Adults are more efficient in creating and transmitting novel signalling systems than children

Vera Kempe
Nicolas Gauvrit
Alison Gibson
Margaret Jamieson

Adults are more efficient in creating and transmitting novel signalling systems than children

Vera Kempe

Abertay University

Nicolas Gauvrit

Université Paris 8 & École Pratique des Hautes Études

Alison Gibson

Abertay University

Margaret Jamieson

Abertay University

Abstract

Iterated language learning experiments have shown that meaningful and structured signalling systems emerge when there is pressure for signals to be both learnable and expressive. Yet such experiments have mainly been conducted with adults using language-like signals. Here we explore whether structured signalling systems can also emerge when signalling domains are unfamiliar and when the learners are children with their well-attested cognitive and pragmatic limitations. In Experiment 1, we compared iterated learning of binary auditory sequences denoting small sets of meanings in chains of adults and 5-7-year old children. Signalling systems became more learnable even though iconicity and structure did not emerge despite applying a homonymy filter designed to keep the systems expressive. When the same types of signals were used in referential communication by adult and child dyads in Experiment 2, only the adults, but not the children, were able to negotiate shared iconic and structured signals. Referential communication using their native language by 4-5-year old children in Experiment 3 showed that only interaction with adults, but not with peers resulted in informative expressions. These findings suggest that emergence and transmission of communication systems is unlikely to be driven by children, and point to the importance of cognitive maturity and pragmatic expertise of learners as well as feedback-based scaffolding of communicative effectiveness by experts during language evolution.

Introduction

Languages are shaped by two sets of constraints: the need to be learnable so they can be transmitted to the next generation, and the need to be expressive to ensure successful communication (Tamariz, 2017). Empirical evidence for this insight comes from experimental semiotics studies of novel signalling systems (Garrod & Galantucci, 2011) which comprise iterated language learning experiments, where the outcome of learning a mini-language by one participant serves as input for the next participant in a chain (Cornish, Smith & Kirby, 2013; Cornish, Dale, Kirby & Christiansen, 2017; Kirby, Cornish & Smith, 2008; Kirby, Tamariz, Cornish & Smith, 2015; Silvey, Kirby, and Smith, 2014; Verhoef, Kirby & Padden, 2011; Verhoef, 2012; Verhoef, Kirby & de Boer, 2014; 2016; Verhoef, Roberts & Dingemanse, 2015), as well as referential communication tasks and signalling games, where multiple participants negotiate meanings of novel signals over several rounds of communicative interaction (Dingemanse, Blasi, Lupyan, Christiansen & Monaghan, 2015; Fay, Ellison & Garrod, 2014; Garrod, Fay, Lee, Oberlander & MacLeod, 2007; Roberts et al., 2015, Selten & Warglien, 2007). These studies have shown that unstructured stimuli become increasingly easier to learn and to use because innovations are shaped by learners' implicit biases towards simpler, more transparent (Jones, Vinson, Clostre, Zhu, Santiago & Vigliocco, 2014; Dingemanse et al., 2015; Roberts et al., 2015) and more compressible (Kirby, Griffiths & Smith, 2014; Tamariz & Kirby, 2015; 2016 Xu & Griffiths, 2010) languages. To date, these types of laboratory experiments have been conducted mainly with adults. It is conceivable that adults, especially when presented with language-like signals, albeit artificial ones, will invoke their considerable meta-linguistic knowledge about what language ought to be like. Yet language is primarily acquired by children who lack this meta-linguistic knowledge and are subject to a range of cognitive constraints that differ from those operating in adults. To gain a better understanding of the generalisability of findings from experimental semiotics, and to explore the role of the cognitive and pragmatic constraints imposed by child learners, this study compares transmission and creation of unfamiliar signalling systems between adults and children.

To predict in what ways children may alter the way consistent and communicatively efficient signalling systems emerge we first need to consider what research in experimental semiotics tells us about how such systems emerge in adults.

The findings can be summarised with respect to three crucial features of language: *iconicity*, *combinatorial structure* and *compositional structure*. *Iconicity* emerges when adult learners are faced with novel signal-meaning mappings, and attempt to exploit transparent links between physical properties of the signals and dimensions of the associated meanings (Dingemanse et al. 2015; Roberts et al., 2015), capitalising either on abundant neonatal cross-modal connections or acquired knowledge about statistical regularities or cross-modal co-occurrences (Spence, 2011). Emergence of iconicity has been demonstrated not just when learners negotiate novel signalling systems during communication but also in simple iterated learning experiments without communication (Jones et al., 2014). Iconic signal-meaning mappings are subsequently aligned and refined during communicative interaction, resulting in conventionalised signals that become increasingly arbitrary (Garrod & Galantucci, 2011; Lister & Fay, 2017).

Emergence of *combinatorial structure* can be demonstrated in iterated learning experiments with novel stimuli that are not linked to referents, e.g. whistle sounds (Verhoef, 2012; Verhoef et al., 2014; 2015), colour sequences (Cornish, Smith & Kirby, 2013), doodles (del Giudice, 2012; Tamariz & Kirby, 2015) or random dot patterns (Kempe, Gauvrit & Forsyth, 2015). As a result of iterations through consecutive cycles of learning, such meaningless stimuli become more systematic and structured as sub-components like pitch contour segments or small colour sequences are recombined to generate potentially unlimited sets, in the same way as phonemes are combined to form morphemes and words in natural languages. When such unfamiliar stimuli are linked to meaning, combinatorial structure can also arise from the pressure to minimise confusion between signals as an increase in the number of signals renders them increasingly difficult to discriminate (Nowak, Krakauer & Dress, 1999; Zuidema & deBoer, 2009), but also from intrinsic signal features such as rapid fading (Roberts & Galantucci, 2012), or limited iconic affordances of the signalling domain (Roberts, Lewandwoski & Galantucci, 2015).

Compositional structure has been shown to emerge when the signals are not only subjected to iterated learning but also used to communicate meaning (Kirby et al., 2015), when meaning spaces undergo expansion (Selten & Warglien, 2007), when communication involves multiple interlocutors in social networks (Raviv, Meyer & Lev-Ari, 2019) or when context-based predictability of referents is low (Winters, Kirby & Smith, 2018). In these situations, sub-components of the signals become

systematically associated with dimensions of the meanings, akin to morpho-syntactic rules in natural languages (del Giudice, 2012; Kirby, Cornish & Smith, 2008; Kirby, Tamariz, Cornish & Smith, 2015; Silvey et al., 2014; Verhoef, Roberts & Dingemanse, 2015).

In the present study we ask if and how these basic results would change when novel signals are learned and used by children. Children differ from adults with respect to cognitive capacities, pragmatic abilities, pre-existing real-world knowledge and prior linguistic experience[1]. It is therefore important to investigate more directly how children create and transmit novel signalling systems in order to gain a better understanding of the underlying constraints operating in this process and the role that children may play in language change, especially in light of claims that diversity of linguistic structure is linked to the proportion of child vs. adult learners of a language and the differences in learning constraints this may impose on the process of language transmission (Lupyan & Dale, 2010; Dale & Lupyan, 2012).

Predictions about what constraints children impose on the emergence of communicatively efficient signalling systems and in what ways these constraints differ from those imposed by adults should address both emergence of iconicity and emergence of structure. Findings from child language development research suggests that such predictions will not necessarily be straightforward: With respect to the emergence of iconicity, the Iconic Bootstrapping Hypothesis (Imai & Kita, 2014) proposes that children benefit from iconic signal-meaning mappings because such mappings are transparent and hence easier to comprehend thereby alleviating the burden of learning. Consequently, children should be more predisposed than adults to capitalise on transparent cross-modal associations between signal features and meaning dimensions. However, the developmental origins of transparent cross-modal association are not clear. A recent meta-analysis of the emergence of the *kiki-bouba*-effect in infancy and early childhood (Ford, Lammertink, Peperkamp, Guevara-Rukoz, Fikkert & Tsuji, 2018) suggested that some cross-model correspondences (e.g. the *bouba*-effect which refers to the association of round shapes with back vowels and voiced consonants) are present early on while others (e.g. the *kiki*-effect which refers to the association of spiky shapes with front vowels and voiceless consonants) tend to emerge over time. This would lead to fairly complex predictions according to which some iconic mappings may be preferred by children while others should more easily accessible to adults based on their greater experience with statistical regularities in the

environment yet the literature at present does not allow us to make predictions with regards to specific age-dependent cross-modal preferences.

Predictions are also inconsistent with respect to the emergence of structure. On the one hand, in accordance with the Less-Is-More-hypothesis (Newport, 1990), children have been credited with superior language learning capabilities precisely because their limited cognitive capacity has been implicated in the injection of structure into language (Hudson Kam & Newport, 2005; 2009; Senghas & Coppola, 2001; Senghas, Kita & Özyürek, 2004; Thompson-Schill, Ramscar, & Chrysikou, 2009). Based on this view, structure should emerge more readily in children than in adults during iterated language learning. However, it is not clear whether cognitive immaturity *per se* aids structure-inducing innovations as attempts to demonstrate experimentally that cognitive limitations lead to superior decomposition (Cochran, McDonald & Parault, 1999) or regularisation of input (Perfors, 2012a) have proved unsuccessful or are open to alternative interpretations (Perfors, 2012b; Rohde & Plaut, 1999; 2003).

On the other hand, considerable evidence suggests that children fail consistently in referential communication tasks (Garrod & Clark, 1993; Glucksberg & Krauss, 1967; Matthews, Lieven, & Tomasello, 2007; Nilsen & Graham, 2009; Sonnenschein & Whitehurst, 1984), especially in unfamiliar referential domains for which easily accessible word labels are not available (Glucksberg, Krauss, & Weisberg, 1966; Krauss & Glucksberg, 1969). Despite attested perspective taking abilities (Liebal, Carpenter, & Tomasello, 2010; Moll, Richter, Carpenter, & Tomasello, 2008) such communicative failure may be the result of children's cognitive limitations, which hinder systematic monitoring of the environment for potential sources of ambiguity and of the informativeness of their own expressions (Rabagliati & Robertson, 2017). Consequently, during creation of novel signalling systems, emergence of systematic and reliable signal-meaning mappings should be less likely in children than in adults. Because of these complex and somewhat contradictory predictions the present study must remain exploratory in nature.

To our knowledge, so far only two other studies have attempted to experimentally study language transmission in children. Firstly, Flaherty and Kirby (2008) compared transmission chains of seven-year-old children and adults learning a small artificial language consisting of bisyllabic pseudo-words denoting objects in a two-dimensional meaning space. The results showed that compositional structure

emerged only in the adults but not in the children. Secondly, Raviv and Arnon (2018) trained 6-12-year-old children and adults to orally reproduce bisyllabic pseudo-words presented as signals for twelve meanings that varied on three dimensions. To encourage productivity, participants were trained on only nine out of the twelve signal-meaning pairs, and subsequently tested on the entire set. While compositional structure did not increase over the course of transmission to reach above-chance levels in either age group, it was nonetheless significantly higher in the adults. These two studies constitute preliminary evidence that emergence of structure is less likely in child learners.

While iterated language learning studies with adults typically present typed input, the studies with children used pseudo-words assembled from syllables in a syllable bank, to accommodate children's still fragile literacy abilities. Producing signals in such a way is a fairly unnatural task. To test the suitability of using more natural stimuli with children, we conducted a pilot study with an oral artificial language denoting a simple meaning space of stars differing in colour, shape and size in iterated language learning with 4-5-year-old children. The results revealed persistent intrusions of familiar words such 'Mummy', 'Daddy' and 'baby', which emerged early in the transmission chains regardless of seed stimuli. Apparently, at this age, children's use of native phonology in pseudo-words is strongly constrained by transfer from prior linguistic knowledge, presumably due to their limited inhibitory control abilities, which preclude suppression of easily accessible, familiar labels (Kahan & Richards, 1986). Thus, to put children and adults on a level-playing field in terms of prior experience with the signals, the present study used an entirely unfamiliar signalling domain that did not resemble natural languages to study iterated language learning (Experiment 1) and creation of novel communication systems (Experiment 2). In addition, to distinguish children's referential ability from their ability to learn novel signals, Experiment 3 examined children's referential communication in their native language. All experiments were approved by the Ethics Committee of the School of Health and Social Sciences at Abertay University.

Experiment 1

The aim of our first experiment was to compare iterated learning of a novel signalling system between adults and children. Although unfamiliar signalling systems, e.g. slide whistles or Leap Motion (Eryilmaz & Little, 2017; Little, Eryilmaz & de Boer,

2017; Verhoef, 2012; Verhoef et al., 2014; 2015; 2016), have been explored before, they either may invoke biases based on pre-existing musical experience or are difficult to execute with small children. We therefore decided to use binary auditory sequences consisting of high and low fixed-length buzzer tones, which are easy and very pleasurable for children to produce. This is an entirely novel signalling domain that has not been explored in the literature. It is unfamiliar to adult and child participants alike, not even resembling music due to its simplicity, yet shares the property of fading with auditory signals of oral language. The binary auditory sequences were combined with eight referents varying in the dimensions of shape, size and brightness (Figure 1). This meaning space had considerable iconic affordances comprising dark and bright *kiki*- and *bouba*-type objects of different size that are known to affect children's word learning during natural language acquisition (Maurer, Pathman & Mondloch, 2006; Ozturk et al., 2013). To keep experiment duration and task demands manageable for small children, we did not include a communication task in addition to the learning task but introduced pressure for expressivity by applying a homonymy filter (Kirby et al., 2008), i.e. by removing ambiguous signals from the transmission process to prevent languages from degenerating into uninformative systems.

