



ELSEVIER

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

Systematicity, but not compositionality: Examining the emergence of linguistic structure in children and adults using iterated learning

Limor Raviv^{a,c,*}, Inbal Arnon^b^a Psychology of Language Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands^b Department of Psychology, Hebrew University, Jerusalem, Israel^c Department of Cognitive Science, Hebrew University, Jerusalem, Israel

ARTICLE INFO

Keywords:

Iterated learning
Cultural transmission
Linguistic structure
Age-related differences
Language learning

ABSTRACT

Recent work suggests that cultural transmission can lead to the emergence of linguistic structure as speakers' weak individual biases become amplified through iterated learning. However, to date no published study has demonstrated a similar emergence of linguistic structure in children. The lack of evidence from child learners constitutes a problematic gap in the literature: if such learning biases impact the emergence of linguistic structure, they should also be found in children, who are the primary learners in real-life language transmission. However, children may differ from adults in their biases given age-related differences in general cognitive skills. Moreover, adults' performance on iterated learning tasks may reflect existing (and explicit) linguistic biases, partially undermining the generality of the results. Examining children's performance can also help evaluate contrasting predictions about their role in emerging languages: do children play a larger or smaller role than adults in the creation of structure? Here, we report a series of four iterated artificial language learning studies (based on Kirby, Cornish & Smith, 2008) with both children and adults, using a novel child-friendly paradigm. Our results show that linguistic structure does not emerge more readily in children compared to adults, and that adults are overall better in both language learning and in creating linguistic structure. When languages could become underspecified (by allowing homonyms), children and adults were similar in developing consistent mappings between meanings and signals in the form of structured ambiguities. However, when homonymity was not allowed, only adults created compositional structure. This study is a first step in using iterated language learning paradigms to explore child-adult differences. It provides the first demonstration that cultural transmission has a different effect on the languages produced by children and adults: While children were able to develop systematicity, their languages did not show compositionality. We focus on the relation between learning and structure creation as a possible explanation for our findings and discuss implications for children's role in the emergence of linguistic structure.

1. Introduction

How does linguistic structure emerge? Under the classic nativist approach, originally formulated in Chomsky (1965), linguistic structure is driven by a set of abstract and language-specific principles, which are both universal and innate, and impact how languages are shaped. An alternative explanation is offered by usage-based theories, suggesting that the kinds of structures we observe in human languages arise from general biases and constraints on individuals' cognitive capacities, such as learning, memory and processing (Tomasello, 2009). Under this view, languages are shaped through the process of *cultural transmission*, where weak individual tendencies become amplified and fixated over time through a repeated cycle of use, observation, and induction (Kirby,

Griffiths & Smith, 2014). This prediction is supported by findings from iterated learning paradigms, which show how the iterative nature of cultural transmission can lead to the creation of linguistic structure over multiple generations without the need to assume strong or language-specific innate biases (Culbertson & Kirby, 2015; Kirby, Cornish & Smith, 2008; Kirby, Smith & Brighton, 2004).

Iterated learning studies simulate the process of cultural transmission by using a diffusion chain paradigm, in which agents (computational or human) are exposed to a target behavior that they need to reproduce. Crucially, the behavior produced by the first agent in the chain becomes the input behavior for the second agent, the behavior of the second agent becomes the input for the third agent, and so on for several "generations" of agents. Mathematical and computational

* Corresponding author at: Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands.

E-mail address: limor.raviv@mpi.nl (L. Raviv).

<https://doi.org/10.1016/j.cognition.2018.08.011>

Received 1 September 2017; Received in revised form 12 August 2018; Accepted 17 August 2018

Available online 13 September 2018

0010-0277/ © 2018 Elsevier B.V. All rights reserved.

models of iterated learning show that the structural properties of artificial languages can be shaped over time to better-fit agents' existing tendencies and predispositions (e.g., Griffiths & Kalish, 2007; Kirby, Dowman & Griffiths, 2007). For example, agents' weak bias against synonymy was amplified over repeated episodes of learning so that an initial lexicon with multiple labels for each item developed unique object labels (Real & Griffiths, 2009).

In a seminal study, Kirby et al. (2008) used such a paradigm with adult participants to show that linguistic structure can emerge over time in an artificial language. Participants were exposed to items that varied across three semantic dimensions (shape, color and motion type), and needed to learn and reproduce novel labels describing these items. The first participant was trained on an artificial language without structure (with random mappings between forms and meanings), and their written output was given as input to the next participant in the chain. The resulting languages were evaluated on their learnability (i.e. how faithfully they were transmitted) and on their structure (how systematic they were). Over ten generations of participants, the languages became easier to learn and developed consistent mappings between meanings and signals. In the first study, languages were transmitted without intervention (allowing homonyms). These languages developed systematicity in the form of structured ambiguities, with small and underspecified lexicons in which items sharing a semantic feature were referred to using the same label. For example, all spiraling items were referred to as “poi”, regardless of their shape or color. Although this study resulted in systematic languages, encoding multiple semantic dimensions using holistic labels meant that the artificial languages lost much of their informativity, and differed dramatically from natural languages in their expressivity. In the second study, homonyms were filtered out during transmission to impose an artificial expressivity pressure and prevent underspecification. The result was that languages developed compositional structure – one of the hallmarks of natural languages: sub-strings were systematically reused to express different features. For example, color was marked with a prefix (e.g., “l-” for blue items vs. “n-” for grey items) and motion was marked with a suffix (e.g., “-plo” for bouncing items vs. “-pilu” for spiraling items). Similar increases in compositional structure and in learnability have since been replicated (Beckner, Pierrehumbert & Hay, 2017), and found for a range of linguistic and non-linguistic behaviors (e.g., drawings, whistles, gestures, visual patterns, for review see Tamariz & Kirby, 2016).

The accumulated findings support the prediction that linguistic structure can emerge through cultural transmission. However, they are limited in one interesting way: they are based only on adult learners. To date, only one study has used iterated learning to compare children and adults on a non-linguistic task (Kempe, Gauvrit & Forsyth, 2015; discussed below), and no published study has examined the emergence of linguistic structure over generations of child learners. The lack of evidence from child learners is problematic for several reasons. First, it limits the conclusions that can be drawn about the role of cultural transmission in the emergence of linguistic structure. The process of transmission is complex, and involves several different components (learning, production, and finally transmission to another agent), all of which may impact the resulting behavior. Learning biases in particular have been argued to shape the emergence of structure in the cultural transmission of language (e.g., Kirby et al., 2004; Smith, 2011). If the emergence of compositional structure over generations is influenced by learners' biases, then similar effects should also be found in children, who are the primary and most prototypical learners of language in real-world transmission. Children's performance is therefore a test case for the hypothesis that typical cross-generational learning can drive the emergence of linguistic structure. Second, adult participants may rely on their extensive and explicit linguistic knowledge when learning an unfamiliar language: adults may have a stronger prior bias in favor of linguistic structure, which can (consciously or not) influence their performance, causing structure to emerge more readily or more rapidly (Cornish, Tamariz & Kirby, 2009). This criticism is consistent with the

wide-spread effects of transfer from individuals' first language (L1) when learning a second language (e.g., White, 2000), as well as with the effect of L1 knowledge on artificial language learning – for instance, L1 phonotactics impact the segmentation of artificial languages (Finn & Hudson Kam, 2008; Siegelman, Bogaerts, Elazar, Arciuli & Frost, 2018). While strong biases are not *necessary* for structure to emerge (e.g., Kirby, 2001; Kirby et al., 2004, 2007; Smith, 2009), agents' existing biases could still influence the nature and rapidness of this process. This idea receives some support from computational models: under certain circumstances, changes in bias strength can impact the speed with which transmission fidelity increases (e.g., Ferdinand & Zuidema, 2008b), and even the resulting structural patterns (e.g., Navarro, Perfors, Kary, Brown & Donkin, 2018; Brochhagen, Franke, & van Rooij, 2016; Morgan & Levy, 2016; Smith, 2011). If adults' experience with their L1 results in a stronger bias for structure, their performance on iterated learning studies might reflect a cognitive bias that is partially the *result* of the evolution of language over time, rather than a bias *responsible* for it. This criticism can be partially overcome by looking at children, who have less explicit meta-linguistic knowledge and are more likely to learn implicitly (Arnon & Ramscar, 2012; Karmiloff-Smith, Grant, Sims, Jones & Cuckle, 1996; Ravid & Malenky, 2001; Ullman, 2001). Finally, examining children's performance in iterated learning is important given the long-lasting debate about their postulated role in the formation and extension of linguistic structure.

Children's role in the formation of grammatical structure has been heavily debated in the language emergence literature. On the one hand, children are claimed to play a special role in the formation of linguistic structure in creole languages. Bickerton's influential Language Bioprogram hypothesis (1984) argues that children, and not adults, are responsible for the formation of grammar in the process of creolization, and that they regularize the language and add structure to it through their reliance on innate linguistic biases. Similar claims have been made in the sign language literature, where children are shown to introduce novel linguistic structures. Studies of deaf children born to hearing parents show that children introduced regularities, like word order, which were not found in their input (e.g., Goldin-Meadow & Mylander, 1998). Research on emerging sign languages further suggests that children have a unique role in making new languages more structured: younger learners (exposed to the developing Nicaraguan Sign Language before the age of ten) produce more structured languages compared to adult learners within the same cohort (Senghas & Coppola, 2001). More generally, children are claimed to add grammatical features (such as linear sequencing) to the language when learning it from a previous cohort (Senghas, Kita & Özyürek, 2004). Additionally, younger children have a stronger tendency to segment and linearize their gestures compared to adolescents and adults (Clay, Pople, Hood & Kita, 2014). Based on these studies, which argue that children create core properties of language, we may predict that children will show similar or even stronger biases for creating structure in linguistic iterated learning. This prediction is supported by the single iterated study that compared children to adults on the same non-linguistic task. Using a visual recall task, Kempe et al. (2015) found that children created more identifiable and less complex visual patterns in comparison to adults, and concluded that structure emerged more readily in child chains. It is also supported by findings showing that children are more likely to regularize compared to adult learners (Hudson Kam & Newport, 2005; 2009; Samara, Smith, Brown & Wonnacott, 2017)

On the other hand, there are reasons to think that structure will emerge less readily in child learners. Children's postulated role in the process of creolization has been challenged by studies showing that it is a slow and multi-generational process (Arends, 1993; Carden & Stewart, 1988), and that complex grammatical structures emerged long before children were acquiring it as their first language (Arends & Bruyn, 1995; Sankoff & Laberge, 1974). These findings suggest that the main innovators in the process of creolization were adult speakers, and argue that children's contribution to the process is, if any, in the selection and

systemization of variation over time. Adults' larger role in this process is also supported by findings showing that agents of language change have some social influence in society, which young children typically lack (Kerswill & Williams, 2000; Labov, 2007; Nettle, 1999; Roberts & Winters, 2012). Given these critiques, and the possible effect of existing L1 knowledge on the bias for structure, we may predict that adults will show a stronger bias for creating structure compared to children. This prediction is supported by several other findings. First, children are generally worse than adults on artificial language learning tasks (Ferman & Karni, 2010; Perry, Axelsson & Horst, 2015), a factor which may impact their ability to create additional structure: children's greater difficulty in learning the input patterns may hinder their ability to add structure to it (we return to this prediction in the discussion). Second, children seem to be more conservative in learning new structures, and generalize less compared to adults under certain conditions (Boyd & Goldberg, 2012; Wonnacott, Brown & Nation, 2017).

