COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 40 (2016) 1969–1994

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12326

Iconicity and the Emergence of Combinatorial Structure in Language

Tessa Verhoef,[a] Simon Kirby,[b] Bart de Boer[c]

[a]*Center for Research in Language, University of California, San Diego*
[b]*School of Philosophy, Psychology and Language Sciences, University of Edinburgh*
[c]*Artificial Intelligence Lab, Vrije Universiteit Brussel*

Received 11 March 2014; received in revised form 11 June 2015; accepted 16 September 2015

Abstract

In language, recombination of a discrete set of meaningless building blocks forms an unlimited set of possible utterances. How such combinatorial structure emerged in the evolution of human language is increasingly being studied. It has been shown that it can emerge when languages culturally evolve and adapt to human cognitive biases. How the emergence of combinatorial structure interacts with the existence of holistic iconic form-meaning mappings in a language is still unknown. The experiment presented in this paper studies the role of iconicity and human cognitive learning biases in the emergence of combinatorial structure in artificial whistled languages. Participants learned and reproduced whistled words for novel objects with the use of a slide whistle. Their reproductions were used as input for the next participant, to create transmission chains and simulate cultural transmission. Two conditions were studied: one in which the persistence of iconic form-meaning mappings was possible and one in which this was experimentally made impossible. In both conditions, cultural transmission caused the whistled languages to become more learnable and more structured, but this process was slightly delayed in the first condition. Our findings help to gain insight into when and how words may lose their iconic origins when they become part of an organized linguistic system.

Keywords: Cultural evolution; Cognitive biases; Iterated learning; Combinatorial structure; Iconicity; Language evolution

1. Introduction

Human speech recombines a small set of meaningless acoustic building blocks into an unlimited set of possible utterances.[1] The use of such *combinatorial structure* is not quite

Correspondence should be sent to Tessa Verhoef, Center for Research in Language, University of California, San Diego, 9500 Gilman dr., La Jolla, CA 92093, USA. E-mail: tverhoef@ucsd.edu

unique—the vocalizations of certain birds, cetaceans, and gibbon species (Berwick, Okanoya, Beckers, & Bolhuis, 2011; Mitani & Marler, 1989; Payne & McVay, 1971) exhibit it as well—but our closest relatives, the great apes appear not to use it. It is therefore highly likely that the latest common ancestor of humans and the other great apes did not use combinatorial vocalizations. An account of the evolution of language must therefore explain how combinatorial speech emerged. Even though (as will be explained below) combinatorial structure has great advantages, for instance when using a large number of signals in the presence of noise (Hockett, 1960), it may not always be a necessary feature of language. Al-Sayyid Bedouin Sign Language (ABSL) is an example of a fully functional, expressive sign language that lacks the clear discrete and combinatorial phonology that other languages have (Sandler, Aronoff, Meir, & Padden, 2011). Perhaps this young sign language has been able to survive up to now with little sublexical combinatorial structure because the manual modality allows for a large degree of iconicity: signals for which the form resembles the meaning they express. It could be the case that iconicity causes the language to be learnable and transmissible even with limited phonological structure. When a system can support a large amount of transparent, holistic mappings, perhaps there is less need for combinatorial structure at the sub-lexical level (Sandler et al., 2011). In this paper, we investigate experimentally whether the potential for iconic form-meaning mappings interferes with the emergence of combinatorial structure in a system of acoustic signals.

Combinatorial structure has the advantage over non-combinatorial signals when signals taken from a limited signaling space are used to communicate in a noisy environment, as pointed out by Hockett (1960) and Nowak, Krakauer, and Dress (1999). As more meanings need to be expressed and new holistic signals carve out the signal space, this space will fill up and signals will become more similar and more easily confused. Using utterances that consist of combinations of a smaller number of signals is a way out of this problem. It has been shown using computer simulations that such signaling systems can arise even in populations where the users of the signals are in no way (consciously or subconsciously) aware of this structure (De Boer & Zuidema, 2010; Zuidema & de Boer, 2009). No active creation by individuals is therefore needed for combinatorial structure to get started in this scenario.

Another advantage of combinatorial speech is that it makes systems of signals more predictable, and therefore easier to learn and to transmit through a learning bottleneck (a situation where learners need to reconstruct a system of which they have only seen a limited number of examples). Different theoretical accounts of how this plays a role in phonology have been proposed (Clements, 2003; Martinet, 1949; Ohala, 1980), all assuming involvement of cognitive biases for detecting, reusing, and preferring certain regularities.