While we expected learnability to increase over the course of transmission in both adults and children, the more interesting question is whether it does so to a similar extent in both age groups and what strategies children and adults use to encode and retrieve the unfamiliar signals. One strategy learners might adopt is to reduce the number of signals in a way that exceeds the cap imposed by the homonymy filter, which would result in a loss of expressivity. Children may be more prone to this strategy given their smaller working memory capacity. Another strategy would be to forge more transparent, iconic links that help to retrieve a signal given its meaning, and children might be more likely to use this strategy according to predictions from the Iconic Bootstrapping Hypothesis (Imai & Kita, 2014). For the meaning space used here, signal dimensions that potentially afford iconic mappings are the auditory features of sequence length, pitch, and number of pitch alternations. These features can be mapped onto the meaning dimensions of size (big vs. small), shape (spiky vs. rotund) and brightness (dark vs. light), resulting in nine possible sound-symbolic mappings. The lack of familiarity with the physical properties of the signals provides a unique opportunity to find out how readily children vs. adults can

exploit these various iconic affordances. However, of the nine possible cross-modal mappings only a subset are attested in the literature on iconicity and cross-modal associations, and these are the ones we will explore in this study:

(a) Given evidence for magnitude symbolism, larger referents might be associated with longer signals (Dingemanse et al., 2015).

(b) Research on cross-modal associations has established associations between pitch height and brightness (Spence, 2011) suggesting that brighter referents might be associated with auditory sequences containing a higher proportion of high-pitched sounds.

(c) Similarly, research on cross-modal associations has also established links between pitch height and shape such that higher pitch is associated with more angular, spiky shapes (Spence, 2011). Consequently, spiky objects might be associated with sequences with a higher proportion of high-pitched sounds.

(d) Studies of whistled tone signals, which are characterised by continuous pitch changes, have indicated that pitch contour can be used to mimic object shape (Verhoef, Kirby & de Boer, 2013). With respect to the referents in this study this means that more pitch alternations within a sequence (e.g. *01010101* as opposed to *00001111*, where *0* represents the high and *1* the low tone) might be used to indicate spiky shapes.

In addition to exploiting iconicity to aid learning and transmission, there are several ways in which learners can also introduce structure: First, learners can attempt to make individual signals easier to remember and reproduce by shortening them (Zipf, 1949) and by reducing their algorithmic complexity. Children's limited working memory capacity should predispose them more than adults towards creating shorter signals but also towards decomposition into smaller sub-components which then can be recombined using simple production algorithms. For example, a binary sequence consisting of high and low tones represented as *0* and *1* that takes the form *010001011* is considerably more complex than the sequence *000111000,* which can be produced according to a simple production rule *"alternate triplets",* or the sequence *01010101,* which simply requires repeating the sub-component *01* four times. This intuition is formally captured by Kolmogorov complexity, which for small binary sequences can be quantified based on the coding theorem method (Gauvrit, Soler-Toscano, Zenil & Delahaye, 2014; Zenil, Soler-Toscano, Delahaye & Gauvrit,

2015). As algorithmic complexity increases with sequence length we provide estimates normalised for sequence length[2].

Secondly, on the level of the entire system, learners can try to systematically map sub-components of the signals and salient dimensions of the meanings thereby inducing compositional structure. Compositional structure can be quantified following the procedure outlined in Kirby et al (2008) where similarities between all possible pairs of signals are correlated with similarities in the associated pairs of meanings: the stronger these correlations the higher the compositional structure. However, such a general measure of compositional structure could obscure the fact that participants may focus on certain dimensions of the meanings but not others, as observed in previous research (Beckner, Pierrehumbert & Hay, 2017). We therefore also compute and analyse correlations between similarities in signals and similarities in various meaning dimensions separately, which will be reported in Appendix 1.

One problem with exploring differences between adults and children in these various manifestations of iconicity and structure is that it necessitates testing a considerable number of dependent variables thereby inflating the possibility of Type-I-errors. Yet this approach is not uncommon in the experimental semiotics literature: iterated language learning studies typically measure a range of outcomes that capture complexities of the emerging systems. This attests to the still exploratory nature of this research and the present study is no exception. We hope that our results will contribute to the formulation of more specific hypotheses that can in the future be tested in more targeted confirmatory studies.

Method

<u>Participants:</u>

Seventy-two adults (51 men, age range 18 to 51 years) and 72 children (42 boys, all primary grade level 1, 2 and 3[3], age range 5 to 7 years) participated in the study. In each age group, participants were assembled into 6 transmission chains of 12 generations each, with the number of generations chosen simply to accommodate all children who were eager to participate in the experiment. The number of generations of adult participants was chosen to match that of the children. Child transmission chains were assembled controlling for grade level within chains, with 2 chains each for primary grade levels 1, 2 and 3. Adult participants were recruited on campus, provided informed consent and were debriefed after the experiment; child participants

were pupils of a Primary school, and received a sticker as reward for participation. Parental consent was obtained for all children.

Materials:

The referents were eight coloured objects (Figure 1) differing in shape (spiky *kiki*-type vs. rotund *bouba*-type), size (2 x 2 cm vs. 4 x 4 cm) and brightness (25% vs. 75% saturation), printed onto 5 x 8 cm laminated cards. All referents also had unique properties that arose from variation in the particular shapes of each individual object as well as in the individual hues. Two 500 ms sine-wave tones (high: 440 Hz = musical note *a*; low: 293.7 Hz = musical note *d*) were synthesised and recorded onto two buzzers (*Learning Resources* Recordable Answer Buzzers) of 9 cm in diameter. Note that pressing the buzzers always generated tones of the same duration making it impossible to modify individual tone duration. Buzzers differed in colour; due to the need to replace the buzzers from time to time their actual colours changed throughout the study.
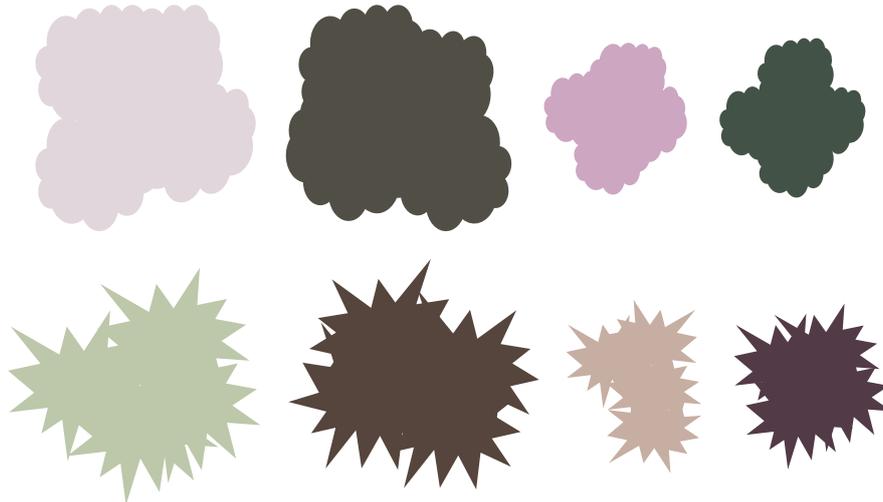
Figure 1: Meanings used in Experiments 1 and 2.

Procedure:

Children and adults received the same child-appropriate instruction, which for adults was prefaced with the comment that they were invited to test a game designed to be suitable for 5-7-year-old children. Participants were told that they would learn a language used by aliens who had no mouth and hence operated buzzers as a means of

communication. Participants were further told they would learn the 'buzzing words' for a set of shapes, and were shown all eight cards. Unbeknownst to the participants, two cards were then withheld according to the criteria described below resulting in a training set of six cards. Training proceeded in incremental fashion such that participants learned signal-meaning pairs one-at-a-time. In the training phase, the experimenter placed the buzzers in front of the participant, shuffled the preselected six cards and showed them to the participant one by one, accompanying the demonstration by buzzing the associated binary sequence of high and low tones at an even pace with one hand, using a prepared script. Participants were asked to repeat the sequence also using only one hand to operate the buzzers, to avoid the overlapping of sounds. Before proceeding to the next card, the experimenter demonstrated the buzzer sequence again, and the participants repeated it for a second time. In the testing phase, participants were shown all eight cards one at a time in randomised order, and were asked to produce the associated buzzer sequence to the best of their ability. If participants felt unsure, they were encouraged to produce what they thought the appropriate alien 'buzzer word' should be, given their acquired knowledge of the alien language. During testing, hand movements and buzzed tone sequences were video-recorded, and then coded for subsequent reproduction by the experimenter during training of the next participant in the chain. Coding accuracy was ensured through double coding.

Generation 1 participants in each chain were trained with random 6-bit or 8-bit binary sequences such that the average length was 7-bits yet participants experienced some length variation. There were six different random seed languages, each used for one child and one adult chain. As indicated above, only six out of the total of eight cards and the associated buzzer tone sequences were presented during training. For the seed languages, these six cards were selected at random. Modifying the procedure introduced by Kirby et al. (2008) to preserve equal size of the training set, we applied a homonymy filter by removing two cards in the subsequent generations using the following criteria, designed to minimise degeneration of languages to maintain expressivity: If two signals had been duplicated by the learner (i.e. the buzzer language contained only six unique signals) one of the two cards associated with each duplicate signal was removed at random; if one signal was duplicated (i.e. the buzzer language contained only seven unique signals), one of the cards associated with the duplicate signal was removed at random, as was one additional randomly chosen card.

As the training set contained six cards, any further duplicate signals that occurred were transmitted to the next generation (and the duplicate signals that were withheld were selected at random), to ensure equal size of the training set for all participants. A single session lasted approximately ten minutes. At the end of each session, adult participants were debriefed; child participants were given the opportunity to choose a reward sticker.

Results

Appendix 2 provides the final generation binary auditory sequences for adult and child chains. Our analyses explored whether children and adults differed in the trajectories of change in expressivity, learnability, signal length, signal structure, and compositional structure, which are depicted in Figure 2. *Expressivity* was operationalised as the number of distinct signals within a language; the more signals there are the more expressive a language is, with fully expressive languages containing the same number of signals as there are meanings. *Learnability*, i.e. transmission fidelity, of each buzzer language was measured as the average normalised Levenshtein edit distance (nLED) between trained and produced signals for each meaning, which constitutes an inverse similarity measure, and was calculated as the number of tone substitutions, insertions and deletions at each position in the sequence required to change one signal into the other, divided by the length of the longer signal. *Signal length* was measured as the total number of high and low tones produced for each meaning. S*ignal structure* was operationalised as length-normalised algorithmic complexity (Gauvrit et al., 2014; Zenil et al., 2014; 2015), which constitutes an inverse measure of the degree of structure of individual signals. Finally, *compositional structure* of each language was computed following the procedure outlined in Kirby et al (2008): For all possible signal pairings, we determined the dissimilarity based on edit distances and correlated this with dissimilarity between the associated meanings based on Hamming distances, which indicate the number of divergent meaning dimensions (out of three: shape, size, brightness). We report standardised scores for these correlations within the distribution of correlations obtained from 10,000 random permutations of all possible pairings using a Monte Carlo process (Mantel, 1967) to determine whether these correlations exceeded the level of chance.

Finally, to estimate *iconicity*, occurrence of each of the hypothesised iconic mappings was quantified by calculating measures of association between differences in the relevant feature of the auditory sequences (e.g. length, proportion of high-pitched tones, number of pitch alternations) and Hamming distances between the target meaning dimensions across all signal pairings using the Mantel-test (Mantel, 1967). Proportion of high-pitched tones was calculated to determine the predominant pitch of an auditory sequence, e.g. sequences like *011* and *001111* were assigned the same proportion of high-pitch tones of 0.33. The rationale behind this relative measure was to capture the overall pitch of a sequence regardless of its length. Similarly, to determine pitch alternations we computed the proportion of all alternations among all tone transitions.

All dependent variables in this and the subsequent experiments were subjected to Growth Curve Analyses (GCAs) with mixed-effect models using the *lmer* and *glmer* functions of the R package lme4 in R version 3.4.3. We included fixed effects of Age Group and Generation as well as their interaction, and a maximal random effect structure (Barr, Levy, Scheepers and Tily, 2013) that included random effects of participants, chains, and of items (i.e. referents) where appropriate, as well as all relevant random intercepts, slopes and slope interactions (Winter & Wieling, 2016)[4]. For all models, p-values were obtained using the Satterthwaite approximation to degrees of freedom provided by the lmerTest package. All fixed effects were centered in order to reduce multicollinearity. Model coefficients and associated t-test for all dependent variables are given in Table 1.

These analyses yielded effects of Generation only for the number of unique patterns and for the edit distances indicating that both of these measures declined over the course of transmission. For edit distances we also found an effect of Age Group suggesting that transmission fidelity was higher in adults[5]. No effects of Generation and Age Group were found for measures of signal length, signal structure, or compositional structure[6].