2. The current study

Importantly, there is currently no experimental data to evaluate these contrasting predictions since no published work has compared the performance of child and adult learners on linguistic iterated learning. To investigate these predictions, we conducted the first large-scale study of iterated language learning that compares children and adults on the same task. We used a novel, child-friendly paradigm that closely resembles previous work with adults. Importantly, we use the same task with both age groups to enable direct comparison between them. We examine changes in the structure and learnability of the languages produced by children and adults over multiple generations with two questions in mind: (1) Will children, like adults, create more learnable and more structured languages over time? And (2) Will learnability and structure increase in child chains in the same way/rate as in adult chains? Given that skills like statistical learning, explicit learning, attention and working memory all improve with age (Arciuli & Simpson, 2011; Anderson, 2002; Gathercole, Pickering, Ambridge & Wearing, 2004), we predict some degree of difference in the overall performance between children and adults. However, we ask whether cultural transmission affected both age groups in a similar way, resulting in similar trends and rate of change. We evaluated child and adult performance in an iterated learning paradigm using two artificial languages with different meaning spaces. The first study used a language with three dimensions of meaning, modeled on Kirby et al. (2008). In the second study, we used an identical procedure but reduced the number of semantic dimensions from three to two to improve participants' learning of the language. For each language, we assessed performance in two filtering methods modeled on Kirby et al. (2008), which allow us to examine cultural transmission effects with and without underspecification. In Experiments 1a and 2a there were no limitations on the number of repeated words that could be transmitted to the next generation, allowing languages to deteriorate over time due to creation of homonyms. In Experiments 1b and 2b we imposed an artificial expressivity filter to prevent the transmission of homonyms across generations, a design which led to the emergence of compositional structure in Kirby et al. (2008). Our filtering procedure differed slightly from that of the original study because of our smaller lexicon (see details in the procedure section).

We focused on children between the ages of seven and twelve years (mean age 8;2). This age range was chosen for both methodological and theoretical reasons. As in Kirby et al. (2008), participants were exposed to written stimuli. Because responses were generated using a written syllable bank, we could only test literate children who were able to recognize the written syllables. We also worried that younger children will not be able to complete the task given their known difficulties with artificial language learning. Since our main goal was to compare the emergence of structure in children and adults using the existing iterated learning paradigm, we did not modify the paradigm further but chose

an age range which could complete it. Importantly, using this age range was also motivated by theoretical considerations: children of these ages were shown to introduce linguistic complexity in studies of developing sign languages (e.g., Senghas & Coppola, 2001), and are perceived as 'young' in previous work looking at children's role in language learning, emergence and change.

3. General method

All experiments use a diffusion chain paradigm, the most common method in iterated learning studies. In this paradigm, the first generation of learners is exposed to a randomly generated language without structure. All subsequent learners are trained on the output produced by the previous learner in the chain. Across all experiments, diffusion chains consisted of ten generations of single participants. Similar to Kirby et al (2008), our task was computerized and participants were tested on both seen and unseen items, simulating a learning bottleneck where learners need to reproduce the entire language after learning only a part of it. The task was modified to make it more child-friendly in the following ways: (1) items were changed from geometric shapes to cartoon figures of aliens; (2) a human experimenter sat next to the participants during learning; (3) the number of items was reduced by half; and (4) we used a syllable bank instead of free typing to reduce the memory load (see Method section for full details).

Across all experiments, we compare the performance of children and adults by looking at the same two parameters used by Kirby et al. (2008)¹:

- (1) **Language learnability:** measured by the differences between the labels participants learned and the labels they produced (mean transmission error), calculated using normalized Levenshtein string distances. The normalized Levenshtein distance between two strings is calculated by counting the minimal number of insertions, substitutions, and deletions of a single character that is required in order to turn one string into the other, and then dividing that number by the number of characters in the longer string. The transmission error between generations reflects participants' accuracy in reproducing the language, with easier languages eliciting fewer mistakes. Thus, there is an inverse relationship between errors and learnability, with a decrease in transmission error over generations indicating an increase in language learnability over time.
- (2) **Linguistic structure:** measured as the consistency in the mapping between meanings and signals in participants' languages, indicating whether similar meanings are being systematically expressed using similar strings. For each generation, a Monte Carlo sample analysis is computed, and the correlation between meanings' distances and strings' distances in a language is compared to the correlations between labels and meanings in 1000 random permutations of the language. The idea behind this test (Mantel, 1967) is to examine how a naïve agent, who has no access to the meanings, would assign labels at random². The result of this analysis is a z-score indicating the statistical significance of the correlation for each language. Following Kirby et al. (2008) and Kirby, Tamariz, Cornish, and Smith (2015), high z-scores indicate that the language has consistent and nonrandom signal-to-meaning mapping to a degree unlikely to arise by chance (specifically, $p < .05$ when the structure score is greater than 1.96). Importantly, this cutoff was only used for demonstrating the existence of structured languages and for descriptive and visualization purposes, but was never used for

¹ We thank Kenny Smith and Simon Kirby, who kindly provided us with their code for these analyses. We used the exact same algorithms to compute structure and learnability scores as used in the original paper (Kirby et al., 2008).

² See Kirby et al. (2008) for a detailed description of this measure.

making any statistical claims about structure: an increase in linguistic structure over time was demonstrated by a significant and continuous increase in z-scores over generations. It is important to note that the structure measure used here and in Kirby et al. (2008) only indicates whether a language is significantly structured or not, but cannot differentiate between different types of linguistic structures (e.g., systematicity in the form of structured ambiguities vs. compositional structure). Once a language has been identified as containing significant structure (with a z-score higher than 1.96), we can examine it and see what type of structure it contains. In all experiments, significantly structured languages could be the result of compositionality and/or the result of structured homonymy. Following the existing literature, the evidence for compositionality and systematicity in significantly structured languages is drawn from particular examples. Given that our meaning space was smaller than in previous studies, the maximal structure score that could be obtained in our experiments was lower than that of Kirby et al. (2008). Our maximal z-score was 7.1 in Experiment 1 and 7.3 in Experiment 2.

Following two recent papers (Beckner et al., 2017; Winter & Wieling, 2016), we analyzed performance using mixed effects regression models instead of t-tests that only compare the first and last generation (as in Kirby et al., 2008). Using regression models has several advantages: we can detect changes in linguistic parameters across the course of all ten generations, rather than just the first and the last; we can examine interactions between age group and time, check for non-linear trends in our data, and better control for differences between chains. These advantages are illustrated in a recent study that used mixed effects models to analyze new data and reanalyze Kirby et al.'s original data (Beckner et al., 2017). The full results of all models are reported in Appendix A, and the code to generate them can be found on <https://osf.io/6bx4q/>. We also provide the analyses using t-tests between the first and the last generation in Appendix B.

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.cognition.2018.08.011>.

All regression models were generated by the `lme4` and `pbkrtest` packages in R (Bates, Maechler, Bolker & Walker, 2018; Halekoh & Højsgaard, 2014; R Core Team, 2016). The p-values reported in these models were obtained using the Kenward-Roger Approximation, which gives more conservative p-values for models based a relatively small number of observations. All the models that examine changes in learnability and linguistic structure included fixed effects for GENERATION NUMBER (centered³), AGE GROUP (children vs. adults, dummy coded with adults as the reference level) and the interaction between GENERATION NUMBER and AGE GROUP. Following Beckner et al. (2017), who found a nonlinear increase in linguistic structure over time, models for structure included both linear and quadratic terms for the effect of GENERATION NUMBER.⁴ These models had the maximal random effects structure justified by the data, including random intercepts for different chains and random by-chain slopes for the effect of GENERATION NUMBER.

4. Experiment 1

This experiment used a language with three dimensions of meaning, similar to Kirby et al. (2008). In the first experiment (Experiment 1a) languages were transmitted over generations without interventions. In the original study, this lack of filtering resulted in adults' languages becoming significantly more structured and easier to learn over time,

³ We centered generation number because it is the most common practice for continuous variables, and is especially important for models with interactions, like the ones we have.

⁴ We used the `poly()` function in R to avoid collinearities between the linear and quadratic terms.

while showing a rapid decrease in the number of unique words in the language. In the second experiment, (Experiment 1b), homonyms were filtered out of the language before transmission to the next participant. Under this filtering method in Kirby et al. (2008), compositional, morphology-like structure emerged in adults - a crucial finding for the iterated learning paradigm. We tested both filtering methods (i.e., with and without filtering) with children and adults using a similar method and procedure.

4.1. Participants

Ninety children (age range: 6.5–12y, mean age: 8:8y, 41 boys and 49 girls) and 40 adults (age range: 21–68y, mean age 33y, 10 men and 30 women) took part in Experiment 1a, comprising a total of four distinct adult chains and nine distinct child chains. Fifty children (age range: 7–10y, mean age: 8:7y, 27 boys and 23 girls) and 50 adults (age range: 18–66y, mean age 34y, 21 men and 29 women) took part in Experiment 1b, comprising a total of five distinct adult chains and five distinct child chains. All child participants were visitors at the Living Lab in the Bloomfield Science Museum in Jerusalem and were recruited for this study as part of their visit in the Israeli Living Lab in exchange for a small reward. In Experiment 1a, half of the adult participants were undergraduate students at the Hebrew University (recruited for this study for credit or a small payment), and half were visitors of the Living Lab in the Bloomfield Science Museum in Jerusalem. To ensure that testing site didn't affect performance in Experiment 1a, we compared the results of the adult chains collected in the university to those collected in the science museum. The results of these analyses are reported in full in Appendix D.⁵ In Experiment 1b, all adult participants were visitors at the Living Lab. All child and adult participants were literate and native speakers of Hebrew (to a degree that enabled proficient use of the syllable bank).

