Iterated Teaching Can Optimise Language Functionality

Vera Kempe (v.kempe@abertay.ac.uk), Kamil Cichon (1301137@abertay.ac.uk)
University of Abertay, Dundee, Scotland

Nicolas Gauvrit (ngauvrit@me.com)
Université Paris 8 & École Pratique des Hautes Études, Paris, France

Monica Tamariz (m.tamariz@hw.ac.uk)
Heriot-Watt University, Edinburgh, Scotland

Abstract

Experimental studies of the cultural evolution of language have focused on how constraints on learning and communication drive emergence of linguistic structure. Yet language is typically transmitted by experts who adjust the input in ways that facilitates learning by novices, e.g. through child-directed speech. Using iterated language learning of binary auditory sequences, we explored how language change is affected by experts' intention to teach the language to novices. Comparison between teaching chains and simple transmission chains revealed that teaching was associated with a greater rate of innovation which led to emergence of more expressive languages consisting of shorter signals. This is the first study to show that during cultural transmission, teaching can modify, and potentially optimise, functional characteristics of language.

Keywords: Teaching; iterated language learning; cultural transmission; algorithmic complexity; compositional structure; combinatorial structure

Introduction

Cultural transmission of knowledge proceeds via the social learning mechanisms of imitation, emulation and teaching Boyd & Richerson, 1985; Richerson & Boyd, 2005). Of these, the adaptive value of teaching has recently received increased attention (Cavalli-Sforza & Feldman, 1981; Kline et al. 2013; Kline, 2015; Csibra & Gergely, 2009), highlighting in particular the ostensive use of language in the transmission of technological knowledge required for production of tools and other cultural artifacts (Caldwell & Millen, 2009; Morgan, Uomini, Rendell, Chouinard-Thuly et al., 2015). In its most general sense, teaching can be defined as any kind of behaviour, intentional or not, that promotes learning by narrowing the range of inferences or behavioural options that another individual can pursue (Kline, 2015). Teaching is especially important for transmission of cognitively opaque cultural traits and traditions, i.e. those whose function is not immediately obvious, thereby contributing to cumulative culture (Mesoudi 2011). While transmission of simpler cultural traits may not benefit from additional teaching (Caldwell & Millen, 2009), once culture becomes more complex, teaching delivers additional benefits for the transmission process (Morgan et al., 2015).

In contrast, studies of the cultural transmission of language have mainly researched how constraints that operate on observational learning drive the emergence of structural features like learnability, expressivity (i.e. lack of ambiguity), combinatorial and compositional structure, that support effective communication (Kirby, Cornish and Smith, 2008; Kirby, Tamariz, Cornish & Smith; 2015; Verhoef, Kirby & de Boer, 2014). These studies use the iterated language learning method whereby the result of learning in one generation of learners serves as input for the next generation. The picture that emerges from these studies is that cognitive capacity constraints in human learners promote compressibility of individual signals and entire languages, and that the requirements of efficient communication drive languages to be expressive. In tandem, these constraints – the need for transmission efficiency and for referential efficiency – lead to emergence of combinatorial linguistic structure, i.e. systematic associations of components of the signal with dimension of meaning (Kirby et al., 2015).

However, research on language development in children has presented substantial evidence that child language learners receive input that is specifically tailored to support learning, in the form of child-directed speech (e.g. Burnham, Kitamura, & Vollmer-Conna, 2002; Kempe & Brooks, 2005; Soderstrom, 2007). While there is considerable debate about whether child-directed speech is universal (e.g. Broesch & Bryant, 2013; Falk, 2004; Schieffelin, 1985), whether it constitutes intentional teaching or whether it predominantly supports affective bonding and emotion regulation (e.g. Singh, Morgan & Best, 2002; Uther, Knoll & Burnham, 2007), functionally it qualifies as a behaviour that not only provides local enhancement by directing the learner's attention to relevant information (Kline, 2015) but also pre-samples the input in a way that can support correct learner inferences about language (Eaves, Feldman, Griffiths & Shafto, 2016). For learning to occur, no ostensive cues or direct feedback are required as long as the statistical properties of the modified input ensure improved learning. In this study, we explore how tailoring the input in such a way for the learner shapes language structure over the course of language transmission.