DV	Intercept	Age Group	Generation	Age Group x Generation
model:				

lmer (DV ~ AgeGroup * Generation + (1 + Generation | Chain))					
unique signals					
	β	6.63	0.05	-0.35	0.08
	t	34.27***	0.23	-2.90*	0.63
compositional structure					
	β	0.10	0.09	-0.04	-0.05
	t	1.20	1.15	-0.51	-0.55
Iconicity – correlations between:					
length – size					
	β	0.64	0.42	0.02	0.05
	t	2.37*	1.54	0.09	0.32
pitch – shape					
	β	0.20	-0.32	0.06	0.13
	t	1.17	-1.91	0.34	0.70
pitch – brightness					
	β	0.04	0.18	-0.02	0.04
	t	0.19	0.94	-0.14	0.20
pitch changes – shape					
	β	-0.13	0.21	0.02	0.08
	t	-0.79	1.31	0.11	0.34
model: lmer (DV ~ AgeGroup * Generation + (1 + Generation|Chain) + (1 + AgeGroup * Generation|Referent))					
length-normalised edit distance					
		0.33	-0.04	-0.04	-0.00
	β	21.49***	-2.48*	-5.17***	-0.11
	t				
signal length					
	β	7.53	-0.99	-0.24	0.10
	t	11.79***	-1.58	-1.15	0.49
algorithmic complexity					
	β	0.09	0.25	0.12	-0.05
	t	0.45	1.28	1.74	-0.76

Table 1: Coefficients for intercepts and fixed effects in the linear mixed-effect models for all dependent variables in Experiment 1. *** p < .001, ** p < .01, * p < .05.
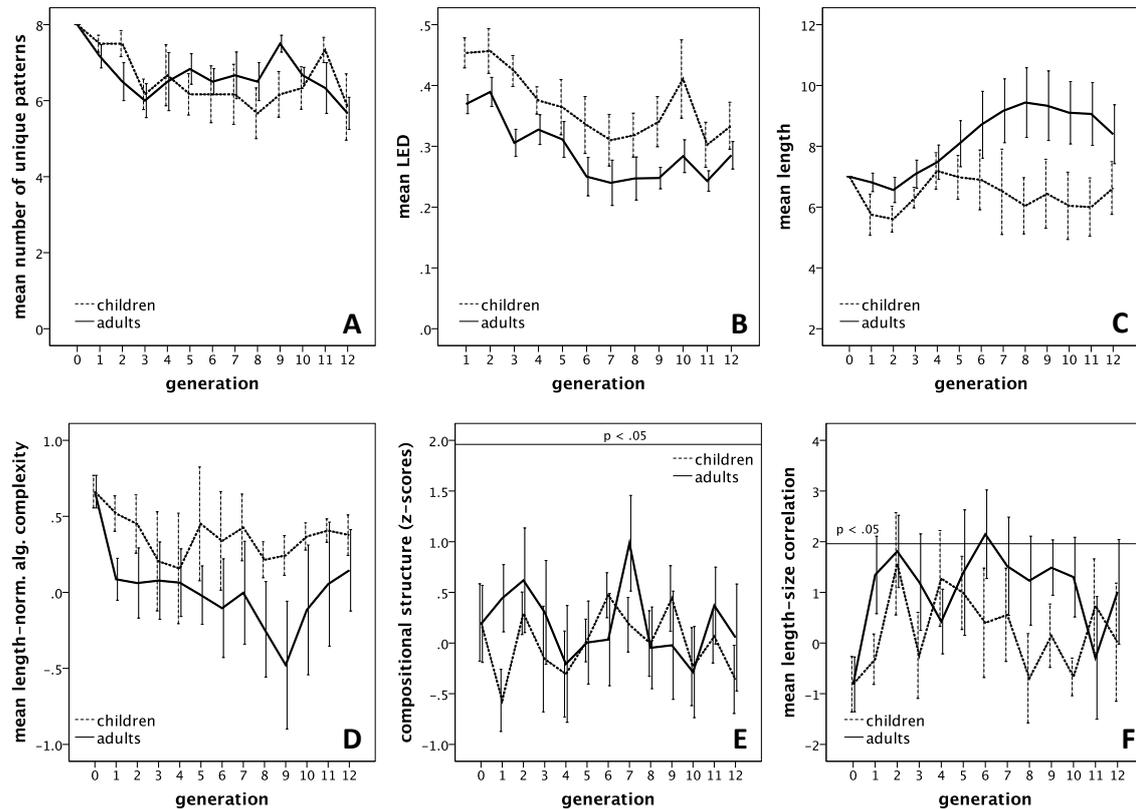

Figure 2: Means for numbers of signals (out of 8) as a measure of degeneration of the language (panel A), length-normalised Levenshtein edit distance (meanLED) between signals denoting the same meanings in consecutive generations as an inverse measure of learnability (panel B), signal length (panel C), length-normalised Kolmogorov complexity as a measure of signal structure (panel D), z-scores of correlations between similarity of signal pairs and meaning pairs as a measure of compositional structure (panel E) and iconic mapping between sequence length and referent size (panel F), for 6 adult and 6 child chains consisting of twelve generations of learners in Experiment 1. Solid lines in panels E and F demarcate the z > 1.96, p < .05 area. Error bars represent ± 1 S.E.M aggregated by participants.

Discussion

Experiment 1 investigated differences between adults and 5-7-year-old children in the iterated reproduction of binary auditory sequences associated with a three-dimensional meaning space with eight referents differing in shape, size and

brightness. We found that the emerging languages degenerated by loosing some of the unique patterns, and became easier to learn over the course of transmission, as indicated by a decrease in edit distance. Although decreasing edit distances indicated significant gains in fidelity of transmission in both children and adults, particularly in the earlier generations, transmission overall tended to be more faithful in adults. Notably, compositional structure and iconicity did not emerge in either age group. Thus, whatever advantages adults may have had in their ability to learn the unfamiliar signals, they were not due to increased structuring of the individual signals nor of the entire languages but most likely just a reflection of their greater working memory capacity that enabled them to retain a larger number of sequences. This is in line with findigs from immediate serial recall of unfamiliar four-colour sequences produced by 6-10-year old children in the SIMON®-game (Mathy, Fartoukh, Gauvrit & Guida, 2016), which demonstrated that age-related increase in memory capacity was not associated with formation of larger chunks, which could potentially introduce greater structure, but with the ability to store more of them.

The null-effects for structure obtained here differ from previous findings comparing iterated reproduction of random dot patterns between adults and children (Kempe et al., 2015), where structure emerged more rapidly in the children. The different results may be due to differences in familiarity: In the visuo-spatial domain, children will only have acquired a simple repertoire of representations such as straight lines or blobs, and hence fall back on these algorithmically simpler structures during reproduction. Adults, on the other hand, are likely to possess a larger repertoire of more complex representations like triangles, diamonds, squares, zigzags, and crosses etc., thus being able to introduce patterns with greater algorithmic complexity. However, sequences of high and low tones are likely to be equally unfamiliar for both age groups so that no such different prior representations can bias children's and adults' reproductions. Under these conditions, adults simply benefit from their larger memory capacity allowing them to retain more unstructured chunks.

Our findings are in line with the study by Raviv & Arnon (2018), where learnability of meaningful signals was also higher in adults than 6 – 12-year-old children. Moreover, as in the present study, compositional structure did not exceed chance levels in that study either although it was significantly higher in adults. There are at least two possible reasons for why structure and iconicity did not emerge reliably in these experiments: First, both in Raviv and Arnon (2018) and in the present

study, the training phase was considerably shorter than in many previous iterated learning studies, a methodological constraint that was necessary to sustain children's attention on the task. A shorter training phase may have simply not provided enough opportunity to explore the iconic and structural potential of the signalling domains. Second, introducing a homonymy filter may not have imposed a sufficiently strong expressivity pressure to avoid ambiguous signals (Flaherty & Kirby, 2008; Carr, Smith, Cornish & Kirby, 2017). A better way of directing learners' attention to the requirement for having unique signals for all meanings would have been to combine iterated learning with a referential communication task, as in Kirby et al. (2015). This, however, would have required administering a fairly complex and elaborate learning and communication procedure to a large number of dyads, which is a considerable logistical challenge for research with small children. As a result, it is not clear whether the failure of iconicity and structure to emerge was a consequence of insufficient pressure for expressivity or whether it reflects general limitations of this signalling domain, either because it is unfamiliar or because it is unsuitable for structuring. To clarify this issue, the next experiment compared children and adults in their ability to exploit iconicity and induce structure when trying to create a novel signalling system for referential communication.

Experiment 2

The goal of this experiment was to examine how children differ from adults in their ability to create an expressive communication system based on an unfamiliar signalling domain like the binary auditory sequences used in Experiment 1. The results will provide insights into the extent to which children are motivated and capable to be sufficiently expressive to create shared signalling systems from scratch.

Method

<u>Participants:</u>

Twenty-four adults (twelve men, overall age range 20 to 26 years) and twenty-four monolingual children with English as native language (sixteen boys, all Primary grade level 3, which in Scotland typically comprises ages 6 to 7) were assembled into 12 same-sex pairs within each age group. Recruitment, consent and reward conditions were the same as in Experiment 1.

Materials:

We duplicated the eight cards used in Experiment 1 to create two identical sets, and used the same buzzers as in Experiment 1. In addition, each interlocutor received a coloured cup to collect reward tokens.

Procedure:

Children and adults were given the same child-appropriate instruction, which for adults was prefaced with the comment that they were invited to test a game designed to be suitable for 6-7-year-old children. Participants were sat on opposite sides of a table with the buzzers placed within equal reach in between them. After brief familiarisation with the cards, they were told that they would be using the buzzer language used by a species of aliens with no mouth, to communicate the identity of the cards to each other, switching between the roles of Director and Matcher. The interaction phase was preceded by a brief training phase, in which participants were exposed to the 'alien language', i.e. to random binary auditory sequences similarly to the ones used at the outset of the iterated learning in Experiment 1. In the training phase, children took turns to repeat the signal provided by the experimenter for each referent such that each child repeated half of the signals. The purpose of this very brief familiarisation was not for the participants to learn the sequences, but simply to expose them to an array of possible sequences to provide them with an understanding of the variation in the signals that could be used.

In the interaction phase, participants were invited to take turns signalling all referents so that their interlocutors were able to identify the cards correctly. For the first half of the adult and child dyads (i.e. dyads 1-6), the Matchers' cards were placed face up on the table by the Experimenter, taking special care that the random, unstructured spatial arrangement of cards would preclude a strategy of signalling cards by location, i.e. avoiding arrangements in rows, squares, rectangles or circles. The Directors then shuffled their cards and placed them face down on the table, drawing from the deck one by one. Placing the Matcher cards on the table had the advantage that participants were continuously reminded of all three relevant meaning dimensions, but had the disadvantage that participants could attempt positional coding despite instructions to avoid this strategy (i.e. indicating referents by the number of buzzes depending on position). To check whether positional coding could have influenced the results, the second half of adult and child dyads (i.e. in dyads 7-12)

underwent a slight change in procedure such that Matchers placed their eight cards away from the Director's view on a cardboard stand resembling an oversized Scrabble tile holder. This prevented positional coding but had the disadvantage that Directors did not have a reminder of the meaning dimensions in front of them. To control for this variation in procedure, the factor of Card Position (visible vs. hidden) was included in all analyses.

Once the cards were in place, Directors selected the first card keeping its face out of the Matcher's view, and signalled the depicted meaning using the buzzers to generate a sequence of tones. The Matcher then picked up the selected card and both participants placed the target cards in the middle of the table in view of the video camera, and received a reward token every time their cards matched. It was stressed that rewards would only be awarded jointly so Directors were encouraged to be cooperative in their signalling. The Matcher then returned the selected card to its previous position before the Director buzzed the sequence denoting the next card. After signals for all eight cards had been communicated, participants switched roles. This interaction continued for a total of five rounds, after which participants counted up their reward tokens. Adults were debriefed and thanked for their participation; children were allowed to choose several different reward stickers. While players operated the buzzers their hands were video-recorded for subsequent coding. Each session lasted between 35-45 minutes.

Results

In addition to the variables examined in Experiment 1, we also analysed the following variables (see Figure 3): (a) a measure of *alignment*, determined as length-normalised Levenshtein edit distance between binary sequences of both participants produced for the same meaning by the interlocutor on the previous turn, (b) *consistency*, defined as length-normalised Levenshtein edit distance between binary sequences produced by the same participant for the same meaning on two consecutive rounds, and (c) *accuracy* of card identifications as an indicator of communicative efficiency. With respect to this latter dependent variable it should be noted that the random sampling of cards without replacement rendered the Matchers' referent choices dependent on their memory capacity, and not just on the Directors' expressivity, as participants could attempt to remember which cards had already been signalled by the Director[7].

All dependent variables were analysed with GCAs using a mixed-effect model with Card Position (visible vs. hidden), Age Group (adults vs. children) and Round (1-5) and their interactions as fixed effects, and Participants nested within Dyads as random effects. To control for the variation of the spatial position of the Matcher cards, we included the factor of Card Position (i.e. matcher cards were either visible or hidden) in all analyses. As in Experiment 1, we included items (i.e. individual meanings) as another random effect where appropriate[8]. The results are given in Table 2.