4.2. Materials

We used three different types of alien figures, appearing in different colors, either alone or in a group. Thus, items varied along three semantic dimensions: alien type (A, B or C), color (blue or red) and plurality (single or plural). Stimuli included all possible combinations of these three semantic dimensions, resulting in 12 items in total (for a full list of the stimuli used in this experiment, see Appendix C). Fig. 1 shows the meaning space structure used in this experiment, with an example on either side:

At the beginning of each diffusion chain, 12 words were randomly selected from a set of 16 novel words that were created beforehand. These 12 words were then randomly assigned as labels to 12 different items, creating the initial language the first participant was trained on. The set of initial words did not contain or resemble any existing words in Hebrew (as judged by a separate sample of native speakers). All the initial words (as well as all words in later languages) were made up of eight syllables, chosen based on Hebrew phonology: “šu”, “gu”, “di”, “ki”, “so”, “mo”, “bal” and “taz”. We included CVC syllables with open vowels, which are common used in Hebrew. Initial words (and all words in later generations) were 2–3 syllables long, and could include repetition of the same syllable. For example, “*momotaz*”, “*kiso*”, “*šuguki*” and “*didi*” were possible labels for the initial language.

⁵ Because there were only two chains in each testing site, the power of this analysis is low and it should be interpreted with caution. While accuracy was higher at the university, there were no differences between the two populations in terms of creating structure, and testing site did not affect the slope of increase for learnability or structure over generations.

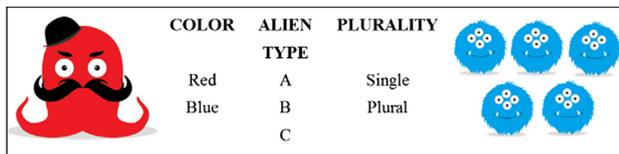


Fig. 1. The three semantic dimensions of items used in the task (color, alien type and plurality), along with an example of two items varying along these dimensions: a single red alien of type C appears on the left, and a group of blue aliens of type A appears on the right. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.3. Procedure

Participants were told they are about to learn an alien language that describes different types of aliens, and that they should try to learn it as best they can. The experiment had three stages: initial exposure, practice and test. Participants were always exposed to a random subset of the target language (SEEN items) during initial exposure and practice. Specifically, participants were trained on eight or nine out of 12 words in the language, but were tested on all 12 items, including UNSEEN items. Note that while adult participants in Kirby et al. (2008) were trained for over 45 min, such long sessions are impractical with children. In our study, participants saw each SEEN item twice (instead of six times in Kirby et al., 2008): once during the initial exposure (including active verbal reproduction) and once during practice (including active written reconstruction).

In the initial exposure phase, participants saw SEEN items (in a randomized order) displayed on the screen together with their label. The experimenter read the label out loud several times and encouraged participants to remember the pairing. Both children and adults were required to reproduce the label aloud before moving on to the next item. During the following practice phase, participants were exposed again to all SEEN items and their labels one by one in a random order. They then had to reconstruct their labels by clicking on syllables from a pre-given syllable bank. The syllable bank appeared at the bottom of the screen, and always contained the same eight syllables (see Materials) in the same order. Participants needed to successfully reconstruct the label by choosing the right syllables in the correct order to correspond with the word that appeared on the screen, and only then they could continue to the next item. Finally, participants completed a test phase: they were presented with a series of items without labels and asked to provide the correct labels according to what they've learned so far, using the same syllable bank. Importantly, the test phase contained all 12 items: the nine SEEN items, as well as the three UNSEEN items.

Transmission was implemented differently depending on the filtering method. In Experiment 1a, languages were transmitted to the next participant without intervention. We took the 12 labels produced by the participant in generation n during the test phase, and randomly selected nine of them (SEEN items) to be the input language for the next participant in generation $n + 1$. In Experiment 1b, homonyms were filtered out of the language before transmission to the next generation, so that the input language for participant $n + 1$ did not contain the same label twice. Our filter was slightly different from that used in Kirby et al. (2008): in the original study there was no limitation on the number of words that can be filtered out from the language, allowing participants to be exposed to relatively small SEEN sets (as little as a third of the language). This method was problematic to use with our smaller meaning space: it could have resulted in participants learning languages with only four or five words. To address this, we set a lower bound on our SEEN set, so it would contain at least eight unique words. As in Experiment 1a, we first randomly selected nine out of the 12 labels produced by participant n during the test phase. If this set included two or more items with the same label, we randomly removed one of those items. If there were still two or more items with the same label in the set, we randomly chose

one of those items and replaced it with one of the remaining unselected unique items in the language. If there were no items with unique labels left in participant n 's productions, we replaced the problematic label with the label produced for the same item in the previous generation $n - 1$. While this method of replacement from generation $n - 1$ could have hindered the accurate transmission over two consecutive generations, it ensured that participants were always exposed to a minimal set of eight items with unique labels while still preserving the general order of generational transmission (generation $n + 1$ can be exposed to relics of generation $n - 1$). Importantly, this replacement was relatively rare and was only implemented in 5% of cases (for both children and adults).

4.4. Experiment 1: Results

We present the results of Experiment 1a and 1b together. Fig. 2 shows the changes in learnability and structure in child and adult chains for both filtering methods (i.e., with or without filtering).⁷ Below we report the results for learnability and linguistic structure in detail. A summary of the effects can be found in Table 1. The full models can be found in Appendix A.

4.4.1. Language learnability

Under both filtering methods (i.e., with or without filtering of homonyms), GENERATION NUMBER had a strong negative effect on transmission errors, with errors significantly decreasing as generations progressed (Exp 1a: $\beta = -0.05$, SE = 0.01, $t = -4.35$, $p < 0.001$; Exp 1b: $\beta = -0.02$, SE = 0.005, $t = -5.01$, $p < 0.001$). That is, languages became easier to learn over time, with both child and adult participants making significantly fewer mistakes in later generations. There was also a significant effect of AGE GROUP, with children making more errors overall compared to adults (Exp 1a: $\beta = 0.18$, SE = 0.04, $t = 3.86$, $p = 0.0013$; Exp 1b: $\beta = 0.12$, SE = 0.03, $t = 3.56$, $p = 0.003$). Crucially, the interaction between AGE GROUP and GENERATION NUMBER was not significant in Experiment 1a ($\beta = 0.007$, SE = 0.01, $t = 0.5$, $p = 0.61$), but was significant in Experiment 1b ($\beta = 0.02$, SE = 0.008, $t = 2.6$, $p = 0.022$). In other words, learnability improved at a similar pace for children and adults when homonyms were allowed (even though children were worse learners in general; Fig. 2a). However, when homonyms were filtered out learnability improved only in adult chains (Fig. 2c).

Both children and adults created more learnable languages when underspecification was allowed (Experiment 1a), replicating previous findings (Kirby et al., 2008). This increase in learnability can be explained by the rapid decrease in the number of distinct words in the languages. Here, the number of unique words dropped to as few as only two or three words in some chains, resulting in very small lexicons. Learnability naturally increased as participants in later generations learned languages with fewer words and were less likely to make mistakes. Indeed, there was a strong correlation between the number of words in the language participants learned and the amount of errors they made, with larger lexicons being associated with more errors in

⁶ We analyze the results of the two filtering methods separately and not in a conjoined model for several reasons. First, this is the type of analyses used in the original paper we are trying to replicate: Kirby et al. (2008) performed separate analyses for the two filtering methods and treated them as separate experiments. We therefore adopt the same conceptualization and use separated analyses. More importantly, the two filtering methods are associated with different structural affordances and communicative pressures, which also lead to a different ranges of possible structure values. Conducting a direct comparison between them is not straightforward given that the no-homonym experiment simply does not allow for structured ambiguities, restricting participants' strategies and creating a different landscape for the emergence of structure. Because our main goal is to compare child and adult performance given the same input and filtering method, all our models analyze child and adult data together.

