Tessa Verhoef

The origins of duality of patterning in artificial whistled languages

Abstract: In human speech, a finite set of basic sounds is combined into a (potentially) unlimited set of well-formed morphemes. Hockett (1960) placed this phenomenon under the term 'duality of patterning' and included it as one of the basic design features of human language. Of the thirteen basic design features Hockett proposed, duality of patterning is the least studied and it is still unclear how it evolved in language. Recent work shedding light on this is summarized in this paper and experimental data is presented. This data shows that combinatorial structure can emerge in an artificial whistled language through cultural transmission as an adaptation to human cognitive biases and learning. In this work the method of experimental iterated learning (Kirby et al. 2008) is used, in which a participant is trained on the reproductions of the utterances the previous participant learned. Participants learn and recall a system of sounds that are produced with a slide whistle. Transmission from participant to participant causes the whistle systems to change and become more learnable and more structured. These findings follow from qualitative observations, quantitative measures and a follow-up experiment that tests how well participants can learn the emerged whistled languages by generalizing from a few examples.

Keywords: cultural evolution, combinatorial structure, language evolution, phonology

Tessa Verhoef: University of Amsterdam, 1012 VT Amsterdam, The Netherlands.
E-mail: t.verhoef@uva.nl

1 Introduction

To determine what distinguishes human communication from animal communication, Hockett (1960) identified thirteen basic design features of language (later expanded to sixteen [Hockett and Altmann 1968]). The one that he listed last was 'duality of patterning' which in part refers to how meaningless sounds are recombined into well-formed words of a language. It is the feature that has been studied the least and it is still unclear how it evolved in language. This combinatorial

structure refers to the systematic ways in which discrete basic building blocks are reused and combined and the presence of learned combinatorial constraints. The specific nature of the building blocks and constraints differ from one language to the other, but they are shared among the members of a language community.

Hockett (1960) proposed that the emergence of combinatorial structure was due to a pressure to keep the signals distinct when the vocabulary grows. As more and more meanings need to be expressed with unanalyzable signals of fixed duration, the signal space for creating such holistic signals fills up and the individual signals become more difficult to distinguish. If there is noise, which limits how accurately signals can be produced and perceived, there is a limit to the number of distinct signals that can be discriminated. This is when combinatorial structure emerges as an advantage for maintaining clear communication with a growing vocabulary. This view has been tested using computer models (Nowak et al. 1999; Zuidema and de Boer 2009). Nowak et al. (1999) have shown that, in the presence of noise, there is an error limit to the number of signals that can be used without the loss of communicative success. This limit can be overcome by combining signals (Nowak et al. 1999). In addition, de Boer (2000) and Zuidema and de Boer (2009) have shown that, when computer agents interact through imitation games, a pressure to keep signals distinct and discriminable will lead to discrete vowel systems (de Boer 2000) and combinatorial organization (Zuidema and de Boer 2009). The agents modeled by Zuidema and de Boer (2009) produce sounds as trajectories through a two-dimensional space. Both holistic sounds and potentially combinatoric sounds are therefore continuous signals with no predefined discrete and combinatorial organization, but such structure emerges in these trajectories. More generally, Abler (1989) draws a parallel between duality of patterning in language and the structure that is found in chemical systems and genetics. He argues that the emergence of each of these structures is attributable to general properties that are necessary to maintain 'self-diversification'. It appears therefore that Hockett may have been right about the mechanism driving the emergence of combinatorial structure. However, recent work offers at least three counterexamples to this hypothesis and suggests that it may not be the only account.

Firstly, Al-Sayyid Bedouin Sign Language (ABSL) is a sign language that shows the emergence of phonological structure (Israel and Sandler 2011; Sandler et al. 2011). Established sign languages have phonological structure with the same features of discreteness and recombination as speech. They have sign lexicons exhibiting discrete sets of location, hand shape and movement features that are recombined into meaningful words and, as in spoken languages, there are constraints on the ways in which features can be combined. ABSL is a young but fully functional and expressive sign language. It has a large vocabulary and a

rich, open-ended meaning space but its combinatorial structure is less discrete than those of established sign languages (Sandler et al. 2011). The ABSL lexicon is characterized by wider variation in sign forms across different signers in the language community. This case brings into question whether emergence of combinatorial structure is necessarily driven by a growing meaning space alone.

In addition, laboratory experiments in which a vocabulary of graphical signals is culturally transmitted over generations of experimental participants (in a diffusion chain) show that even when there is only a tiny vocabulary, structure emerges rapidly (del Giudice et al. 2010; del Giudice 2012). This is likely due to constraints on learning that are built into the experiment rather than of constraints on discriminability.

Finally, there are examples of vocal systems in animal communication that show predictable patterning and reuse of song parts. These are structured in way that is very similar to combinatorial structure in human language. Examples are the songs of humpback whales (Payne and McVay 1971) and the songs of certain bird species (Doupe and Kuhl 1999). Obviously the complex semantics we find in human language is absent in these cases but as in the earlier examples, the involvement of a pressure for discriminability with a growing vocabulary in its emergence seems unlikely.

The work discussed in this article investigates the emergence of combinatorial structure in speech by studying the transmission of a different kind of sound system, artificial whistled languages. The method that is being used is experimental iterated learning. Preliminary results of this experiment have been reported in Verhoef et al. (2011). In this paper, we present additional data and a more extensive analysis of the experiment.

1.1 Experimental iterated learning

The methods that are used in the current study stem from findings that show the importance of viewing language as a complex adaptive dynamical system (Steels 1997). Language is the result of influences from three different time scales: individual learning, cultural evolution and biological evolution. The complex interactions between these time scales have been studied extensively with the use of computer simulations. Many of the simulations have focused on the role of cultural evolution in the way languages are shaped. Typically, in these simulations computer agents interact with each other and cultural evolution is studied by simulating conventionalization and regularization of languages through social coordination (Steels 1997; de Boer 2000; Zuidema and de Boer 2009) and/or iterated learning (Kirby and Hurford 2002; Smith et al. 2003). The

iterated learning model simulates a process of (repeated) acquisition and repro-duction, where an agent learns its (linguistic) behaviour through observation of this behavior by another agent that acquired it in the same way. This process and the social coordination that occurs in the interactions between the agents leads to the establishment of shared communication systems and regularization of such systems.