The importance of cognitive adaptations (language-specific or not) and how they are involved in the emergence of combinatorial structure in (modern) human language can be investigated experimentally. In earlier experiments (Verhoef, 2012; Verhoef, Kirby, & de Boer, 2014; Verhoef, Kirby, & Padden, 2011), it was shown that structure in acoustic signals emerges as a result of cultural transmission and cognitive biases and from a human

tendency to reuse and modify learned building blocks rather than from a pressure to use the available signaling space as effectively as possible. The experiment used an iterated learning approach (Kirby, Cornish, & Smith, 2008) in which participants learn a set of signals that was learned and reproduced by an earlier participant. The set of signals to be learned contained only 12 signals. This number of signals was so small that limits of the signaling space could not cause confusion. Nevertheless, combinatorial structure did emerge, and the way in which it emerged clearly showed a gradual increase in the re-use and systematic modification of building blocks. The procedure of the experiment was very similar to the one followed in this paper, but one crucial aspect of linguistic communication was missing: The signals the participants had to learn did not have meaning. However, meaning *can* influence the form of a signal—many languages have iconic signals in which the form resembles the meaning.

Iconicity can manifest itself in many different ways in language. It involves classes of words where, for instance, the shape, complexity, sound, or some other characteristic of the meaning expressed is mimicked or iconically represented in the form of an utterance. Examples have been identified as "ideophones," "mimetics," or "expressives" and the phenomenon is often called sound-symbolism (Hinton, Nichols, & Ohala, 1994) for spoken languages. As Cuskley and Kirby (2013) describe, *conventional sound symbolism* refers to the statistical correspondences between certain clusters of similar forms and meaning classes, where sub-lexical elements are systematically used for a certain semantic domain. *Sensory sound symbolism* describes words that phonetically imitate the sound their referent makes, such as "bang" or "buzz" (which are called "onomatopoeia"), or words that cross-modally imitate other characteristics of the referent, for instance based on vision, temporal structure, touch, taste, smell, or other domains (Cuskley and Kirby, 2013; Dingemanse, 2012). The role of iconicity in language acquisition and processing has indicated a positive relation (Perniss, Thompson, & Vigliocco, 2010). It has been shown, for instance, that in the context of a lexical decision task non-arbitrary form-meaning pairs are processed faster than arbitrary form-meaning pairs (Bergen, 2004) and that sound-symbolic mappings help young children in acquiring new words (Imai, Kita, Nagumo, & Okada, 2008). Moreover, it has been found that parents use sound-symbolic words in their infant-directed speech more often than in adult-to-adult conversations (Imai et al., 2008).

With the use of experiments where participants learn novel nonsense words for abstract shapes, it has been shown that participants are better able to learn and reproduce the right words if these words are matched with the shapes in a way that is congruent with a known sound-symbolic bias (Nielsen & Rendall, 2012). Sound-symbolic mappings in language have been connected to cross-modal mappings in the human brain (Ramachandran & Hubbard, 2001; Simner, Cuskley, & Kirby, 2010). There appear to be many cognitive biases in cross-modal perception that are shared by humans. The bouba/kiki effect is one famous example that shows a strong preference to relate sharp shapes to the name "kiki" (or "takete") and round shapes to the name "bouba" (or "baluma") (Ramachandran & Hubbard, 2001). Many mappings have been investigated and identified, especially in the visual-auditory domain (Hubbard, 1996; Ward, Huckstep, & Tsakanikos, 2006), but also

for instance relating taste to speech sounds (Simner et al., 2010). Such shared biases have been argued not only to aid language processing and acquisition (Perniss et al., 2010) but also to play an important role in the evolution of language by forming a starting point for the initial emergence of grounded speech (Ramachandran & Hubbard, 2001).

On the other hand, some studies show that iconicity does not always convey a learning or processing advantage. For instance, very young children have more difficulty interpreting some iconic mappings (Tolar, Lederberg, Gokhale, & Tomasello, 2008) and arbitrary mappings have the advantage when acquiring word meanings in context (Monaghan, Christiansen, & Fitneva, 2011). Another example is that in tip-of-the-finger (the sign language analogue to tip-of-the-tongue) experiences signers do not necessarily remember the most iconic part of a sign first. For instance, Thompson, Emmorey, and Gollan (2005) describe how the sign for Switzerland in American Sign Language has a movement that depicts the cross of the Swiss flag, but this part of the sign was not more likely to be remembered at first than other, non-iconic, dimensions.