If teaching leads to modification of the input that promote correct inferences about language then we can make predictions about the directions in which iterated teaching will change the structure of the emerging system, compared to simple transmission that is only constrained by cognitive limitations of the learner. Previous iterated language learning studies have highlighted a number of structural features that emerge because languages need to be both learnable and communicatively efficient. We predict that emergence of these features should be facilitated under conditions of teaching: First, if teaching accommodates learnability constraints one would expect transmission fidelity to improve even faster in chains where teaching takes place, compared to simple transmission. Second, as languages evolve to be efficient signalling systems individual signals acquire combinatorial structure by which smaller meaningless subcomponents are recombined to improve signal discriminability, a feature that improves transmission across noisy channels (Verhoef, 2012; Verhoef et al., 2014; Roberts, Lewandowski & Galantucci, 2015) and should be enhanced through teaching. Third, in order to be communicatively efficient, languages also need to maintain expressivity by avoiding under-specification and ambiguity, another feature we expect to emerge faster under conditions of teaching. Fourth, languages become more systematic and self-similar, a property that, akin to phonotactic rules, supports learnability by reducing combinatorial freedom. This feature should also be enhanced by teaching. Finally, emerging compositional structure serves to systematically link meaningless components of signals to underlying dimensions of their meanings. Although compositional structure requires communicative pressure in addition to language transmission (Kirby et al., 2015; Nowak & Baggio, 2016), to isolate the effect of teaching, and in the interest of feasibility, we decided to start our exploration with a simple transmission study that did not impose communicative pressure on learners. Despite the lack of communicative pressure, it is still conceivable that teachers modify the input so as to enhance compositional structure in order to highlight this functional aspect of language.

Method

The present experiment compares transmission of an 'alien' language along chains of learners where each learner's output generated during testing is faithfully represented as input to the learner in the next generation (Simple Transmission condition) with transmission of the language through chains of learners who, after training, are asked to teach the language to the learner in the next generation (Teaching condition). Teaching in this set-up constitutes demonstration of the language to the next learner without any verbal explanation or instruction. The crucial question is whether teachers modify the input when presenting the language to the next learner in a way that goes beyond those modifications that are due to constraints on learning and reproduction.

Participants: Sixty undergraduate students were recruited at the University library for participation in a transmission chain study. Participants were assigned numbers corresponding to their slot in one of six chains, and were called into the test area when the previous participant had finished the training phase. Performance of these participants was compared to that of sixty other participants who had been tested in a different simple transmission study conducted earlier (Kempe, Gauvrit, Gibson & Jamieson, submitted). The 'alien' language used in this experiment consisted of high and low tones assembled into 6 or 8-tone sequences. This signalling system was developed to eliminate any familiarity with signals that could bias learners towards preferences for specific aspects of combinatorial structure.

Materials: Two 500 ms sine-wave tones (high: 440 Hz = musical note *a*; low: 293.7 Hz = musical note *d*) were synthesised and recorded onto differently coloured answer buzzer of 9 cm diameter each. The fixed tone duration made it impossible to modify the length of the tones if pressing the buzzers for longer periods of time thus eliminating duration as a property of the signals. Seed languages consisted of random sequences of high vs. low tones, which were either six or eight tones long. These binary sequences instantiating the 'words' in the 'alien language' were paired with eight coloured objects differing in shape (spiky 'kiki'-type vs. fluffy 'bouba'-type), size (2 x 2 cm vs. 4 x 4 cm) and brightness (25% vs 75% saturation) (see Figure 1), which were printed on laminated cards sized 5 x 8 cm. All objects also had unique properties due to differences in specific shapes and hues.

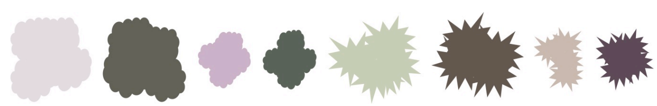

Figure 1: Meanings associated with the signals (binary sequences) in the 'alien buzzer language'.