More recently, in response to criticisms about the simplicity of the cognitive behavior of agents in these models, computer simulations were translated into studies that could be done with human participants in the laboratory (Galantucci 2005). A variety of experimental designs have been studied since then, ranging from interactive game-strategic (e.g. Galantucci 2005; Scott-Phillips et al. 2009) or Pictionary tasks (e.g. Garrod et al. 2007), to human iterated learning (e.g. Kirby et al. 2008). Scott-Phillips and Kirby (2010) provide an overview of many of these experiments and findings.

Kirby et al. (2008) first introduced the experimental iterated learning method while studying the emergence of compositional syntax. The idea of this tech-nique is to create a chain of learners in which the outcome of the learning of one participant is used as the input for the next person (Kirby et al. 2008). A key insight gained from computational studies of iterated learning is that lan-guages adapt to the transmission bottleneck and only structures that are learn-able, predictable and transmissible will survive (Kirby and Hurford 2002; Smith et al. 2003). In these experiments, as the language passes through the minds of learners, the system is expected to adapt to the learning biases, expectations and constraints of the learners (Deacon 1997; Kirby and Hurford 2002; Christian-sen and Chater 2008; Griffiths et al. 2008). For example, it has been applied suc-cessfully to show the emergence of compositional structure (Kirby et al. 2008), color terms (Dowman et al. 2008), predictability in plural marking (Smith and Wonnacott 2010) and in other category or function learning tasks (Griffiths et al. 2008).

The emergence of sub-lexical combinatorial structure has rarely been studied with participants in the laboratory. Galantucci et al. (2010) used an interactive game setting involving a visual communication medium with different levels of rapidity of fading. Emerged combinatorial structure was found to be stronger in signaling systems that were subjected to more rapid fading. del Giudice et al. (2010) studied the emergence of combinatorial structure with the experimental iterated learning method in systems of graphical signals. The work presented in this paper extends del Giudice et al.'s findings to the auditory modality by study-ing the emergence of structure in a whistled language. The next section describes the methods, followed by a discussion of the results. Then, a second experiment is presented which builds on the results from the first experiment.

2 Methods

In the experiment participants had to learn an artificial language and reproduce it from memory. These reproductions were used as input for the next participant. This process created four chains with ten generations of transmission through learning and reproduction. The languages our participants learn are in many ways more limited than real human languages. There are for instance no meanings in the language. This abstraction from the complexity of semantics allows us to closely investigate meaningless phonological structure on a basic level. The sounds of these languages are produced with a slide whistle (see Figure 1). Since hearing humans already have discrete and combinatorial organization in the sounds they produce when speaking, we wanted to introduce a novel auditory 'speech' apparatus that would involve less interference from previous experience with speech. The slide whistle is an intuitive and easy to use, non-linguistic instrument that turned out to be perfect for this purpose. The plunger can be used to adjust the pitch of the whistle sound.

Fig. 1: Slide whistle.

2.1 Initial whistle set

In all four chains the first participant received the same initial whistle set as learning input. The whistles in this set were selected from a database of whistles that were collected from nine participants in a pilot study. These participants were asked to freely produce and record a number of whistles. A set of signals was selected in which a wide range of whistle techniques (e.g. slides, siren-like, staccato) were used such that the total set of whistles did not exhibit any observable combinatorial structure. A few examples of these whistles are shown in Figure 2, plotted as pitch tracks on a semitone scale using Praat (Boersma 2001).

2.2 Procedure

The participants completed four rounds of learning and recall. In the learning phase they were exposed to all signals one by one, and asked to imitate each
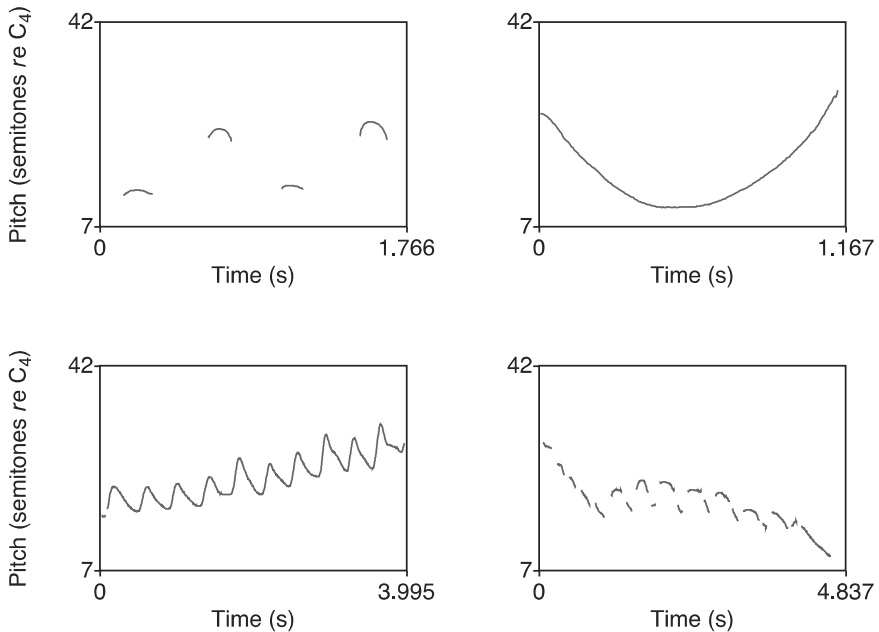
Fig. 2: Examples from the initial whistle set. Whistles are plotted as pitch tracks on a semitone scale.

sound with the slide whistle immediately. After this, a recall phase followed in which they reproduced all twelve whistles from memory. The input stimuli consisted of the output that the previous participant produced in the fourth and final recall round (or of the initial input set for the very first participant in a chain).

In the recall phase, participants were prevented from reproducing the same whistle twice. The user interface of the experiment automatically compared each new whistle produced to all other whistles already accepted and stored during that phase and it rejected that whistle if it was too similar to any other. In this case, the participant was asked to produce another whistle. The requirement of reproducing twelve unique whistles provides an artificial pressure for expressivity, which would normally result naturally from the need to express distinct meanings. In Kirby et al. (2008) an expressivity constraint was used effectively as well to prevent underspecification. To determine how close two whistles are, a whistle distance measure was defined. This measure is a weighted combination of several separate measures: the Dynamic Time Warping (DTW) (Sakoe and Chiba 2003) distance between the two pitch tracks (D_p), the DTW distance between the two intensity tracks (D_i), the difference in the number of segments

(separated by silent pauses) (D_s), the difference in variation of segment duration (D_{sd}) and the difference in variation of pitch (D_{pv}). These separate measures were scaled to have approximately a maximum value of one so that they could be given a weighted share in the final measure. For the final measure they were then combined following: $0.5D_p + 0.2D_i + 0.2D_s + 0.05D_{sd} + 0.05D_{pv}$. Data collected in a pilot study was used to create this measure. Participants in this pilot were all asked to imitate the same set of 10 whistles and the dataset created from these responses was used to find the set of weights that resulted in the highest whistle recognition score. The rejection threshold was set at a low value of 0.06 because it was not supposed to influence the outcome of the recall phase in any way other than to reject doubles.