The objective of the experiment described here is to investigate how the (potentially iconic) relation between form and meaning influences the emergence of combinatorial structure. Two conditions were studied: one in which the use of iconic form-meaning mappings is possible and one in which the use of iconic form-meaning mappings is experimentally made impossible. This is expected to provide insights into the possible role of iconicity in the emergence of combinatorial structure since it may reveal whether in a situation that allows for more iconicity the emergence of combinatorial structure may be delayed.[2]

The aim of the experiment is to study the development of sets of acoustic signals that are associated with meanings and transmitted from person to person. Participants learn and reproduce signal-meaning pairs. The signals a person is exposed to are the reproductions of earlier participants in the experiment. This creates what is called a transmission chain (Kirby et al., 2008; Smith, Kalish, Griffiths, & Lewandowsky, 2008). This method is used because it provides a good model of cultural transmission, which has been shown to play an important role in the emergence of structure (Christiansen & Kirby, 2003; Kirby, Dowman, & Griffiths, 2007; Kirby & Hurford, 2002; Zuidema, 2003). The experiment uses images that present unfamiliar objects that have as little obvious structure as possible in order to prevent participants from using culturally learned conventions that may be associated with more familiar stimuli (such as when using pictures of animals, for example). The signals are acoustic, so the participants are required to make a mapping between visual meanings and acoustic signals, similar to what is needed when using spoken language. However, the signals are produced with a slide whistle (see Fig. 1) in order to minimize the influence of existing (shared) linguistic knowledge.

The work is related to an experiment described by Roberts, Lewandowski, and Galantucci (2015) in which it is investigated whether combinatorial structure is influenced by (a lack of) iconicity. They find that iconicity results in signals with less combinatorial structure. However, their work differs from that presented here in that it uses social coordination (negotiation of communicative conventions through repeated interaction

Fig. 1. Plastic slide whistle from the brand Grover Trophy.

between two participants) rather than transmission from generation to generation. Social coordination is an important factor in the emergence of language, but it tends to favor rapid conventionalization and simplification of individual signs (Garrod, Fay, Rogers, Walker, & Swoboda, 2010). In the absence of generation turnover, dyadic interaction leads to greater communicative success, but it does not necessarily result in the emergence of system-wide structure (Kirby, Tamariz, Cornish, & Smith, 2015). Transmission from generation to generation therefore seemed more suitable for our purpose as a mechanism for studying the emergence of combinatorial structure. Also, Roberts et al.'s (2015) work makes use of a graphic signaling system (Galantucci, 2005). Although this signaling system makes it impossible to use existing symbols or drawings, nevertheless both signals and meanings in their experiment exist in the same (the visual) modality. As mentioned before, here we investigate mappings in which the signals are acoustic and meanings are visual.

2. Methods

Participants are asked to learn and reproduce whistled signals with a slide whistle. These signals are presented as names for objects they see on a computer screen. There were 12 whistled signals in the training set in total. The meanings in this study are part of a set of unusual objects that look like possible mechanical parts, but that are novel objects for which there are no conventional names in existing languages. This helped to prevent people from mimicking characteristics of the words they know for the objects, for instance the syllable structure, in the whistled signals. The objects were selected as a subset of those created by Smith, Smith, and Blythe (2011) and were slightly modified to reduce the structure in the meaning space: All objects are colored blue (transformed with a blue filter) and can therefore not be grouped by their color. They also do not share shapes or parts and are not structured in any other obvious way. The meanings themselves have structure in the sense that they are complex objects with sometimes many different parts, but what is meant here is that there is no systematic structure between the items in the set, making it difficult to identify similarities or group items in the set into categories. Since this experiment attempts to investigate the emergence of sub-lexical combinatorial structure, the recombination of meaningless sounds into words, a meaning space with minimal structure is desirable. A few examples of objects that were used are shown in Fig. 2.

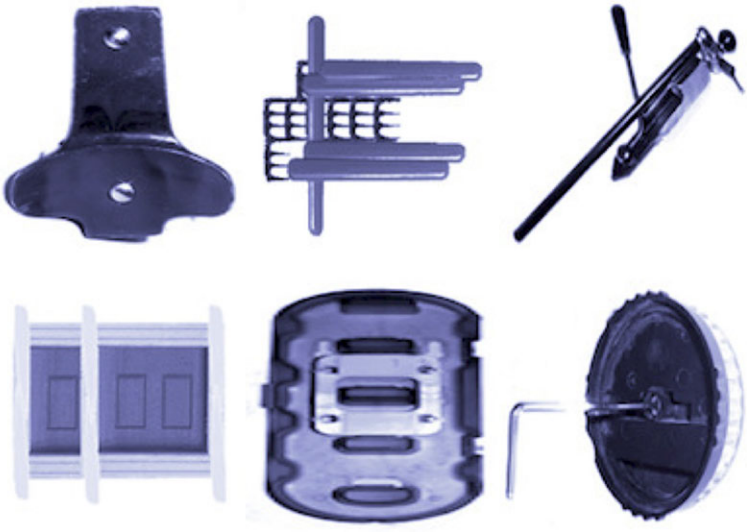

Fig. 2. Examples of novel objects used in the experiment. These objects were created by Smith et al. (2011) and were slightly modified. To reduce potential categorization according to colors in the meaning space, all objects are in blue tone (transformed with a blue filter).