DV		Inter-cept	CardPos	AgeGroup	Round	Card Pos x Age Group	CardPos x Round	AgeGroup x Round	3-way
model: DV ~ CardPosition * AgeGroup * Round + (1+Round|Dyad/Participant))									
unique patterns									
	β	7.21	-0.21	-0.06	-0.04	-0.08	0.09	-0.04	-0.05
	t	52.87***	-1.54	-0.47	-0.58	-0.56	1.33	-0.58	-0.80
compositional structure:									
	β	0.27	-0.19	-0.39	0.18	0.12	-0.18	0.12	0.08
	t	2.66*	-1.89	-3.86***	2.34*	1.13	-2.28*	1.54	1.06
Iconicity – correlations between:									
length - size									
	β	1.33	-0.43	-0.60	0.50	0.22	0.01	-0.54	-0.05
	t	5.23***	-1.66	-2.36*	2.63*	0.85	0.06	-2.84*	-0.28
pitch - shape									
	β	0.27	-0.15	0.04	0.05	0.05	0.29	0.03	-0.30
	t	1.06	-0.60	0.17	0.28	0.18	1.60	0.19	-1.64
pitch - brightness									
	β	-0.15	0.19	-0.02	-0.16	-0.35	0.08	0.19	0.14
	t	-0.77	0.95	-0.13	-0.94	-1.76	0.47	1.11	0.81
pitch change - shape									
	β	-0.07	-0.06	0.26	0.10	-0.04	-0.07	0.17	-0.03
	t	-0.49	-0.41	1.93	0.82	-0.27	-0.57	1.35	-2.18

		model: DV ~ CardPos * AgeGroup * Round + (1 + Round|Dyad/Participant) + (1 + Experiment * Age * Round|Item))						
accuracy (logit)									
	β	-1.00	0.13	-0.40	0.11	0.11	-0.01	-0.10	-0.02
	z	-8.66***	1.22	-4.48***	1.51	1.30	-0.08	-1.34	-0.29
signal length									
	β	6.89	-0.01	0.62	0.13	-0.08	-0.09	0.34	-0.00
	t	27.16***	-0.57	3.55***	1.00	-0.49	-0.70	2.75**	-0.01
LEDown									
	β	0.38	-0.00	0.02	-0.03	-0.01	0.01	0.01	-0.01
	t	35.00***	-0.01	2.24*	-3.43**	-1.14	0.84	2.34*	-2.08*
LEDother									
	β	0.41	0.00	0.02	-0.00	-0.00	0.01	0.02	0.0-
	t	33.33***	0.09	2.12*	-0.19	-0.53	0.52	2.36*	1.981
algorithmic complexity									
	β	0.26	0.02	0.17	-0.15	0.04	-0.03	0.03	0.03
	t	4.03**	0.37	3.01**	-3.70**	0.95	-0.88	0.91	0.95

Table 2: Coefficients for intercepts and fixed effects in the linear mixed-effect models for all dependent variables in Experiment 2 (LEDown: length-normalised Levenshtein edit distance to previous signal used for a given meaning by the same participant, LEDother: length-normalised Levenshtein edit distance to signal used for the same meaning by the other participant in the preceding round, alg. comp.: length-normalised algorithmic complexity of individual signals), *** p < .001, ** p < .01, * p < .05.

Only one dependent variable, compositional structure, showed an interaction between Card Position and Round, which we explored in separate analyses for both conditions. When cards were visible, there was an effect of AgeGroup, $\beta = -0.28$, $t = -2.30$, $p = .039$, confirming greater compositional structure in adults. When cards were hidden, there was also an effect of Age Group, $\beta = -0.51$, $t = -3.03$, $p = .011$, as well

as an effect of Round, $\beta = 0.37$, $t = 3.19$, $p < .003$, indicating that while higher in adults, compositional structure also increased over consecutive rounds of the interaction. This indicates that despite our attempts to prevent positional coding in the condition where cards were visible, some participants might have attempted it despite instructions to the contrary, which may have reduced occurrences of compositional innovations in later rounds. However, despite these effects it is important to reiterate that most z-scores were below 1.96, indicating that even in the adults, compositional structure did not emerge reliably.
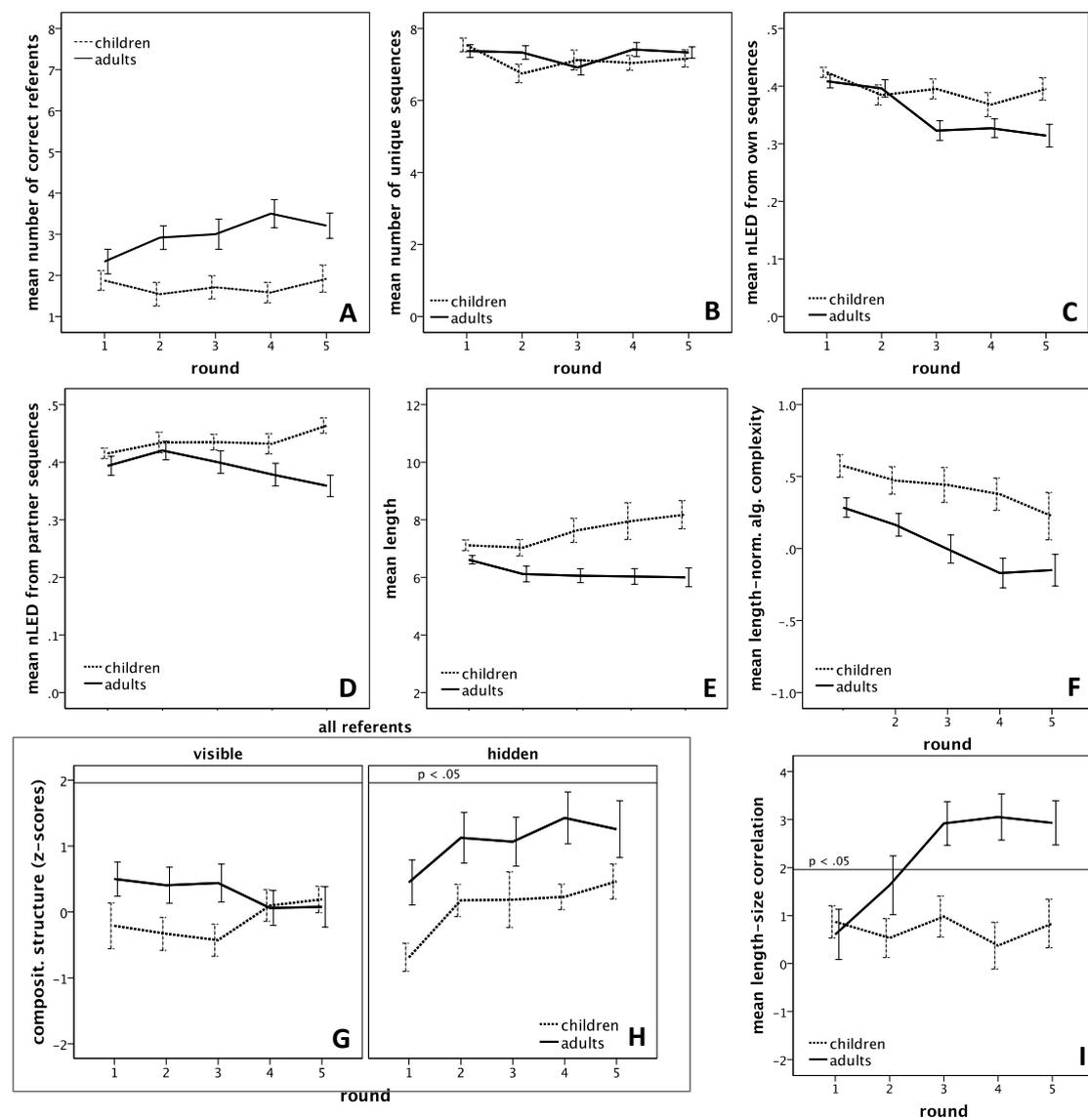

Figure 3: Number of correctly identified referents (panel A), number of different signals (out of 8) as an inverse measure of degeneration of the language (panel B),

length-normalised Levenshtein edit distance from own binary sequences produced on the previous round (panel C) and from binary sequences produced by interlocutor on the previous turn (panel D), binary sequence length (panel E), length-normalised algorithmic complexity of binary sequences as a measure of signal structure (panel F), z-scores of correlations between similarity of sequence pairs and meaning pairs indicating compositional structure as a function of Card Position (panels G and H), and iconic mapping between sequence length and referent size measured as z-scores of length-size correlations (panel I) for 12 child and 12 adult dyads interacting over 5 rounds in Experiment 2. Solid lines indicate z = 1.96, p = .05.   Values represent means across dyads; error bars represent ± 1 S.E.M aggregated by participants.

Discussion

In Experiment 2, pairs of adults and children communicated over five rounds using binary auditory sequences to signal a set of eight meanings differing in shape, size and brightness. Not unexpectedly, we found that adults outperformed the children in terms of communicative success. As expected for communicative interactions (Kirby et al., 2015), languages did not degenerate or shorten over repeated rounds. We also found that over the course of repeated interactions, individual signals were reproduced more consistently, especially in the adults who more frequently than the children reused the signals they had created previously. In contrast to Experiment 1, communicative interaction over five rounds led to increasingly structured individual signals, as indicated by a reduction in algorithmic complexity in both age groups, which in general was lower in the adults than in the children. This suggests that more frequent production of the same signals compared to Experiment 1 can drive an increase in signal structure, which, in turn, facilitates consistency of reproduction, as more structured sequences are easier to remember. As expected, without the learnability pressures that arise from inter-generational transmission (Kirby et al., 2015) interlocutors did not spontaneously introduce sufficient compositional structure that would have correlations between similarities in signals and meanings exceed chance levels, even though overall compositional structure was higher in adults.

Most notably, we observed that alignment between interlocutors increased in the adults but not in the children, as evidenced by the decreasing edit distance to signals produced by the interlocutor on the previous turn. Even at the end of the interaction children took little notice of the signals produced by their interlocutor,

while adults increasingly matched their interlocutors' productions. Note that simply imitating the interlocutor's most recent signals does not lead to alignment; interlocutors need to align signals with the corresponding meanings. There are three ways in which such alignment can be achieved. First, participants can memorise and reproduce the individual signals produced by their interlocutors, a strategy that is taxing on memory resources and therefore favours adults. Secondly, participants can discover and reproduce any systematicity in their interlocutor's way of linking signals to meanings, i.e. compositional structure. Our findings showed greater compositionality in adults, but not in children, in line with findings from Raviv and Arnon (2018), even if compositional structure still did no exceed chance. This age effect was mainly carried by the dimension of shape. Finally, alignment can be achieved through discovery and sharing of transparent, iconic links between signals and meanings. We found that adults, but not children, established such links by consistently and reliably associating bigger referents with longer sequences. The observation that magnitude symbolism emerged as the preferred cross-modal association lends lending further support to the notion that links between linguistic features that differ in magnitude (e.g. pitch or word length) and meaning dimensions that express quantity (e.g. size or complexity) appear to be fundamental and perhaps universal (Dingemanse et al., 2015). However, it should be pointed out that mapping signal length onto size *per se* is not sufficient to achieve alignment. For example, the hypothetical sequences *000000* and *111111* are perfectly matched in length but bear no resemblance in terms of their structure and are therefore not aligned. The fact that edit distance to interlocutor signals, the inverse measure of alignment, decreased in the adults suggests that they pursued other ways of linking signals to meanings in addition to relying on magnitude symbolism. Misalignment in children, then, may have been either due to a lack of understanding that mappings between signal and meaning need to be shared, inability to discover cross-modal links or inability to monitor form and communicative efficiency of their own and their interlocutor's productions, a task that is difficult for children of this age group, as it requires considerable processing capacity (Rabagliati & Robertson, 2017).

Two factors may be responsible for why sound-symbolic mappings between signal length and referent size did not emerge in the children, counter to predictions from the Iconic Bootstrapping Hypothesis (Imai & Kita, 2014). First, children's processing limitations may restrict their capacity to systematically scan the novel

signalling space for iconic affordances, and to monitor their own signals as well as interlocutor responses. Secondly, children may lack the fundamental insight that novel signals need to be informative, in the same way in which monolingual children often lack the meta-linguistic understanding that the same meanings can be expressed unambiguously by different signals in different languages (Bialystok, 1986). To distinguish between these options – the cognitive cost associated with using a novel signalling system vs. the lack of insight that signals need to be informative – Experiment 3 tested children's expressivity in a communication task that allowed them to use their native language.

Experiment 3

To establish whether children are capable of the expressivity that would be needed to create novel communication systems when not hampered by lack of familiarity with the signalling domain we examined pre-schoolers' referential communication with peers and adults in their native language. While there is evidence that children are able to monitor knowledge states of their addressees (Bahtiyar & Küntay, 2009; Liebal, Carpenter, & Tomasello, 2010; Moll, Richter, Carpenter, & Tomasello, 2008; Nadig & Sedivy, 2002; O'Neill, 1996), communication is generally unsuccessful when children do not have labels for referents readily available (Garrod & Clark, 1993; Glucksberg & Krauss, 1967; Glucksberg et al., 1966; Krauss & Glucksberg, 1969; Kahan & Richards, 1986; Kraus & Glucksberg, 1969; Nilsen & Graham, 2009; Sonnenschein & Whitehurst, 1984). There is also evidence that children show limited understanding of communicative failure (Robinson & Robinson, 1978) and require adult scaffolding to repair communicative breakdown (Matthews et al., 2007), which otherwise presents problems in peer interaction up until about 11-12 years of age (Garrod & Clark, 1993; Girbau, 2001). A second goal of Experiment 3 was therefore to explore how children adjust their communication strategies in response to such scaffolding to shed light on the conditions under which children are able to succeed in negotiating meaningful signalling systems.