Before the participants started the experiment, they were asked to sign an informed consent form and to fill out a questionnaire about their background information. Then, the task was explained, both by the experimenter and in written form on the screen. The participants were given a chance to ask questions and to familiarize themselves with the whistle before we started the task. After completing the final recall phase, the participants were asked to provide feedback about their own performance and experience and they were paid 10 euro or 10 dollars in cash as compensation for their time.

2.3 Participants

Forty participants took part in this study. This resulted in four parallel chains of ten generations of learning and reproduction. All participants were recruited among university students from either the University of California San Diego, or the University of Amsterdam, ranging in age from 18 to 32 (mean age of 22). Twenty-six were female. Each chain contained either three or four male participants.

2.4 Hypothesis

Based on earlier work on (human) iterated learning where it has been shown that this mechanism can lead to the emergence of compositional structure (Kirby et al. 2008; Kirby and Hurford 2002) and combinatorial structure in systems of graphical signals (del Giudice et al. 2010), we expected to find an increase in the amount of structure in the whistle systems at the end of each transmission chain as well. In addition, we expected to find an increase in the learnability of the set of signals as the chain progresses, because the sound systems change to become optimized

for learnability. When the system is more structured, participants are expected to learn faster and perform better.

3 Results

In this section the qualitative results are presented first, showing the development of the sound systems from generation to generation. This will give insight into the kind of structure that emerged and how this happened. Second, quantitative data are presented, introducing a measure of combinatorial structure and showing how the learnability of the whistled systems changed over the course of each chain.

3.1 Qualitative observations

Remembering twelve distinct whistles after limited exposure is not an easy task judging by the fact that participants usually do not recall all of them correctly. They make mistakes and in their productions they overgeneralize the structure that they think is present in the examples. This results in an introduction of whistles that are related in form to the other learned whistles, which can be observed as an increase of the presence of mirrored, combined or repeated versions of existing whistles. Mirroring happens in two ways. Sometimes an existing pattern is mirrored in the movement pattern in such a way that all slides up are transformed into slides down and the other way around. Others are mirrored in time in such a way that existing patterns stay intact, but their order of appearance within the produced whistle is mirrored. An example of this is shown in Figure 3a. In generation 8 there is a whistle with two quick level notes repeated three times from high to low pitch and in generation 9 a version is created in which the order is reversed and the same pattern is repeated from low to high pitch. Combining happens when two patterns that existed in distinct whistles are concatenated to form a new whistle. An example of this is shown in Figure 3b, where the short-short-long level notes pattern is used twice in a new whistle, combined with the two rising slides pattern from another whistle. Repetition involves the simple repeating of the same pattern more than once in a new whistle and this is shown in Figure 3c. Consequently, caused by these specific recall-strategies people use, over time, the signals share more and more elements, yielding increased combinatorial structure.

Another example of such cumulative increase of structure over generations is shown in Figure 4. In this example, a combination of mirroring, repetition and
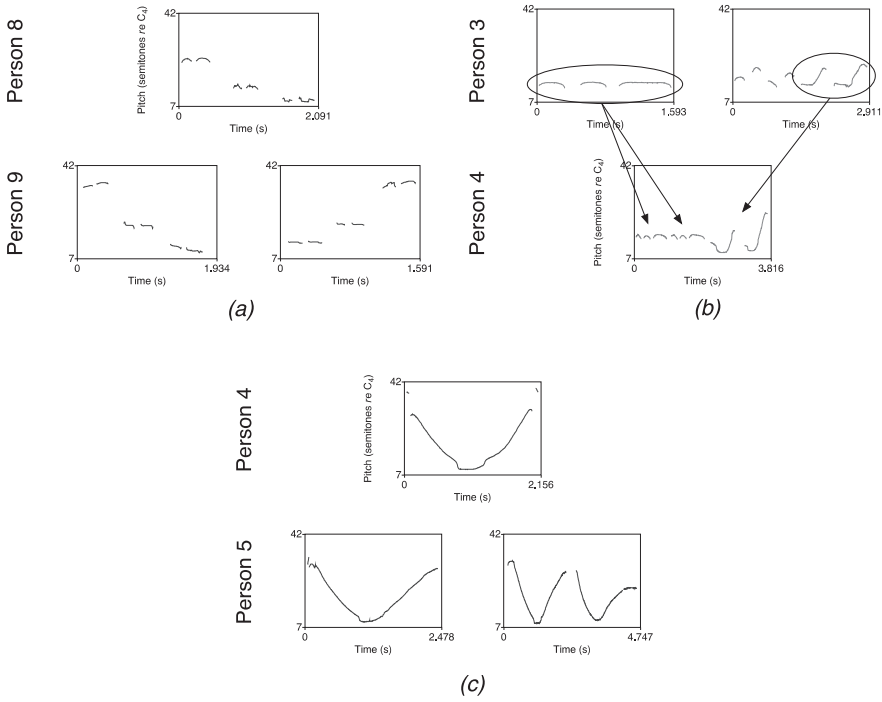
Fig. 3: Examples from different chains where patterns are introduced through mirroring (a), combining (b) and repetition (c) in the last whistle recall phase of consecutive generations.

borrowing results in a predictable system that is stable and persists after its innovation. In the productions of person four there is not yet another whistle that resembles the one shown here with two falling slides, but in generation five a mirrored version of this whistle appears. Then, in generation six one of these is borrowed and combined into a new whistle. This one may have been interpreted by person seven as having meant to be a repetition of the falling slide element present in the original two, because suddenly a version with three falling slides appears as well as a mirrored version with three rising slides which fills a gap in this regular system.