2.1. Procedure

Participants were told they had to learn an alien language: 12 words for alien space ship parts. The words of this language were produced with the use of a slide whistle. Instructions on the task were given both in spoken and written form and there was time for participants to ask questions in case anything was not yet clear. The written instructions can be found in the supplementary material, section S.1. Before the actual experiment started, participants signed an informed consent form and completed a background questionnaire. After this, they were given some time to practice using the slide whistle. During the experiment, they completed three rounds of learning and recall. The first two learning phases were also followed by a guessing game phase before the recall phase. In the learning phase the objects and their corresponding whistle were presented one by one in a random order, and participants recorded an imitation of the whistle. In the recall phase a panel was shown with a button for each object and the participant had to choose each of the objects once to record the right whistle for it from memory. In the guessing phase the whistles were played one by one in a random order and for each whistle the participant had to choose the right object from a panel. This was done with half of the whistle-object pairs after the first learning phase and with the other half after the second. The guessing phase was meant to encourage people to keep paying attention to the mapping between whistle sounds to objects. After the last recall phase, participants were asked to complete a post-participation questionnaire and there was a debriefing.

The whistles from the last recall phase were used as training input for the next participant. The sounds were first normalized to have the same intensity value in order to prevent large differences in loudness in the sounds participants were exposed to. There were two different conditions in the experiment. In the "intact" condition, the next participant is exposed to the output of the previous participant exactly as it was produced. The mapping from signals to objects is kept intact. In the "scrambled" condition, the output of the previous participant is altered before it is given to the next person. The produced form-meaning mappings are broken down by replacing the set of objects and randomly pairing the produced whistles to these new objects between consecutive generations. In this way, if any iconic relations were to emerge in the sets, they would only be helpful for the participants in the intact condition. For the scrambled condition, any semantics-related structure is broken down in between the transmission steps and only the signal sets stay intact. Fig. 3 illustrates the two conditions.

Transmission continued from person to person until there were eight generations in each chain and four chains per condition. The entire procedure took place inside a soundproof booth and it took approximately 60 min in total. In section S.2 of the supplementary material, a screenshot of the user interface that was used for this experiment is shown.

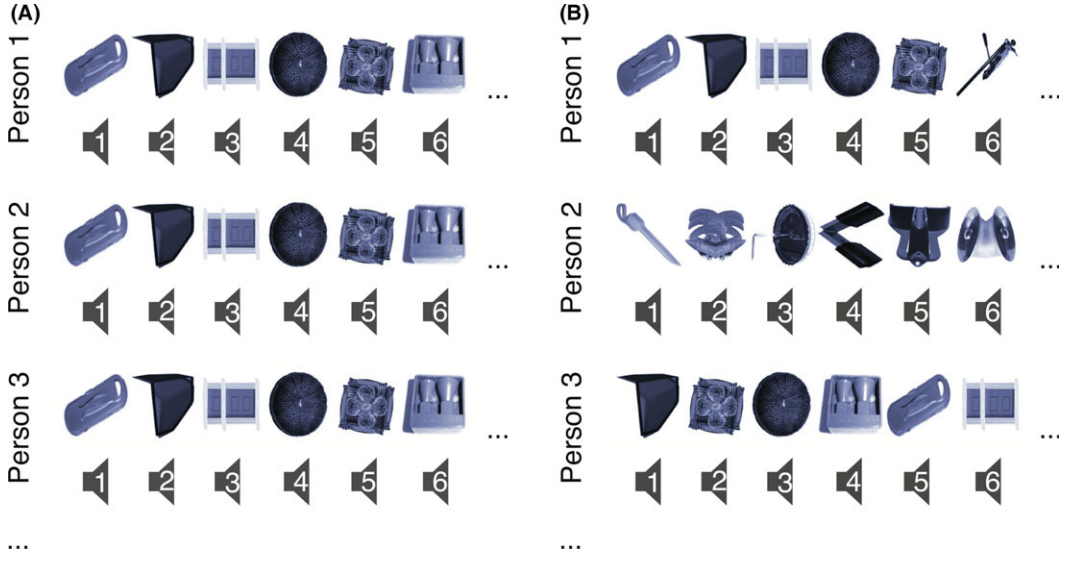

Fig. 3. In the intact condition (A) the next person in a chain is exposed to the exact pairs of whistles and objects that the previous person created. In the scrambled condition (B) the next person in a chain is exposed to the exact set of whistles that the previous person created, but from one person to the other the set of objects is replaced and the whistles are randomly paired with the objects. Two sets of 12 objects were alternated and each was used every other generation so that the odd-numbered generations saw one set, and the even-numbered generations the other set.

