COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 44 (2020) e12877

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12877

Individual Differences in Learning Abilities Impact Structure Addition: Better Learners Create More Structured Languages

Tamar Johnson,[a,b] Noam Siegelman,[b,c] Inbal Arnon[b]

[a]*Centre for Language Evolution, University of Edinburgh*
[b]*Department of Psychology, The Hebrew University of Jerusalem*
[c]*Haskins Laboratories*

Received 1 December 2019; received in revised form 24 May 2020; accepted 17 June 2020

Abstract

Over the last decade, iterated learning studies have provided compelling evidence for the claim that linguistic structure can emerge from non-structured input, through the process of transmission. However, it is unclear whether individuals differ in their tendency to add structure, an issue with implications for understanding who are the agents of change. Here, we identify and test two contrasting predictions: The first sees learning as a pre-requisite for structure addition, and predicts a positive correlation between learning accuracy and structure addition, whereas the second maintains that it is those learners who struggle with learning and reproducing their input who add structure to it. This prediction is hard to test in standard iterated learning paradigms since each learner is exposed to a different input, and since structure and accuracy are computed using the same test items. Here, we test these contrasting predictions in two experiments using a one-generation artificial language learning paradigm designed to provide independent measures of learning accuracy and structure addition. Adults (N = 48 in each study) were exposed to a semi-regular language (with probabilistic structure) and had to learn it: Learning was assessed using seen items, whereas structure addition was calculated over unseen items. In both studies, we find a strong positive correlation between individuals' ability to learn the language and their tendency to add structure to it: Better learners also produced more structured languages. These findings suggest a strong link between learning and generalization. We discuss the implications of these findings for iterated language models and theories of language change more generally.

Keywords: Language learning; Language evolution; Linguistic structure; Statistical learning; Individual differences; Artificial language learning

Correspondence should be sent to Inbal Arnon, Psychology Department, Hebrew University, Mount Scopus, Jerusalem 9190501, Israel. E-mail: inbal.arnon@mail.huji.ac.il

T. Johnson and N. Siegelman contributed equally to this work.

1. Introduction

How did structure emerge in language? This foundational question has been the center of much controversy and debate. Within nativist approaches (e.g., Chomsky, 1972; Pinker, 1984), the emergence of structure is understood to reflect innate and domain-specific principles. More recently, however, studies have focused on the possible role of cultural transmission in the emergence of structure—where weak individual learning biases become amplified over time, leading to more structured languages (see Kirby, 2017; Kirby, Griffiths, & Smith, 2014 for a review). This effect was first demonstrated experimentally by Kirby, Cornish, and Smith (2008), who developed an iterated learning paradigm showing that linguistic structure can emerge from non-structured input over the course of transmission across "generations" of learners. These results provide important evidence for the claim that structure can emerge as a result of repeated transmission between learners. Much work since has extended this work to examine the conditions under which structure is more or less likely to emerge. Thus, some studies focus on social factors, demonstrating that group size (e.g., Raviv, Meyer, & Lev-Ari, 2019; Vogt, 2007) and communication dynamics (Kirby, Tamariz, Cornish, & Smith, 2015) affect the emergence of linguistic structure. Other studies focus on the age of the learners, with opposing views on the role of children and adults in the emergence of structure (e.g., Kempe, Gauvrit, Gibson, & Jamieson, 2019; Lupyan & Dale, 2010; Raviv & Arnon, 2018a; Senghas & Coppola, 2001).

These studies highlight the way group-level characteristics impact structure emergence (e.g., group size) and suggest that individuals may differ in their tendency to add structure. However, very little work to date has systematically investigated individual differences in structure emergence. This is at least partially related to the structure of experimental iterated learning paradigms. In these paradigms, the input of each learner differs (because it is the output of the previous learner), and it is hard to distinguish learning from structure addition (due to inherent confounds; see below). Most computational models also assign all learners similar learning biases, assuming that such variance in learning, if it exists, is unrelated to structure emergence (e.g., Griffiths & Kalish, 2007; Reali & Griffiths, 2009; Kirby et al., 2015, but see Navarro, Perfors, Kary, Brown, & Donkin, 2017). Overall, then, individual differences in structure emergence have been understudied, and it is unclear whether they exist and if so—what underlies such variation.

Here, we ask whether structure emergence can be predicted by individuals' learning abilities: Do better learners add more (or less) structure? This question is motivated by studies comparing child and adult regularization patterns. When faced with unconditioned (or "free") variation, children seem to regularize their input more than adults (Hudson & Newport, 2001; 2005; see Newport, 2020 for review), possibly because their worse learning abilities prevent them from learning the correct input distribution (Kam & Newport, 2009). In another study, children and adults showed different regularization strategies, with children increasing use of the dominant form, and adults making it lexically conditioned such that each form was used with certain lexical items (Samara, Smith, Brown, &

Wonnacott, 2017). This predicts that learning abilities will be negatively associated with an individual's tendency to add structure (even within an age group): Individuals who manage to replicate their input will not need to add structure to it while those who fail to learn it will create more structured systems. The opposite relation is predicted by a different study comparing the generalization of a novel linguistic construction in children and adults (Boyd & Goldberg, 2012), where younger children were more conservative than older children and adults. This was attributed to their lower ability to detect patterns in their input: Without learning the input, it is claimed, structure cannot be added. From this perspective, individuals' learning abilities are predicted to be positively correlated with their tendency to add structure.

The existence of a positive relation between learning and structure addition receives some support from a recent iterated learning study of child and adult learners (Raviv & Arnon, 2018a). In this study, children and adults were assigned to separate age-based chains (separate chains for children and adults). While the languages of both children and adults showed increased learnability (reduction in transmission error with generation), only adults showed evidence for the emergence of structure. The authors propose that the lack of structure in the child chains may reflect their difficulty with learning the input they were exposed to. This interpretation is supported by a significant positive correlation between transmission error (taken as a measure of learning), and structure score within individuals for both children and adults. That is, better learners created more structured languages. While suggestive, this study was not designed to directly assess the relation between learning and generalization and consequently had several methodological limitations that restrict the conclusions that can be drawn. First, each participant was exposed to input with varying levels of structure (since they were from a different generation and chain), meaning it cannot dissociate the effect of individuals' tendencies from that of the input structure itself. Second, and importantly, the measurement of learning and added structure were confounded, since they were calculated on the same test trials: Specifically, all calculations were based on all test items, seen (old) and unseen (new). This was done because there were very few unseen items: Each participant only saw only three unseen items at test in Raviv and Arnon (2018a). To examine the relation between learning and structure addition, a design is needed where the two are not inherently linked.

The current study sets out to address these limitations and directly test the link between learning abilities and structure creation at the individual level: Are better learning abilities associated with more structure addition? Operationally, we ask if participants who showed better learning of the input produce languages with a higher degree of structure—that is, a higher degree of systematicity between meanings and forms (this is the common definition of structure in the ILM literature, see, e.g., Kirby et al., 2008, albeit a definition that does not fully capture the complexity of natural languages, see Section 5). To test this, we used an artificial language learning paradigm similar to the one typically used in iterated learning studies, where each participant is exposed to part of a language, has to learn it, and then produce the full language. Importantly, and in contrast with typical iterated learning studies, we used a one-generation paradigm where all participants are exposed to the same language, ensuring identical input for all participants, and the output is assessed after one generation. Another important difference is that the initial language is not random, but has a certain degree of structure. Since we are not interested in showing that structure can emerge from no structure, but in the link between structure and learning in individuals, our starting point is a probabilistically structured language (with a certain degree of structure but with room to make it more regular). Such a language allows us to see how well participants learn their quasi-regular input, and whether they add structure to it in their own productions. Another important feature is that, unlike Raviv and Arnon (2018a), the language was divided into an equal number of seen and unseen items. This stands in contrast to typical iterated learning studies where participants are exposed to most of the language, with only a small number of items used to test generalization, a design that cannot be used to compute separate measures for learning and structure addition. In the current study, we used a larger feature space (four shapes, four colors, and three manners of motion, total possible combinations = 48), with half of the possible items appearing during learning and all of them appearing during testing. We used only the seen items to calculate learning scores and only the unseen items to calculate structure addition, creating two independent measures.

