Systems from Sequences: an Iterated Learning Account of the Emergence of Systematic Structure in a Non-Linguistic Task

Hannah Cornish (hannah@ling.ed.ac.uk)

Kenny Smith (kenny@ling.ed.ac.uk)

Simon Kirby (simon@ling.ed.ac.uk
University of Edinburgh, School of Philosophy, Psychology and Language Sciences,
Language Evolution and Computation Research Unit, 3 Charles Street
Edinburgh, EH8 9AD, UK

Abstract

Systematicity is a basic property of language and other culturally transmitted behaviours. Utilising a novel experimental task consisting of initially independent sequence learning trials, we demonstrate that systematicity can unfold gradually via the process of cultural transmission.

Keywords: Iterated Learning; Cultural Evolution; Sequence Learning

Introduction

Language, like many other culturally-transmitted aspects of human behaviour, works as a system: individual words or phrases do not behave independently of one another but instead form part of a mutually reinforcing system of conventions. This is at the heart of what we mean when we talk about the grammar of language - grammar exists only to the extent that individual utterances are non-independent, and related to one another in systematic ways. This is so obviously true of language, and other cultural systems such as music, that it hardly seems to stand in need of explanation. But where does this basic fundamental property come from? Why are individual utterances not independent? After all, they arguably are for all other systems of communication in nature. Even when we discount the fact that non-human animals can only convey a finite set of meanings, many animal communication signals are gradable and related to one another only as much as they are produced by the same vocal apparatus (Fitch, 2010). This stands in stark contrast to human language, resting as it does on a system of infinitely reusable discrete signals[1].

We propose that the answer lies in the nature of cultural transmission. In systems like language which are transmitted by *iterated learning*[2], the most transmissible behaviours are those that are most learnable (Smith, Kirby, & Brighton, 2003). Cultural evolution tends therefore to maximise learnability (Christiansen & Chater, 2008; Kirby, Cornish, & Smith, 2008). One way to increase learnability of a set of behaviours is for those behaviours to behave systematically, so that learning one will increase the ease with which others will be learned. In other words, we should expect cultural evolution to create systems of dependence between previously independent learned behaviours.

In this paper we present an experimental paradigm in which we can observe the cultural evolution of such systematicity in a task which involves many initially independent learning trials. Our task is purposefully non-linguistic, but designed to have relevant similarities with language. Specifically, it is a simple immediate sequence-recall task based around the *Simon Game*. This was a children's electronic game developed by Milton-Bradely in 1978 with four coloured illuminated buttons arranged on its surface in a circle. These buttons lit up to display a random sequence and the player's goal was to repeat this sequence back immediately. This task has a number of useful properties for our purposes. Firstly, although it is clearly non-linguistic, thereby making it unlikely that participants will bring any language-specific biases to it, the task nevertheless involves sequence learning, which is highly relevant to the linguistic domain (Misyak, Christiansen, & Tomblin, 2010; Christiansen, Conway, & Onnis, 2012). Secondly, the task is overtly one in which each sequence acts as an independent task. The player can be scored on their learning of each sequence immediately after recall. This is in contrast to a typical artificial language learning task (e.g., Gomez & Gerken, 2000) which might involve learning a set of sequences for recall at a later stage.

Our question is the following: given this kind of simple independent sequence recall task, will cultural transmission nevertheless lead to the evolution of systematicity in the set of sequences? In effect, can an implicit system-wide learning effect exert influence on the evolution of the set of sequences? To test this, we create an iterated version of the Simon Game in which the sequences produced by one participant in the task become the sequences that the next participant in the experiment is exposed to. We start with a set of 60 random sequences, and observe whether these sets evolve in such a way to make learning easier, and whether they do so by becoming more systematically structured.

Methods

The experiment utilises a diffusion chain paradigm, a technique used widely amongst researchers investigating cultural

---

[1] An obvious exception to this are species capable of vocal learning - particularly song-birds and cetaceans. We suspect it is no coincidence that these species (a) have combinatorial signalling systems, and (b) make use of cultural transmission. See Feher et al. (2009) for an example of the important role culture has in bird-song.

[2] This refers to a particular type of learning where a behaviour is acquired by observing another who also acquired that behaviour the same way (Kirby & Hurford, 2002).

transmission (e.g., Mesoudi & Whiten, 2008), whereby all learners (apart from those in the initial generation) are trained on the output of previous learners.