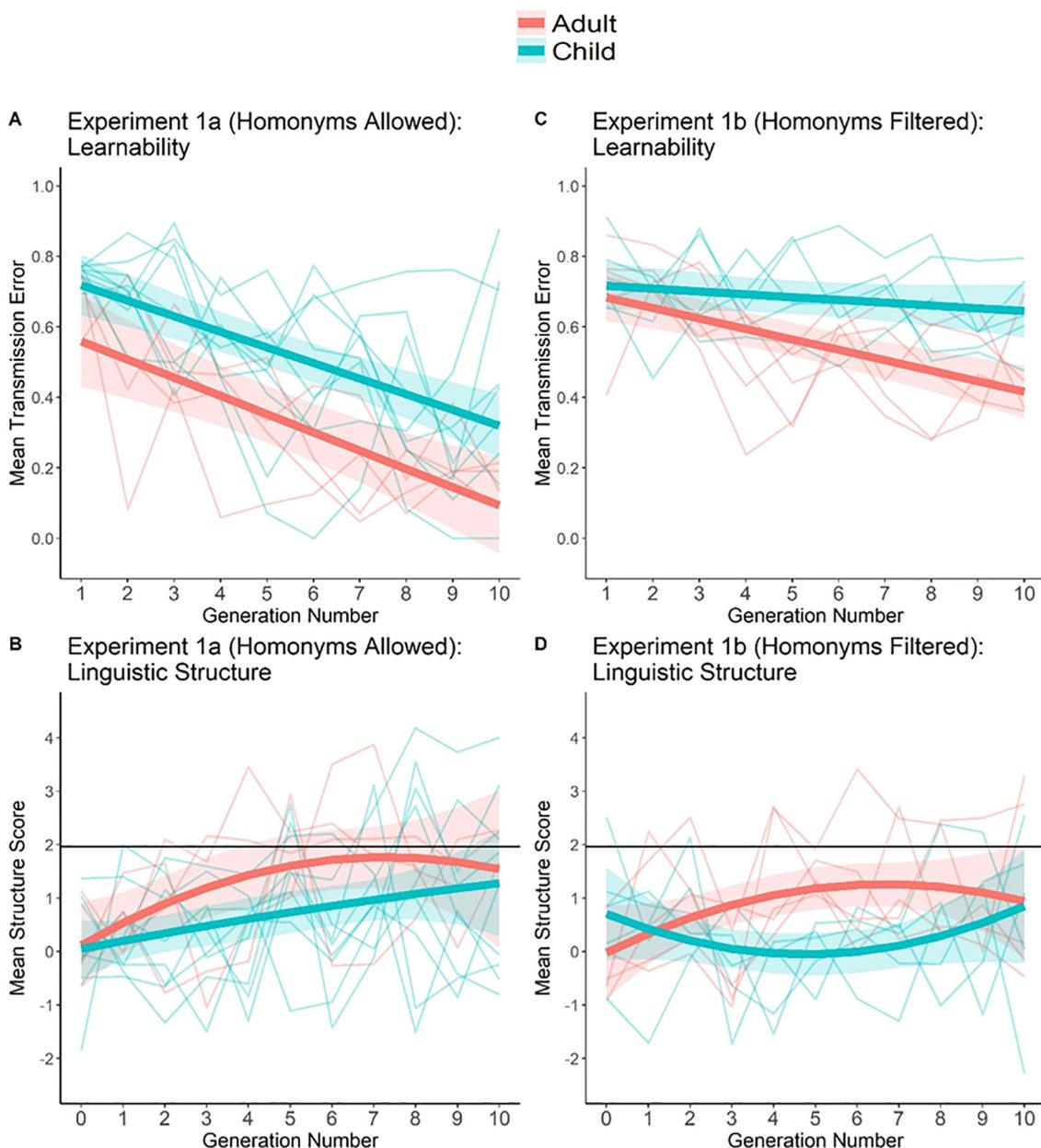


Fig. 2. Changes in learnability and linguistic structure over the course of ten generations in Experiment 1. Thin lines represent the raw data from child chains (in blue) and adult chains (in red). Thick lines represent the reported models' estimates by age group, and their shadings represent the reported models' standard errors. A decrease in mean transmission error over generations indicates an increase in language learnability over time, and an increase in structure score over generations reflects an increase in linguistic structure over time. The black line plotted in panels B and D shows the 95% confidence interval so that results above this line demonstrate that there is a nonrandom mapping between signals and meanings in the language. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

both child ($r = 0.69$, $t(88) = 9$, $p < 0.001$) and adult chains ($r = 0.66$, $t(38) = 5.4$, $p < 0.001$). That is, children and adults made fewer mistakes as the number of unique words in the language dropped.

When homonyms were filtered out and languages always contained the same number of unique words (Experiment 1b), adults still created more learnable languages over time (presumably thanks to the creation of some linguistic structure), while children did not. Even though children in this filtering method did not show an *increase* in learnability over generations (i.e., they did not make fewer mistakes over time), they did show evidence of learning. To demonstrate that children understood the task and were trying to reproduce their input, we manually analyzed children's errors in this experiment. We evaluated the similarity between children's errors and the labels they heard, and

found that 47% of errors were *label mismatch errors*, where children used a previously heard label with the wrong object (i.e., using the label *dikiso* for one object even though it appeared with a different label during training). In these cases, children remembered the learned label in its exact form, but had difficulty associating it with the right object. 10% of errors were *mispronunciation errors*, where children produced labels that differed in only one syllable from the correct label (e.g., *dikigu* instead of *dikiso*). In these cases, children correctly remembered the mapping between labels and items, but had difficulty reproducing the exact form. An additional 32% of children's errors were *mixed labels* errors that differed by only one syllable from *some* previously heard label, even if not the correct one (that is, a mix of *mismatch* and *mispronunciation* errors, or a blend of previously heard labels; e.g., using

Table 1
Summary of effects in Experiment 1.

Experiment 1	Experiment 1a (Homonyms allowed)		Experiment 1b (Homonyms filtered out)	
	Children	Adults	Children	Adults
	Learnability increase	Yes 	Yes 	No
Linguistic Structure increase	Marginal 	Marginal 	No 	No

the label *dikiso* for one object when the label *dikigu* appeared with a different one during training). In total, 89% of children’s errors had a systematic relation to what they heard during training. Only 11% of errors were complete innovations, where children produced labels that bore no relation to their input. This analysis shows that children in Experiment 1b were indeed engaged with the task in the expected way, and that they were both reproducing and changing their input in non-random ways (despite not showing a cumulative change over generations).

4.4.2. Language structure

The effect of generation on structure emergence was different in the two filtering methods. When homonyms were allowed (Experiment 1a), linguistic structure marginally increased over generations in a linear fashion ($\beta = 5.3$, $SE = 2.85$, $t = 1.88$, $p = 0.08$), with participants in later generations creating marginally more structured languages. AGE GROUP had a significant effect on performance in this experiment ($\beta = -0.58$, $SE = 0.2$, $t = -2.2$, $p = 0.045$), with adults creating significantly more linguistic structure compared to children. There was no significant interaction between AGE GROUP and GENERATION NUMBER, indicating that even though adults’ languages were more structured overall, time affected structure similarly across age groups (Fig. 2b). When we examined linguistic structure in individual languages, we found that systematic, non-random mappings between meanings and labels were accomplished by both children and adults multiple times when homonyms were allowed. 16 out of 90 child languages (18%) and 15 out of 40 adult languages (37%) were significantly structured (defined as languages with z-scores higher than 1.96). As in Kirby et al. (2008), both children and adults created languages with structured ambiguities where homonyms were systematically associated with differences in one of the semantic dimensions. For example, in one child chain the final language converged to three distinct words representing alien type, regardless of color and quantity: all aliens of type A were called “*didi*”, all aliens of type B were called “*balgu*” and all aliens of type C were called “*šuki*” (Fig. 3). Similar structure emerged in adults’ chains. In a different child chain, systematic structure emerged in generation 8, and was transmitted flawlessly to the last two participants. This language converged to just two distinct words representing alien color, regardless of type and quantity: “*ditaz*” for all red aliens and

	Alien A	Alien B	Alien C	
Red	didi	balgu	šuki	Single
	didi	balgu	šuki	Plural
Blue	didi	balgu	šuki	Single
	didi	balgu	šuki	Plural

Fig. 3. A significantly structured child language in generation 10 in Experiment 1a.

	Alien A	Alien B	Alien C	
Red	ditaz	ditaz	ditaz	Single
	ditaz	ditaz	ditaz	Plural
Blue	balšu	balšu	balšu	Single
	balšu	balšu	balšu	Plural

Fig. 4. A significantly structured child language in generations 8 to 10 in Experiment 1a.

“*balšu*” for all blue aliens (Fig. 4).

Yet when homonyms were filtered out (Experiment 1b), the effect of GENERATION NUMBER was not significant. That is, structure did not increase over time in this filtering method in either children or adults, unlike the findings of Kirby et al. (2008). Interestingly, the interaction between AGE GROUP and the quadratic term of GENERATION NUMBER was significant ($\beta = 5.7$, $SE = 2.6$, $t = 2.1$, $p = 0.049$), suggesting that there were different curves of change in structure for children and adults: while structure changed concavely in adult chains, it changed convexly in child chains (Fig. 2d). However, neither group showed a significant increase in structure overall. As in Experiment 1a, AGE GROUP significantly affected performance ($\beta = -0.6$, $SE = 0.2$, $t = -2.5$, $p = 0.023$) with adults displaying more structure than children.

Even though there was no increase in linguistic structure over time when homonyms were filtered out, there were several cases of systematic languages: 12 out of 50 adult languages were significantly structured (24%), but only 4 out of 50 child languages were significantly structured (8%). While these languages contained irregularities and were still partly ambiguous, it seems that morphology-like structure was beginning to emerge. For example, in one adult chain the final language contained a systematic suffix “*ki*” representing plurality, and a unique label “*šumo*” for blue items (Fig. 5). Additionally, red items were almost always referred to using a distinct affix “*bal*”. Despite its ambiguity (due to coding only some semantic dimensions), this language does exhibit compositionality. In child chains, the evidence for compositionality was rarer. In one child chain, there seemed to be 3 distinct words representing each alien type regardless of quantity: most aliens of type A were called “*šugu*”, most aliens of type B were called “*tazmo*” and most aliens of type C were called “*tazšu*” (Fig. 6). Aliens of type B were also marked for their color using a suffix “*gu*” for red items and a suffix “*šu*” for blue items. Similarly, aliens of type C may be marked for plurality (“*mo*” for single aliens and “*gu*” for multiple aliens).

	Alien A			Alien B			Alien C			
Red	bal	mo		bal	taz		bal	shu		Sg
	bal	taz	ki	di	mo	ki	bal	mo	ki	Pl
Blue	shu	mo		shu	mo		shu	mo		Sg
	shu	mo	ki	shu	mo	ki	bal	mo	ki	Pl

Fig. 5. A significantly structured adult language in generation 10 in Experiment 1b.

	Alien A			Alien B			Alien C			
Red	shu	mo	gu	taz	mo	gu	taz	shu	mo	Sg
	shu	gu		taz	gu		taz	shu	gu	Pl
Blue	shu	gu		taz	mo	shu	taz	mo	shu	Sg
	shu	gu		taz	mo	shu	taz	shu	gu	Pl

Fig. 6. A significantly structured child language in generation 8 in Experiment 1b.

4.5. Experiment 1: Discussion

When homonyms were not filtered out and underspecification was possible (Experiment 1a), children and adults behaved similarly: their languages became marginally more structured over time and significantly easier to learn as generations progressed, at a similar pace and using the same strategy of structured ambiguities. By reducing the number of distinct labels in the language to as little as two or three words, children and adults were able to create simple and systematic lexicons in which a small number of homonyms were consistently assigned according to some semantic dimension (e.g., alien type). Significantly structured languages (with non-random signal-to-meanings mapping) emerged in both child and adult chains on multiple occasions, but adults were more likely to create structured languages overall. Importantly, despite adults' overall superiority in this non-filtering method (making fewer mistakes and creating more structured languages overall), the effects of cultural transmission on the languages' structure and learnability were similar in both age groups.