To remain compatible with Experiments 1 and 2, we compared children's referential communication about nameable referents varying on three dimensions, so that informative expressions required production of a set of three modifiers. The need to produce modifiers created a situation that would discourage referential pacts (Brennan & Clark, 1996) so that unlike other studies (e.g. Krauss and Glucksberg,

1969), our stimuli allowed us to quantify informativeness of referring expressions directly by counting up the mentions of star attributes. If children are unable to produce informative, unambiguous referential expressions under these circumstances this would suggest that their inferior performance in transmitting and negotiating meaningful signals in Experiments 1 and 2 was not just due to lack of familiarity with the signalling domain but a reflection of their more general pragmatic limitations that impair communicative expressivity.

Method

Participants:

Thirty monolingual children with English as native language (13 boys) aged between 4;0 and 5;1 years (mean age 4;5 years) were tested in a nursery. Note that the age group was 1-2 years below Experiments 1 and 2. This was a pragmatic decision owed to the fact that it was only possible to recruit children with their mothers in a nursery but not in a school setting, and the implications for the comparison with Experiments 1 and 2 will be discussed below. Ten children (1 boy) were paired with their mother; the other children were paired with each other based on play preferences reported by nursery staff. Written parental consent was obtained for all children.

Materials:

In order to present children with a readily nameable object for which attributes would then have to be specified in the children's native language, we used stars as referents rather than the unfamiliar objects from Experiments 1 and 2, but preserved the size of the meaning space and the number of meaning dimensions. We created four identical sets of laminated cards sized 5 x 8 cm depicting five-pointed stars on white background. Within each set, stars varied on the dimensions of size (big vs. small), colour (red vs. blue) and shape ('spiky' vs. 'fat', depending on whether the angles of the points were acute or obtuse), resulting in eight different stars per set.

Procedure:

Participants were informed that they would play four rounds of a game that required them to instruct each other to find specific cards. The number of rounds was reduced to accommodate the fact that children were slightly younger than in Experiment 2. In the child-adult dyads, the child was always given the role of Director (i.e. the one to

describe the cards) for the first half of the interaction. In the child-child dyads, roles were assigned at random. In all dyads, roles were reversed in the second half of each round.

Because performance in referential communication tasks depends not only on the ability to appreciate the interlocutor's point of view but also on task demands associated with processing the meaning dimensions that may distinguish referents (Bishop & Adams, 1991), we thought it important to make sure participants attended to the relevant identifying features of the stars. The experimenter therefore first presented each participant with minimal pairs of stars differing only on one dimension, accompanied by questions about the three critical dimension (e.g. 'Show me the big / small one.' etc.). Both adults and children were successful at selecting the correct stars under these conditions. Next, both participants were seated at a table at opposite sides of a barrier, and were given two sets of cards each, placed face up on the table in no particular order. The experimenter then selected four of the Director's cards according to a pre-determined pseudo-randomised order such that all four cards were always different, and placed them in a row in front of the remaining 12 cards. The rationale behind having participants select from two, rather than just one, set of cards was to discourage identification of cards by elimination (i.e. to prevent participants from realising that if the big, red, spiky star had already been selected, the only other big red star would have to be the fat one, rendering the shape attribute over-informative). Using two sets ensured that all three attributes were always required to be informative (i.e. even if one big, red, spiky star had already been mentioned, there was still another one).

The Director was then asked to describe the four cards in the order they were laid out so that the Matcher would be able to find the described targets in their set of cards. The Matcher was asked to identify the cards based on the Director's description and to arrange them in a row. No restrictions were imposed on the amount and type of communication. If children were reluctant to provide verbal descriptions, one experimenter, seated behind the Director, would gently encourage them to do so with prompts like 'Tell Mummy / X what the star is like.' After the four cards had been described and identified by the Matcher, the barrier was lifted to enable Director and Matcher to see whether the selected cards matched, and the experimenter to take a photograph of the cards. All children received a yellow smiley-sticker for each correctly matched card. The barrier was then replaced, and the cards were returned to

the overall pool, before the roles of Director and Matcher were reversed. Next, the experimenter helped the Director to select four different cards, which again were placed in a row in front of the other cards, before the game continued in the same way as before the role reversal. The subsequent check of the matched cards concluded one round of the game. Participants then completed the remaining three other rounds of the game so that each participant served as Director four times.

Coding:

Audio recordings were transcribed and two coders independently coded Directors' mentions of the three relevant features. Discrepancies arose only in 2.6 % of cases, and were resolved by jointly re-coding the utterances. To establish informativeness of the Director's utterances we coded, for each star, whether Directors mentioned each of the features (colour, shape and size) on their first utterance that was produced without any prior communication from the Matcher, as a measure of informativeness of Directors' spontaneous expressions before receiving any feedback or clarification requests. We also coded the number of features that were mentioned over the course of an entire exchange pertaining to one card. We did not code whether the Matcher selected the correct card because accuracy is not a suitable measure of referential efficiency due to differences in comprehension ability between adults and children. Because the interaction was unconstrained, the information provided by the Director could be distributed over a number of conversational exchanges that the dyad engaged in until the Matcher selected a star. We therefore also coded the number of turns the Director took and whether utterances produced on those turns were volunteered by the Director, i.e. produced without any prompting or clarification requests from the Matcher. This measure enabled us to determine to what extent children provided information on their own accord without feedback from their addressee.

Results

Our analyses had to take into account that performance was dependent on role position: Participants acting as Directors second could produce more informative expressions because they may have gained prior communicative experience when performing the role of Matcher. Because the adults always acted as Directors second, a comparison of this role position would thus entail comparing performance between adults and children which would reveal trivial results given the adults' undoubtedly

superior performance in the task. We therefore only analysed the data for the children who acted as Directors first, interacting either with a peer or with their mother. We analysed the following dependent variables, depicted in Figure 4:

(a) *Informativeness of the first referring expression* indicates how much information children produced on their very first utterance on their own accord, before the addressee provided any feedback. Because all stars were present twice in the entire set (see Procedure), Directors always had to mention all three attributes on every trial, i.e. describe a star as '*big, red and spiky*'. Informativeness was therefore quantified as number of mentioned modifiers, i.e. referent features (out of three).

(b) *Informativeness of the entire communicative episode*. The number of features produced over the course of an entire communicative episode does not just indicate the child's referential competence but also depends on the feedback from the Matcher. It is therefore a measure of the communicative efficiency that children can achieve as a part of a dyad, and is instructive about the communicative strategies that children adopt in response to Matcher feedback.

(c) Number of turns within a communicative episode pertaining to a single referent is a measure of the overall amount of communication that took place.

 (d) Number of volunteered turns indicates how many turns children took without prompting or clarification requests from their interlocutor. After the children had provided their initial descriptions, subsequent turns could have been either volunteered if the addressee did not request further information, or could have been prompted by specific (e.g. 'What colour is the star?') or general (e.g. 'Tell me more!') clarification requests. The number of unprompted (volunteered) turns is a measure for children's attempts to provide information spontaneously, and how it changes over the course of interaction.

We performed GCAs using linear mixed effect models with Addressee (child vs. adult) and Round (1 to 4) as centered fixed effects and crossed random effects of participants and items (i.e. the eight star types). We used maximal random effect structure as permitted by the data (see Supplementary Materials at DOI 10.17605/OSF.IO/RQPD5)[9]. We also fitted models with a fixed effect of Round to the data from each group separately when there was an interaction between Addressee and Round.

The results, provided in Table 3, showed that at the outset of the interaction, the 1.2 out of three features produced by children interacting with a peer were not

significantly different from the 1.3 features produced by children interacting with an adult, and there was no significant increase over the course of the interaction. However, there was a dramatic difference in how much information children were able to provide as a result of the interaction: At the end of a communicative episode, children interacting with a peer showed no improvement mentioning on average still only 1.4 features, compared to the dramatic improvement resulting in an average of 2.7 features provided by children interacting with adults. This difference was due to feedback and clarification requests from the adult as indicated by the significantly larger average number of 3.0 turns taken in child-adult dyads, compared to only 1.3 turns in child-child-dyads. The analysis also showed that the number of volunteered turns increased only in the children who interacted with adults, suggesting that these children were learning from the interaction with the adult to provide more information on their own accord.

Dependent Variable		Intercept	Addressee	Round	Addressee x Round
		model: lmer (DV ~ Addressee * Round + (1 + Round|Participant) + (1 + Addressee * Round |Item)			
# features 1st utterance	β	1.66	-0.15	0.15	-0.02
	t	14.84***	-1.39	2.03	-0.35
# features overall	β	2.02	-0.64	0.12	-0.04
	t	19.76***	-6.22***	2.73*	-0.92
# turns	β	2.13	-0.82	-0.19	0.07
	t	8.26***	-3.17**	-2.23	0.85
# volunteered turns	β	1.36	-0.12	0.09	-0.18
	t	16.35***	-1.42	1.36	-2.52*
child- adult	β	1.48		0.27	
	t	11.39***		2.50*	

| child- child | β | 1.25 | | -0.08 |
| | t | 11.38*** | | -0.10 |

Table 3: Coefficients for intercepts and fixed effects in the linear mixed-effect models for all dependent variables in Experiment 3, *** $p < .001$, ** $p < .01$, * $p < .05$.
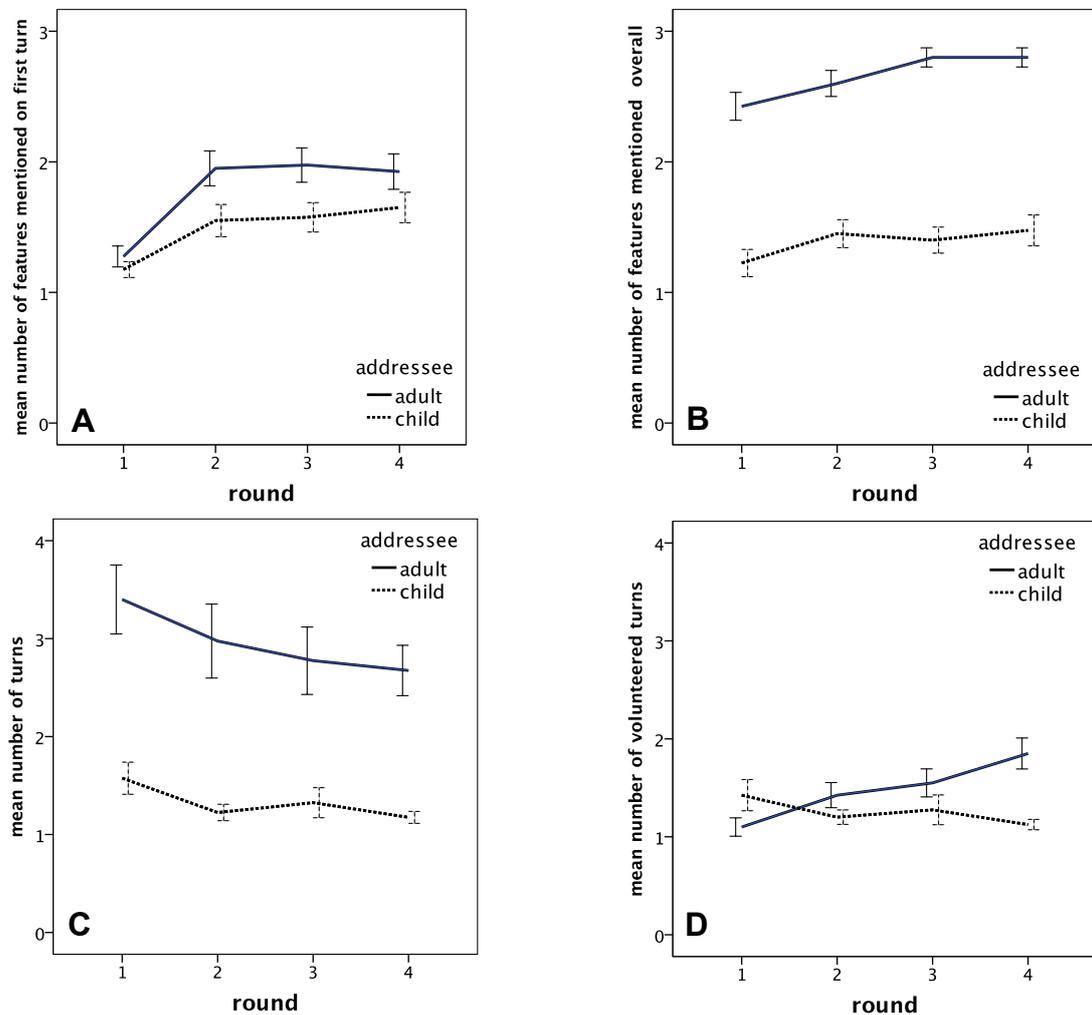

Figure 4: Mean number of features per referent (out of 3) mentioned on first turn (panel A) and during entire conversational episode (panel B), mean number of turns taken per referent (panel C) and mean number of volunteered turns per referent (panel D) for children addressing either a peer or an adult (their mother). Error bars represent ± 1 S.E.M aggregated by participants.