Figure 5 shows a fragment of a whistle set that emerged in one of the chains in generation ten. This set exhibits a set of basic building blocks. There are short level tones and falling-rising slides and these elements are systematically reused and combined. The whistles differ from each other in the amount of short level tones they start and end with and for each there is often a version mirrored in order as well. In addition, the set has become more constrained, for instance in
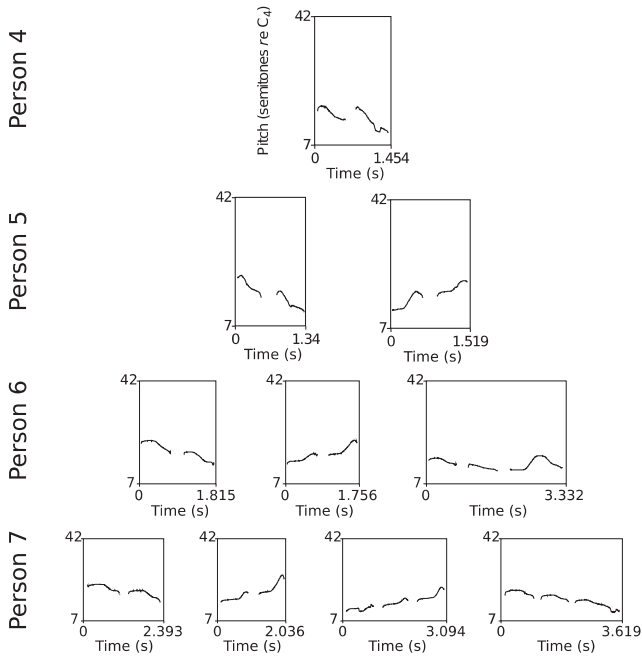
Fig. 4: An example of cumulative mirroring, repetition and borrowing. Person 5 mirrors the whistle from the previous set, then person six borrows one of the two in a new whistle and finally this new whistle becomes generalized to fit the pattern of the original two, but repeated. This predictable system is stable towards the end of the chain. The whistles are plotted as pitch tracks on a semitone scale.
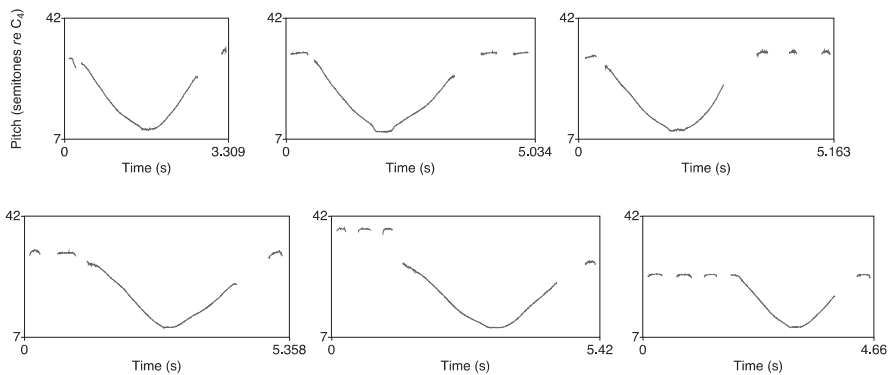

Fig. 5: Fragment of the whistles plotted as pitch tracks in the last set of a chain. Basic elements can be identified that are systematically recombined.

the number of falling-rising movements per segment. In the initial set (see Figure 2) there were whistles with several falling-rising movements in one segment, but this has reduced to a maximum of two movements in the last generation of this chain. Another constraint is the fact that in this set all segments with slides start with a falling tone and there is no longer any version that is mirrored in pitch. Note that this is specific for this particular chain; in other chains rising-initial patterns did occur. Qualitatively, basic elements and systematic recombination can be observed in all four emerged sets of whistles, but the elements and constraints on the way they are combined differ from one chain to the other.

In summary, the qualitative observations indicate that the whistle sets indeed start to exhibit more combinatorial structure towards the end of the experimental chains, suggesting that the emergence of such structure occurred as a result of repeated learning and reproduction.

3.2 Quantitative results

In order to quantify the observations that were made in the previous subsection, the increase of combinatorial structure was measured over generations in the chains. In addition, to determine whether the learnability of the systems increases over generations, the recall errors of all participants in their last recall phase were measured.

For the quantitative analysis a different whistle distance measure was used than the one described in Section 2.2. After the data was collected and the results were analyzed qualitatively, we found that participants paid more attention to the movements of the plunger and the contours of the whistles than to the precise acoustic realization (on which the first distance measure was based). People seemed to remember and classify the sounds according to the plunger 'gestures' they make to produce them. A building block of a certain movement can be performed just as 'big' with the plunger on the bottom of the whistle, with lower pitch, as with the plunger on the top, with higher pitch. But in terms of pitch, this results in a much bigger difference when it is produced at higher pitch than at lower pitch, because of the non-linear relation between the scales of pitch and plunger movement of the whistle. So if acoustical features are used, distances between plunger gestures tend to be overestimated in the high pitch range, while they are underestimated in the low pitch. This finding may be interesting in the context of the debate on what the basic units of speech perception are and we will come back to this in the discussion section.

For the new measure the pitch tracks are first transformed into sequences of plunger positions (from approximately 3 cm to 21 cm) following equation (1),

where l is the length in cm between the mouthpiece and sliding stopper, c is the speed of sound at body temperature (35,000 cm/s) and f is the measured frequency in Hz. These new tracks represent the actual movements the participants made, and the distance between two whistles is the Derivative Dynamic Time Warping (Keogh and Pazzani 2001) distance between two movement-tracks.

$$l = \frac{c}{4f} \tag{1}$$

Note that we did use the acoustics-based measure during data collection for rejecting whistles that were too similar. This could raise a concern that this may have had implications for the kinds of whistles that were rejected, but we expect that this had a negligible influence. The threshold for whistle rejection was carefully chosen to be very low, so that we would never accidentally reject a whistle that was similar to another one, but differed only in one aspect. It is therefore more likely that very similar whistles in the higher pitch region were accidentally accepted than that any whistle was wrongly rejected.