Procedure: After signing a consent form, participants in both conditions were told they would learn an 'alien' language used by a species of aliens that had no mouth and therefore used buzzers for communication. Participants were then shown six of the eight cards one at a time to familiarise them with the 'alien' objects. Training proceeded in an incremental fashion: In the Simple Transmission condition, participants were given a demonstration of the binary buzzer sequence for each card, and were asked to repeat it. Demonstration and practice were then immediately repeated resulting in incremental training consisting of two consecutive trials per card, before proceeding to the next card. Order of cards was randomised for each participant by shuffling the cards. After training, participants were shown the cards one at a time, and asked to produce the 'alien buzzer words' to the best of their abilities. Their responses were videotaped, coded and then presented unaltered to the next participant. To prevent the languages from degenerating, a 'homonymy filter' (Kirby et al., 2008) was applied by which up to two identical sequences (i.e. homonyms) were withheld and only six items were presented during training. In case of no homonyms, two cards were withheld at random; in case of just one pair of homonyms, the card corresponding to one of the homonyms was withheld at random along with one other randomly chosen card. This manipulation was used to encourage productivity and to prevent languages from degenerating into ambiguous systems.

In the Teaching condition, chains were also seeded with six out of eight cards, to maintain compatibility with the Simple Transmission condition. In Generation 1, training proceeded in exactly the same way as in the simple transmission condition. However, from Generation 1 onwards, after training, participants were asked to 'teach' the language to the next participant in the chain who was called into the testing area at that time. When teaching, participants were instructed not to provide any verbal comments or instructions but to simply demonstrate the buzzer sequences twice to the next participant, allowing them to repeat the sequence after each demonstration.

Results

Participant buzzer responses were videotaped and coded for further analyses. Inter-coder reliability, determined for 17% of trials, was 94%. All dependent variables were analyzed using Growth Curve Analyses (GCA). To see whether trends were linear or tended to level off we included a quadratic term of Generation following Beckner, Pierrehumbert & Hay (2017). Thus, our model contained fixed effects of Condition (Simple Transmission vs. Teaching) and linear and quadratic effects of Generation, and random intercepts of Chains as well as random slopes of Generation (Winter & Wieling, 2016), resulting in a model of the structure *Condition* + *Generation* + *Generation*^{2} + *Condition*Generation* + *Condition*Generation*^{2} + *(1|Chain)* + *(0+Generation|Chain)* + *(0+Generation*^{2}*|Chain)*. In all cases, the quadratic model provided a better or the same fit to the data compared to the linear model as determined by likelihood-ratio tests.

Expressivity: Languages in the teaching condition contained fewer homonyms than languages in the simple transmission condition. The outputs in the transmission condition were more prone to degenerate into underspecified, more ambiguous languages, as indicated by an interaction of Condition with the linear, β = -0.44, t = 2.50, $p < .05$, and the quadratic effect of Generation, β = 0.05, t = 2.78, $p < .01$. These findings are depicted in Figure 2. Note that in this and all subsequent figures error bars correspond to one S.E.M.

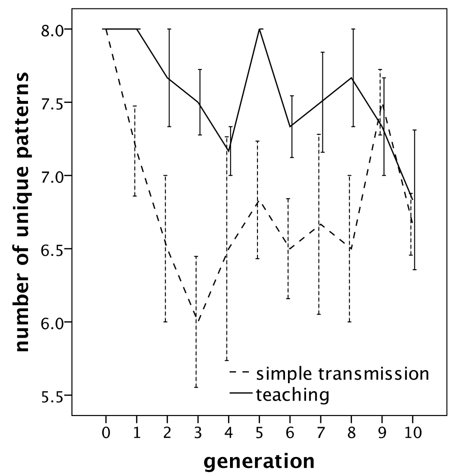

Figure 2: Number of unique sequences (out of 8) in Simple Transmission and Teaching chains.

Transmission Accuracy: We used length-normalised Levenshtein edit distance (LED) as an inverse measure of transmission accuracy. LED decreased faster in the Simple Transmission condition (Figure 3), as evidenced by a significant interaction between Condition and the linear effect of Generation, β = -0.03, t = -2.13, $p < .05$. In other words, when participants were asked to teach they introduced more innovations than when they simply tried to reproduce the binary sequences.

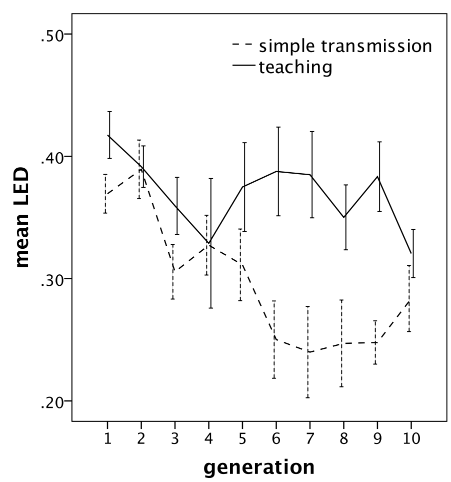

Figure 3: Mean length-normalised Levenshtein edit distance in Simple Transmission and Teaching chains.