We make several predictions. First, if there are differences in structure addition across individuals, then we should see variability in our adult participants, despite being exposed to the same input. Second, we predict that this variability will be predicted by individual differences in learning abilities. Thus, we predict a correlation between the learning measure (calculated on seen items) and the structure addition measure (calculated on unseen items). Given previous findings (Raviv & Arnon, 2018a), we predict this to be a positive correlation such that participants who show better learning will also add more structure to the language, in line with accounts that see the identification of the underlying structure of a language as a pre-requisite to its systematization (Boyd & Goldberg, 2012). Alternatively, however, if it is the *difficulty* in learning the language which drives its generalization (Hudson Kam & Newport, 2005), we will see a non-positive correlation, most likely a negative correlation, where participants who struggle in learning the language (did not learn the seen items well) would add more structure to it, presumably to make it simpler to use and learn. Importantly, such a negative correlation is theoretically possible in our paradigm: Bad learners could, in principle, create structured languages (see concrete examples below, under *Measures of learning and added structure*).

We conduct two studies to test these competing hypotheses. In Experiment 1, the input language had a high level of regularity (83% regularity), whereas in Experiment 2 the input language was less regular (67% regularity). We compared these two levels of structure to examine the relation between learning abilities and structure emergence across different initial languages. To preview our findings, we show that in both conditions individual differences in learning were strongly and positively correlated with the tendency to add structure to the language.

2. Experiment 1

2.1. Methods

2.1.1. Participants

In all, 48 students at the Hebrew University participated in the study (30 females, 18 males, mean age: 24.6 years) for course credit or payment. Data from one participant were removed due to a technical error.

2.1.2. Materials and design

The artificial language consisted of 48 object–label pairs, varying across three semantic dimensions: shape, color, and manner of motion. There were four different shapes (star, heart, cross, and moon), four different colors (black, yellow, blue, and red), and three manners of motion (bouncing, spiral, and straight-line motion). Their combination resulted in the full set of $4 \times 4 \times 3 = 48$ possible objects. All corresponding labels were three syllables long, with each syllable corresponding to a different feature: The first syllable marked the shape of the object (with four different syllables in this set), the second marked its color (four different syllables), and the third its motion (three different syllables).

Half of these items were used during learning and test (SEEN items), whereas the rest of the items were only used in the test phase (UNSEEN items). The input language had a regularity level of 83%. This was achieved by pairing every instance of a feature with a syllable in a probability of 83%. Effectively, five out of six objects with the same feature (e.g., the heart shape) shared the same syllable in the relevant position in the initial language (e.g., the first syllable of labels, marking shape). For example, the regular label *li-pe-re* referred to a blue heart moving in a straight line (*li*-heart, *pe*-blue, *re*-straight-line motion). Of the 24 seen items, heart-shape objects appeared with the initial syllable *li* five out of six times (83%): We call these regular pairings. The sixth occurrence of a heart-shape object did not start with *li*, but with a syllable that was usually paired with another shape (e.g., *ka* usually marking the moon shape). We call these irregular pairings. See Appendix for the full list of object–label pairs.

As mentioned above, this artificial language differs from the one in typical iterated language studies in multiple respects. First, to enable comparison of individuals' learning and tendency to add structure, all participants were exposed to identical input (i.e., the same 24 object–label pairs). Second, the initial language already had substantial amount of structure (here—83% regularity per object–label pairing), unlike typical iterated learning studies where the initial language is random on average. Third, the test included large sets of familiar (SEEN) and novel (UNSEEN) trials, to produce separate (and reliable) measures of learning and added structure at the individual level.

2.1.3. Procedure

The task had two parts: an exposure phase, followed by a test phase. During exposure, participants were exposed to object–label pairings from the SEEN set. On each trial, they saw a short video of an object moving on the screen, with the object's label written at the bottom of the screen. Participants were asked to say the label out loud, and their productions were recorded. This was done to increase the task's interactivity and to ensure that participants are engaged. Participants were exposed to each object–label pairing from the SEEN items set four times during the training phase, resulting in 96 trials (24×4), presented in a random order. During the test phase, participants were exposed to all 48 objects in the language: the 24 SEEN items and the additional 24 UNSEEN items. On each trial, participants saw a short video of the object and were asked to produce a matching label according to the language from the exposure phase. Following Raviv and Arnon (2018a), participants produced labels using a syllable bank: All 11 possible syllables were presented on the screen and participants clicked on them to produce a three-syllabic label. The order of test trials was randomized for each participant (with SEEN and UNSEEN test items interleaved).

2.1.4. Measures of learning and added structure

Two separate measures were calculated based on the responses of each participant during the test phase: (a) an accuracy measure and (b) an added structure measure. Note that accuracy scores were calculated based on test trials with SEEN items only, whereas added structure scores were calculated based on performance on UNSEEN items. To calculate accuracy, each of the 24 SEEN trials was given a score between 0 and 1, where each correct syllable within the label scored 1/3 of a point (e.g., the label "*li-ni-fa*" for an object whose correct label was "*li-ni-ze*," resulted in item accuracy of 2/3). Note that accuracy was calculated by comparing participants' productions of SEEN items to the exact labels they were exposed to during learning: Specifically, irregular syllables were scored as incorrect if they were produced with the dominant syllable instead (see error analysis below). Coding was done using a script that compared each selected syllable to the target syllable.

The added structure measure (ΔZ) was operationalized as the difference between the structure score of the produced object–label pairings in the UNSEEN items (Z_{produced}) and the structure score of the input language (Z_{input}). Structure scores were calculated following the procedure in Kirby et al. (2008) where label–object pairings are compared to fully random label–object pairings using a bootstrapping technique, resulting in a Z-score where higher values represent more structured output (see Kirby et al., 2008 for details). Note that the initial language was the same for all subjects and had a structure score of $Z_{\text{input}} = 8.42$. To re-iterate, structure scores were calculated over UNSEEN items only. Thus, ΔZ expresses the change to the language structure (either positive or negative) that does not inherently result from participants' accuracy. A $\Delta Z = 0$ reflects similar level of structure in input and output languages, negative scores indicate a reduction in structure, and positive values indicate an increase in structure.