However, when underspecification was prevented by imposing an artificial filter on the transmission of homonyms (Experiment 1b), children and adults differed in their performance. First, only adults' languages became easier to learn as generations progressed, while children showed no evidence for a decrease in errors over time (though their errors were systematic and did demonstrate that they were trying to reproduce their input). Second, adults outperformed children again by making fewer mistakes and creating more linguistic structure overall. Crucially, when homonyms were filtered out, linguistic structure did not increase over time in this filtering method for neither children nor adults, although there were rare occasions of significantly structured languages emerging in some chains. The results of Experiment 1b highlight adults' better performance in this paradigm.

Importantly, we cannot draw conclusions about child-adult differences from this experiment since our results did not replicate the Kirby et al. (2008) findings for adults' languages did not become more structured, and compositionality did not emerge. These results may reflect the greater difficulty participants had in learning the language in our study due to its' shorter exposure period. While participants in Kirby et al. (2008) received six exposures to the input language during the learning phase, our participants received only a third of that exposure. It is possible that when languages were not allowed to simplify through a reduction in the number of words, they were too complex for both children and adults to learn properly within the allocated exposure time, so that child-adult differences may have been masked by the difficulty of the language.

To address this, we ran a second set of studies (Experiments 2a and 2b) using the same paradigm, but with a simplified language that includes only two semantic dimensions: alien type and motion. In light of children's documented difficulty with learning multiple features of novel items (Perry et al., 2015), we simplified the meaning space by reducing the number of features participants needed to attend to. We replaced aliens' color and plurality features with motion, creating more salient and event-like scenes in which a given alien would move on a screen in a certain way (e.g., bouncing, spiraling). We chose to simplify the language by reducing the number of semantic dimensions rather than by reducing the number of words since we wanted the language to be large enough to introduce a pressure for developing compositionality. If the number of labels was reduced even further, participants could possibly remember all labels without any need to introduce structure. We also gave children an additional exposure to each label when homonyms were filtered out (Experiment 2b) to compensate for their greater difficulty with artificial language learning. We did not increase child exposure when homonyms were allowed (Experiment 2a), since children showed increased learnability in this filtering method even with the more complex language used in Experiment 1a. In sum, Experiment 2 included two modifications that were meant to improve participants' learning: a simplified meaning space, and

additional exposure for children in the more difficult no-homonym filtering method (Experiment 2b).

5. Experiment 2

This experiment was modeled on Experiment 1, and also included two filtering methods: one without filtering homonyms where underspecification is allowed (Experiment 2a), and one where homonyms are filtered out of the language (Experiment 2b). We used a meaning space with only two semantic dimensions to make the language easier to learn for both children and adults. In addition, children in Experiment 2b received an additional exposure to all items.

5.1. Participants

Fifty children (age range: 7–12y, mean age: 9y, 29 boys and 21 girls) and 50 adults (age range: 20–31y, mean age 24y, 21 men and 29 women) took part in Experiment 2a, comprising a total of five distinct adult chains and five distinct child chains. Fifty children (age range: 7–12y, mean age: 9:2y, 24 boys and 26 girls) and 50 adults (age range: 20–31y, mean age 23.5y, 16 men and 34 women) took part in Experiment 2b, comprising a total of five distinct adult chains and five distinct child chains. All child participants were visitors at the Bloomfield Science Museum in Jerusalem and were recruited for this study as part of their visit in the Israeli Living Lab in exchange for a small reward. All adult participants were undergraduate students at the Hebrew University, recruited for this study for credit or a small payment. All participants were native and literate speakers of Hebrew.

5.2. Materials

We used three different types of alien figures that moved in different ways on the screen. Thus, items varied along two semantic dimensions: alien type (A, B or C) and motion type (bouncing, spiraling, diagonal or straight line). Stimuli included all possible combinations of these two semantic dimensions, resulting in 12 possible items (for a full list of the stimuli used in this experiment, see Appendix C). Fig. 7 shows the meaning space structure used in this task, with an example on either side:

As in Experiment 1, at the beginning of each diffusion chain 12 words were randomly drawn from a closed set of 16 novel words and were randomly assigned as labels to all 12 items to create the initial language on which the first participant was trained.

5.3. Procedure

Experiment 2a was identical to Experiment 1a. Experiment 2b was identical to Experiment 1b, with one modification to the exposure phase: children (but not adults) received an additional exposure to all SEEN items. This was done by repeating the initial exposure phase (including active verbal reproduction), resulting in a total of three exposures to each word in the language.

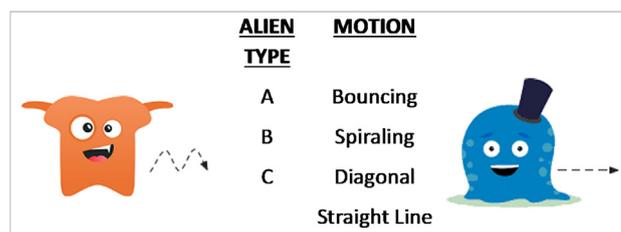


Fig. 7. The two semantic dimensions of items in the task (alien type and motion), along with an example of two items varying along these dimensions: a bouncing alien of type A appears on the left, and an alien of type B moving in a straight line appears on the right.

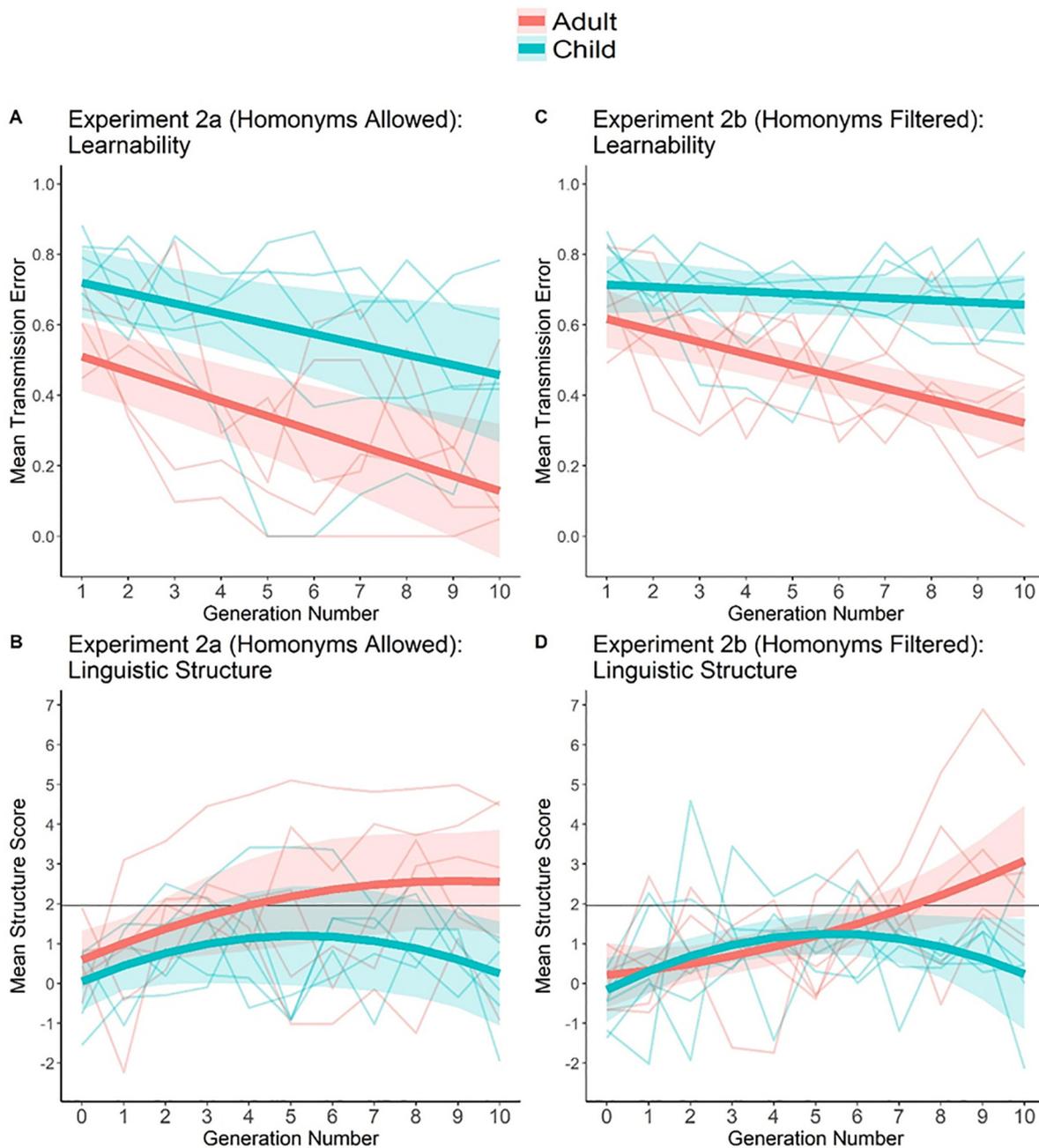


Fig. 8. Changes in learnability and linguistic structure over the course of ten generations in Experiment 2. Thin lines represent the raw data from child chains (in blue) and adult chains (in red). Thick lines represent the reported models' estimates by age group, and their shadings represent the reported models' standard errors. The black line plotted in panels B and D shows the 95% confidence interval so that results above this line demonstrate that there is a nonrandom mapping between signals and meanings in the language. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.4. Experiment 2: Results

Fig. 8 shows the changes in learnability and structure in child and adult chains for both experiments⁷. Below we report the results for learnability and linguistic structure in detail. A summary of the effects can be found in Table 2.

⁷ The y-axes in Fig. 8b and 8d have a different range from that of the equivalent plots in Fig. 2 because structure scores were higher in Experiments 2a and 2b. We did not match the y-axes across the two figures because doing so made it harder to see the differences between children and adults in Fig. 2.

5.4.1. Language learnability

As in Experiment 1, GENERATION NUMBER had a strong negative effect on transmission error in both experiments, with errors significantly decreasing over generations (Exp 2a: $\beta = -0.04$, SE = 0.01, $t = -4.1$, $p = 0.0012$; Exp 2b: $\beta = -0.03$, SE = 0.006, $t = -4.8$, $p < 0.001$).

Crucially, and as we found in Experiment 1, the interaction between AGE GROUP and GENERATION NUMBER was not significant when homonyms were allowed (Exp 2a: $\beta = 0.013$, SE = 0.01, $t = 0.9$, $p = 0.37$), but was significant when they were filtered out (Exp 2b: $\beta = 0.026$, SE = 0.009, $t = 2.7$, $p = 0.016$). That is, learnability improved at a similar pace across age groups when languages were allowed to deteriorate (Experiment 2a; Fig. 8a), but seemed to improve only in the adult chains when underspecification was prevented (Experiment 2b;

Table 2
Summary of effects in Experiment 2.