Self-similarity: Average pairwise length-normalised LED between all pairs of sequences in a language served as an inverse measure of within-language similarity. This self-similarity increased (i.e. LED decreased) overall as indicated by a main effect of Generation, β = -0.06, t = -3.80, $p < .001$. The significant quadratic term suggests that increase of self-similarity was mainly due to the drop from the seed language and levelled off in subsequent generations (Figure 4).

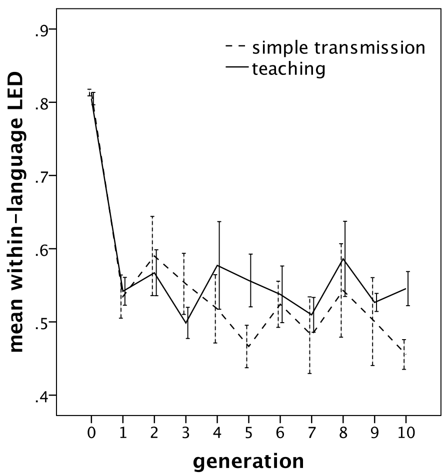

Figure 4: Mean inverse self-similarity (within-language LED) in Simple Transmission and Teaching chains.

Length: Sequences had started out with an average length of 7 tones in the seed language at Generation 0. In the Teaching condition, sequences remained of roughly the same length, which was significantly shorter than in the Simple Transmission condition, $\beta = -0.81$, $t = -2.60$, $p < .05$. The interaction between Condition and the linear effect of Generation, $\beta = 0.66$, $t = 2.46$, $p < .05$, confirmed that sequence length increased only during simple transmission (Figure 5).

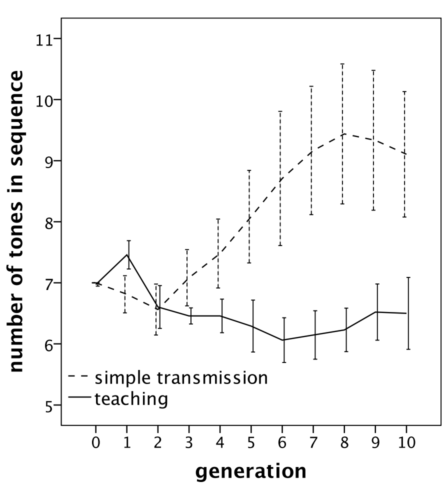

Figure 5: Mean sequence length in Simple Transmission and Teaching chains.

Combinatorial structure: Structure of individual signals was operationalised as algorithmic complexity, using an estimate developed for short binary strings based on the coding theorem method (Gauvrit, Soler-Toscano, Zenil & Delahaye, 2014; Zenil, Soler-Toscano, Delahaye & Gauvrit, 2015). This measure provides an inverse estimate for the amount of structure of a given sequence relative to the variation in structure possible for all sequences of the same length (Figure 6). It captures the intuition that sequences like *adadadad* or *aaaadddd*, where *a* represents the high and *d* represents the low note, are more structured than sequences like *aadadddda*. GCA did not yield any significant effects although the interaction between Condition and the linear effect of Generation, $\beta = -0.13$, $t = -1.82$, $p = .08$, fell short of significance, suggesting that there may have been a trend for algorithmic complexity to decrease somewhat during simple transmission.

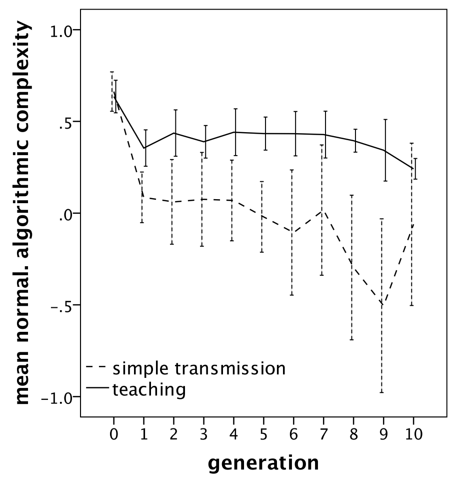

Figure 6: Mean length-normalised algorithmic complexity in Simple Transmission and Teaching chains.