Discussion

Our final experiment examined informativeness of modifying expressions used to describe a meaning space of the same size and complexity as used in Experiments 1

and 2. The findings showed that 4-5-uear old children have extremely limited ability to produce informative and unambiguous referential expressions in peer interaction, in accord with previous findings (e.g. Krauss & Glucksberg, 1969). Despite the fact that the signalling domain – their native language – was familiar to them, children typically mentioned less than half of the defining features in their first, unprompted utterances. Only children interacting with adults, but not children interacting with peers, ended up supplying the required information, as illustrated in (1). Children interacting with peers, on the other hand, produced considerably less informative expressions than were needed to identify the cards. Thus, at this age, children require feedback and scaffolding for communication to be efficient yet their peers do not provide such feedback: The small amount of turns in the child-child dyads indicates that peer interlocutors did not request the missing information when they found themselves on the receiving end of uninformative referential expressions, as illustrated in (2). This shows that in order to develop communicative efficiency, children require scaffolding through clarification requests and feedback that only adults can provide (Matthews et al., 2007). As a result of such scaffolding, children learn to volunteer the required information on their own accord.

(1) Child: ...

Adult: *What colour is it?*

Child: *Erm – red.*

Adult: *And is it a big star or a small star?*

Child: *Erm - a big star.*

Adult: *And is it a fat one or a skinny one?*

Child: *A skinny one.*

(2) Child 1: *A skinny one.*

Child 2: ...

Child 1: (whispers) *A skinny one.*

Child 2: ...

Child 1: *Skinny.*

Child 2: ...

It could be argued that the results of Experiment 3 are not relevant as a potential explanation for the inability to negotiate a novel signalling system because the children tested in this experiment were about one to three years younger than the children in Experiments 1 and 2. However, the literature on children's difficulties with communicating about unfamiliar referents in their native language suggests that development of referential communication is long and protracted, and certainly not concluded until children reach middle childhood (e.g. Garrod & Clark, 1993; Krauss & Glucksberg, 1969). The findings of Experiment 3 therefore support the notion that it is the lack of pragmatic insight necessary for efficient communication, rather than the lack of familiarity with novel signalling domains, that prevented children from learning and creating novel communication systems in Experiments 1 and 2.

General Discussion

In this study, we found that iterated learning of an entirely novel signalling system did not readily lead to emergence of iconicity or structure, at least not under limited learning conditions and without a strong pressure to communicate. As a result, adults outperformed children in terms of learning and transmitting the system faithfully based mainly on their more mature cognitive abilities, which enabled them to memorise more of the unfamiliar signals. However, when the novel signals were used for referential communication, adults, but not children, developed shared systems by exploiting iconic mappings between referent size and signal length as well as by inducing signal structure and, to some extent, compositional structure. One possible reason for why compositional structure was not as pronounced in the adults to reliably exceed chance levels is that lack of familiarity with the signalling system may have prevented adults from applying their meta-linguistic knowledge of linguistic structure. This invites some caution with respect to findings based on language-like stimuli reported elsewhere, which may bias participants towards use of meta-linguistic knowledge. It is also possible that binary auditory sequences may just be very difficult to learn, not the least because of their fading nature. However, the crucial finding is that under communicative pressure adults were able to align their productions and to negotiate a system based on iconicity whereas children were not. Experiment 3 found that even when given the opportunity to use their native language children were unable to produce informative expressions that would enable interlocutors to identify the intended meanings. This points to substantial pragmatic

limitations in children as one of the possible causes for their inability to negotiate meaning and to produce motivated signs, be it iconic or compositional ones. Children only attained communicative efficiency when adult expert language users provided scaffolding through feedback and clarification questions to elicit missing information. While these findings clearly call for further research that explores factors like length of training or frequency of signal use, they so far provide no indication that children are more likely than adults to introduce the kinds of innovations that could drive the emergence of structure during cultural transmission and use of language. Below we provide further suggestions for what may prevent children from creating and transmitting novel communication systems, what this implies for the role of children in language change, and what conditions may allow children to overcome those limitations.

The nature of children's limitations in creating and transmitting novel signalling systems

When learning and using a novel signalling system, children did not compensate for their lower cognitive capacity by imposing structure more readily than the adults. Thus, our results do not suggest that 'less is more' (Newport, 1990) with respect to emergence of structure during language transmission. Instead, they are in line with evidence that larger processing capacity benefits language processing and language learning in general (e.g. Daneman & Merikle, 1999; Braine, Brody, Brooks, Sudhalter, Ross, Catalano & Fisch, 1990; Brooks, Braine, Catalano, Brody & Sudhalter, 1993; Snow & Hoefnagel-Höhle, 1978). We suggest several pathways by which more mature cognitive abilities may actually benefit creation and transmission of novel signalling systems.

One possibility is that adults have a better insight into the pragmatic requirement that interlocutors need to be able to map signals onto meanings, which is likely to drive what Lister & Fay (2017) termed the creation of 'motivated signs'. This insight may be linked to mentalising abilities that arise from a more mature Theory of Mind. In contrast, children may not yet understand that motivated signs are needed in the first place, by assuming that meaning is an intrinsic property of a signal rather than a negotiated convention.

Another possibility is that children appreciate the need for motivated signs but have insufficient means to create them – either because they lack more sophisticated

aspects of Theory of Mind necessary to predict which signal-meaning mappings are transparent enough for an addressee to recover, or because they have insufficient processing capacity to monitor whether their own and their addressee's signals express the newly established meanings in a consistent way. Thus, to create motivated iconic signal-meaning mappings, interlocutors need to systematically scan the signalling domain for iconic affordances, and monitor the communicative efficiency of the selected iconic mappings. Although children are sensitive to statistical regularities in the linguistic form from a very early age, linking novel forms to meanings is cognitively very demanding (Naigles, 2002), even if the signalling domain affords considerable potential for iconic mappings. However, while the Iconic Bootstrapping Hypothesis suggests that children benefit from iconicity in early word learning (Imai & Kita, 2014), this does not necessarily imply that children are also more predisposed towards using iconicity in innovative ways in novel signalling systems, an ability that may require considerable processing capacity, especially when it relies on careful appraisal of the structure of the meaning space – a demand that may have limited ecological validity in heavily contextualised day-to-day communicative interactions (Bishop & Adams, 1991).

Finally, children may be less adept at aligning their own signals with those of their interlocutor, as our findings suggest. Lister & Fay (2017) propose that behavioural alignment drives cognitive alignment such that when interlocutors copy signals produced by their communication partners this will gradually lead to the establishment of shared cognitive representations. However, behavioural alignment requires monitoring of the match between one's own and the interlocutor's productions, to compute an error term in cases of mismatch, and to adjust subsequent productions to minimise this error term. Given the evidence for children's limitations in monitoring the informativeness of their own and their interlocutor's expressions (Rabagliati & Robertson, 2017), it is not surprising that behavioural alignment seems to be difficult for children.

The role of children's processing limitations in language change

The finding that motivated and structured signal-meaning mappings did not emerge in 4-7-year-old children challenges the idea that processing limitations make children important agents of language evolution and language change (Bickerton, 1981; Dale & Lupyan, 2012; Hudson Kam & Newport, 2005; 2009; Lumsden, 1999). One

domain in which a role for children has been demonstrated empirically is the spontaneous emergence of grammatical structure in the home sign systems of deaf children (Goldin-Meadow, 2003), and the observation that younger cohorts of deaf children were responsible for introducing structure into the newly emerging Nicaraguan Sign Language (Senghas & Coppola, 2001; Senghas, Kita & Özyürek, 2004). These younger cohorts tended to grammaticise holistic gestures, either by linearising simultaneously occurring elements of gestural signs, or by restricting the meaning of certain aspects of the signs, such as using spatial location to denote a specific type of co-reference. How can these observations be reconciled with children's inability to induce structure and create motivated signs in the present experiments?

One important difference between our findings and the findings from Nicaraguan Sign Language is that the signalling domain of gesture was already familiar to the sign-learning children, as they all had used home sign systems before encountering sign language for the first time. Pre-existing gestural representations may have supported decomposition of gestural input. If such easily decomposable signals are then transmitted from cohort to cohort of learners, just like artificial languages are transmitted in iterated learning studies (e.g. Kirby et al., 2008; 2015), structure is likely to emerge. Consequently, emergence of structure in subsequent cohorts cannot be taken as evidence that it was the children's age and the associated limited processing capacity, rather than their position in the transmission chain, which is the driving factor.

Our findings are more in line with socio-phonetic observations and associated models of diachronic change which suggest that linguistic innovations are introduced and sustained by cognitively more mature agents, specifically, female adolescents and young adults (Labov, 2001). Children, on the other hand, at least up to the age of five, display distributions of form variants that are similar to their caregivers', and only start to engage in vernacular reorganisation later in childhood (Kerswill & Williams, 2000), reaching peak usage of novel variants by approximately age 17 (Cedergren, 1988). Such a trajectory of change is not confined to phonetic changes but extends to morphological, syntactic and discourse-pragmatic features (Tagliamonte & D'Arcy, 2009) as well. On this account, young children are conservative learners unlikely to introduce or take up innovation beyond what it is they are able to retain from the input (Tomasello, 1992). Moreover, children's alleged propensity to regularise does not

necessarily lead to language change either: Agent-based simulations of factors that govern spread of linguistic innovations within social networks also demonstrate a very limited role of regularising, as opposed to probability-matching, biases (Pierrehumbert, Stonedahl & Daland, 2014). These simulations show that in the face of inconsistent input where traditional forms and innovations coexist, the latter are only able to spread through the population if agents adopt decision rules that are closer to the probability matching, rather than the dichotomising end of the continuum of decision-making biases. Thus, some of the evidence points in the direction of adults as the innovators responsible for emergence and change of linguistic structure.

Overcoming children's limitations

Experiment 3 demonstrated that children's expressivity improved dramatically through scaffolding provided by expert language users (Matthews et al., 2007). This points to an important caveat in studies exploring iterated language learning and referential communication: Inter-generational transmission of language does not occur between adults, who are expert language users, as simulated in previous studies (e.g. Kirby et al., 2008; 2015), nor does it occur between children, who are cognitively constrained novices, as examined in this study. Instead, language is transmitted from adults to children, and a vast body of research on the nature of child-directed speech (for reviews see Golinkoff, Can, Soderstrom & Hirsh-Pasek, 2015; Saint-Georges, Chetouani, Cassel, Apicella, Mahdhaoui, et al., 2013; Soderstom, 2007) has documented that adults adjust their communication to accommodate children's cognitive and linguistic limitations, although the universality of this adjustment is debated (e.g. Cristia, Dupoux, Gurven & Stieglitz, 2017; Lieven, 1994; Schieffelin & Ochs, 1989). Moreover, not only do adults pre-select the input provided to children in a manner that may aid learning (Ferguson, 1964; Fernald, Taeschner, Dunn, et al., 1989; Eaves, Feldman, Griffiths & Shafto, 2016) but, as Matthews et al. (2007) and Experiment 3 of this study suggest, they also provide feedback that teaches children how to monitor the communicative efficiency of their productions.

Recently, the role of accommodation to novices in the cultural transmission of language has started to receive some attention in empirical (Atkinson, Smith & Kirby, 2018) and computational (Frank & Smith, 2018) research. This work shows that adult experts modify their language for the benefit of adult novices, and, thus, constitutes a simulation of foreigner-directed rather than child-directed speech. The findings

presented here suggest that input enhancement and scaffolding of communication strategies for child learners may be an important factor that complements the constraints imposed by the nature of the signalling domain, the cognitive capacity of the learner and the social environment in which communication takes place. Given the growing interest in how teaching as a universal behaviour (Kline, 2015) may have shaped cultural evolution in general (Csibra & Gergely, 2009), future research should attempt to integrate insights from the vast body of research on the role of child-directed communication in language development with those from research on the creation and transmission of novel signalling systems to add another interesting facet to our understanding of how human language might have evolved.

Notes:

[1] It has been argued that imposing a transmission bottleneck, i.e. presenting learners with only a sub-set of the signal-meaning mappings during learning (Smith, Kirby & Brighton, 2003), can be taken as a simulation of the role of child learners in this process (Vogt, 2005). The underlying rationale is that children have to induce structure based on much more limited input than adults. Yet poverty of the stimulus is intrinsic to any language learning situation, regardless of age. Thus, merely imposing a transmission bottleneck is insufficient to study the constraints imposed by child learners.

[2] Note that this operationalisation of signal structure is different from combinatorial structure as described in the Introduction, i.e. from the extent to which the signals within an entire signalling system (i.e. a language) consist of re-combinations of smaller sub-components (e.g. doubles, triplets, quadruplets), which can be taken as a loose analogy to phonemes in natural languages. For other novel signalling systems, combinatorial structure has been estimated either through measures of entropy (Verhoef, 2012; Verhoef et al., 2014), through a Form Recombination Index which captures how often a sub-component of a signal is re-used in other signals of that language (Roberts et al., 2015), by applying compression algorithms (Cornish et al., 2013) or by computing Associative Chunk Strength (Cornish et al., 2017). The first two methods rely on individual components being easily discriminable, e.g. through periods of silence in auditory signals or through interruptions in continuous lines. This is not possible in our case: Even though binary auditory sequences can be decomposed into smaller units like doubles or triplets, we are unable to determine how an individual learner might have decomposed a given sequence, e.g. whether the sequence *000111* (where 0 represents the high and 1 the low tone) was decomposed as *00-01-11* or *000-111*. The third measure, size of the entire language after compression, is suitable to compare amount of structure across languages with identical numbers of signals but is heavily influenced by degeneration such that languages with fewer signals will be more compressible thereby obscuring degree of combinatoriality. Finally, Associative Chunk Strength is uninformative given the limited number of possible bigrams contained in binary sequences. So even though inverse algorithmic complexity is not a measure of combinatorial structure for entire languages because we cannot tell to what extent the sub-components or chunks from

any given signal were also used in the other signals, it still provides a proxy for how efficiently learners were able to chunk individual signals.