Participants

In total, 40 participants (mean age 21y, 11m; females = 25) were recruited from the University of Edinburgh's graduate employment service, to take part in a visual memory experiment involving sequences of flashing coloured lights. Each participant was allocated at random into one of four different chains (A, B, C, D), consisting of ten generations each. All learners received £7 remuneration for taking part.

Procedure

The task itself was simple: participants were shown a light sequence on a touch-screen tablet device, and then asked to immediately reproduce it by tapping the sequence back. The layout of the Simon Game is shown in Figure 1. Once a complete sequence has been input (Figure 1.a), immediate accuracy feedback is given (Figure 1.b). Participants could then request another sequence.

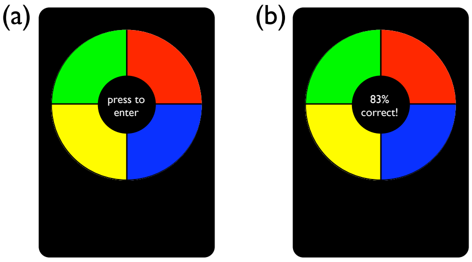

Figure 1: A diagram showing the layout of the Simon Game: (a) participants see a sequence on the screen, and are asked to immediately reproduce it; (b) feedback is then given on the task.

In all, participants were trained and tested on 60 different sequences, seeing each sequence once in each of two rounds in random order, making 120 exposures in total. In order to catch obvious mistakes in sequence entry, if any participant submitted a sequence of length 6 or shorter, this was rejected by our software, and the target sequence would reappear at a random point later in the player's round for them to reattempt. The 60 sequences produced in the second round were collected to be used as training stimuli for the next learner in the chain.

Initial Sequences

Although subsequent learners were trained on the output of the previous learner, the four initial participants were trained on a set of sequences that adhered to the following properties: (i) the length of each sequence was 12; (ii) each sequence consisted of 3 flashes of each colour (red, blue, green, yellow); (iii) these colours appeared in random order. This resulted in a set of 60 sequences which had no structure.

Results

The sequences were analysed in order to determine (1) whether the individual sequences would adapt to become easier to learn over time, and (2) whether individual sequences would co-evolve together to form a collective system. In order to assess these effects, we look at quantitative measures of learnability and structure, along with an additional measure examining the degree of divergence between the four chains into specific lineages. We also qualitiatively examine some of the evolved sequences at the ends of the chains, and note some striking structural regularities.

Learnability

To determine the learnability of a sequence set at a given moment in time, we first need a measure of how accurately each sequence is reproduced. For this we calculated the inter-generational error using the Levenshtein (1966) edit-distance between each target sequence and response from the participant, normalised for length of sequence (Kirby et al., 2008)[3]: we count the minimum number of insertions, deletions and substitutions required to turn one sequence (input) into another (output), dividing this by the length of the longer sequence. From the normalised edit distance of each individual sequence, we then calculate the average error of the sequence set. For consistency with our later analysis we converted this into mean similarity (1-error), shown in Figure 2 below.

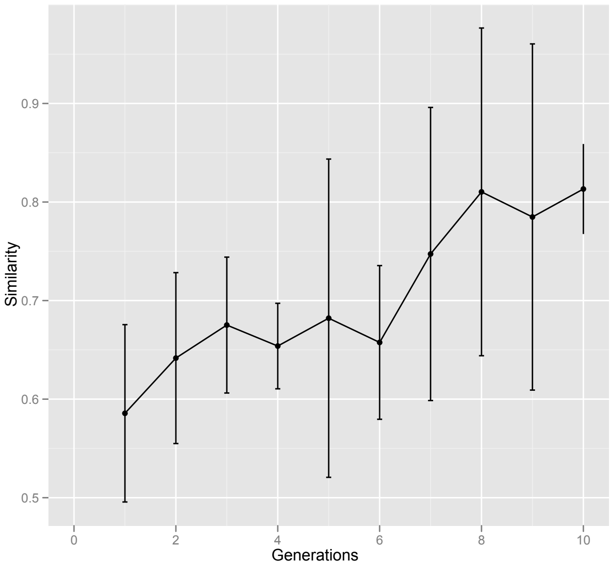

Figure 2: Graph showing the average mean similarity score of sequences in each set over generations. Error-bars represent the 95% confidence intervals across the four chains, here and throughout.

---

[3] The same metric was used to give feedback to participants, but was first recast into a similarity score by computing 1-error, and then given as accuracy by percentage. An error score of 0.17 translates to a similarity score of 0.83, and appeared to participants as 83% correct.