To compute the combinatorial structure, we make use of the fact that more combinatorial structure implies more efficient coding and compressibility. When there is more combinatorial structure, it means that a set of signals can be reconstructed by combining a smaller number of basic building blocks and thus the set is more compressible. The information-theoretic measure of entropy (Shannon 1948) is used to measure this. To compute entropy for a set of whistles, the whistles were divided into segments, taking silences within a whistle as segment boundaries. Then, using all segments that occur in the set of twelve whistles, average linkage agglomerative hierarchical clustering (Duda et al. 2001) was used to group together those segments that were so similar that they could be considered the same category or building block. Clustering continued until there was no pair of segments left with a distance smaller than 0.08. Equation (2) from Shannon (1948) was then used to compute entropy, where p_i is the probability of occurrence of building block i.

$$H = -\sum p_i \log p_i \tag{2}$$

Figure 6 shows this measure for the four chains, with the generations on the horizontal axis and 0 referring to the initial set. A significant cumulative decrease in entropy was measured using Page's (1963) trend test (L = 1427, m = 4, n = 10, p < 0.001), excluding the artificially inserted initial set (with this set included it is also significant (L = 1853, m = 4, n = 11, p < 0.001)). This implies an increase of structure and predictability as well as more efficient coding and compressibility.
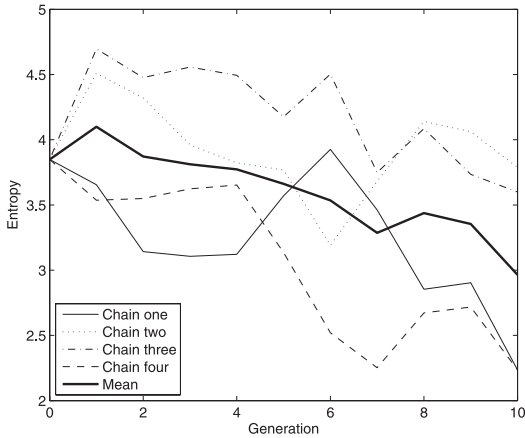
Fig. 6: Entropy of the whistle sets over generations for all four chains, demonstrating that the combinatorial structure increases.

To determine the recall error in the last recall phase for each participant, the whistles of their learned input and recalled output were compared. This was done by pairing the whistles of the two sets in all possible ways, then computing the distance between the paired signals and taking the lowest sum of distances, which belongs to the most optimal pairing. Figure 7 shows this measure for the four chains, with the generations again on the horizontal axis. A significant
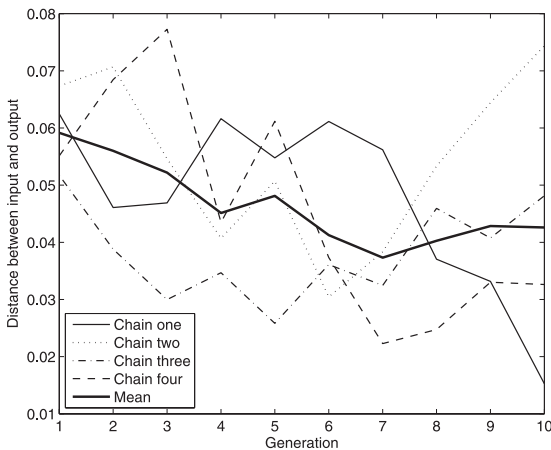

Fig. 7: Recall error on the whistle sets over generations for all four chains, demonstrating that the whistle systems evolve through cultural transmission and become more learnable.

cumulative decrease in recall error was measured using Page's (1963) trend test (L = 1318, m = 4, n = 10, p < 0.05), implying an increase of learnability and reproducibility of the whistle sets over generations.

To summarize, the quantitative analysis confirms the qualitative observation that combinatorial structure increases over generations as well as the hypothesis that the whistle sets become more learnable through the process of cultural transmission.

4 Analysis of whistle sets in a perceptual category learning game

Zuidema and de Boer (2009) introduced a distinction between two kinds of combinatorial structure that can be identified when studying systems of signals. The first kind is what they call 'superficial combinatorial structure' and this refers to combinatorial structure that can be identified when a system is analyzed by an outside observer, but the users of the system are not necessarily aware of this structure. The second kind is called 'productive combinatorial structure' and this refers to the structure that users of the system are cognitively aware of and that is actively used in production, perception and learning. The results that were presented in the previous section show both qualitatively and quantitatively that a system of auditory signals gains (superficial) combinatorial structure and becomes more learnable when it is transmitted culturally. What we have not shown quantitatively yet is that the learners actively use the combinatorial structure in a way that Zuidema and de Boer (2009) would call productive. Note that their definition does not require signal production before a system can be considered to have 'productive combinatorial structure'. It involves awareness of the structure in production as well as perception and learning. Although it seems unlikely given the combination of qualitative and quantitative results, it may still be the case that the experimenters and measures only observe the structure as outside observers. In addition, the fact that an increase in learnability of the system was measured does not necessarily mean that it has become more learnable because of the increased structure and cognitive ease that comes with it. An alternative explanation may be that only the individual whistles have evolved to become easier to imitate and that therefore only articulatory constraints made the set more reproducible.

To test the possibilities for human productive use and awareness of the structure that seems to be present in the emerged whistle sets, and to identify whether cognitive constraints may indeed have been involved in shaping these sets, a

Fig. 8: Screenshot of the UFO game, created by Jelle Zuidema and Vanessa Ferdinand (http://www.webexperiment.nl/).

separate experiment was conducted. In this experiment, the stimuli that were used came from the languages that resulted at the end of chains one and four in the whistle experiment described in Section 2. The aim was to test if human participants, that are exposed to a few examples of such an emergent whistle language, can decide for other examples if they belong to the set or not. For the design of this experiment we used the existing UFO game paradigm[1]. In this game, two species of aliens exist: good aliens and bad aliens. The player's goal is to save the good aliens and kill the bad ones. A screenshot of the game is shown in Figure 8. First, participants are exposed to the language of the good aliens in the familiarization phase and they practice to save the spaceships of these aliens. Second, they practice shooting spaceships on a few empty ships. Third, UFOs fly by in the combat phase and participants have to decide whether they are good or bad according to the sounds they make and kill or save them accordingly. Last, they see their final score.

Two conditions were created, differing in which individual whistle sounds from the two emergent languages were part of each alien species' language. In the 'intact' condition, each of the two alien species languages consisted of a complete emergent whistle language. This means that one alien species had a vocabulary consisting of all twelve sounds produced by the last person in chain one (of the iterated learning experiment described in Section 2) and the other alien species used those from chain four. In the 'mixed' condition, each alien species had six

---

1 The UFO game was created by Jelle Zuidema and Vanessa Ferdinand (http://www.webexperiment.nl/) and they kindly allowed me to use and adapt their code.