Experiment 2	Experiment 2a (Homonyms allowed)		Experiment 2b (Homonyms filtered out)	
	Children	Adults	Children	Adults
	Learnability increase	Yes 	Yes 	No
Linguistic Structure increase	Yes 	Yes 	No 	Yes

Fig. 8c). We also found that AGE GROUP affected performance in both experiments, with children making significantly more errors than adults (Exp 2a: $\beta = 0.26$, $SE = 0.08$, $t = 3.2$, $p = 0.007$; Exp 2b: $\beta = 0.2$, $SE = 0.03$, $t = 5.09$, $p < 0.001$).

Similar to Experiment 1a and to Kirby et al (2008), the languages of both children and adults in Experiment 2a were characterized by a rapid decrease in the number of unique words, with the number of words dropping to as few as only two words in certain chains. In Experiment 2a, the number of words in the language was strongly correlated with transmission error for both children ($r = 0.72$, $t(48) = 7.2$, $p < 0.001$) and adults ($r = 0.56$, $t(48) = 4.7$, $p < 0.001$), so that learning languages with fewer words elicited fewer mistakes. To ensure that children were engaged with the task in the expected way and that they were attempting to reproduce their input, we analyzed children’s errors in the harder scenario where homonyms were filtered out (Experiment 2b) in the same way we did for Experiment 1b. Here, we found that 63% of the errors were label mismatches, 9% were mispronunciations errors, 24% were mixed labels, and only 4% of errors included complete innovations. That is, learning was better in Experiment 2b (as predicted given the simpler meaning space and the additional exposure), so that almost all of children’s errors were closely linked to their input. Even though learnability did not increase overall, this analysis shows that children’s errors were not random, but had a systematic relation to their input.

5.4.2. Language structure

In Experiment 2a, we found a significant effect of GENERATION NUMBER, with structure increasing linearly over time for both age groups ($\beta = 6.4$, $SE = 2.6$, $t = 2.4$, $p = 0.028$; Fig. 8b), replicating the effect found in Kirby et al. (2008). The interaction between GENERATION NUMBER and AGE GROUP was not significant (linear term: $\beta = -5.85$, $SE = 3.6$, $t = -1.5$, $p = 0.13$; quadratic term: $\beta = -1.6$, $SE = 3.2$, $t = -0.5$, $p = 0.62$), though a visual inspection of the data (Fig. 8d) suggests the effect was stronger in adults. Our modified paradigm led to a significant increase in structure that was not found in Experiment 1, and that is also reflected in the larger effect sizes for the effect of GENERATION NUMBER in this version ($\beta = 5.38$ for Experiment 1a vs. $\beta = 6.49$ for Experiment 2a). The effect of AGE GROUP was marginal ($\beta = -1.16$, $SE = 0.6$, $t = -1.85$, $p = 0.08$), with adults creating marginally more linguistic structure compared to children. That is, children seem to show a similar (albeit weak) increase in structure as adults when homonymity was allowed. Significantly structured languages (with consistent mappings between meanings and signals) were accomplished by more than half of the adult learners (54% of adult languages) and by 18% of children. While the evidence for children creating structure is not particularly strong, looking at these significantly structured languages confirmed that when structure was present, homonyms were not assigned at random, and that children were able to create languages with

structured ambiguities along one semantic dimension (more commonly, Alien type) similar to the adults in this experiment.

A different pattern emerged when homonyms were filtered out (Experiment 2b): structure increased linearly over generations ($\beta = 9.5$, $SE = 2.5$, $t = 3.6$, $p < 0.01$), but this effect was qualified by a significant interaction between GENERATION NUMBER and AGE GROUP (linear term: $\beta = -8.2$, $SE = 3.6$, $t = -2.2$, $p = 0.044$; quadratic term: $\beta = -6.1$, $SE = 2.6$, $t = -2.3$, $p = 0.038$), suggesting that structure increased only in the adult chains, but not in the child chains (Fig. 8d). As in Experiment 2a, the effect of AGE GROUP was marginal ($\beta = -0.6$, $SE = 0.3$, $t = -1.9$, $p = 0.07$), meaning adults showed marginally more structure compared to children overall. Together, these findings replicate the increase in structure found in Kirby et al. (2008) in adults, and document the lack of it in child learners.

When homonyms were filtered out, 34% of adult languages were significantly structured, and we expected these languages to show compositional structure in which sub-elements are systematically re-used to encode different meanings. This was indeed the case, with one adult chain even showing a “perfect” compositional system, which was fully expressive and encoded all dimensions of meaning (Fig. 9). This language was also transmitted flawlessly from generations 8 to 10. In this case, each alien type has a unique affix (“gu” for Alien A, “di” for Alien B, “šu” for Alien C) and each motion type has a unique suffix (“šu” for bouncing, “di” for diagonal, “bal” for spiraling and zero marking for movement in a straight line). In addition, all words in the language share an internal systematic structure in which the second syllable is always “ki”. Interestingly, this language also presents clear combinatorial structure, in which the same syllable expresses alien type when it is in an initial position, but expresses motion type when it is in a final position (e.g., “šu”, “di”). Children created fewer structured languages in this filtering method compared to adults (18%), and these were scattered across different generations, suggesting that even when structure was introduced, it was often not well transmitted. Since we did not restrict the production of homonyms during test (only the transmission of those homonyms to the next generation), many of the significantly-structured child languages still contained structured ambiguities. For example, one child language in generation 3 (Fig. 10) showed a unique affix for alien type as found in Experiment 1a and 2a (“šu” for Alien A, “tazgu” for Alien B and “taz” for Alien C), and also contained a few partly-systematic suffixes for some motion types, seemingly grouping together bouncing and stright line motions (marked with “so” or “ki”) vs. diagonal and spiral motions (marked with “di” or “taz”). Traces of this structure persist in the following three generations of children, yet were not picked up by the next child, who dramatically regresses the language back to a random state.

Alien A			Alien B			Alien C			
gu	ki	shu	di	ki	shu	shu	ki	shu	Bouncing
gu	ki	di	di	ki	di	shu	ki	di	Diagonal
gu	ki		di	ki		shu	ki		Straight line
gu	ki	bal	di	ki	bal	shu	ki	bal	Spiral

Fig. 9. A significantly structured adult language in generations 8 to 10 in Experiment 2b.

Alien A			Alien B			Alien C			
shu	mo		taz	gu	ki	taz	so	taz	Bouncing
shu	di		taz	gu	so	taz	taz		Diagonal
shu	ki		taz	gu	taz	taz	so	gu	Straight line
shu	di		taz	gu	di	taz	taz		Spiral

Fig. 10. A significantly structured child language in generation 3 in Experiment 2b.

5.5. Experiment 2: Discussion

Simplifying the meaning space by reducing the number of semantic dimensions allowed us to replicate the findings of Kirby et al (2008) for adults, and to observe a somewhat different pattern for children. Both children and adults were capable of creating systematic languages when underspecification was possible (Experiment 2a), but only adults created compositional languages when this strategy was not available (Experiment 2b). These results are in line with the findings of Experiment 1, but include two major differences with respect to linguistic structure: (1) The increase in linguistic structure was significant when homonyms were allowed for both age groups; and (2) Adults, but not children, now showed the expected increase in structure when homonyms were filtered out.

When languages were allowed to deteriorate (Experiment 2a), learnability and structure increased similarly in both populations (though the increase in structure may be weaker in children). Yet when underspecification was prevented (Experiment 2b), children and adults differed in their performance on both measures. Despite being given additional exposure in this filtering condition, children still showed no increase in learnability or in linguistic structure over time. This experiment now replicates Kirby et al. (2008) in showing the emergence of compositional structure with adult participants (see Fig. 9 for an example of a fully expressive compositional language), while highlighting the lack of such emerging compositionality in child languages.

6. General discussion

Over the past decade, many studies have used iterated language learning paradigms to explore how learning biases impact linguistic structure. This paradigm has generated exciting empirical findings supporting the feasibility of cultural transmission as an explanation for how linguistic structure came about. However, these studies have focused only on adult learners, and no study to date has examined similar trends with children. Given that children, and not adults, are the primary language learners in real-life transmission, this gap in the literature limits the generality of the findings obtained from the paradigm. Moreover, comparing child and adult performance can help evaluate the contrasting predictions regarding children's ability to introduce linguistic structure compared to adults. To address these issues, we compared children and adults' performance in an iterated language learning task, asking if children also show an increase in linguistic structure over time, and if they differ from adults' in the effect of cultural transmission on the emerging languages. We conducted a series of four studies using a novel child-friendly paradigm, comparing the learnability and structure of the artificial languages created by children and adults in two experimental scenarios: when homonyms were allowed and when they were filtered out during transmission (similar to Kirby et al., 2008). Taken together, our results reveal interesting differences between children and adults: while both children and adults were capable of introducing structured ambiguities, only adults showed evidence of introducing compositional structure.

The first study examined performance on a language with three dimensions of meaning. We found that when homonyms were allowed (Experiment 1a), the languages of both children and adults became more learnable over time, and both groups showed a marginal increase in structure. In the crucial experiment where homonyms were filtered out before transmission (Experiment 1b), only adults' languages showed an increase in learnability, and neither group showed a significant increase in linguistic structure. While children and adults differed in their performance, it is hard to draw clear conclusions from this first study since adult learners did not show the expected increase in structure when homonyms were not allowed (Kirby et al., 2008). This may have been caused by the shorter exposure time in our study, making the languages too complex to learn. The second study explored this explanation by using a simplified language with only two

dimensions of meaning, and giving children one additional exposure to the language. Adults now showed the expected pattern, with structure and learnability increasing in both filtering conditions. Children, in contrast, showed no increase in structure when homonyms were filtered out. These results replicate Kirby et al. (2008) with adults and suggest that children differ from adults in their ability to create compositional structure.