As Figure 2 indicates, the sequence sets become easier to learn over time: reproduced sequences become more similar to their targets. In order to determine whether this cumulative increase in similarity was significant, we ran Page's (1963) *L* trend test. This reveals a significant increase in similarity over generations, both when including (L=1469, m=4, n=10, $p<.0001$) and excluding (L=1074, m=4, n=9, $p<.0001$) the initial set of sequences, which had not been produced by participants.

Structure

One possible explanation for the increase we see in learnability could simply be that early participants are forgetting parts of each sequence, leading to the sequences eventually becoming short enough to be more easily reproduced by later learners. In order to assess this claim, we examined the average length of sequences across each chain for any signs of change. Figure 3 confirms that that there was no significant reduction in sequence length over the course of the experiment. Given that length is in fact highly stable across each generation, some other feature of the sequences must be responsible for their increase in learnability.

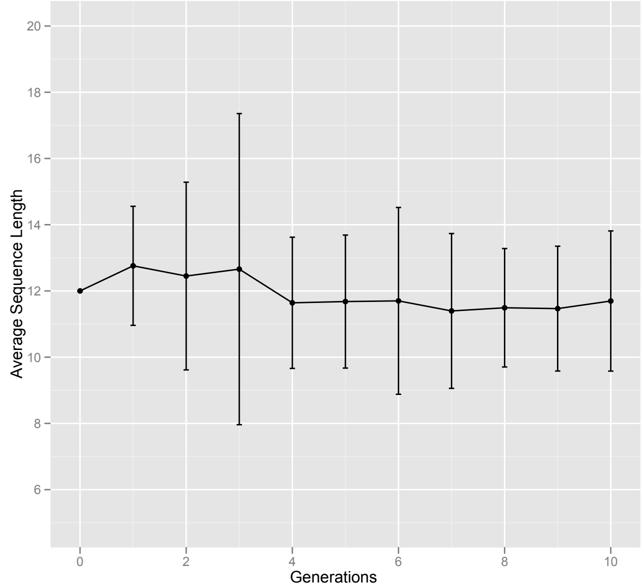

Figure 3: Graph showing the average length of each sequence by generation. Sequence length remains stable throughout the experiment, ruling out a simplistic explanation for the improvements to performance in Fig. 2.

The other possibility is that the sequences have become structured in some way. In order to determine whether this is the case or not, we examined the composition of the sequences in each set, using two different metrics. The first is a measure of *dispersion*, which looks at how similar each sequence is to other sequences within that set. This is calculated using the same distance metric as before, this time comparing the distance of each sequence from all other sequences within that generation, rather than across generations between target and reproduction. This figure, when averaged over all pairs of sequences, returns the amount of dispersion within the set at a given generation. Figure 4 shows that over time, individual sequence sets lose variation as the sequences within them begin to resemble one another (L=1980, m=4, n=11, $p<.0001$). This could happen if, for instance, smaller subsequences come to be shared across whole sequences within the set.

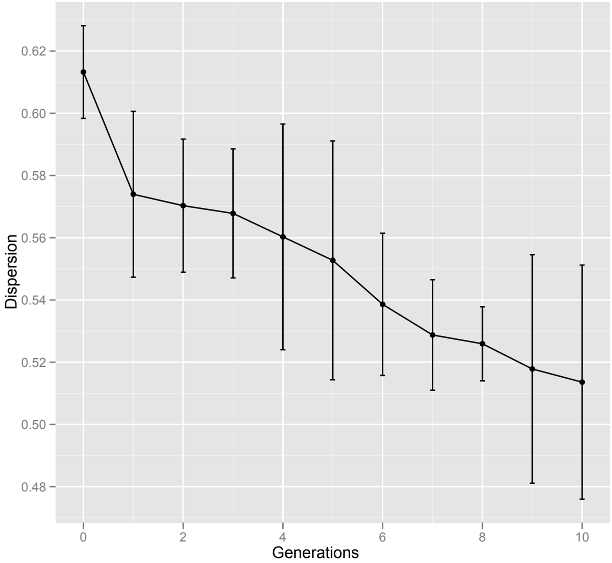

Figure 4: Graph showing the mean of the normalised dispersion score of the four sequence sets over generations. Dispersion decreases cumulatively over generations, indicating that sequence sets are becoming more self-similar over time.