2.2. Initial input sets

Two separate initial whistle sets were constructed. Each set was used as the starting point for half of the chains in each condition. The whistles were taken from a database of whistles that were collected during a pilot experiment. These whistle sounds were created by people who were asked to freely record a number of whistle sounds. In this way, a large database of different whistle sounds was created. The two initial sets were constructed so that they would exhibit minimal combinatorial structure, determined using the entropy measure for quantifying combinatorial structure (see section 2.5 on measures below). Sets of 12 whistles were generated randomly from the database until two sets were found with no overlap, and which had a comparable and relatively high measured entropy (4.18 and 4.28). Fig. 4 shows the two sets of 12 whistles plotted as pitch tracks on a semitone scale using Praat (Boersma & Weenink, 2013).

2.3. Reproduction constraint

Experiments that involve iterated learning without a pressure for expressivity tend to result in systems of signals with under-specification, in which the same word is used for many different meanings (Kirby et al., 2008). To prevent this from happening here, a reproduction constraint was used. When a participant produced a whistle for an object that was too similar to another whistle that had already been produced for another object, the participant was told that this whistle had already been produced and was asked to redo the recording. Because participants tend to remember whistles in terms of the movement they make with the whistle plunger, the whistles were compared using a distance measure based on plunger position reconstructed from the recorded sound. The distance measure was a linear combination of different measures, combined as follows:

$$D_{tot} = 0.3 \times D_m + 0.6 \times D_{md} + 0.2 \times D_i + 0.05 \times D_d \tag{1}$$

where D_m is the Dynamic Time Warping (DTW) (Sakoe & Chiba, 1978) distance between the two plunger position tracks which were computed from the pitch tracks using:

$$l = \frac{c}{4f} \tag{2}$$

where l is the plunger position (in cm), c is the speed of sound at body temperature (35,000 cm/s), and f is the frequency measured in Hertz. D_{md} is the Dynamic Time Warping distance between the derivatives (Keogh & Pazzani, 2001) of the movement tracks, D_i is the DTW distance between the two intensity tracks, D_d is the difference in duration, computed using the following equation