As noted above, the main goal of the current work is to estimate the correlation between the learning and added structure measures, in order to adjudicate between two possible theories. Per one theoretical hypothesis—learning as a pre-requisite for generalization (Boyd & Goldberg, 2012)—we expect a positive relation between learning and added structure: Better learners will generalize the quasi-regularities they learned, creating highly structured outputs, whereas poorer learners will create languages with more arbitrary links between forms and meanings (because they failed to learn the regularities). In contrast, a second hypothesis posits that *poorer* learners will create languages with a high degree of structure. We derive this second hypothesis from the findings of Hudson Kam and Newport (2005) and Kam and Newport, (2009) showing that children are more likely than adults to eliminate free variation and regularize their input. This could reflect a greater difficulty with learning (and reconstructing) the variation. That is, difficulty with learning could be driving generalization. This could happen in several different ways. One such situation is a case where poorer learners produce languages that are fully structured, but use a different set of syllable–feature pairings than those in the input language (e.g., if *li* was used to mark the heart shape in the input language, non-learners would now use it to mark circular motion). In this situation, the accuracy of non-learners on seen items would be low (because they did not match the input mappings), but they would produce highly structured languages. Another situation is a case where poorer learners *simplify* the language, by reducing the overall number of syllables while maintaining the distinction between different semantic features within each dimension. For example, instead of having four syllables correspond to different patterns of motion, and four other syllables marking shape, non-learners would use the same four syllables to mark both sets (e.g., use *li* to mark heart when it appears in the first position, and use the same syllable *li* to mark the color red when it appears in the second position). This, too, would result in a highly structured language. Note that this behavior is possible because in the miniature language syllable position indicates semantic feature (the first syllable marks shape, the second color, and the third manner of motion). Importantly, in both cases, poor learners would have high added structure scores. This will then lead to either a negative correlation between learning and added structure (if good learners reproduce the irregularity of the input, their productions will be less structured than those of the bad learners) or, at a minimum, a zero-order correlation (in a case where good learners also create regularized outputs).

2.2. Results and discussion

2.2.1. Learning of the input language

Mean accuracy was $M = 48\%$ ($SD = 0.23$), meaning that on average, participants labeled correctly about half of the syllables across SEEN items. This value is significantly higher than chance (i.e., random choice of 1/11 syllables; $t(46) = 11.21$, $p < .001$[1]). Importantly, we observed substantial interindividual variability in participants' accuracy, with performance ranging from 0.04 to 0.85 (Fig. 1).

To better understand what participants learned and what was the source of their errors, we conducted an error analysis by examining participants' productions of each syllable during test. We found that participants were less accurate in producing irregular (40% correct) compared to regular pairings (54%; *paired-t*$(46) = 7.00$, $p < .001$). Examining errors in irregular pairings revealed that many of these were due to incorrect productions

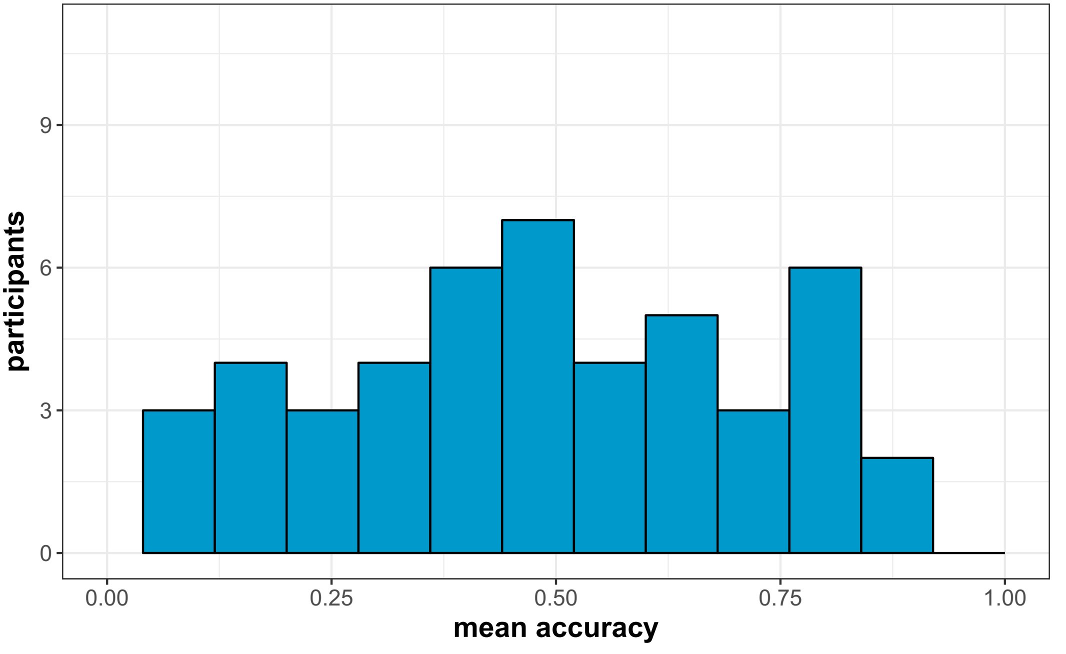

Fig. 1. Distribution of accuracy scores.

of the irregular syllable, with 91% of errors in the irregular syllable. The majority of these errors (58%) were generalization errors, where the irregular syllable was replaced with the dominant form (e.g., labeling the yellow-moon-circles as "*ka-ni-ze*," instead of the original irregular label *"gu-ni-ze,"* when *ka* is the dominant form for moon-shaped objects). Overall, these results show that perhaps unsurprisingly, learning is affected by regularity: Regular items were learned more easily, and irregular items were often "corrected" to the dominant form. An additional error analysis showed that participants did learn the association between position and semantic dimension: Participants usually choose syllables from the appropriate set of labels even when making a mistake (69% of errors for SEEN items). For example, erroneous syllables in the first position—which marked shape—most often came from the set of syllables describing shapes ("*gu*," "*ka*," "*be*," and "*li*"), rather than syllables corresponding to color or motion.

Structure score analyses: We next calculated an added structure score (ΔZ) for each participant to quantify the difference between the produced language (Z_{produced}) and the initial language (where $Z_{\text{input}} = 8.42$). To re-iterate, produced structure scores were computed on UNSEEN items only. Somewhat surprisingly, we found that the mean added structure score was *negative* ($M_{\Delta Z} = -2.14$, $SD = 5.22$), meaning that on average participants produced languages with lower levels of structure compared to the input. This result differs from the basic findings of iterated language studies, where participants add structure to the language (see Experiment 2 and Section 5). Critically for our purposes, we observed substantial interindividual variability in the tendency to add (or reduce) structure (ΔZ ranging from -9.68 to 7.39; see Fig. 2).

Correlation of learning and added structure: To examine whether individuals' tendency to add structure to the language relates to their learning abilities, we examined the correlation between participants' accuracy (on SEEN items) and added structure levels (ΔZ on UNSEEN items). Strikingly, the correlation between the two measures was near perfect, $r = .91$, $p < .001$ (95% CI: [0.84, 0.95]; Fig. 3). This extremely high correlation unequivocally shows that participants who learned the input language better produced more structured output.