In order to explore this idea further, the second measure looks at *compression*. This is related to the notion of Kolmogorov complexity (Kolmogorov, 1963), and is essentially a measure of how easy it is to compress data into a smaller representation. If a dataset contains repetitions (redundancy), then the algorithm can exploit that by creating a shorter representation to substitute for the larger one, and thus the size of the file can be reduced. We tested this directly by computing the compression ratio (size of the file after compression/size of the file before compression)[4] in order to assess how much structure was present in each sequence set (Fig. 5).

Figure 5 demonstrates that there is a decrease in the compression ratio over time (L=1964.5, m=4, n=11, $p<.0001$). This shows that the sequences are becoming structured, and further supports the idea that those sequences produced later on in the chains have become fractionated into smaller higher-frequency units which repeat within sequence sets.

---

[4] To do this, we used tools from the Zlib library: www.zlib.net

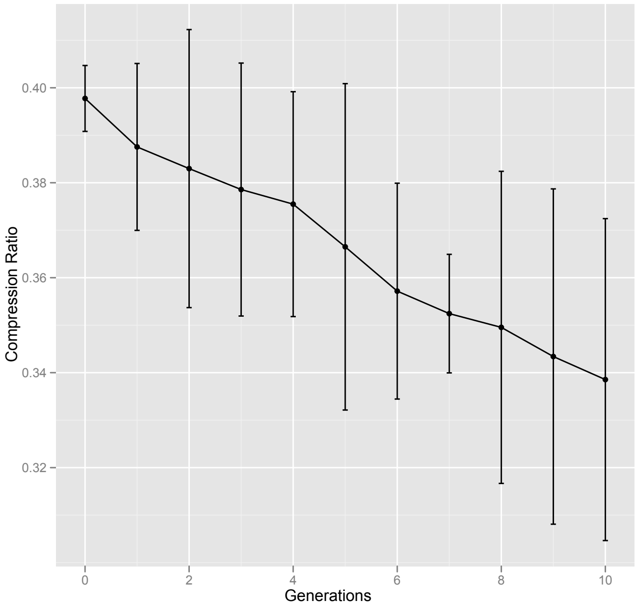

Figure 5: A graph showing the mean compression ratio over generations. Sequence sets become more compressible over time, indicating that there is more structure in later sequences than in earlier sequences.

Identifiability

The decrease in dispersion scores and compression ratios across our four chains could be due to a universal bias pulling all of the sequences towards a similar (structured) attractor. If this were the case, we would expect to find that a given sample of sequences drawn from within a chain, at a certain generation, would look fairly similar to any other sample of sequences drawn from any other chain at that same generation. In order to to determine whether there are in fact different types of structural patternings in our data, organised across the different lineages, or just one kind of structural patterning shared amongst all chains, we used a measure of lineage divergence (referred to as *identifiability*) taken from Matthews, Roberts, and Caldwell (2012). This determines (for each sequence) the within-group similarity and the across-group similarity, and then calculates a proportion: [within-group similarity/(within-group similarity + across-group similarity)]. This returns a value between 0 and 1, where values above 0.5 indicate higher overall within-group similarity, and values below 0.5 indicate higher across-group similarity.

Figure 6 shows, the initial sequences (generation 0) are not identifiable as coming from their particular chain. This is to be expected due to the fact that they were all randomly constructed according to the same procedure. However, sequences do begin to diverge into separate lineages. A one-sample Wilcoxen test confirms that the within-group identifiability of sequences from all chains produced by participants (generations 1-10) were significantly higher than our expected chance level of 0.5 (Z=-5.86, N=40, $p<.0001$). Furthermore, our trend analysis reveals that this effect is cumulatively increasing over time (L=1901, m=4, n=11, $p<.0001$), such that sequences drawn from later generations are significantly more likely to resemble sequences from within their own lineage, than from those of any other lineage.

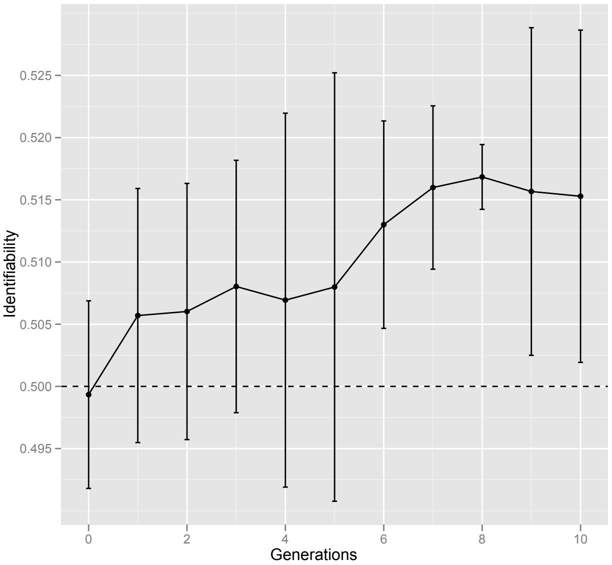

Figure 6: Graph showing how the average identifiability of each sequence increases over time. The dashed line represents chance levels.

