowing of process), maybe brings up to adult level, maybe somewhere in between adult and child baseline (most likely)

If we find sharp differences (how define?? What looking for? Qualitative??? In complexity??? In general shape of graph??), that is evidence that children are fundamentally different at learning language (or other similar memory tasks, as language learning really is learning patterns at the beginning) compared to adults. Why is this? We don’t really know… maybe

* Adults and kids think of chunks differently
* Kids just like different patterns
* Adults make qualitatively different errors, like translating errors vs children making new patterns and one is easier to correct than the other

**Methods**:

The 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, a \*description of turk\*, and therefore users likely completed the task on a computer. All child data was collected with children ages 6-8 at the Museum of Science and Industry in Hyde Park, Chicago, on an iPad. This age range was primarily chosen due to the practicality of running a complicated and difficult task with children. 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 et al., 2016). Adults were compensated with $0.50 for their participation, and children were given their choice of stickers. Parents of the children in the study completed an additional child information sheet about the child’s language experiences and home environment.

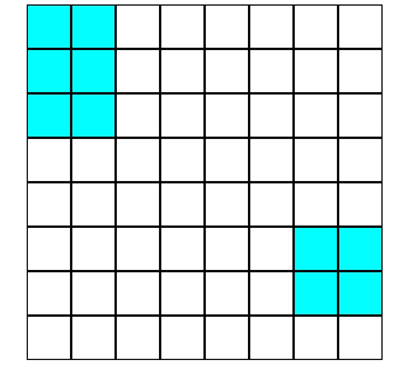
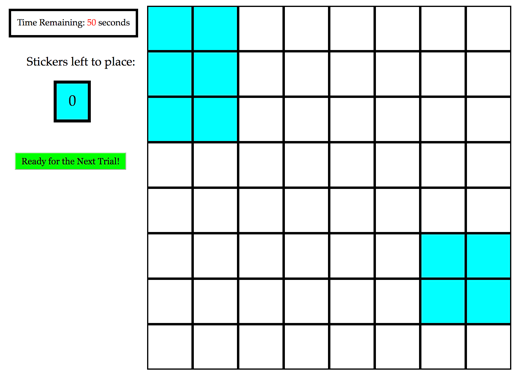
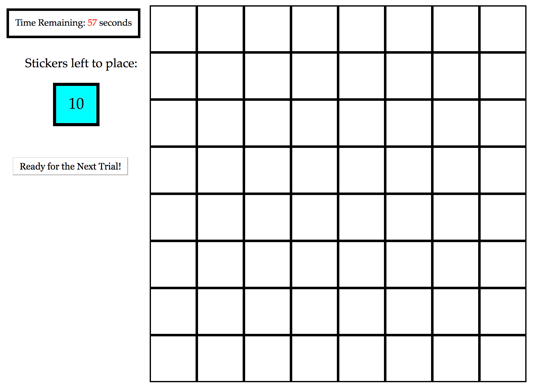
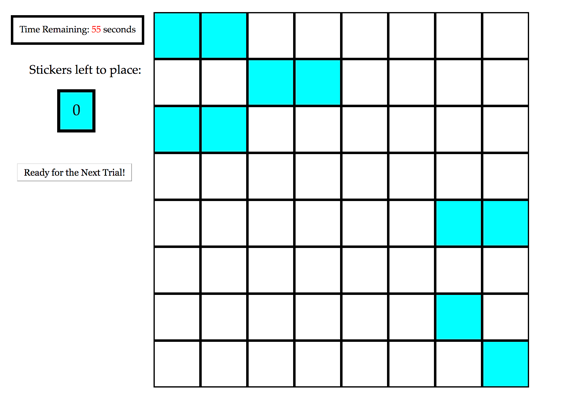
Subjects in both baseline conditions and those who were assigned to be “children” in the dyad conditions were told that in this task, they would be re-creating patterns which consisted of 10 cells colored in (“target blocks”) on an 8x8 grid. There were 4 practice trials. After a consent screen, subjects first viewed a trial with two grids on the screen – a target grid and a blank grid. They were told to make the blank grid match the target grid exactly, and were unable to progress to the subsequent trial until the grids matched exactly. Following this trial, participants were either told or read that they would see a target grid appear on the screen for 10 seconds, followed by a picture which displayed for 3 seconds (a visual mask, meant to clear any lingering visual effects from the target grid, which was a bunch of scribbles), followed by a blank grid where they were given 60 seconds to re-create the target grid. Subjects could click on any cell in the grid to have it change color, and could also remove any block which they placed. On the input grid screen, there was a counter that varied based on the number of blocks a participant had placed in order to help the subject place exactly 10 blocks, as well as a timer. Subjects had 60 seconds to complete each trial, and an audio cue reminded them when they had only 10 seconds left to complete their pattern. There were additional audio cues such as sparkling sounds, encouragement, etc. throughout the task to keep it engaging for children. After 3 practice trials, participants were told that the study was ready to start. The subject’s performance on the practice trials was used to determine whether their trial data were viable and would be used to read into the next participant in the chain. If the subject scored less than 75% accuracy on the last 2/3 scored practice trials (the targets on the practice trials were significantly easier than the collected data targets), their data would be marked as “unavailable” to the next user in the chain. There were 6 data collection trials.

Those in the “adult” condition of either dyad task were given an almost identical 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, adults in the dyad condition viewed the same target grid as the children, but instead of seeing a blank input grid, they saw a grid that already had 10 elements filled in – the elements that the previous child had submitted. The participant could then click and unclick the elements and change their positions. There was no “reset” button on these input grids, so they reflect participants first memory instincts.