These findings have implications for our understanding of the role of different learners in the process of cultural transmission and in the emergence of structure. For starters, they argue caution in generalizing from adult iterated learning studies to learners in general. Across experiments, children differed from adults in their ability to both learn the language and add structure to it. Regardless of the underlying source of this difference (which we discuss below), its existence highlights the difficulty of extending claims based on adult findings to child learners (i.e. adult biases may not reflect those of all learners), and underscores the difficulty of characterizing the process of cultural transmission using only adult learners. Naturally, cultural transmission does not involve only learning, but also production and transmission: each of which could have a different effect on the emergence of structure depending on the agent who is producing/transmitting. While children may not be the prototypical transmitters (in the sense that their language is not readily adopted by adult speakers; e.g., Kerswill & Williams, 2000; Nettle, 1999), they are indeed the prototypical learners. Similarly, while adults are the typical transmitters of languages, they do not usually acquire the language they transmit as adults. This makes the performance of both children and adults relevant for evaluating the iterated learning paradigm and its outcomes, and stresses the need to seriously consider that cultural transmission may impact the emergence of structure differently in children and adults.

The lack of compositional structure in our child chains could stem from several different sources. Children may have weaker biases for structure (Smith et al., 2017), an explanation consistent with their overall lower structure scores in our experiments. Children's structural biases could also differ from those of adults, such that given the same data, they will make different structural innovations. For example, children may weigh word length more than adults (e.g., prefer shorter labels), show different preferences for what to mark as a prefix or suffix, have different preferences for sequential ordering of affixes, or differ in their reliance on morphology vs. syntax to convey novel structural relations. Our data cannot be used to evaluate such differences directly because the type of structure that could emerge was restricted to single words (so only morphological structure could emerge), and was further limited by various aspects of the experimental design (such as the use of a syllable bank which prevented novel sounds from being introduced, and the restriction that words could only be two or three syllables long). That is, using this paradigm and measures does not allow us to detect differences in the kind of biases used by children and adults, even though such differences may be present in real-life emergence situations. The idea that learner age can impact the strength and kind of structural biases finds some support in computational simulations of language change using child and adult agents, where assuming age-related differences in learning affected the spread of innovations in a community (Ke, Gong & Wang, 2008) and the prevalence of complex inflectional morphology (Dale & Lupyan, 2012).

An additional (and probably complementary) explanation is that the difference in the introduction of structure is related to how well children and adults learned the language. In all four experiments, adults significantly outperformed children in learning the artificial languages despite having the same (or less) exposure. The lack of compositional structure in child chains may be related to children's inferior learning, and reflect a deep connection between learning and generalization. Put differently, children may have similar structural biases as adults, but may simply be worse at learning the original input, and hence limited in the structure they can add to it. This idea is compatible with the only other iterated learning study that compared children and adults on a

non-linguistic task: Kempe et al. (2015) found that when children and adults had similar learning curves (i.e., showing similar transmission accuracy), children were able to add structure to visual patterns (and even did so more than adults). This idea that learning the input is crucial for generalization is also compatible with findings from artificial language learning studies, where children generalize more when the distribution or the variability of exemplars is changed to enhance learning (Boyd & Goldberg, 2012; Wonnacott, Boyd, Thomson, & Goldberg, 2012; Wonnacott et al., 2017). Of particular relevance, Boyd & Goldberg (2012) showed that children do not generalize an abstract construction if they have not learned the relevant pattern in the original input. In their words: “A pattern must be implicitly recognized in order to be generalized” (p. 476). Based on these findings we can predict that children’s difficulty in learning the artificial language may have affected their ability to regularize and introduce structure.

This prediction can be directly examined by looking at the relation between learning and structure in our data. In most iterated learning studies, languages are evaluated on the basis of two measures: how learnable they are (reflected by transmission error) and how structured they are (reflected by the correlation between meanings and labels). The relation between those two measures is typically not investigated, even though it can be highly informative for understanding how learning affects the accumulation of linguistic structure. If learning impacts the introduction of structure, there should be a significant relationship between the two measures: better learning (fewer mistakes) should be associated with more structure.⁸ We tested this prediction in the data from all four experiments by using mixed effects regression models where the dependent variable was the structure score, and the fixed effects were TRANSMISSION ERROR (centered), AGE GROUP (children vs. adults, dummy coded with adults as the reference level) and the interaction between them (see full models in Appendix A). We also included random intercepts and random slopes for the effect of TRANSMISSION ERROR with respect to different generations. The results are striking and consistent: TRANSMISSION ERROR was a significant predictor of linguistic structure in all four experiments (Exp 1a: $\beta = -2.3$, $SE = 0.86$, $t = -2.6$, $p = 0.009$; Exp 1b: $\beta = -2.4$, $SE = 1.1$, $t = -2.18$, $p = 0.03$; Exp 2a: $\beta = -3.68$, $SE = 0.89$, $t = -4.13$, $p < 0.001$; Exp 2b: $\beta = -5.74$, $SE = 1.1$, $t = -5.1$, $p < 0.001$). Importantly, the interaction between TRANSMISSION ERROR and AGE GROUP was not significant in any experiment, suggesting that children and adults showed the same pattern.

These results indicate that participants who showed better learning of the input language also created more linguistic structure during the test. They also highlight an important parallel between children and adults: both showed the same relationship between how well they learned and how much linguistic structure they created. These analyses support the link between learning and structure creation and suggest that children’s lack of structure was related to their inferior learning. That is, participants who did not show sufficient learning were less likely to introduce novel structure. This link is maintained in a one-generation paradigm. In a series of studies, Johnson, Siegelman, and Arnon (in preparation) exposed adult participants to a semi-structured language with the same meaning space used in Experiment 1 (three dimensions: color, shape, plurality) where each object was assigned a label with three syllables, so that each of the syllables probabilistically predicted a semantic dimension (e.g., the syllable ‘ka’ was predictive of red, but only 84% of the time). Importantly, learning was assessed separately for SEEN and UNSEEN items ($N = 24$ for each), providing an independent measure for learning and structure. In line with our prediction, there was a high correlation between accuracy on SEEN items and the amount of added structure for UNSEEN items. Taken together, the results point to the importance of examining the relationship

⁸ This should be true as long as learning is not perfect: if the input is reproduced in full, then no structure is added.

measures of learning and structure in iterated learning paradigms and predict that similar patterns can be found in existing iterated learning data, and in other artificial language learning settings. As long as participants do not reproduce the language as is, those who show better learning should also introduce more structure.

Importantly, this study is only a first step in examining child performance using linguistic iterated learning, and as such, has several limitations. The most obvious one has to do with how well children learned: even though our error analyses showed that children were engaged in the task and attempted to reproduce their input, there was considerable variation in how well they learned. While some children were able to faithfully reproduce the language, others displayed relatively poor learning even when languages were structured. As we suggest above, the lack of compositional structure may stem from children’s overall difficulty in learning the artificial language given the short exposure. We are currently testing this prediction by increasing children’s exposure and by introducing real communication between pairs of participants in the child chains, which leads to compositionality in adults (Kirby et al., 2015). An additional issue has to do with the ages we looked at. Since literacy is required in our child-friendly paradigm, it is still not suitable for testing younger children, which limits the generality of our conclusions. It is possible that younger children (under six) will show similar structural biases to adults, and differ from the older children tested in our studies. This possibility is unlikely given younger children’s greater difficulty with artificial language learning (Ferman & Karni, 2010), but can be tested by modifying the existing paradigm to allow for verbal rather than written production. Finally, the structure scores in our studies were somewhat lower than what was found in Kirby et al. (2008). The most structured language in Kirby et al. (2008) had a z-score of 13, while the most structured language in our study (Fig. 9) had a structure score of 6.88. Lower structure scores were also found in Beckner et al. (2017), where most structured languages had z-scores between 3 and 6. Importantly, the lower scores in our study were mainly driven by the smaller lexicon we used, which inherently leads to lower possible structure scores given that z-scores are dependent on the number of observations. Interestingly, this difference in structure scores may also be related to the L1 of the participants in our sample: all previous iterated language learning studies have been conducted with English learners, while our studies were done with L1 speakers of Hebrew, a language with non-concatenative morphology. Because the use of a syllable bank prevented Hebrew speakers from using non-concatenative morphology (since sounds within syllables cannot be changed), it may better fit English speakers’ existing morphological biases. We are currently investigating this possibility by using the exact same language with English speakers to see if they will generate languages with higher structure scores. Such a finding would strengthen the concern that existing linguistic knowledge impacts the emergence of linguistic structure in these paradigms. Further research is needed to investigate the effect of prior linguistic knowledge on linguistic outcomes in iterated language learning studies.

The current findings also have implications for the debate on the relative role of children and adults in the emergence of linguistic structure. On the one hand, children regularize more than adults in some artificial language learning studies (e.g., Hudson Kam & Newport, 2005, 2009), and were shown to have a special role in creating core linguistic properties in the emerging Nicaraguan sign language (Senghas & Coppola, 2001; Senghas et al., 2004). On the other hand, children seem to generalize less than adults in other artificial language learning studies (e.g., Boyd & Goldberg, 2012), and data from creole languages suggests that they play a smaller role in the process of creolization (e.g., Arends & Bruyn, 1995). In the current study, children played a smaller role than adults in creating structure as well. These seemingly contradictory predictions and findings may be reconciled by (a) differentiating between different types of linguistic structures, and (b) distinguishing between introducing structure and reducing variation. The emerging sign language and creole literatures have typically

focused on different aspects of grammar. The emerging Nicaraguan Sign Language literature tends to look at children's role in the emergence of grammatical properties such as linear sequencing, segmentation and spatial modulations, which involve re-ordering or breaking up existing linguistic elements. The creole literature, on the other hand, has mainly focused on children's role in the development of comparative and recursive structures such as embedded clauses, which require the introduction of novel linguistic elements such as complementizers. Given this difference, talking about children's role in the emergence of language structure as a whole may obscure differences between different aspects of language: children and adults may play different roles in the emergence of different grammatical properties.

In particular, child learners may play a larger role in systematizing the language, but a smaller role in introducing novel regularities. The experimental literature on child and adult regularization patterns provides some support for this idea. Interestingly, children seem to regularize more than adults only in the sense of eliminating variation, but not in the sense of introducing predictable or novel conditioning: when faced with unpredictable variation in artificial languages, children tend to systematize the language by over-producing one form, while eliminating the other (Hudson Kam & Newport, 2005, 2009; Schwab, Casey & Goldberg, 2018). Adults, on the other hand, tend to preserve the variation and regularize it by making the different forms lexically or semantically conditioned (Samara et al., 2017). In our paradigm, children showed a similar tendency to introduce systematicity by eliminating variation and over-using one form in a semantically conditioned way. When homonyms were allowed, children took one label for a given item, and used it for other items that shared some semantic dimension with the original item, eliminating all other labels. While children were capable of introducing such systematic conditioning into the language, they did not "regularize" it in the same way adults did when homonyms were filtered out (i.e., by introducing novel compositional structure). That is, it seems like "adults regularise, while children systematise" (Hudson Kam & Newport, 2009; p. 31). This does not mean that children do not play a unique role in introducing linguistic structure, but rather suggests that the scope of that role depends on the linguistic properties in question.