We also examined the added structure values—and their relation to learning scores—separately for participants with higher and lower learning abilities. To do so, we split participants into "good" and "poor" learners based on a median split. We found that "good" learners added structure to the language on average, $\Delta Z = 2.07$ (a value significantly higher than 0, $p = .02$), meaning that their languages were more structured than the input language, whereas "poor" learners reduced structure, $\Delta Z = -6.17$ (significantly lower than 0, $p < .001$). Importantly, we observed significant positive correlations between learning and added structure within each of these two sub-groups: $r = .84$ and .65 for "good" and "poor" learners, respectively (both $p < .001$). These results again demonstrate the strong relation between learning and added structure, and show that the correlation between learning and added structure holds across learners: The degree to which the input language was learned impacted how much structure will be added.[2]

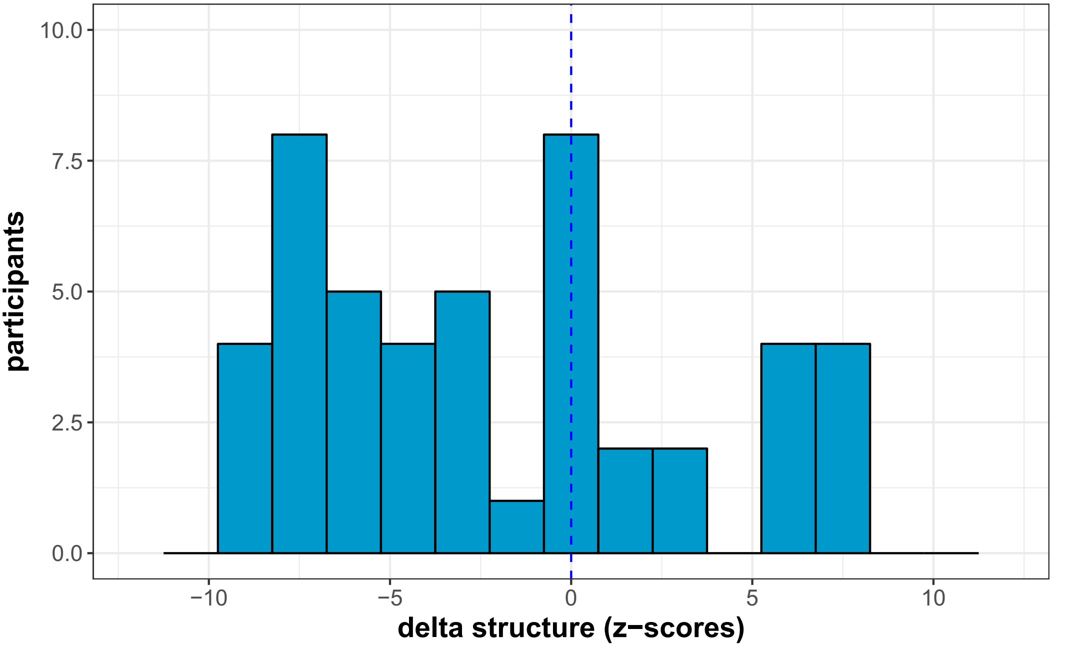

Fig. 2. Distribution of the added structure scores.

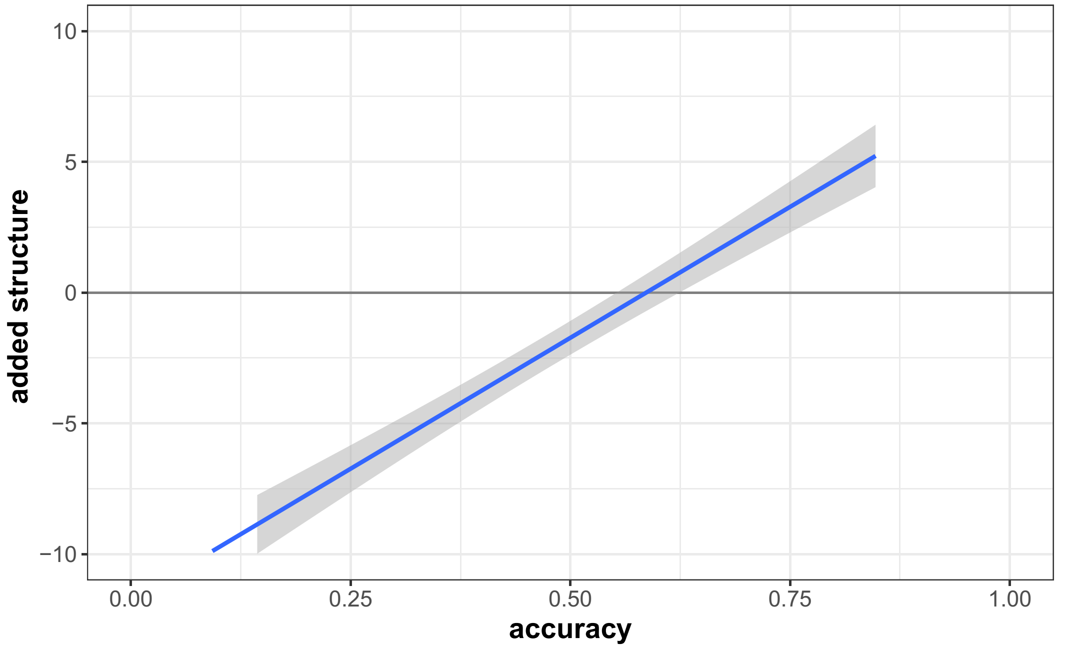

Fig. 3. Added structure (ΔZ) as a function of accuracy scores.

3. Experiment 2

The results of Experiment 1 provide evidence that (a) even among a relatively homogeneous sample of university students, not all learners add (or reduce) the same amount of structure to an artificial language, and importantly that (b) these individual differences can be traced back almost fully to individual's learning abilities. These results point to a strong link between learning and the tendency to add structure. We next conducted an additional experiment, with two goals in mind. First, given the novelty of the findings, we want to replicate them in an additional group of subjects. Second, we want to evaluate their generalizability by changing the characteristics of the initial language. In particular, we ask whether a similar correlation will be found when the initial language is less structured compared to the highly regular input in Experiment 1. Does the correlation hold when the language is harder and less structured to begin with?

3.1. Methods

3.1.1. Participants

In all, 47 students at the Hebrew University who did not take part in Experiment 1 participated in this study (29 females, 18 males, mean age: 24.5 years).

3.1.2. Design, materials, and procedure

The design uses a similar miniature language as that of Experiment 1, with 48 object–label pairings (24 SEEN and 24 UNSEEN). The initial language in this experiment had, however, a level of regularity, set to 67% (instead of 83% in Experiment 1). This was done by having only four out of six instances of each semantic feature marked by the same syllable (instead of 5/6). Irregular labels were again created by replacing the regular syllable with a syllable from the same semantic dimension. Note that because each object could have up to one irregular syllable, the labels for most objects had an irregular pairing: 22/24 SEEN items (due to 2 irregular tokens in each of the 11 semantic features). This input language had a structure score of $Z = 3.8$.

3.2. Results

3.2.1. Learning of the input language

Mean accuracy score was $M = 29\%$ ($SD = 13\%$), with considerable individual differences (Fig. 4). Mean accuracy was again better than chance ($t(46) = 10.27$, $p < .001$) but significantly lower than in Experiment 1 ($t(92) = 4.66$, $p < .01$), reflecting the lower structure (and increased difficulty) of this language.