$$D_d = \frac{|\log(d_1/d_2)|}{\log(d_1 + d_2)} \tag{3}$$

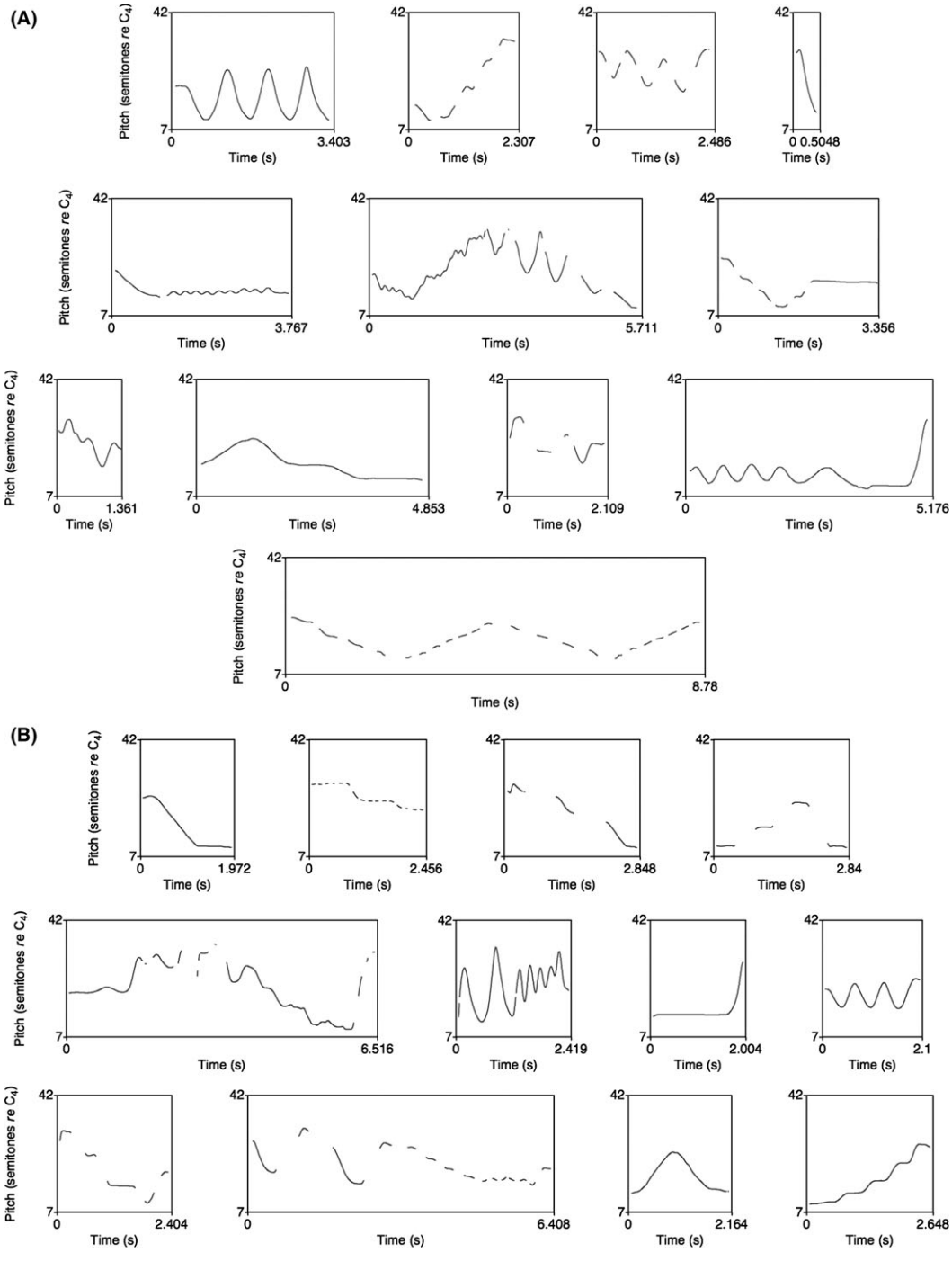

Fig. 4. The initial whistle sets used in the experiment, plotted as pitch tracks on a semitone scale.

where d_1 and d_2 are the lengths of the sampled movement tracks (at 500 samples per second).

Data collected in the pilot study were used to create this measure and to determine the weights on each of the separate parts. The participants in the pilot all imitated 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 measure was therefore based on human judgment of what should be considered the same whistle. The distance below which two whistles were considered the same was set at a relatively low value (0.02). In this way, participants could still produce relatively similar whistles and it would not influence the outcome of the recall phase in any way other than to reject doubles. This was effective, since after all data were collected, 70% of all participants were never asked to redo their recording and on average it happened only 0.6 times per participant within the entire duration of the experiment. This prevented the initial introduction of accidental doubles well enough to prevent the emergence of systems in which the same signal is used several times and variation was preserved much better than without the constraint. In earlier pilots we did with no constraint, the final whistle set often showed the reuse of the same whistle up to five times in the same set and most whistles were used at least twice. This is not the case in the results presented below with the constraint in place.

2.4. Participants

In total, 64 participants took part in the experiment. They were divided over eight transmission chains, four in each condition. Participants were recruited from the University of Amsterdam community through posters and e-mail invitations. All participants were between the ages of 19 and 41 years old; 43 were female and 21 male. In each chain either two or three men participated. They were compensated for their time with a cash payment of 10 euros.

2.5. Measures

In order to evaluate the outcome of the experiment, we defined quantitative measures to test whether the transmitted systems become more learnable (measured as a decrease of recall error) and more structured. The recall error was calculated as the sum of the derivative DTW distances (Keogh & Pazzani, 2001) of all signals, comparing the reconstructed plunger movements of input and output whistles that referred to the same meaning. The use of DTW helps to compensate for small errors in timing. Worse recall will result in larger distances. Here, the DTW distance between two sequences was computed using the original method described in Sakoe and Chiba (1978), using their step pattern Symmetric P1. For the computation of derivative DTW, the same implementation for DTW was used, but the input signals were the derivatives of the signals computed in the way described by Keogh and Pazzani (2001). The signals all had different durations, so in order to normalize for the differences in the lengths of the signals, the DTW

distance was divided by the sum of the lengths of the signals as in Sakoe and Chiba (1978). More details about the implementation and data pre-processing are given in the supplementary material section S.4.

The measure of combinatorial structure makes use of the notion of entropy (Shannon, 1948) from information theory and is based on the idea that a set with more combinatorial structure is composed of fewer basic building blocks that are more widely reused and combined. Such sets are more compressible and therefore have lower information entropy. To compute entropy for a set of whistles, the whistles were divided into segments. Then, using all segments that occur in the set of 12 whistles, (average-linkage) agglomerative hierarchical clustering (Duda, Hart, & Stork, 2001) was used to group together those segments that were so similar (according to the distance measure described above) 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. Shannon's (1948) information entropy was then used to compute entropy:

$$H = -\sum_i p_i \log p_i \tag{4}$$

where p_i is the probability of occurrence of building block i. Note that the entropy measure gives a lower value in case of more structure.