Good aliens   Bad aliens

Condition 1:

Condition 2:
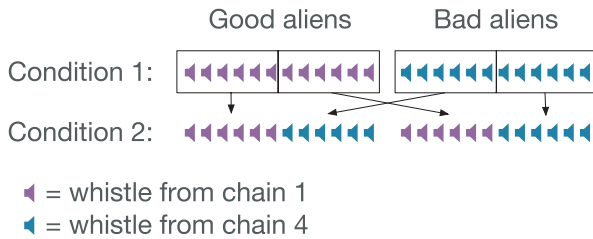
◀ = whistle from chain 1
◀ = whistle from chain 4

Fig. 9: Two experimental conditions: (1) the 'intact' condition, where each of the two alien species languages consisted of an intact emergent whistle set from the last generation of chain 1 and 4 of the experiment described in Section 2. (2) the 'mixed' condition, where mixing sounds from both sets created the two languages.

sounds in their language from chain one and six sounds from chain 4, breaking up the emergent whistle languages from the iterated learning experiment. This is illustrated schematically in Figure 9. We selected the languages of chains one and four because, as can be seen in Figure 6, these were the two chains that resulted in emergent languages exhibiting the most combinatorial structure and their measured amount of structure was very similar. In the intact condition it was alternated whether the good aliens used sounds from chain one or four. In the mixed condition, two different ways of breaking up the languages from the two chains was alternated as well.

With this design we could ask the question: can participants generalize and use the combinatorial structure in the emerged whistle languages to classify new aliens as good or bad and save or kill them accordingly? In the familiarization phase, participants are exposed to six out of the twelve sounds that the good aliens use. In the mixed condition, they are exposed to three sounds originating from each of the two chains. In the combat phase they are tested on all sounds of both species, including the ones of the good aliens they had never heard before. If the participants are aware of a potential structure in the sounds and use it productively, they should perform better on the whistles they never heard before in the intact condition. The mixed condition, where the two emergent languages are broken up, should give participants much less evidence about potential rules, building blocks or constraints in the languages to generalize from. In the first condition, if structure is present in the emergent languages from the iterated learning experiment, participants should be able to generalize and classify the identity of UFOs with an accuracy above the baseline of random guessing.

The experiment was conducted as an online game for which participants were recruited through Facebook. Ten participants completed the game in the mixed condition and eleven in the intact condition. Their ages ranged from 22 to

	Condition:	
	Intact	Mixed
Median score	47	23.5

Table 1: Results of the category learning experiment. The median score of correct classification was 47 in the intact condition and 23.5 in the mixed condition. There is a significant difference between the distributions of the two groups (Mann–Whitney U = 55, n1 = 11, n2 = 10, P < 0.001).

50 (mean age of 29). There were twelve male participants and six of them participated in each condition.

To analyze the results, for each participant it was determined how many of the sounds that they had never heard in the familiarization phase they were able to classify as belonging to good or bad aliens correctly. In total there were 54 new items in the combat phase (twelve sounds from the bad aliens and 6 from the good alien that were never heard before, each appearing 3 times). The results are shown in Table 1. In the intact condition, the median score of correct classification was 47 and in the mixed condition it was 23.5. There is a significant difference between the distributions of the two groups (Mann–Whitney U = 55, n1 = 11, n2 = 10, P < 0.001). The expected baseline score in the case of random guessing would be 27 (54 * 0.5). The intact group scored well above this baseline and the mixed group slightly below it.

By means of this perceptual category learning game it has been shown that there is structural evidence available in the emergent whistle languages and learners use it to distinguish between distinct languages. Following the definitions proposed by Zuidema and de Boer (2009), the observed combinatorial structure could be concluded to be of the productive type. Human participants are aware of the regularities and they are able to use it in perception and recognition.

5 Discussion

The experiments described in this paper show that experimental iterated learning can cause an artificial whistled language to become organized in a way that is reminiscent of how speech sounds and signs in sign languages are organized. By observing the emerged whistle languages qualitatively, we identified the presence of basic building blocks and a systematic recombination of these building blocks. Quantitatively, we measured a significant cumulative increase of combinatorial structure and a significant cumulative decrease of recall error, indicating

that the whistle systems become gradually more learnable. With the use of an additional perceptual categorization game, we showed that humans are aware of the observed combinatorial structure and that they use it productively. It is therefore unlikely that it is merely superficial combinatorial structure that we observed and it supports the hypothesis that cognitive constraints cause a culturally transmitted system to become more structured and more learnable.

What seems to be driving the emergence of structure in this experiment relates to predictability and compressibility of the system of signals. Without structure, there are no constraints that help to decide which whistles could be part of the system. This makes the signal space theoretically unrestricted and unpredictable. Combinatorial structure limits possibilities and allows learners to focus only on the variations that are linguistically relevant. By developing from a holistic system, in which virtually everything is possible within the limits of the modality, towards a discrete and combinatorial system, in which only a few elements can be used and these elements can only be combined in restricted ways, the system becomes more predictable. It has been argued that languages are generally organized to be predictable and Smith and Wonnacott (2010) have shown that the process of iterated learning can cause a linguistic system to lose unpredictable variation. In addition, the principle of measuring predictability in a linguistic system has also been applied to explain the existence and learnability of complex morphological systems in real languages (Ackerman et al. 2009). Moreover, 'Maximal Utilisation of Available Distinctive Features' (Ohala 1980) and 'feature economy' (see Clements 2003) have been proposed as organizational principles of sound systems for speech. These principles focus on maximizing the efficient reuse of distinct features to make up a system of sounds: 'features used once in a system tend to be used again' (Clements 2003). The more economical the sound inventory, the more compressible and predictable the system will be as a whole. These examples all highlight the clear tendency toward efficient coding in languages, but they may differ in the assumptions about where such a tendency comes from. As argued in Verhoef and de Boer (2011) the structure found in the emerged whistle languages resembles principles of efficient coding and this supports the view that efficient coding may emerge as a result of constraints on learning in cultural evolution. It has been proposed that the combinatorial nature of sound organization represents a general tendency to organize linguistic input into a number of categories that are then generalized maximally. The same cognitive mechanisms are therefore expected to be involved both in sound organization and other levels of grammar (Clements 2003). It appears that the underlying cognitive mechanisms are not specific to the auditory domain, and probably not even specific to language, because strikingly similar results are found in both visual (del Giudice et al. 2010) and auditory (Verhoef et al. 2011; Verhoef and de

Boer 2011; Verhoef et al. 2012) versions of the experiments which both involve non-linguistic signaling devices.