Error analyses again showed that participants made more errors on irregular (93% of errors) compared to regular pairings (7%) for SEEN items. Note though, that it is hard to

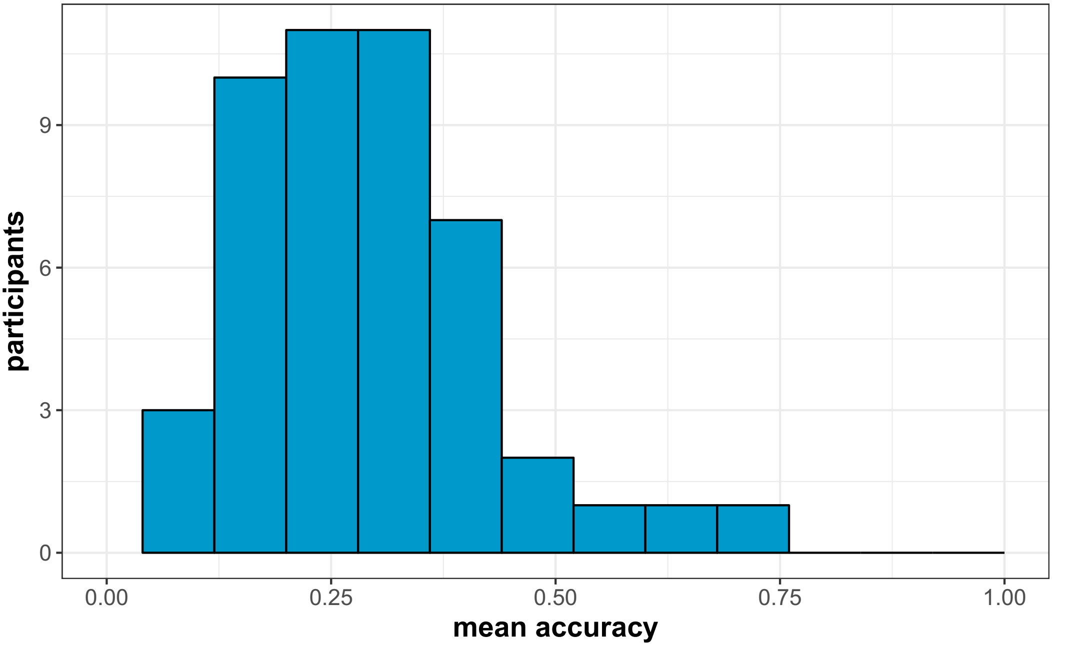

Fig. 4. Distribution of accuracy scores in Experiment 2.

accurately estimate performance for regular items due to their small number (2/24 items). As in Study 1, a large proportion of the errors for irregular items occurred in irregular syllables (84%), 36% of which were generalization errors where the irregular syllable was replaced with the dominant form. As in the previous study, participants showed learning of the association between syllable position and semantic feature (with 60% of errors using a syllable from the same set). Overall, then, the patterns of errors mirror that of Experiment 1.

3.2.2. Structure score analyses

We again calculated for each subject an added structure measure (ΔZ) by computing the difference between the structure in UNSEEN items ($Z_{produced}$), and the structure level of the initial language (Z_{input} = 3.8). As in Experiment 1, mean ΔZ was (slightly) negative ($M_{\Delta Z} = -0.72$ $SD = 3.63$), indicating that participants produced languages with lower level of structures compared to the input. Importantly, we again observed substantial individual differences (Fig. 5), with participants' added structure scores ranging from −5.09 to 8.96.

3.2.3. Correlation of learning and added structure

Critically, the correlation between accuracy and added structure was again positive, strong, and highly significant: $r = .62$, $p < .001$ (95% CI: [0.40, 0.77]; Fig. 6), suggesting

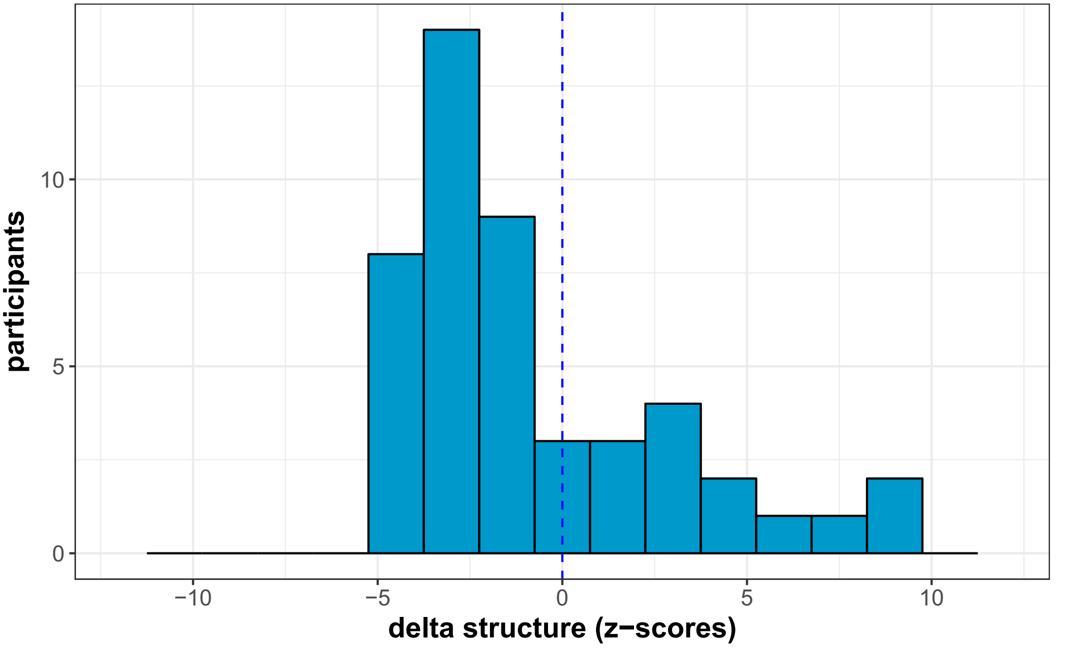

Fig. 5. Distribution of the added structure (ΔZ) scores across participants in Experiment 2.

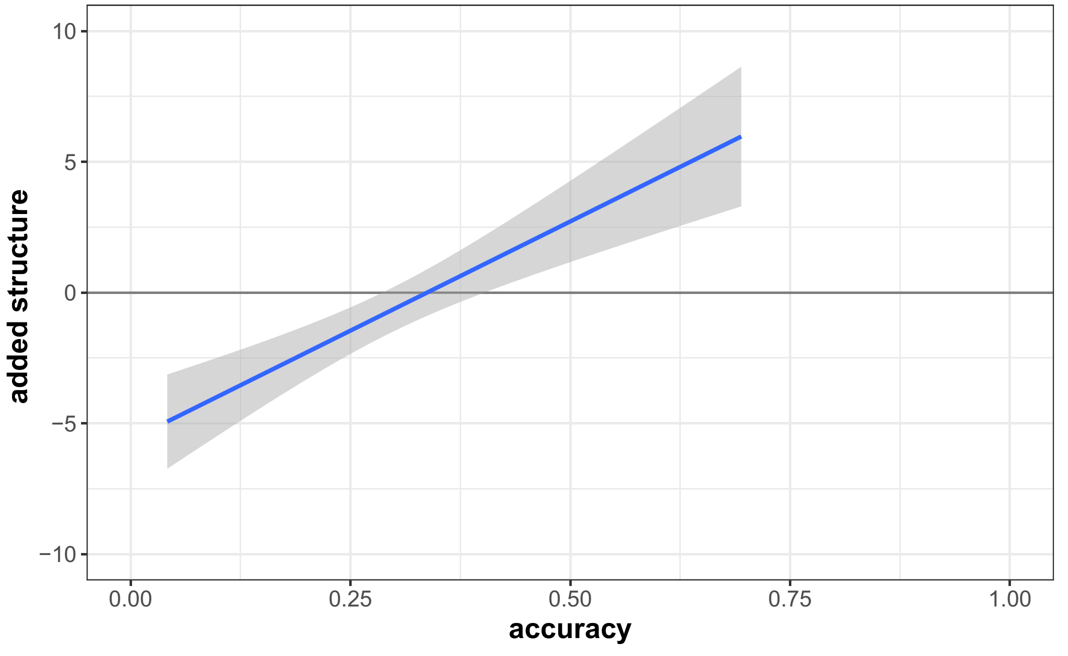

Fig. 6. Added structure (ΔZ) as a function of learning in Experiment 2.