Compositional Structure: To determine compositional structure for each language, we calculated the Pearson product-moment correlations between differences in the three meaning dimensions of all meaning pairs and differences between associated signals pairs within each language, using 10,000 iterations of a Monte Carlo process to obtain a standardised score. This measure remained below the value associated with $p = .05$, and did not differ between conditions and generations indicating that no compositional structure had emerged (Figure 7).

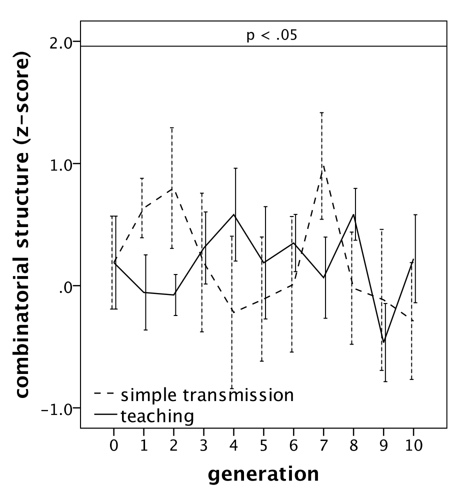

Figure 7: Mean compositional structure in Simple Transmission and Teaching chains. The horizontal line indicates $z = 1.96$, $p = .05$.

Discussion

We compared six teaching chains with six simple transmission chains to explore the effect of teaching during cultural transmission of language. As this was an exploratory study, we did not include a requirement to engage in referential communication. Thus, it was not unexpected that compositional structure did not emerge (Kirby et al., 2015), and we found no evidence that teachers would introduce it spontaneously.

What we found was that although transmission accuracy increased overall, it was significantly lower in the Teaching condition, counter to our expectations. Thus, considerably more innovations were introduced into the signals when participants were asked to teach rather than just to reproduce what they had learned. We suggest that these innovations served to stabilise certain features of the languages that the teachers considered crucial. The most notable change from Simple Transmission performance achieved through innovation in the Teaching condition was to maintain expressivity of the language: The number of different sequences within the taught languages remained high thus preventing these languages from degenerating by accumulating homonyms. This is an interesting result because previous research had demonstrated that without a strong incentive to communicate, capacity constraints of the learners drive languages towards under-specification and ambiguity (Kirby et al., 2015). What our findings suggest is that teaching can override this tendency, presumably due to strong biases about the functional destination of language, which is to be expressive, i.e. referentially efficient. It is noteworthy that the expressivity advantage in the Teaching condition arose even though we applied a homonymy-filter in the Simple Transmission condition to prevent the languages from degenerating. Without this filter, languages in the Simple Transmission condition would have accumulated even more homonyms (Kirby et al., 2008), presumably further deviating from the Teaching condition.

Another feature that remained stable in the Teaching condition was sequence length: Teachers managed to maintain sequence length at around the original 7 tones while in the Simple Transmission condition, sequence length increased dramatically. Stabilising or even reducing length is a strategy that can ensure learnability and transmission accuracy by keeping the form of signals within the limits imposed by working memory constraints. As this brevity constraint operated only in the Teaching condition it may reflect a cooperative adjustment on the part of the teacher designed to aid the learner.

We observed little further increase of self-similarity of languages beyond an initial gain following exposure to the initial random binary sequences. Self-similarity can be thought of as a measure of systematicity that is somewhat akin to phonotactic rules. If teachers attempted to resolve the trade-off between expressivity of the language and brevity of the signals, they would be more likely to use the full space of distinct binary sequences of shorter lengths, which restricts opportunity to achieve self-similarity. In line with this conjecture, the trend towards self-similarity was less pronounced in the Teaching condition, although the difference between conditions did not reach statistical significance.

We also had hypothesised that teaching would lead to a faster increase in combinatorial structure to improve transmission efficiency. For the binary sequences used as signals in this study, introducing combinatorial structure would entail establishing subcomponents (e.g. *ad* or *aad*) that can be recombined using operations like repetition or mirroring, as in strings like *adadad* (-0.75 [numbers in parentheses are the associated values of length-normalised algorithmic complexity) or *aadaadaad* (-1.43). In contrast, complex strings like *ddaaad* (1.43) or *ddaaaadda* (2.02) do not contain combinations of discernible subcomponents. According to our hypothesis, the Teaching condition should have given rise to more sequences of the former than the latter type. However, our data showed exactly the opposite trend: Although not significantly so, length-normalised algorithmic complexity tended to be higher in the Teaching condition, indicating less combinatorial structure than in the Simple Transmission condition. One possible explanation for this finding is that when trying to produce as many different sequences as possible while maintaining brevity of the signals, teachers sample more densely from the distribution of shorter sequences thereby inevitably utilising more sequences of higher complexity. On the other hand, when the brevity constraint is relaxed, learners in the Simple Transmission condition may produce longer sequences yet processing capacity limitations will force them to settle for more structured ones, which are made up of a limited repertoire of subcomponents.