7. Conclusions

We conducted four experiments to test the effect of cultural transmission on the emergence of linguistic structure in children and adults. To date, no study has examined the emergence of linguistic structure in child learners, though children's performance is crucial for validating previous findings and evaluating contrasting predictions on their role in the emergence of structure. We used a modified linguistic iterated learning paradigm with two filtering methods similar to Kirby et al. (2008) to ask whether children, like adults, will also show an increase in linguistic structure and language learnability over time. We found that children differed from adults in their ability to introduce structure to the language: while children were able to create systematic, underspecified languages with structured ambiguities similar to those created by adults when homonyms were allowed, they did not develop compositionality over time when homonyms were filtered out of the language. Importantly, children showed worse learning than adults in all our experiments. Our findings are the first to document differences in the effect of cultural transmission on children and adults' artificial languages. These differences could reflect meaningful variation in learning biases, and may point to a deeper link between learning and generalization. Children's more limited learning capacities may have hindered their ability to generalize and introduce compositional structure, highlighting the importance of learning in generalization. This prediction is supported by the relation we find between transmission error and structure scores: participants who learned better also introduced more linguistic structure. This study is a first and much needed step in addressing the paucity of data on child learners in iterated language learning studies,

and highlights the need to conduct additional studies comparing child and adult learners using child-friendly tasks.

Acknowledgments

We wish to thank Noam Siegelman, Louisa Bogaerts, Amie Fairs and Antje Meyer for their helpful comments, and Kevin Stadler and Vera Kempe for helpful suggestions and discussions. We also want to thank the research assistants at the Living Lab in the Bloomfield Science museum (Tamar Johnson, Maya Enisman, Amira Laufer; Ayelet Gertsovski; Ruth Goldberg), the museum staff, and the children and parents who participated in the studies. The research was funded by an Israeli Science Foundation grant to the second author (grant number 584/16).

References

- Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology*, 8(2), 71–82.
- Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: the role of age and speed of stimulus presentation. *Developmental Science*, 14(3), 464–473.
- Arends, J. (1993). Towards a gradualist model of creolization. *Atlantic Meets Pacific*, 371–380.
- Arends, J., & Bruyn, A. (1995). Gradualist and developmental hypotheses. In Pidgins and creoles: An introduction (Vol. 15, pp. 111–120). John Benjamins Publishing.
- Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, 122(3), 292–305.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2018). *Package 'lme4'*. Vienna: R Foundation for Statistical Computing.
- Beckner, C., Pierrehumbert, J. B., & Hay, J. (2017). The emergence of linguistic structure in an online iterated learning task. *Journal of Language Evolution*, lxx001.
- Bickerton, D. (1984). The language bioprogram hypothesis. *Behavioral and Brain Sciences*, 7(02), 173–188.
- Boyd, J., & Goldberg, A. (2012). Young children fail to fully generalize a novel argument structure construction when exposed to the same input as older learners. *Journal of Child Language*, 39, 457–481.
- Brochhagen, T., Franke, M., & van Rooij, R. (2016). Learning biases may prevent lexicalization of pragmatic inferences: a case study combining iterated (Bayesian) learning and functional selection. In Proceedings of the 38th annual conference of the cognitive science society (pp. 2081–2086).
- Carden, G., & Stewart, W. A. (1988). Binding theory, bioprogram, and creolization: Evidence from Haitian Creole. *Journal of Pidgin and Creole Languages*, 3(1), 1–67.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: M.I.T. Press.
- Clay, Z., Pople, S., Hood, B., & Kita, S. (2014). Young children make their gestural communication systems more language-like: Segmentation and linearization of semantic elements in motion events. *Psychological Science*, 25(8), 1518–1525.
- Cornish, H., Tamariz, M., & Kirby, S. (2009). Complex adaptive systems and the origins of adaptive structure: What experiments can tell us. *Language Learning*, 59(s1), 187–205.
- Culbertson, J., & Kirby, S. (2015). Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Frontiers in Psychology*, 6, 1964.
- Dale, R., & Lupyan, G. (2012). Understanding the origins of morphological diversity: The linguistic niche hypothesis. *Advances in Complex Systems*, 15(03n04), 1150017.
- Ferdinand, V., & Zuidema, W. (2008b). Language adapting to the brain: a study of a Bayesian iterated learning model. Inst. for Logic, Language and Computation, Preprint Series PP-2008- 53. University of Amsterdam.
- Ferman, S., & Karni, A. (2010). No childhood advantage in the acquisition of skill in using an artificial language rule. *PLoS One*, 5(10).
- Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, 108(2), 477–499.
- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, 40(2), 177.
- Goldin-Meadow, S., & Mylander, C. (1998). Spontaneous sign systems created by deaf children in two cultures. *Nature*, 391(6664), 279–281.
- Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, 31(3), 441–480.
- Halekoh, U., & Højsgaard, S. (2014). A kenward-roger approximation and parametric bootstrap methods for tests in linear mixed models—the R package pbkrtest. *Journal of Statistical Software*, 59(9), 1–32.
- 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.
- Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, 59(1), 30–66.
- Johnson, Siegelman, & Arnon (in preparation). The importance of learning in generalization: Better learners add more linguistic structure to artificial languages.
- Karmiloff-Smith, A., Grant, J., Sims, K., Jones, M. C., & Cuckley, P. (1996). Rethinking metalinguistic awareness: Representing and accessing knowledge about what counts as a word. *Cognition*, 58(2), 197–219.

- Ke, J., Gong, T., & Wang, W. S. (2008). Language change and social networks. *Communications in Computational Physics*, 3(4), 935–949.
- Kempe, V., Gauvrit, N., & Forsyth, D. (2015). Structure emerges faster during cultural transmission in children than in adults. *Cognition*, 136, 247–254.
- Kerswill, P., & Williams, A. (2000). Creating a new town koine: Children and language change in Milton Keynes. *Language in Society*, 29(01), 65–115.
- Kirby, S. (2001). Spontaneous evolution of linguistic structure—an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, 5(2), 102–110.
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31), 10681–10686.
- Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, 104(12), 5241–5245.
- Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, 28, 108–114.
- Kirby, S., Smith, K., & Brighton, H. (2004). From UG to universals: Linguistic adaptation through iterated learning. *Studies in Language*, 28(3), 587–607.
- Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141, 87–102.
- Labov, W. (2007). Transmission and diffusion. *Language*, 344–387.
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27, 209–220.
- Morgan, E. & Levy, R. (2016). Frequency-dependent regularization in iterated learning. In Roberts, S. G., Cuskley, C., McCrohon, L., Barceló-Coblijn, L., Fehér, O., & Verhoef, T. (Eds.). *The Evolution of Language: Proceedings of the 11th international conference (EVOLANG 2016)*.
- Navarro, D. J., Perfors, A., Kary, A., Brown, S. D., & Donkin, C. (2018). When extremists win: Cultural transmission via iterated learning when populations are heterogeneous. *Cognitive Science*, 5, 5–50.
- Nettle, D. (1999). Using social impact theory to simulate language change. *Lingua*, 108(2), 95–117.
- Perry, L. K., Axelsson, E. L., & Horst, J. S. (2015). Learning what to remember: Vocabulary knowledge and children's memory for object names and features. *Infant and Child Development*.
- R COore Team (2016). *A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ravid, D., & Malenky, A. (2001). Awareness of linear and nonlinear morphology in Hebrew: A developmental study. *First Language*, 21(61), 25–56.
- Real, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, 111(3), 317–328.
- Roberts, S., & Winters, J. (2012). Social structure and language structure: The new nomothetic approach. *Psychology of Language and Communication*, 16(2), 89–112.
- Samara, A., Smith, K., Brown, H., & Wonnacott, E. (2017). Acquiring variation in an artificial language: Children and adults are sensitive to socially conditioned linguistic variation. *Cognitive Psychology*, 94, 85–114.
- Sankoff, G., & Laberge, S. (1974). On the acquisition of native speakers by a language. *Pidgins and Creoles: Current Trends and Prospects*, 73–84.
- Schwab, J. F., Casey, L. W., & Goldberg, G. A. E. (2018). When regularization gets it wrong: Children over-simplify language input only in production. *Journal of Child Language*, 1–19.
- Senghas, A., & Coppola, M. (2001). Children creating language: How Nicaraguan Sign Language acquired a spatial grammar. *Psychological Science*, 12(4), 323–328.
- Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science*, 305(5691), 1779–1782.
- Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, 177, 198–213.
- Smith, K. (2011). Learning bias, cultural evolution of language, and the biological evolution of the language faculty. *Human Biology*, 261–278.
- Smith, K., Perfors, A., Fehér, O., Samara, A., Swoboda, K., & Wonnacott, E. (2017). Language learning, language use and the evolution of linguistic variation. *Philosophical Transactions of the Royal Society B*, 372(1711), 20160051.
- Smith, K. (2009). Iterated learning in populations of Bayesian agents. In 31st Proceedings of the annual meeting of the cognitive science society (pp. 697–702).
- Tamariz, M., & Kirby, S. (2016). The cultural evolution of language. *Current Opinion in Psychology*, 8, 37–43.
- Tomasello, M. (2009). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, 4(2), 105–122.
- White, L. (2000). Second language acquisition: From initial state to final state. In J. Archibald (Ed.). *second language acquisition and linguistic theory* (pp. 130–155). Blackwell.
- Winter, B., & Wieling, M. (2016). How to analyze linguistic change using mixed models, growth curve analysis and generalized additive modeling. *Journal of Language Evolution*, 1(1), 7–18.
- Wonnacott, E., Boyd, J. K., Thomson, J., & Goldberg, A. E. (2012). Input effects on the acquisition of a novel phrasal construction in 5-year olds. *Journal of Memory and Language*, 66(3), 458–478.
- Wonnacott, E., Brown, H., & Nation, K. (2017). Skewing the evidence: The effect of input structure on child and adult learning of lexically based patterns in an artificial language. *Journal of Memory and Language*, 95, 36–48.