that even after changing the level of regularity of the input language, individuals' tendency to add structure is highly correlated with their ability to learn it. Nonetheless, this correlation was significantly weaker compared to Experiment 1 (r = .62 vs. r = .91; Z = 3.75, p < .001), suggesting that this association is even stronger in a more regular environment. We return to this point in Section 5. As in Experiment 1, we next looked at the added structure values of "good" and "poor" learners separately (based on a median split). Good learners had a numerically (but not statistically) positive mean added structure score (ΔZ = 0.8, p = .35), whereas "poor" learners significantly reduced the structure of the language (ΔZ = 2.19, p < .001) Importantly, we again found significant positive correlations between structure and learning abilities among the two groups: r = .54 (p = .007) for "good" learners, and 0.45 (p = .02) for "poor" learners.[3]

4. General discussion

We set out to investigate the relation between learning abilities and structure addition in iterated learning paradigms. While recent work documents the influence of group-level properties on structure emergence (such as group-size, e.g., Vogt, 2007), little work to date has asked whether individuals vary in their tendency to add structure, and if so—what underlies this variation. This question is also of relevance for understanding what kind of learners/speakers drive language change, an issue that has been heavily debated. In two studies, we explore the link between learning the input language and adding structure to it at the individual level. To do so, we use a one-generation learning task, where all participants are exposed to the same semi-structured initial language, with an equal (and large) number of seen and unseen items. This allows us to have independent measures of learning (accuracy on seen items) and structure addition (computed based on the unseen items). The results of our two experiments are unequivocal: We observed a strong positive correlation between individuals' ability to learn the language and their tendency to add structure to it. This positive correlation held for languages with varying degree of structure, and for better and poorer learners within each study. That is, across initial structure levels and across learning skills, individuals who were better at learning the input language also produced more structured output languages.

Our results highlight the fact that even within a homogeneous group of subjects (i.e., a group of university students in a similar age range, tested under the same experimental conditions), there is extensive variability in participants' ability to learn—and consequently, generalize—the regularities embedded in their input. This observation has important implications for iterated language studies in particular and for models of language change more broadly. As for the former, ILM studies typically have only one individual learner per generation, a methodological choice that implicitly assumes similar structure emergence tendencies across learners. The current results suggest that learning and generalization are inherently linked, and that variation in learning accuracy leads to variation in structure addition. Methodologically, this means that the overall change in the language structure in iterated learning tasks may also reflect individual learner tendencies in ways that are not currently modeled. Theoretically, it suggests that better learning leads to more structure addition: Participants who struggle in learning the input language do not make it easier to learn, but instead make it noisier (less structured). This raises the interesting possibility that chains of "good learners" will result in more structured languages than chains of "bad learners." The link between variation in learning and structure is also relevant for other one-generation artificial language learning paradigms, where participants' tendency to change variable input is examined (e.g., Culberston & Newport, 2015; Culberston, Smolensky, & Legendre, 2012). Here also, little is known about individual differences, and here also, learners' tendency to regularize their input (measured as an increase in a form's frequency relative to the input) may be related to their ability to learn or remember the input language.

Turning to implications for theories of language change more broadly, the positive correlation between learning and structure addition is consistent with a theoretical framework that views learning as a pre-requisite for generalization. Such a framework can help explain differences between children and adults in the emergence of structure in laboratory-based investigations (Raviv & Arnon, 2018a), and their differential role in processes of emergence and change more generally. There is ongoing controversy about the role of child versus adult learners in processes of emergence and change, with previous studies focusing on qualitative differences between the two age groups. Thus, on the one hand, studies of emergent sign languages highlight children's unique role in the creation (or increase) of linguistic structure, a pattern attributed to innate knowledge no longer available to adults (Senghas & Coppola, 2001). At the same time, studies of creolization question children's role in processes of change because of their lower social prestige relative to adult speakers (e.g., Arends & Bruyn, 1994). From a learnability perspective, the difference between children and adults can be seen as one of degree, not kind: If the ability to introduce structure is related to learning, then children will be more or less likely to add structure depending on how well they learned the input language. That is, instead of assuming an overarching difference between children and adults we can make the prediction that children's tendency to add structure will be related to their success in learning the input. The idea that children's ability to extract regularities improves with age is supported by recent findings from the statistical learning literature, showing that statistical learning improves during development (Arciuli & Simpson, 2012; Raviv & Arnon, 2018b; Shufaniya & Arnon, 2018). Interestingly, this same literature also illustrates substantial individual variability in learning abilities within an age group, and its impact on language learning (e.g., Arciuli & Simpson, 2012; Siegelman & Frost, 2015; see Siegelman, Bogaerts, Christiansen, & Frost, 2017 for review). Combining these two sets of findings—the clear link between structure addition and learning, and the improvement of detecting regularities with age—leads to the prediction that adults, as a group, will add more structure to language than children will, especially when faced with learning complex input (a prediction consistent with the results of Raviv & Arnon, 2018a and Boyd & Goldberg, 2012). To re-iterate, our framework suggests that differences in structure addition between children and adults are rooted in quantitative differences in their learning abilities. That is, we predict that learning abilities will be a better predictor than age in explaining variability in structure addition (i.e., good child learners should look like adults while bad adult learners should look like kids).

One seemingly unintuitive observation in the current study is that participants created outputs that were on average *less* structured than the input they were exposed to, in contrast with the general finding of an overall increase in structure across diffusion chains (e.g., Kirby et al., 2008). This structure reduction may reflect two unique properties of our design: the use of only one generation and not multiple ones, and the fact that the initial language did have substantial structure. The iterative design is often seen as necessary for the emergence of structure (Kirby et al., 2008, 2015). The reduction in structure we found supports this claim, or, at a minimum, shows that additional factors are required for structure to increase (see Raviv et al., 2019, for discussion). Our language also differed from typical iterated learning studies where the initial language has no structure. It is possible that beyond a given level of structure, learners tend to reduce the structure (on average), to converge on languages that are neither too predictable nor complex. Importantly, the lack of overall increase was also impacted by individual variation: Good learners did add structure to the language. When we split participants into good and bad learners (based on the median), we see that "good learners" did add structure to the language (as evidenced in their positive ΔZ scores; see Figs. 3 and 6), whereas the "bad learners" did not. Further work is needed to understand what underlies the group-level reduction in structure, by manipulating both the levels of input structure and the communication dynamics. An additional open question has to do with the difference in correlation between studies 1 and 2: Why was the correlation between added structure and learning abilities significantly weaker in the lower level of regularity (67%) compared to the more regular condition (83%)? It is reasonable to think that less structured languages lead to greater learning difficulty, resulting in a larger number of individuals whose productions have a very small degree of structure. That is, in lower levels of regularity, which are characterized by more difficult learning conditions, a larger number of individuals produce less constrained languages. The output of those participants will be more random, adding unexplained variability to both measurements (accuracy and structure), and consequently, to their correlation. If this is indeed the case, we predict even lower correlations when learning difficulty will be increased (e.g., languages with lower levels of regularity, and/or with a larger number of items to learn).

One potential limitation of the current study is its ecological validity: How similar is the structure we examined to the complex structure found in natural language? While artificial languages do not do justice to the complexity of natural language, they nevertheless provide valuable information about learning mechanisms and biases. Such paradigms do not simulate the complexity of natural language, but instead isolate specific properties to examine their impact on learning. In the current study, we follow a long tradition of artificial language studies that present learners with probabilistic structure and ask how they change it. This paradigm has been used to uncover new insights on a range of question, among them the emergence of structure (e.g., Kirby et al., 2008); the emergence of case-marking systems (Fedzechkina, Jaeger, & Newport, 2012); and the possibly universal learning biases that underlie the frequency of harmonic alignment across the worlds' languages (Culbertson et al., 2012).