According to Hockett (1960), an increase in the number of meaningful elements drove the emergence of combinatoriality because the limits of the signal space were reached and no more distinguishable holistic signals could be added. However, as mentioned in the introduction, the high functionality but still emerging combinatorial structure evident in ABSL shows that a language can have a large open-ended meaning space without needing highly constrained combinatorial signals (Sandler et al. 2011). This suggests that combinatorial structure is not only the result of pressures from vocabulary growth and signal dispersion. The data presented here shows that, in absence of meaningful referents, combinatorial structure still emerged in sets of whistles that were culturally transmitted. This happened in a very small vocabulary of only twelve whistles and long before the signal space had been fully covered. Apparently, a good reason to have combinatorial structure, even for a very simple system, is that a system with such structure is easier to learn and reproduce. In line with earlier findings on the dynamics of iterated learning, the whistles that fit the structure and conform to people's cognitive biases are more likely to be preserved from generation to generation in cultural evolution. Combinatorial structure therefore potentially emerges within a gradual cultural evolutionary process. This provides an additional explanation for the origins of combinatorial structure, suggesting that the theory that was proposed by Hockett (1960) may not be-the-only factor involved. The fact that different parallel chains result in whistle languages that are recognizably different in terms of the specific rules, building blocks and constraints (as follows from the UFO game results), supports the view that the structure is conventionalized and emerges through cultural evolution.

Another interesting insight, which was not the aim of this study, but followed from our analysis, is the fact that the basic elements that are recombined in the whistles seem to be comprised of articulatory movements rather than acoustic information. The observed structure could not be quantified reliably when the distance measure was based on acoustic patterns. Justified by the observation that participants were paying more attention to the plunger movements than to acoustic cues, a measure based on plunger gestures was used, which did result in a successful quantification of combinatorial structure. This finding is potentially interesting in light of the discussion on whether the basic units of speech perception are acoustic or gestural (e.g. Browman and Goldstein 1986). Galantucci et al. (2006) provide a modern evaluation of the motor theory of speech perception and they review evidence in favor of gesture-based theories. They also point out that the theory has been very well received outside the field of speech research, but has been less popular within the field and reviews of evidence against gesture

based hypotheses have been written in response (e.g. Massaro and Chen 2008). Evidence from comparative data with birds has been used to present both arguments in favor of (Williams and Nottebohm 1985) and against (Kluender et al. 1987) gesture-based theories. In short, controversy prevails surrounding these ideas, but our data seem to support the motor theory of speech perception.

A possible concern with the current results, if we were to compare them to real human languages, involves the lack of meaning conveyed by the whistled signals. With the experimental design described in this paper, we abstracted away from full human semantic complexity by not having an explicit meaning connected to the whistles. The system is not entirely meaningless though, because the requirement of reproducing twelve unique whistles provides an artificial pressure for expressivity, which would normally result naturally from the need to express distinct meanings. The requirement that participants have to retrieve the whistles from memory also often makes them 'label' the whistles as for instance: 'the one with many up and down movements' or 'the very first whistle I learned'. Moreover, once the whistles evolve towards sharing features, people tend to categorize them. The combinatorial structure makes it possible to remember the whistles as subsets, such as 'the ones that all start with one slide down' or 'the ones that only have slides up'. This adds meaning implicitly and makes learning and recall of the whole set of whistles easier. This first investigation of combinatorial structure in a set of whistles without referents was necessary to be able to control for effects of semantics such as iconicity or compositional structure. With such influences present it would be harder to distinguish whether the emerging structure relates to the structure of the meaning space or whether they are truly meaningless units being recombined. In addition it would be harder to know what drove the emergence of the structure. Given the current results that show combinatorial structure can emerge without complex semantics, an interesting next step is certainly to include meanings in a follow-up study. In our current ongoing work we try to address this issue and a preview of this is described below.

Another concern with our results, if we were to consider this experiment as a reconstruction of language evolution, would be that obviously we use modern human participants who have modern cognitive adaptations unlike our ancestors. This problem is shared among all language evolution research that makes use of human participants. As Scott-Phillips and Kirby (2010) pointed out, the results of this type of work should not be interpreted as an attempt to reconstruct the emergence of linguistic structure, or, in this case of duality of patterning, but as a method to shed light on what mechanisms may be involved in this emergence. The current work is meant to illustrate how human cognitive biases influence a sound system when it is transmitted over generations and what role these biases play in the maintenance of combinatorial structure. However, it may in fact

be likely that our ancestors already had the cognitive abilities that are needed to deal with combinatorial structure. Research with cotton-top tamarins (Hauser et al. 2001) for instance has shown that at least some non-human primates have simple abilities for sound segmentation. The ability to find regularities in sound systems is therefore likely to be much older than the evolutionary split between humans and other apes.

A question that remains in the research we report here is what might happen if whistle words were linked to meaningful objects and what the influence would be of iconicity on the development of combinatorial structure. Currently, a follow-up study is being conducted that involves a new version of the experiment in which the whistle signals do refer to meanings. The meanings in this study will be a subset of unusual objects that were created by Smith et al. (2011) and were slightly modified. The objects look like possible mechanical parts, but they are novel objects for which we do not have words in existing languages. The objects do not share colors, shapes or parts and are not structured in any other obvious way. The emergence of semantics-related structure should therefore be limited, although it is expected that iconicity will still play a role. To investigate how iconic form-meaning mappings influences the emergence of combinatorial sub-lexical structure, two conditions are studied: one in which the use of iconic form-meaning mappings is possible (and is expected to be used) and one in which the use of iconic form-meaning mappings is experimentally made impossible by scrambling the form-meaning mapping at each change of generation and to use different objects between consecutive generations. This should provide insight into the possible role of iconicity in the delayed emergence of duality of patterning in ABSL and could tell us whether a situation that allows for more iconicity, which is generally the case for sign languages as compared to spoken languages, can 'survive' without the emergence of combinatorial structure longer.

So far we have only scratched the surface and many questions remain to be answered, but the application of the experimental iterated learning paradigm to questions of the origins of duality of patterning has already created valuable insights. Future experiments are expected to provide answers to more specific matters by unraveling the nature of combinatorial structure in auditory, graphical and gestural systems and this will hopefully reveal how duality of pattering has emerged.