Qualitiative analysis

The quantitative results all point towards the fact that our independent sequences are gradually becoming structured as a collective, and that they do so in ways which are specific to different lineages. What then might some examples of these systematic structures look like? Figure 7 shows a sample of sequences that came from chain A.

As Figure 7.a shows, the initial set of sequences contain very little obvious structure. By generation 10 however (Figure 7.b), a common pattern has emerged. In fact, of the 60 sequences in this set, just over half of them begin with an initial alternation pattern of red-yellow-red-yellow, or red-red-yellow-yellow. This is frequently followed by a cyclical pattern - moving around the Simon board in either a clock-wise or anti-clockwise direction from a given starting point (usually red in this case) - which can itself be repeated to extend the sequence. This kind of structure lends itself easily to being analysed into hierarchically arranged sub-parts, containing non-adjacent dependencies. Figure 7.c shows one such possible analysis of this kind[5].

---

[5] This analysis is intended to be illustrative rather than definitive, and is based on the sample of 6 sequences shown here, not the whole set. As we can see in Figure 2 the sequence sets are not fully stable and are still undergoing change, making a more detailed structural analysis difficult.

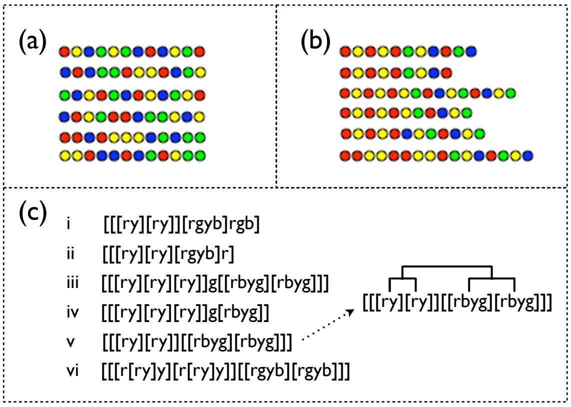

Figure 7: Some examples of sequences from chain A: (a) a sample of six random sequences at generation 0; (b) those same sequences at generation 10; (c) sequences at generation 10 again, bracketed to highlight their nested hierarchical structure. This bracketing can be used to generate tree-structures (as shown in example v) which more clearly demonstrate the nature of the system.

As to be expected from the identifiability results however, the way in which the other chains are organised is noticeably different, both to the statistical measures employed earlier, and to the human eye. These different styles can be easily contrasted visually in Figure 8, again by drawing a sample of six strings from the set to illustrate general structural regularities in the final generations.

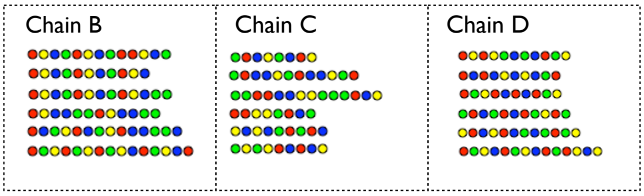

Figure 8: Some examples of sequences from the final generations of chains B, C and D. There are clear qualitative differences between the sequences across these different chains, and from chain A in Figure 7 above.

In contrast to the distinctive alternation found in chain A, chain B favours the more cyclical patterns, and shows a very dominant tendency to begin all sequences with a red (92%). Chain C also places restrictions on the identity of the first colour: green is most common (48%), followed by yellow (30%) and red (22%), but never blue. It also contained the highest proportion of sequences with two or more of the same colour adjacent to one another (58%: as compared to A 28%, B 35% and D 23%). Finally, Chain D seems to prefer triplets and alternations as reusable sub-sequences.

In summary, the qualitative analysis of the chains reveals that sequence-sets can become systematic in multiple ways. Some commonalities do exist across lineages - for instance, a strategy of repeated alternation of two colours was present in all chains. However, (i) the frequency with which a given strategy was employed, (ii) where it was employed (sequence-initially, sequence-medially, or sequence-finally), and (iii) with what particular colour combinations it was employed, all varied, contributing to the development of a unique 'profile' for each lineage.