The structure learners were exposed to in our study can be described as a mix of morphological and lexical variation. Learners were presented with a quasi-regular mapping between forms and meanings, such that one form usually, but not always, marked the same semantic feature (e.g., red was usually marked with the syllable "*mo*," but appeared with the syllable "*fa*" instead in a minority of cases, at the same time, the syllable "*mo*" marked a different semantic feature in a minority of cases). While the irregularity we implemented does not have a precise parallel in natural language (i.e., three syllables each exhibiting probabilistic variation)—a limitation we acknowledge—there are real-world parallels for similar forms of variation, especially (but not only) in situations of language emergence and change. One parallel can be found in morphological variation, when the same morpheme is used to express multiple meanings (i.e., the -s morpheme in English). From the perspective of the learners, such variation is initially probabilistic, until the relevant context of use is acquired. Another example is the probabilistic use of morpho-phonological markers (e.g., -a at the end of the word usually, but not always, reflects feminine gender in Spanish), which learners have to acquire. Both morphological and lexical variation occur more often in situations of language emergence and change (see Kroch, 1994), which are the ones we set out to simulate. In language change, the use of specific morphemes becomes probabilistic as they change their meaning or function. For example, in Hebrew, there is an ongoing change in gender agreement for feminine forms, which results in the concurrent use of feminine and masculine pronouns for plural female human entities (Levon, 2012). Here, as in our study, learners have to deduce which is the dominant (and hence more generalizable) form. A more direct parallel to the kind of variation we tested is found in emerging sign languages. Lexical variation is documented in emerging sign languages, where different labels are used for the same object in a probabilistic manner, including for items similar to the ones we used, like color terms (e.g., Ergin, 2017; Meir & Sandler, 2019; Stamp et al., 2014). That is, the irregularity we implemented does not have a precise parallel in natural language, but its different components do. While artificial language learning paradigms do not simulate the complexity of natural language, they can nevertheless illuminate aspects of language learning that are hard to test in the wild. Their power lies in their simplicity: They allow us to isolate certain characteristics of natural languages (here, quasi-regular compositional structure), to assess how they are learned.

To conclude, our study is the first to examine the link between individuals' tendency to add structure and their learning abilities. Across initial learning conditions, we observe a strong positive correlation between added structure and learning skills. This, we believe, suggests that the agents of language change are those who are capable of learning the quasi-regularities in their environments, and using them to add further structure to the language.

Acknowledgments

We thank Kenny Smith for sharing the structure calculation code and for helpful discussions. We thank the two anonymous reviewers for their comments on previous versions of this paper. The research was funded by an Israeli Science Foundation grant to the third author (grant number 584/16).

Open Research badges

This article has earned Open Data badge. Data are available at https://osf.io/qtm8g/.

Notes

1. For interpretability, we report *t* tests when comparing performance to chance-level and in the error analysis below. We did ensure that our results hold also when using generalized mixed-effect models, which are less susceptible to statistical errors in analysis of categorical data (Jaeger, 2008). Indeed, the results were qualitatively similar: All significant effects remained significant.

2. A related question is whether the correlation between accuracy and added structure holds when disregarding participants who do not show evidence of learning. To examine this, we calculated an individual chance level (i.e. the number of syllables a given participant needs to produce correctly to show above-chance level at the individual level according to a binomial distribution; see e.g., Siegelman, Bogaerts, & Frost, 2017; estimated at 12/72 syllables). We found that the vast majority of participants did show evidence of learning: 42/47. Moreover, the strong positive correlation between accuracy and added structure remained highly significant even after excluding the five at-chance participants: $r = .91$, $p < .001$. These findings show that the relation between added structure and accuracy is not driven by at-chance subjects.
3. We again found a significant positive correlation between accuracy and added structure when looking only at participants who performed above-chance at the individual level ($n = 39$): $r = .56$, $p < .001$.

References

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science*, *36*(2), 286–304.

Arends, J., & Bruyn, A. (1994). Gradualist and developmental hypotheses. *Pidgins and Creoles: An Introduction*, *15*, 111.

Boyd, J. K., & Goldberg, A. E. (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*(3), 457–481.

Chomsky, C. (1972). Stages in language development and reading exposure. *Harvard Educational Review*, *42* (1), 1–33.

Culbertson, J., & Newport, E. L. (2015). Harmonic biases in child learners: In support of language universals. *Cognition*, *139*, 71–82.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*(3), 306–329.

Ergin, R. (2017). Central taurus sign language: A unique vantage point into language emergence. Doctoral dissertation. Tufts University.

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, *109*(44), 17897–17902.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, *31*(3), 441–480.

Hudson, C. L., & Newport, E. (2001). Regularization during creolization: The learning of inconsistent linguistic input. 26th Annual Boston University Conference on Language Development, Boston, MA.

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.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

Kam, C. L. H., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*(1), 30–66.

Kempe, V., Gauvrit, N., Gibson, A., & Jamieson, M. (2019). Adults are more efficient in creating and transmitting novel signalling systems than children. *Journal of Language Evolution*, *4*(1), 44–70.
Kirby, S. (2017). Culture and biology in the origins of linguistic structure. *Psychonomic Bulletin & Review*, *24*(1), 118–137.
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., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28*, 108–114.
Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.
Kroch, A. (1994). Morphosyntactic variation. In K. Beals, J. Denton, R. Knippen, L. Melnar, H. Suzuki, & E. Zeinfeld (Eds.), *CLS 30, Vol. 2: The parasession on variation in linguistic theory* (pp. 180–201). Chicago: CLS.
Levon, E. (2012). Gender, prescriptivism, and language change: Morphological variation in Hebrew animate reference. *Language Variation and Change*, *24*, 33–58.
Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS One*, *5*(1), e8559.
Meir, I., & Sandler, W. (2019). Variation and conventionalization in language emergence. *Language Contact, Continuity and Change in the Genesis of Modern Hebrew*, *256*, 337.
Navarro, D. J., Perfors, A., Kary, A., Brown, S., & Donkin, C. (2017). When extremists win: On the behavior of iterated learning chains when priors are heterogeneous. *Cognitive Science*, *42*, 847–852.
Newport, E. L. (2020). Children and adults as language learners: Rules, variation, and maturational change. *Topics in Cognitive Science*, *12*, 153–169.
Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
Raviv, L., & Arnon, I. (2018a). Systematicity, but not compositionality: Examining the emergence of linguistic structure in children and adults using iterated learning. *Cognition*, *181*, 160–173.
Raviv, L., & Arnon, I. (2018b). The developmental trajectory of children's auditory and visual statistical learning abilities: Modality-based differences in the effect of age. *Developmental Science*, *21*(4), e12593.
Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Compositional structure can emerge without generational transmission. *Cognition*, *182*, 151–164.
Reali, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, *111*(3), 317–328.
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.
Senghas, A., & Coppola, M. (2001). Children creating language: How Nicaraguan sign language acquired a spatial grammar. *Psychological Science*, *12*(4), 323–328.
Shufaniya, A., & Arnon, I. (2018). Statistical learning is not age-invariant during childhood: Performance improves with age across modality. *Cognitive Science*, *42*(8), 3100–3115.
Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160059.
Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods*, *49*(2), 418–432.
Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, *81*, 105–120.
Stamp, R., Schembri, A., Fenlon, J., Rentelis, R., Woll, B., & Cormier, K. (2014). Lexical variation and change in British sign language. *PLoS One*, *9*(4).
Vogt, P. (2007). Group size effects on the emergence of compositional structures in language. European Conference on Artificial Life (pp. 405–414). Berlin, Heidelberg: Springer.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Supplementary Material. Raw data is available at: https://osf.io/qtm8g/