[3] Consent conditions did not allow us to gain access to the records of the children to ascertain their exact ages but all children were within the normal age range expected for their grade level.

[4] Because primary grade level was controlled within chains we were also able to explore developmental trajectories within the children by performing separate GCAs with the fixed effects of Grade Level (1st vs. 2nd vs. 3rd) and Generation and their interactions for the children only, which did not yield any significant effects (see Supplementary Materials at DOI 10.17605/OSF.IO/RQPD5).

[5] Following Beckner et al. (2017), we also explored the shape of the trajectory of generational change by including a quadratic term of Generation to capture potential non-linearity, but only when the effect of Generation was significant. For the number of unique patterns, the model *AgeGroup + Generation + Generation^2 + AgeGroup : Generation + AgeGroup : Generation^2 + (1 + Generation + Generation^2 | Chain)* yielded a better overall fit than the linear model in Table 1, $\chi^2 = 17.78$, df = 5, p = .003, and confirmed the declining linear trend over Generations, $\beta = -0.35$, t = 2.72, p = .02. However, neither the quadratic term nor the interactions with Age Group showed significant effects based on t-tests, all p's > .11; it therefore seems warranted not to make strong claims about whether the declining number of unique patterns levelled off in later generations or not. For the length-normalised Levenshtein edit distance, we fitted the model *AgeGroup + Generation + Generation^2 + AgeGroup : Generation + AgeGroup : Generation^2 + (1 + Generation + Generation^2|Chain) + (1 + AgeGroup + Generation + Generation^2 + AgeGroup : Generation + AgGroupe : Generation^2|Referent)* which failed to converge. Uncorrelating the intercept from the slopes and slope interactions by Referents showed an improved model fit to the model with the linear term of Generation (see Table 1), $\chi^2 = 32.71$, df = 1, p < .001, and yielded main effects of the linear, $\beta = -0.04$, t = -4.38, p = .001, and the quadratic, $\beta = 0.03$, t = 2.82, p = .018, term of Generation, while the effect of Age Group was now no longer significant, p = .09. This suggests that the decline in edit distance levelled out in both age groups thereby presumably somewhat reducing the difference between children and adults as both groups reached a similar asymptote in learnability.

[6] To estimate the extent to which signal structure represented individual semantic dimensions, we analysed within-category measures of structure. For all pairs of signals and their associated meanings we computed signal similarity based on Levenshtein edit distance and Pearson product-moment correlations with differences in just one dimension of interest – size, shape or brightness – using the Mantel test (Mantel, 1967) as described for overall compositional structure. There were no effects of Generation or Age Group on these within-category measures of structure (see Appendix 1).

[7] As in Experiment 1, we also tested whether any of the individual meaning dimensions were given preference when linking individual meaning dimensions to signals. This was again computed over correlations between similarity in the binary sequences and similarity in referent size, shape, and brightness separately using the Mantel-test (Mantel, 1967). These analyses showed an effect of Age Group on shape-based structure indicating that there was more structure associated with the dimension of shape in the languages produced by adults, as well as an overall increase in size-based structure, even though none of the dimension-specific signal-meaning correlations exceeded the level of chance (see Appendix 1).

[8] As in Experiment 1, we only included a quadratic term of Round to examine non-linear trends over the five rounds when the linear effect of Round was significant. In cases where the more complex models containing the quadratic term did not converge, we eliminated slope interactions by items until the models converged. The dependent variables that showed significant effects of Round were overall compositional structure, size-based structure, length-size correlation, and edit distance from own previous pattern. Adding the quadratic term of Round as well its interaction with Age Group provided a significantly better fit of the model to the data only in the case of size-based compositional structure, $\chi^2 = 10.97$, df = 4, p = .026, even though the quadratic effect of Round did not reach significance, p = .150.

[9] In this experiment, we only modelled linear effects of Round as having only four rounds of interaction made the inclusion of a quadratic term in analogy to the previous experiments superfluous.

References:

Atkinson, M., Smith, K. & Kirby, S. (2018) Adult learning and language simplification. *Cognitive Science.* Advance online publication.

Bahtiyar, S., & Küntay, A. C. (2009). Integration of communicative partner's visual perspective in patterns of referential requests. *Journal of Child Language, 36(03)*, 529-555.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68(3),* 255-278.

Bartlett, F. C. (1932). *Remembering.* Oxford, UK: Macmillan.

Beckner, C., Pierrehumbert, J. B., & Hay, J. (2017). The emergence of linguistic structure in an online iterated learning task. *Journal of Language Evolution, 2*(2), 160-176.

Bialystok, E. (1986). Factors in the growth of linguistic awareness. *Child Development,* 498-510.

Bickerton, D. (1981). *Roots of language.* Ann Arbor, MI: Karoma.

Braine, M. D., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J. A., Catalano, L., & Fisch, S. M. (1990). Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Language, 29(5),* 591-610.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22(6)*, 1482-1493.

Brooks, P. J., Braine, M. D., Catalano, L., Brody, R. E., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language, 32(1),* 76-95.

Carr, J. W., Smith, K., Cornish, H., & Kirby, S. (2017). The cultural evolution of structured languages in an open-ended, continuous world. *Cognitive Science, 41*, 892-923.

Cedergren, H. (1988). The spread of language change: Verifying inferences of linguistic diffusion. *Language spread and language policy: Issues, implications, and case studies*, 45-60.

Cochran, B. P., McDonald, J. L., & Parault, S. J. (1999). Too smart for their own good: The disadvantage of a superior processing capacity for adult language learners. *Journal of Memory and Language, 41(1),* 30-58.

Cornish, H., Smith, K., & Kirby, S. (2013). Systems from sequences: An iterated learning account of the emergence of systematic structure in a non-linguistic task. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 340-345).

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2017). Child-Directed Speech Is Infrequent in a Forager-Farmer Population: A Time Allocation Study. *Child Development.*

Csibra, G. & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13(4),* 148–153.

Dale, R., & Lupyan, G. (2012). Understanding the origins of morphological diversity: The linguistic niche hypothesis. *Advances in Complex Systems, 15*(03n04), 1150017.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3(4),* 422-433.

Del Giudice, A. (2012). The emergence of duality of patterning through iterated learning: precursors to phonology in a visual lexicon. *Language and Cognition, 4(4),* 381–418.

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

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

Eaves Jr., B. S., Feldman, N. H., Griffiths, T. L., & Shafto, P. (in press). Infant-directed speech is consistent with teaching. *Psychological Review.*

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48(1),* 71–99.

Eryilmaz, K., & Little, H. (2017). Using leap motion to investigate the emergence of structure in speech and language. *Behavior Research Methods, 49*(5), 1748-1768.

Fay, N., Ellison, M., & Garrod, S. (2014). Iconicity: From sign to system in human communication and language. *Pragmatics & Cognition, 22(2),* 244-263.

Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist, 66(6, Part 2)*, 103-114.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16(03)*, 477-501.

Flaherty, M., & S. Kirby (2008) Iterated language learning in children. *Proceedings of the 7th International Conference (EVOLANG7).* Vol. 12.

Fort, M., Lammertink, I., Peperkamp, S., Guevara‐Rukoz, A., Fikkert, P., & Tsuji, S. (2018). Symbouki: a meta‐analysis on the emergence of sound symbolism in early language acquisition. *Developmental Science,* e12659.

Frank, S. & Smith, K. (2018) A model of linguistic accommodation leading to language simplification. *Proceedings of the 40th Annual Meeting of the Cognitive Science Society,* 390-395.

Galantucci, B., Garrod, S., & Roberts, G. (2012). Experimental semiotics. *Language and Linguistics Compass, 6(8),* 477-493.

Garrod, S., & Clark, A. (1993). The development of dialogue co-ordination skills in schoolchildren. *Language and Cognitive Processes, 8(1),* 101-126.

Garrod, S., & Galantucci, B. (2011). Experimental Semiotics: A Review. *Frontiers in Human Neuroscience, 168(6)*, 161.

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of representation: where might graphical symbol systems come from*? Cognitive Science, 31(6),* 961-987.

Gauvrit, N., Soler Toscano, F., Zenil, H., & Delahaye, J.-P. (2014). Algorithmic complexity for short binary strings applied to psychology: A primer. *Behavior Research Methods, 46(3),* 732-744.

Girbau, D. (2001). Children's Referential Communication Failure The Ambiguity and Abbreviation of Message. *Journal of Language and Social Psychology, 20(1-2)*, 81-89.

Glucksberg, S., Krauss, R. M., & Weisberg, R. (1966). Referential communication in nursery school children: Method and some preliminary findings. *Journal of Experimental Child Psychology, 3,* 333–342.

Goldin-Meadow, S. (2003). *The resilience of language: What gesture creation in deaf children can tell us about how children learn language.* New York: Psychology Press.

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby) talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science, 24(5),* 339-344.

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.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B, 369(1651),* 20130298.

Jones, J. M., Vinson, D. P., Clostre, N., Zhu, A. L., Santiago, J., & Vigliocco, G. (2014). The Bouba effect: Sound-Shape Iconicity in Iterated and Implicit Learning. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society,* Quebec City, Canada: Cognitive Science Society (pp. 2459-2464).

Kahan, L. D., & Richards, D. D. (1986). The effects of context on referential communication strategies. *Child Development, 57(5),* 1130-1141.

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., 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.

Kline, M. A. (2015). How to learn about teaching: An evolutionary framework for the study of teaching behavior in humans and other animals. *Behavioral and Brain Sciences, 38*, e31.

Köymen, B., Schmerse, D., Lieven, E., & Tomasello, M. (2014). Young children create partner-specific referential pacts with peers. *Developmental Psychology, 50(10)*, 2334.

Krauss, R. M., & Glucksberg, S. (1969). The development of communication: Competence as a function of age. *Child Development,* 255-266.

Labov, W. (2001). Principles of linguistic change Volume 2: Social factors. *Language in Society*. Oxford: Blackwell.

Lewis, M. L., & Frank, M. C. (2016). The length of words reflects their conceptual complexity. *Cognition, 153,* 182-195.

Lewis, M., & Frank, M. C. (2015). Conceptual complexity and the evolution of the lexicon. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society,* Austin, TX: Cognitive Science Society (pp. 1138-343).

Liebal, K., Behne, T., Carpenter, M., & Tomasello, M. (2009). Infants use shared experience to interpret pointing gestures. *Developmental Science, 12,* 264–271.

Lieven, E. V. M. (1994). Crosslinguistic and crosscultural aspects of language addressed to children. In C. Gallaway & B. J. Richards (Eds.), *Input and interaction in language acquisition* (pp. 56–73). Cambridge: Cambridge University Press.

Lister, C. J., & Fay, N. (2017). How to create a human communication system: A theoretical model.

Little, H., Eryılmaz, K., & de Boer, B. (2017). Signal dimensionality and the emergence of combinatorial structure. *Cognition, 168*, 1-15.

Lumsden, J. S. (1999). Language acquisition and creolization. In M. Degraff (Ed.), *Creolization, diachrony, and language acquisition* (pp. 129–157). Cambridge, MA: MIT Press.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS one, 5(1)*, e8559.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research, 27,* 209–220.

Mathy, F., Fartoukh, M., Gauvrit, N., & Guida, A. (2016). Developmental abilities to form chunks in immediate memory and its non-relationship to span development. *Frontiers in Psychology*, 7.

Matthews, D., Lieven, E., & Tomasello, M. (2007). How toddlers and preschoolers learn to uniquely identify referents for others: A training study. *Child Development, 78*(6), 1744-1759.

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: sound-shape correspondences in toddlers and adults. *Developmental Science, 9(3),* 316-322.

Moll, H., Richter, N., Carpenter, M., & Tomasello, M. (2008). Fourteen- month-olds know what ''we" have shared in a special way. *Infancy, 13*, 90–101.

Monaghan, P. (2014). Age of acquisition predicts rate of lexical evolution. *Cognition, 133(3),* 530-534.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science, 13(4),* 329-336.

Naigles, L. R. (2002). Form is easy, meaning is hard: Resolving a paradox in early child language. *Cognition, 86(2),* 157-199.

Naigles, L. R. (2002). Form is easy, meaning is hard: Resolving a paradox in early child language. *Cognition, 86(2),* 157-199.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science, 14(1)*, 11–28.

Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology, 58(2),* 220-249.

Nowak, I., & Baggio, G. (2016). The emergence of word order and morphology in compositional languages via multigenerational signalling games. *Journal of Language Evolution,* lzw007.

Nowak, M. A., D. C. Krakauer & A. Dress. 1999. An error limit for the evolution of language. *Proceedings of the Royal Society of London Series B-Biological Sciences 266(1433),* 2131–2136.

O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development,* 659-677.

Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology, 114(2),* 173-186.

Perfors, A. (2012). Probability matching vs over-regularization in language: Participant behavior depends on their interpretation of the task. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 845-850).