Discussion

We have presented an experiment in which participants attempt to immediately recall visually presented sequences. The sequences that participants produce become the sequences which subsequent participants try to recall. In this way, we create lineages of sequences in an experimental simulation of cultural evolution. These lineages are potentially independent of each other, since the initial set of sequences are generated at random and participant responses are gathered immediately after each sequence.

The effect of cultural evolution in the experiment is that the sequences become easier to recall correctly. In other words, errors introduced by participants are in the direction of easier sequences. How is this achieved? We see that the set of sequences at each generation becomes self-similar, suggesting that the sequences are not operating independently any more. This conclusion is confirmed if we look across separate chains in the experiment: the sequences are more similar *within* a chain, and less similar *across* chains. Additionally, the set of sequences at each generation in the experiment becomes more compressible, as system-wide structure starts to emerge.

The systematic structure in sequences shows tantalising evidence of hierarchy, although a deeper analysis will have to await further analytic tools being applied. For example, in some chains we see the emergence of pairs of pairs of colours. We also see a pattern in which some sequences are "doubled" versions of others in the set. So, for example, the sequence prefix "rryyrryy" in chain A matches the prefix "ryry". It is tempting to suggest that this provides evidence of a grammar with centre embedding of the form $(A^nB^n)^m$, although such an analysis is premature without further probing of the way in which these sequences are processed. Building on work such as Christiansen and Ellefson (2002), recent attempts have been made to tease apart the different cognitive mechanisms at work when processing non-adjacent dependencies resembling these sequences in language (Vries et al., 2012; Christiansen et al., 2012). This may provide further clues as to why these particular structures emerge in this study. Likewise, studying the process of emergence itself, using iterated sequence learning tasks in the laboratory as we have done, may help us better understand the way these learning and processing biases shape behaviour at the population level.

Conclusion

A hallmark of complex culturally transmitted behaviours in humans, such as language and music, is their systematic structure. Instances of behaviour do not behave independently, but form part of a system of mutually reinforcing conventions. Here we show that such systematic structure can emerge in an experimental task through the process of cultural transmission even when the task is designed to minimise the influence of domain specific biases and with no explicit reward for treating behaviours as co-dependent. We propose that this result is suggestive of a similar process operating in the origins of behaviours like language in humans. Cultural evolution favours transmissible behaviours. A solution to the challenge of becoming more transmissible is for behaviours to form part of a system, thus increasing their learnability. Language, our most systematic suite of behaviours, bears the hallmark of just such a process of cultural optimisation.

Acknowledgments

Hannah Cornish is supported by a British Academy Post-Doctoral Fellowship.

References

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Science*, *31*(5), 489-509.

Christiansen, M. H., Conway, C. M., & Onnis, L. (2012). Similar neural correlates for language and sequential learning: Evidence from event-related brain potentials. *Language and Cognitive Processes*, *27*, 231-256.

Christiansen, M. H., & Ellefson, M. (2002). Linguistic adaptation without linguistic constraints: the role of sequential learning in language evolution. In A. Wray (Ed.), *Transitions to language* (p. 335-358). Oxford: Oxford University Press.

Feher, O., Wang, H., Saar, S., Mitra, P., & Tchernichovski, O. (2009). De novo establishment of wild-type song culture in the zebra finch. *Nature*, *459*, 564-568.

Fitch, W. T. (2010). *The evolution of language*. Cambridge: Cambridge University Press.

Gomez, R. L., & Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Science*, *4*(5), 178-186.

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 of the United States of America*, *105*, 10681-10686.

Kirby, S., & Hurford, J. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (p. 121-148). Cambridge: Cambridge University Press.

Kolmogorov, A. N. (1963). On tables of random numbers. *Sankhyā Series A.*, *25*, 369-375.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics-Doklady*, *10*, 707-710.

Matthews, C., Roberts, G., & Caldwell, C. A. (2012). Opportunity to assimilate and pressure to discriminate can generate cultural divergence in the laboratory. *Evolution and Human Behavior*, *33*, 759-770.

Mesoudi, A., & Whiten, A. (2008). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society B*, *363*, 3477-3488.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science*, *2*(138-153).

Page, E. B. (1963). Ordered hypotheses for multiple treatments: A significance test for linear ranks. *Journal of the American Statistical Association*, *58*(301), 216-230.

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

Vries, M. H. de, Petersson, K., Guekes, S., Zwitserlood, P., & Christiansen, M. H. (2012). Processing multiple non-adjacent dependencies: evidence from sequence learning. *Philosophical Transactions of the Royal Society B*, *367*, 2065-2076.