Appendix: A

Stimuli of Experiment 1 (83% regularity). Regular SEEN items appear in white background; irregular SEEN items in light-blue background; and UNSEEN items in gray

Number	Label	Object's Shape	Object's Color	Object's Manner of Motion	SEEN or UNSEEN	Irregularity
1	gu ni ze	moon	yellow	rounds	SEEN	Shape
2	li pe re	heart	blue	straight	SEEN	None
3	li pe tu	star	blue	bounce	SEEN	Shape
4	gu fa ze	star	black	bounce	SEEN	Motion
5	ka fa tu	moon	red	bounce	SEEN	Color
6	ka pe tu	moon	blue	bounce	SEEN	None
7	ka fa re	moon	black	straight	SEEN	None
8	ka fa ze	moon	black	rounds	SEEN	None
9	ka mo re	moon	yellow	straight	SEEN	Color
10	be mo tu	cross	red	bounce	SEEN	None
11	be mo re	cross	red	rounds	SEEN	Motion
12	be pe ze	cross	blue	rounds	SEEN	None
13	be fa tu	cross	black	straight	SEEN	Motion
14	be ni re	cross	yellow	straight	SEEN	None
15	ka ni tu	cross	yellow	bounce	SEEN	Shape
16	li mo ze	heart	red	rounds	SEEN	None
17	li ni ze	heart	blue	rounds	SEEN	Color
18	be fa re	heart	black	straight	SEEN	Shape
19	li ni re	heart	yellow	straight	SEEN	None
20	li ni tu	heart	yellow	bounce	SEEN	None
21	gu mo tu	star	red	bounce	SEEN	None
22	gu mo ze	star	red	rounds	SEEN	None
23	gu pe ze	star	blue	rounds	SEEN	None
24	gu pe re	star	black	straight	SEEN	Color
25	ka mo re	moon	red	straight	UNSEEN	None
26	ka mo ze	moon	red	rounds	UNSEEN	None
27	ka pe re	moon	blue	straight	UNSEEN	None
28	ka fa tu	moon	black	bounce	UNSEEN	None

(continued)

Table A1. (continued)

Number	Label	Object's Shape	Object's Color	Object's Manner of Motion	SEEN or UNSEEN	Irregularity
29	ka ni tu	moon	yellow	bounce	UNSEEN	None
30	be mo re	cross	red	straight	UNSEEN	None
31	be pe re	cross	blue	straight	UNSEEN	None
32	be pe tu	cross	blue	bounce	UNSEEN	None
33	be fa tu	cross	black	bounce	UNSEEN	None
34	be fa ze	cross	black	rounds	UNSEEN	None
35	be ni ze	cross	yellow	rounds	UNSEEN	None
36	li mo re	heart	red	straight	UNSEEN	None
37	li pe tu	heart	blue	bounce	UNSEEN	None
38	li fa tu	heart	black	bounce	UNSEEN	None
39	li fa ze	heart	black	rounds	UNSEEN	None
40	li ni ze	heart	yellow	rounds	UNSEEN	None
41	gu mo re	star	red	straight	UNSEEN	None
42	gu pe re	star	blue	straight	UNSEEN	None
43	gu fa ze	star	black	rounds	UNSEEN	None
44	gu ni ze	star	yellow	rounds	UNSEEN	None
45	ka pe ze	moon	blue	rounds	UNSEEN	None
46	li mo tu	heart	red	bounce	UNSEEN	None
47	gu ni re	star	yellow	straight	UNSEEN	None
48	gu ni tu	star	yellow	bounce	UNSEEN	None

Stimuli of Experiment 2 (67% regularity). Regular SEEN items appear in white background; irregular SEEN items in light-blue background; and UNSEEN items in gray

Number	Label	Object's shape	Object's color	Object's manner of motion	SEEN or UNSEEN	Irregularity
1	gu ni ze	moon	yellow	rounds	SEEN	Shape
2	li fa re	heart	blue	straight	SEEN	Color
3	li pe tu	star	blue	bounce	SEEN	Shape
4	gu fa ze	star	black	bounce	SEEN	Motion
5	ka fa tu	moon	red	bounce	SEEN	Color
6	be pe tu	moon	blue	bounce	SEEN	Shape
7	ka fa ze	moon	black	straight	SEEN	Motion
8	ka mo ze	moon	black	rounds	SEEN	Color
9	ka mo re	moon	yellow	straight	SEEN	Color
10	be ni tu	cross	red	bounce	SEEN	Color
11	be mo re	cross	red	rounds	SEEN	Motion
12	li pe ze	cross	blue	rounds	SEEN	Shape
13	be fa tu	cross	black	straight	SEEN	Motion
14	be pe re	cross	yellow	straight	SEEN	Color

(continued)

Table A2. (continued)

Number	Label	Object's shape	Object's color	Object's manner of motion	SEEN or UNSEEN	Irregularity
15	ka ni tu	cross	yellow	bounce	SEEN	Shape
16	li mo tu	heart	red	rounds	SEEN	Motion
17	li ni ze	heart	blue	rounds	SEEN	Color
18	be fa re	heart	black	straight	SEEN	Shape
19	li ni re	heart	yellow	straight	SEEN	None
20	gu ni tu	heart	yellow	bounce	SEEN	Shape
21	gu mo re	star	red	bounce	SEEN	Motion
22	gu mo ze	star	red	rounds	SEEN	None
23	ka pe ze	star	blue	rounds	SEEN	Shape
24	gu pe re	star	black	straight	SEEN	Color
25	ka mo re	moon	red	straight	UNSEEN	None
26	ka mo ze	moon	red	rounds	UNSEEN	None
27	ka pe re	moon	blue	straight	UNSEEN	None
28	ka fa tu	moon	black	bounce	UNSEEN	None
29	ka ni tu	moon	yellow	bounce	UNSEEN	None
30	be mo re	cross	red	straight	UNSEEN	None
31	be pe re	cross	blue	straight	UNSEEN	None
32	be pe tu	cross	blue	bounce	UNSEEN	None
33	be fa tu	cross	black	bounce	UNSEEN	None
34	be fa ze	cross	black	rounds	UNSEEN	None
35	be ni ze	cross	yellow	rounds	UNSEEN	None
36	li mo re	heart	red	straight	UNSEEN	None
37	li pe tu	heart	blue	bounce	UNSEEN	None
38	li fa tu	heart	black	bounce	UNSEEN	None
39	li fa ze	heart	black	rounds	UNSEEN	None
40	li ni ze	heart	yellow	rounds	UNSEEN	None
41	gu mo re	star	red	straight	UNSEEN	None
42	gu pe re	star	blue	straight	UNSEEN	None
43	gu fa ze	star	black	rounds	UNSEEN	None
44	gu ni ze	star	yellow	rounds	UNSEEN	None
45	ka pe ze	moon	blue	rounds	UNSEEN	None
46	li mo tu	heart	red	bounce	UNSEEN	None
47	gu ni re	star	yellow	straight	UNSEEN	None
48	gu ni tu	star	yellow	bounce	UNSEEN	None