This pattern of results shows that when teaching, which in this study entailed knowingly serving as input-generating models, participants changed their behaviour to adjust the input so as to constrain learner hypotheses in accordance with their own tacit knowledge about how languages function. Specifically, they were negotiating a trade-off between referential efficiency and transmission efficiency by introducing innovations that allowed them to generate unique sequences for each meaning, to prevent languages from degenerating into under-specified systems, while at the same time facilitating transmission fidelity by stabilizing the length of these sequences. It can be argued that the biases that shaped this teaching behaviour reflect participants' knowledge about the functionality of language as it was acquired through their native language use and, thus, these biases may not be informative about of the role teaching may have played in language evolution. However, learners in the simple transmission condition had access to exactly the same knowledge yet without the motivation to teach those biases were overridden by the drive towards compressibility. Thus, whatever the origins of the knowledge about the functionality of language are, our findings suggest that this knowledge affects teaching.

To summarise, the results of this study support the idea that teachers modify the input to learners in ways that reflect their biases about the functional utility of a cultural trait. Applying this idea to the study of the cultural evolution of language means that theories of language transmission need to include teaching into the suite of transmission mechanisms under consideration. We hope that our findings will inspire more detailed explorations of the role of teaching in the cultural evolution of language in the future.

References

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

Boyd, R. & Richerson, P. J. (1985). *Culture and the evolutionary process.* Chicago: University of Chicago Press.

Broesch, T. L., & Bryant, G. A. (2015). Prosody in infant-directed speech is similar across western and traditional cultures. *Journal of Cognition and Development, 16(1),* 31-43.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science, 296(5572),* 1435-1435.

Caldwell, C. & Millen, A.E. (2009) Social learning mechanisms and cumulative cultural evolution: Is imitation necessary? *Psychological Science, 20,* 1478-83.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution: A quantitative approach.* Princeton: Princeton University Press.

Csibra, G. & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Science, 13,* 148-153.

Eaves Jr, B. S., Feldman, N. H., Griffiths, T. L., & Shafto, P. (2016). Infant-directed speech is consistent with teaching. *Psychological Review, 123(6),* 758.

Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioural and Brain Sciences, 27*, 491–541.

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

Kempe, V. & Brooks, P. J. (2005). The role of diminutives in the acquisition of Russian gender: Can elements of child-directed speech aid in learning morphology? *Language Learning, 55, Supplement: The Best of Language Learning,* 139-176.

Kempe, V., Gauvrit, N., Gibson, A. & Jamieson, M. (submitted) Adults are more efficient in creating and transmitting novel signalling systems than children.

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., 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 behaviour in humans and other animals. *Behavioural and Brain Sciences, 38,* E31.

Kline, M. A., Boyd, P. & Henrich, J. (2013). Teaching and the life history of cultural transmission in Fijian villages. *Human Nature, 24,* 351-374.

Mesoudi, A. (2011). *Cultural evolution: How Darwinian theory can explain human culture and synthesize the social sciences.* Chicago and London: University of Chicago Press.

Morgan, T. J. H., Uomini, N. T., Rendell, L. E., Chouinard-Thuly, L., Street, S. E., Lewis, H. M., ... & Whiten, A. (2015). Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nature Communications, 6*.

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

Richerson, P. J., & Boyd, R. (2005). *Not by genes alone.* Chicago: University of Chicago Press.

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.

Schieffelin, B. B. (1985). The acquisition of Kaluli. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition: Vol. 1: The data* (pp. 525–594). Hillsdale, NJ: Lawrence Erlbaum Associates.

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' listening preferences: Baby talk or happy talk? *Infancy, 3(3),* 365-394.

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

Uther, M., Knoll, M., & Burnham, D. (2007, January). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication, 49(1)*, 2–7.

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

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.

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.

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*.