There are several different ways in which the signals can be segmented to describe the discretization of the signal space. Segments can, for instance, be separated by short silences (pauses in the air stream). In this case, pauses are reliable indicators for where one segment ends and another one begins. However, not all sets have many whistled signals with pauses; sometimes they are unbroken movements that differ from one another only in the number of falling and rising parts. Here, changes in plunger movement direction would be better segment boundaries. It would be too subjective to determine the segment boundaries for each whistle set by hand; therefore, three separate types of segmentation were implemented. Each of these was applied to each set of whistles (corresponding to a generation in a chain), and the resulting entropy values were compared to determine which one would most likely have been the right one for a particular signal set. The idea behind this is that each different segmentation type assumes the existence of a system, where the way signals get segmented is a rule that presumably also plays a role when people process the data. Computing the entropy value of several different possible ways of segmenting allows us to better approximate what rule people may have been using. The set of segments that resulted in the shortest description of the whole signal set was assumed to be the best approximation. Therefore, the segmentation type that resulted in the lowest entropy value was selected. The first type of segmentation used the silences as segment boundaries. The second type used the minima and maxima in the plunger movement track as segment boundaries and the third used the points of maximal velocity. More details about segmenting can be found in supplementary material section S.4.

3. Results

This section describes both a quantitative and qualitative analysis of the development of learnability, structure, and iconicity across the data in all chains. First, the learnability is investigated by computing how well participants were able to recall the set of whistle-object pairs they had to remember. Then, the development of combinatorial structure is measured and compared over generations. Finally, the role of iconicity is assessed. How the quantitative measures relate to what is going on in the emerging whistle sets is illustrated with qualitative observations from different transmission chains. Section S.3 of the supplementary material contains the complete transmission chains that resulted from this experiment with whistle signals plotted as pitch tracks. Details on the implementation of the analysis and the signal preprocessing steps can be found in supplementary material section S.4.

3.1. Recall error

To measure whether the sets of whistle-object pairs became easier to learn and reproduce, the measure for recall error was used (as defined in section 2.5). Fig. 5 shows the values for this measure of recall error for the four chains in both conditions, with increasing generations on the horizontal axis. The mean over the four chains for each condition is plotted with the standard errors. A significant decrease in recall error was measured using Page's (1963) trend test for the intact condition ($L = 729$, $m = 4$, $n = 8$, $p < .01$) as well as for the scrambled condition ($L = 738$, $m = 4$, $n = 8$, $p < .01$), which means that there is an increase in the learnability and reproducibility of the form-meaning pairs over generations in both conditions.

One perhaps surprising finding in the comparison between the two conditions may be that the average recall error seems to be higher at each generation for the intact condition. A linear trend analysis of variance on the recall error for exact pairs with generation and condition as factors in a 2×8 mixed design ANOVA reveals that there is indeed a main effect of condition, $F(1, 48) = 19.53$ ($p = 5.63 \times 10^{-5}$), as well as a main effect of generation, $F(7, 48) = 2.35$ ($p = .037$) (in accordance with the result of Page's trend test) and no interaction between generation and condition. A post hoc Tukey's HSD test showed that recall error is significantly higher in the intact condition across generations as compared to the scrambled condition. Given the expectation that iconicity would lead to more transparent and more learnable systems, one might expect to see the reversed pattern. Perhaps the iconic signals are in general more complex and therefore harder to reproduce precisely. Previous experiments using a Pictionary game interaction task (e.g., Fay, Garrod, & Roberts, 2008) have revealed that through repeated interactions, signals become both less iconic and lose complexity. Looking at the average signal duration and the average number of up and down movements in the signals, no difference between the two conditions could be found, though. This issue will be addressed in more depth in the discussion section.

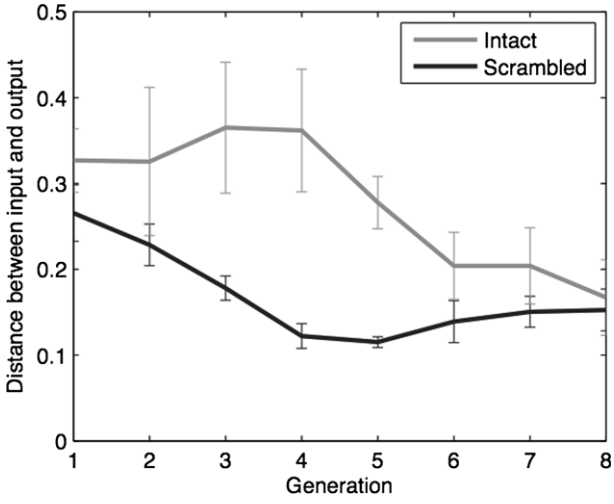

Fig. 5. Recall error over generations in both conditions, showing the mean and standard error. Recall error decreases significantly in both conditions.