Chains consisted of 6 generations, and 20 separate chains were run during each condition of the study. Each chain began with the same initial target grid. The participant in generation1 received the initial grid as their target, and each subsequent generation received the previous generation’s inputs as their targets. In the dyad tasks, a generation consisted of a child, who re-created the target grid, and an adult, who received the same target grid as well as the child’s input grid as their grid to edit. The adult’s final input was used as the target grid for the subsequent generation.

The 8x8 grids for data collection were generated randomly using Excel’s random number generator. A 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. All initial patterns remained constant across all four conditions of the study; this allows a consistent comparison of aggregate data. Aside from the first practice trial, participants never received feedback on their responses.

Step 2 in Baseline or child conditions in dyad task

Example resultant practice grid

Step 2 in adult conditions of dyad task

Step 1: Participant sees target grid for 10 seconds

**Results—Adult Baseline & Turk Dyads:**

**Analysis Measures:**

**% Accuracy**

Percent accuracy was calculated as the number of targets out of 10 which were placed in exactly the same location on the target and input grids. This measure does not account for the degree of error in targets which may have been displaced by one position versus many.

**Algorithmic Complexity**

\*insert some stuff from joe’s thing here\*

**Chunking**

Our measure of chunking calculated the number of discrete parts in a grid pattern, where a part was defined as when two or more cells shared an edge (this does not include blocks which were connected diagonally). CITE GAUVRIT

**Edge Length**

Edge Length was calculated as the combined perimeter of the blocks or block chunks. For example, two blocks which were not in contact would have an edge length of 8, whereas two blocks which shared an edge would have an edge length of 6. CITE GAUVRIT

**Earth-Mover Distance**

The \*\*\*\* Earth-Mover Distance calculated the number of moves from one pattern to another. This is the only measure which accounts for different degrees of changes, where we would believe if someone shifted a pattern over by one unit that is a lesser error than creating an entirely new pattern.

**Identifiability** (to be calculated in the future)

Identifiability is a measure of whether the restultant grid patterns were more similar to the other patterns within a chain, or the patterns in the same generation in a different chain. It helps to see if each chain created a unique and diverging set of patterns (or language), or if the initial conditions were more important in determining the end state of the language.

**Qualitative**

**Combined Results:**

RMD File! !! !

**Adult Baseline:**

**Turk Dyad:**

**Changes to Method:**

* Added extra training trial to change inclusion requirements with children between adult baseline, first 8 child baseline participants and turk dyad conditions +onward
* In the adult baseline condition, those who did not meet inclusion requirements were removed post-data collection, and data collection was repeated; whereas this happened in real-time in the turk dyad condition + onward

**Changes to Hypotheses:**

* If all graphs linear, can talk about slope
* If graphs asymptote; can talk about slope and point they approach

**Future Directions:**

* Run with younger children (pre-critical period?)
* Look at developmental effects w/children
* Run conditions out more generations to see if complexity asymptotes over time
* In order to do qualitative analysis, run turk task where people just make patterns and see what they make—use those as the standards
* Run conditions to look at how manipulating the authority of the original grid affects the changes people make
  + if people think grid is from child or adult
  + if the grid actually has more or fewer errors (like the explore/exploit task you saw at CogSci)

**Current Conclusions:**

**Significance & Implications (OLD):**

The completion of the proposed study will contribute to literature on language acquisition and language change. Importantly, the dynamic contributions of both parents and children are often overlooked in studies of language acquisition, as many isolate either child-directed language (CDL) or child-produced language, instead of looking at them simultaneously.

The theories explored in this study are not only relevant to language development but could possibly be extended to other domains of development, including social, emotional, and moral development. Do caregivers play the same important role in teaching their children cultural values? If children do not practice a cultural value “correctly”, or at all, and parents do not reconstruct this mistake, the culture may also simplify or change (Henrich, 2004), but would it change in the same way as a language system?

This study attempts to show not only how caregivers are necessary to help reconstruct an individual child’s errors, but also how they are pivotal on a large-scale to protect language as a whole from simplifying to disuse. Thus, correcting a child’s errors over time serves a dual purpose: to aid the child, and to protect the language system.

**Issues/Questions:**

* Based on memory & learning research, should we expect the complexities to asymptote?
  + Is there a particular level of complexity that is optimal for learning?
* Generally, how does memory research interact with this study? What’s the interface? What does the field inform about what’s happening here?
* How do we measure the “qualitative” difference between child and parent grids? It feels like we will not be able to tell the whole story just using our measures of complexity
  + Looking at different types of strategies used (ex diagonal vs chunking vs iconics)
  + Hard code this?
  + We want some way to tell if the errors parents and kids are making are different
* What’s happening with earth-mover distance? Why does it look so weird when the other measures make sense?
* Right now we ran 6-8 year olds because of practicality. Is this a bad thing for us – should we really have been working to run kids who are pre-critical period? Or does it not matter because this is a novel language? What’s the justification for using any given age range in this study?
* Our Earth-Mover Distance measure is very inconsistent with all the other measures of complexity; it also doesn’t make much intuitive sense—what was happening? What does this mean? Should we be concerned?
* Is it a problem that we don’t have age/gender demographic data for the adult-adult dyad participants? I don’t have any demographic hypotheses, but maybe someone somewhere would…

References

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

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. Trends in cognitive sciences, 19(10), 603-615.

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 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), e8559.

Newport, E., Gleitman, H., & Gleitman, L. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style.

Senghas, A. (2003). Intergenerational influence and ontogenetic development in the emergence of spatial grammar in Nicaraguan Sign Language. Cognitive Development, 18(4), 511- 531.

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: Statistical Mechanics and its Applications*, *404*, 341- 358.