Perfors, A. (2012a). When do memory limitations lead to regularization? An experimental and computational investigation. *Journal of Memory and Language, 67*(4), 486-506.

Perry, L. K., Perlman, M., Winter, B., Massaro, D. W., & Lupyan, G. (2017). Iconicity in the speech of children and adults. *Developmental Science.*

Pierrehumbert, J. B., Stonedahl, F., & Daland, R. (2014). A model of grassroots changes in linguistic systems. *arXiv preprint* arXiv:1408.1985.

Rabagliati, H., & Robertson, A. (2017). How do children learn to avoid referential ambiguity? Insights from eye-tracking. *Journal of Memory and Language, 94,* 15-27.

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Compositional structure can emerge without generational transmission. *Cognition*, *182*, 151-164.

Raviv, L., & Arnon, I. (2018). Systematicity, but not compositionality: Examining the emergence of linguistic structure in children and adults using iterated learning. *Cognition*, *181*, 160-173.

Roberts, G., Lewandowski, J., & Galantucci, B. (2015). How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition, 141,* 52-66.

Robinson, E. J., & Robinson, W. P. (1978). The roles of egocentrism and of weakness in comparing in children's explanations of communication failure. *Journal of Experimental Child Psychology, 26(1),* 147-160.

Rohde, D. L., & Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition, 72(1),* 67-109.

Rohde, D. L., & Plaut, D. C. (2003). Less is less in language acquisition. In Quinlin, P. (Ed.) *Connectionist modelling of cognitive development.* Hove, UK: Psychology Press, pp. 189-231.

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., ... & Cohen, D. (2013). Motherese in interaction: at the cross-road of emotion and cognition? (A systematic review). *PloS one, 8(10),* e78103.

Samuelson, L. K., & Smith, L. B. (2005). They call it like they see it: Spontaneous naming and attention to shape. *Developmental Science, 8(2),* 182-198.

Schieffelin, B., & Ochs, E. (1983). A cultural perspective on the transition from prelinguistic to linguistic communication. In R. M. Golinkoff (Ed.), *The transition from prelinguistic to linguistic communication* (pp. 115–132). Hillsdale, NJ: Erlbaum.

Selten, R., & Warglien, M. (2007). The emergence of simple languages in an experimental coordination game. *Proceedings of the National Academy of Sciences, 104(18),* 7361-7366.

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.

Silvey, C., Kirby, S., & Smith, K. (2015). Word meanings evolve to selectively preserve distinctions on salient dimensions. *Cognitive Science, 39(1),* 212-226.

Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial Life, 9*(4), 371-386.

Snow, C. E., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development, 49(4),* 1114-1128.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review, 27(4),* 501-532.

Sonnenschein, S., & Whitehurst, G. J. (1984). Developing referential communication: A hierarchy of skills. *Child Development, 55,* 1936–1945.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics, 73(4),* 971-995.

Tagliamonte, S. A., & D'Arcy, A. (2009). Peaks beyond phonology: Adolescence, incrementation, and language change. *Language, 85(1)*, 58-108.

Tamariz, M. (2017). Experimental Studies on the Cultural Evolution of Language. *Annual Review of Linguistics*, (0).

Tamariz, M., & Kirby, S. (2015). Culture: copying, compression, and conventionality. *Cognitive Science, 39(1)*, 171-183.

Tamariz, M., & Kirby, S. (2016). The cultural evolution of language. *Current Opinion in Psychology, 8*, 37-43.

Tamariz, M., Roberts, S. G., Martínez, J. I., & Santiago, J. (2017). The interactive origin of iconicity. *Cognitive Science*.

Thompson-Schill, S. L., Ramscar, M., & Chrysikou, E. G. (2009). Cognition without control when a little frontal lobe goes a long way. *Current Directions in Psychological Science, 18(5)*, 259-263.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development.* Cambridge University Press.

Verhoef, T. (2012). The origins of duality of patterning in artificial whistled languages. *Language and Cognition, 4(4),* 357–380.

Verhoef, T., Kirby, S., & Boer, B. (2016). Iconicity and the emergence of combinatorial structure in language. *Cognitive Science, 40(8),* 1969-1994.

Verhoef, T., Kirby, S., & de Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics, 43*, 57-68.

Verhoef, T., Kirby, S., & Padden, C. (2011). Cultural emergence of combinatorial structure in an artificial whistled language. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 483-488).

Verhoef, T., Roberts, S. G., & Dingemanse, M. (2015). Emergence of systematic iconicity: transmission, interaction and analogy. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 2481-2486). Cognitive Science Society.

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.

Winters, J., Kirby, S., & Smith, K. (2018). Contextual predictability shapes signal autonomy. *Cognition*, *176*, 15-30.

Xu, J., & Griffiths, T. L. (2010). A rational analysis of the effects of memory biases on serial reproduction. *Cognitive Psychology, 60(2)*, 107–126.

Zenil, H., Soler-Toscano, F., Delahaye, J.-P., & Gauvrit, N. (2015). Two-dimensional Kolmogorov complexity and an empirical validation of the coding theorem method by compressibility. *PeerJ Computer Science, 1:e23*, https://dx.doi.org/10.7717/peerj-cs.23.

Zipf G. (1949) *Human behavior and the principle of least effort.* New York: Addison-Wesley.

Zuidema, W., & De Boer, B. (2009). The evolution of combinatorial phonology. *Journal of Phonetics, 37(2)*, 125-144.

Appendix 1: Analysis of category-specific measures of structure based on the individual dimensions of shape, size and brightness.

DV		Intercept	Age Group	Generation	Age Group x Generation
model: lmer (DV ~ AgeGroup * Generation + (1 + Generation | Chain))					
Compositional structure:					
shape					
	β	0.01	0.05	-0.05	-0.05
	t	0.11	0.63	-0.45	-0.39
size					
	β	0.21	0.11	0.01	0.04
	t	2.30	1.27	0.07	0.50
brightness					
	β	0.08	-0.04	-0.02	-0.07
	t	1.14	0.52	0.24	-1.00

Table 4: Coefficients for intercepts and fixed effects in the linear mixed-effect models for within-cateogory compositional structure in Experiment 1. [***] $p < .001$, [**] $p < .01$, [*] $p < .05$.
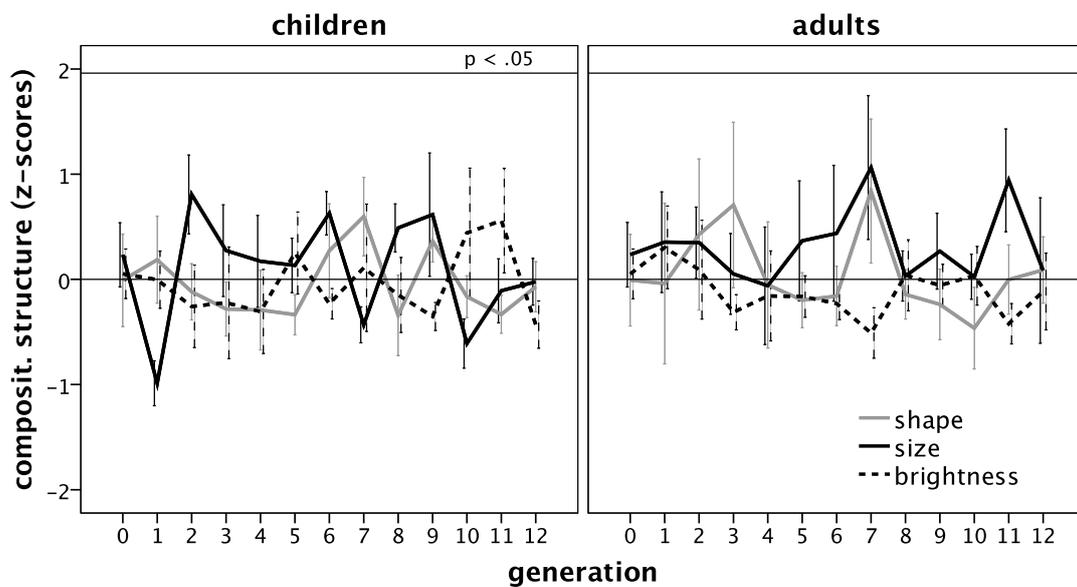
Figure 5: Category-specific compositional structure for shape, size and brightness over 12 generations of transmission in Experiment 1. Error bars represent ± 1 S.E.M aggregated by participants.

DV		Inter-cept	CardPos	AgeGroup	Round	CardPos x AgGroup	CardPos x Round	AgeGroup x Round	3-way
model: DV ~ CardPosition * AgeGroup * Round + (1+Round|Dyad/Participant))									
Compositional structure:									
shape									
	β	0.13	-0.14	-0.39	0.02	0.08	-0.06	0.12	0.02
	t	1.11	-1.15	-3.29**	0.17	0.69	-0.60	1.34	0.16
size									
	β	0.26	-0.05	0.19	0.21	-0.01	-0.12	0.02	-0.04
	t	2.63*	0.52	-1.95	2.59*	-0.10	-1.41	0.21	-0.47
brightness									
	β	-0.01	-0.08	0.03	0.03	0.09	-0.08	0.03	0.14
	t	-0.08	-0.72	0.29	0.32	0.80	-0.81	0.28	1.40

Table 5: Coefficients for intercepts and fixed effects in the linear mixed-effect models for within-cateogory compositional structure in Experiment 2. *** $p < .001$, ** $p < .01$, * $p < .05$.
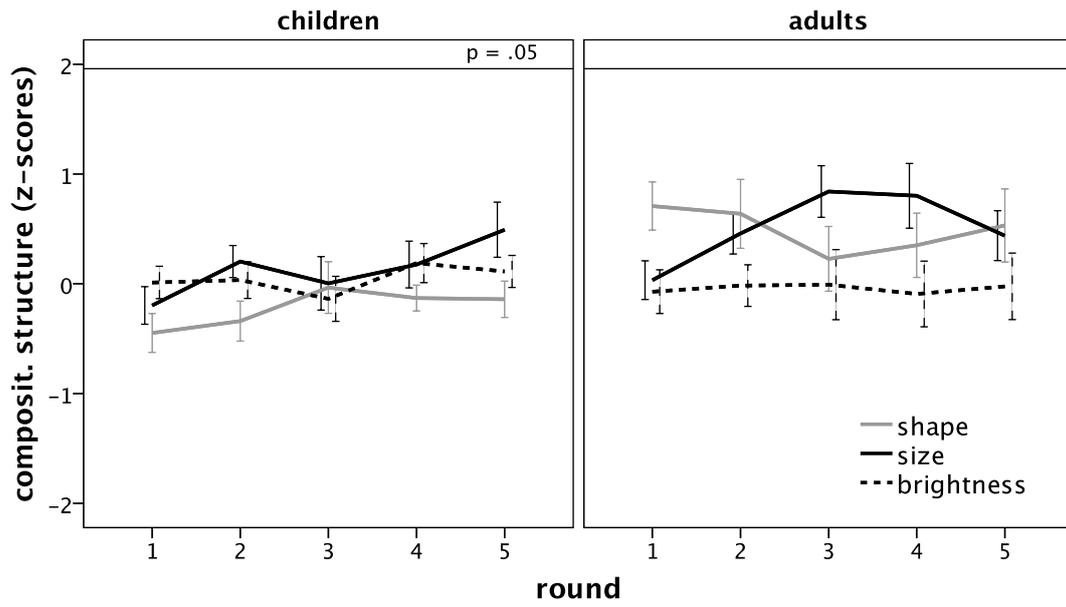
Figure 6: Category-specific compositional structure for shape, size and brightness over 5 rounds of interaction in Experiment 2. Error bars represent ± 1 S.E.M aggregated by participants.

Appendix 2: Final generation patterns in Experiment 1 for chains of adults and children with Primary School grade level in parentheses. 0 – high tone, 1 – low tone.

			adult chains					
brightness	**size**	**shape**	**1**	**2**	**3**	**4**	**5**	**6**
dark	big	fluffy	0001010	010101010011	1011101	101101	001010	1100110011000
light	big	fluffy	0010110	10101010101100	100010	00111010	01001	1100110011000
dark	small	fluffy	0011010	01010101	100010	0011100011010	01001	001100111
light	small	fluffy	0011010	010101010011	10110	11011001	0011010	0011000
dark	big	spiky	0011010	101010101100	101101	00111000110	011010	001100110011000
light	big	spiky	001010	101010100	1011101	000110101	0011010	1100110011000
dark	small	spiky	0010110	10101010	101001	00111000110	001011	0011001100111
light	small	spiky	0010101	010101010011	101101	00110101	011010	0011000
			child chains					
			1 (P1)	**2 (P1)**	**3 (P2)**	**4 (P2)**	**5 (P3)**	**6 (P3)**
dark	big	fluffy	001100	1010	1111000010101	0001111	1100101	0101001
light	big	fluffy	101100	1010	01010101	00011111	0010101	10010
dark	small	fluffy	0011000	0101	0101100	00001111	10101	01001
light	small	fluffy	0010001	0101	110010	000011111	10010	10101
dark	big	spiky	01001	0101	1010011	000011111	101101	100101

light	big	spiky	001100	0101	0101000111010111	000111	10010110	10101
dark	small	spiky	1010100	1010	010110011001010011	00001111	10110	101001
light	small	spiky	001001	0101	0101100	000111	10110	101001