References

Abler, W. L. 1989. On the particulate principle of self-diversifying systems. *Journal of Social and Biological Structures* 12(1). 1–13.

Ackerman, F., J. P. Blevins & R. Malouf. 2009. Parts and wholes: Patterns of relatedness in complex morphological systems and why they matter. In J. P. Blevins & J. Blevins (eds.), *Analogy in grammar: Form and acquisition,* 54–82. Oxford: Oxford University Press.

Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5(9/10). 341–345.

Browman, C. P. & L. Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3. 219–252.

Clements, G. N. 2003. Feature economy in sound systems. *Phonology* 20. 287–333.

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

Deacon, T. W. 1997. *The symbolic species: The co-evolution of language and the brain.* New York: WW Norton & Co.

de Boer, B. 2000. Self-organization in vowel systems. *Journal of Phonetics* 28(4). 441–465.

del Giudice, A. 2012. The emergence of duality of patterning through iterated learning. *Language and cognition.* [This volume].

del Giudice, A., S. Kirby & C. Padden. 2010. Recreating duality of patterning in the laboratory: a new experimental paradigm for studying emergence of sublexical structure. In A. D. M. Smith, M. Schouwstra, B. de Boer & K. Smith (eds.), *The evolution of language: Proceedings of the 8th international conference*, 399–400. Hackensack, NJ: World Scientific Press.

Doupe, A. J. & P. K. Kuhl. 1999. Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience* 22. 567–631.

Dowman, M., J. Xu & T. L. Griffiths. 2008. A human model of color term evolution. In A. D. M. Smith, K. Smith & R. F. i Cancho (eds.), *The evolution of language: Proceedings of the 7th international conference*, 421–422. Hackensack, NJ: World Scientific Press.

Duda, R. O., P. E. Hart & D. G. Stork. 2001. *Pattern recognition*. New York: A Wiley-Interscience.

Galantucci, B. 2005. An experimental study of the emergence of human communication. *Cognitive Science* 29. 737–767.

Galantucci, B., C. A. Fowler & M. T. Turvey. 2006. The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review* 13(3). 361–377.

Galantucci, B., C. Kroos & T. Rhodes. 2010. The effects of rapidity of fading on communication systems. *Interaction Studies* 11. 100–111.

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

Griffiths, T., M. Kalish & S. Lewandowsky. 2008. Theoretical and empirical evidence for the impact of inductive biases on cultural evolution. *Philosophical Transactions of the Royal Society: Biological Sciences* 363(1509). 3503–3514.

Hauser, M., E. Newport & R. Aslin. 2001. Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition* 78(3). B53–B64.

Hockett, C. 1960. The origin of speech. *Scientific American* 203. 88–96.

Hockett, C. F. & S. Altmann. 1968. A note on design features. In T. A. Sebeok (ed.), *Animal communication; Techniques of study and results of research*, 61–72. Bloomington: Indiana University Press.

Israel, A. & W. Sandler. 2011. Phonological category resolution in a new sign language: A comparative study of handshapes. In R. Channon & H. van der Hulst (eds.), *Formational units in sign languages*, 177–202. Berlin: Mouton de Gruyter.

Keogh, E. J. & M. J. Pazzani. 2001. Derivative dynamic time warping. *First SIAM International Conference on Data Mining (sdm2001)*. 1–11.

Kirby, S., H. Cornish & K. Smith. 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. & J. Hurford. 2002. The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (eds.), *Simulating the evolution of language*, 121–148. New York: Springer Verlag.

Kluender, K. R., R. L. Diehl & P. R. Killeen. 1987. Japanese quail can learn phonetic categories. *Science* 237(4819). 1195–1197.

Massaro, D. W. & T. H. Chen. 2008. The motor theory of speech perception revisited. *Psychonomic Bulletin and Review* 15(2). 453–457.

Nowak, M., D. Krakauer & A. Dress. 1999. An error limit for the evolution of language. *Proceedings of the Royal Society of London* 266. 2131–2136.

Ohala, J. J. 1980. Moderator's introduction to symposium on phonetic universals in phonological systems and their explanation. *Proceedings of the 9th International Congress of Phonetic Sciences*. 181–185.

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.

Payne, R.S. & S. McVay. 1971. Songs of humpack whales. *Science* 173(3997). 585–597.

Sakoe, H. & S. Chiba. 2003. Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on* 26(1). 43–49.

Sandler, W., M. Aronoff, I. Meir & C. Padden. 2011. The gradual emergence of phonological form in a new language. *Natural Language and Linguistic Theory*. 29(2). 503–543.

Scott-Phillips, T. C. & S. Kirby. 2010. Language evolution in the laboratory. *Trends in cognitive sciences* 14. 411–417.

Scott-Phillips, T. C., S. Kirby & G. R. S. Ritchie. 2009. Signalling signalhood and the emergence of communication. *Cognition* 113(2). 226–233.

Shannon, C. E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27. 379–423, 623–656.

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

Smith, K., A. D. M. Smith & R. A. Blythe. 2011. Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science* 35. 480–498.

Smith, K. & E. Wonnacott. 2010. Eliminating unpredictable variation through iterated learning. *Cognition* 116. 444–449.

Steels, L. 1997. The synthetic modeling of language origins. *Evolution of Communication* 1(1). 1–34.

Verhoef, T. & B. de Boer. 2011. Cultural emergence of feature economy in an artificial whistled language. In E. Zee & W. Lee (eds.), *Proceedings of the 17th international congress of phonetic sciences,* 2066–2069. Hong Kong: City University of Hong Kong.

Verhoef, T., B. de Boer & S. Kirby. 2012. Holistic or synthetic protolanguage: Evidence from iterated learning of whistled signals. In T. C. Scott-Phillips, M. Tamariz, E. A. Cartmill & J. R. Hurford (eds.), *The evolution of language: Proceedings of the 9th international conference (evolang9)*, 368–375. Hackensack, NJ: World Scientific.

Verhoef, T., S. Kirby & C. Padden. 2011. Cultural emergence of combinatorial structure in an artificial whistled language. In L. Carlson, C. Hölscher & T. Shipley (eds.), *Proceedings of*

the 33rd annual conference of the cognitive science society, 483–488. Austin, TX: Cognitive Science Society.

Williams, H. & F. Nottebohm. 1985. Auditory responses in avian vocal motor neurons: A motor theory for song perception in birds. *Science* 229(4710). 279–282.

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

Acknowledgement