3.2. Combinatorial structure

To investigate whether the sets of whistles gradually become more structured after a number of transmissions, the entropy measure was applied to the current data. Fig. 6 shows the development of entropy for the four chains in both conditions, where 0 refers to the initial whistle set. Again, the mean over the four chains for each condition is plotted with the standard error. The significance of the decrease in entropy was established using Page's (1963) trend test for the intact condition ($L = 728$, $m = 4$, $n = 8$, $p < .01$), excluding the artificially inserted initial set (with this set included it is also significant, $L = 992$, $m = 4$, $n = 9$, $p < .05$), as well as for the scrambled condition ($L = 712$, $m = 4$, $n = 8$, $p < .05$), excluding the artificially inserted initial set (with this set included it is also significant, $L = 1{,}033$, $m = 4$, $n = 9$, $p < .001$). These findings imply that the process of iterated learning in both conditions caused structure to emerge. Independent of the objects to which the whistles refer, there is an increase of structure and predictability and the whistles become internally more efficiently coded.

Looking at examples from individual chains, it can be observed how such structure develops. Whistles were introduced that were clearly related in some way to the form of whistles that already existed in the set. For instance, mirrored versions (flipped vertically), combinations of existing whistles, repetitions of the same pattern within a whistle, or whistles with similar shapes but different whistle manners (e.g., smooth vs. staccato) appeared.

Fig. 7 shows an example from one of the chains in the intact condition. In this example, one whistle from generation three seems to be used as an example for two new whistles in the next generation: one with one "bump" and another with two. In generation five

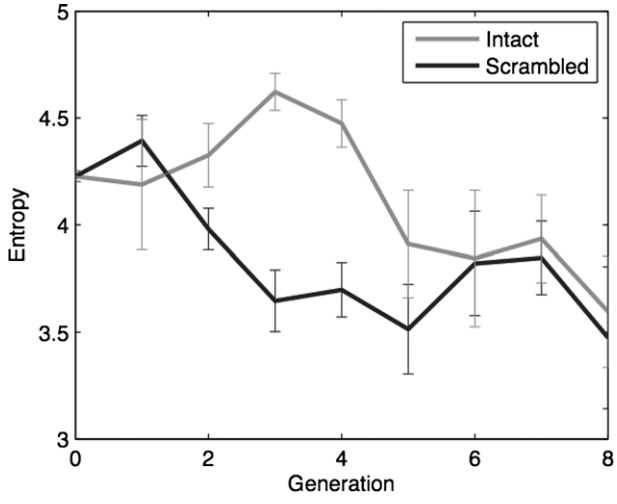

Fig. 6. Entropy of the whistle sets over generations in both conditions, showing the mean and standard error. Entropy decreases significantly in both conditions. This suggests that the combinatorial structure increased over generations.

the "two-bump" whistle starts to be reused and combined with another pattern, and in generation six both the one-bump and two-bump whistles are being reused, mirrored (flipped vertically), and recombined more widely. An existing whistle with several up and down movements is even segmented into two parts, where the first part is again the two-bump whistle.

To examine the final result of these gradual changes in the chains, we can look at the set of whistles produced by the eighth and last participant in a chain. Fig. 8 shows a fragment of such a set from the scrambled condition and here we can identify a clear combinatorial structure. There is a set of building blocks (short level notes, falling-rising slides, rising-falling slides and falling or rising slides) and these are reused and combined in a systematic way to create the whistles in the set. For some of the whistles, there is another version that is mirrored vertically and a pattern of short notes of alternating pitch height seems to be a recurring theme. The set has become very constrained as well, for instance, in terms of the complexity of the falling-rising patterns and the overall variation in the type of building blocks that are left.

From these examples, it becomes clear that conventionalized rules and systems emerged as the whistle sets were transmitted. This fact is corroborated when looking at the development of the number of segments per whistle in each set over generations. The standard deviation of the number of segments in each set of whistles significantly decreases over generations in both the intact ($L = 1{,}024$, $m = 4$, $n = 9$, $p < .01$) and mixed ($L = 1{,}023$, $m = 4$, $n = 9$, $p < .01$) condition. This shows that the whistles within a set become less varied, perhaps more similar and more uniform over time. This is not due to a simple overall reduction in number of segments. When looking at either the median or mean number of segments per set, this does not significantly decrease over

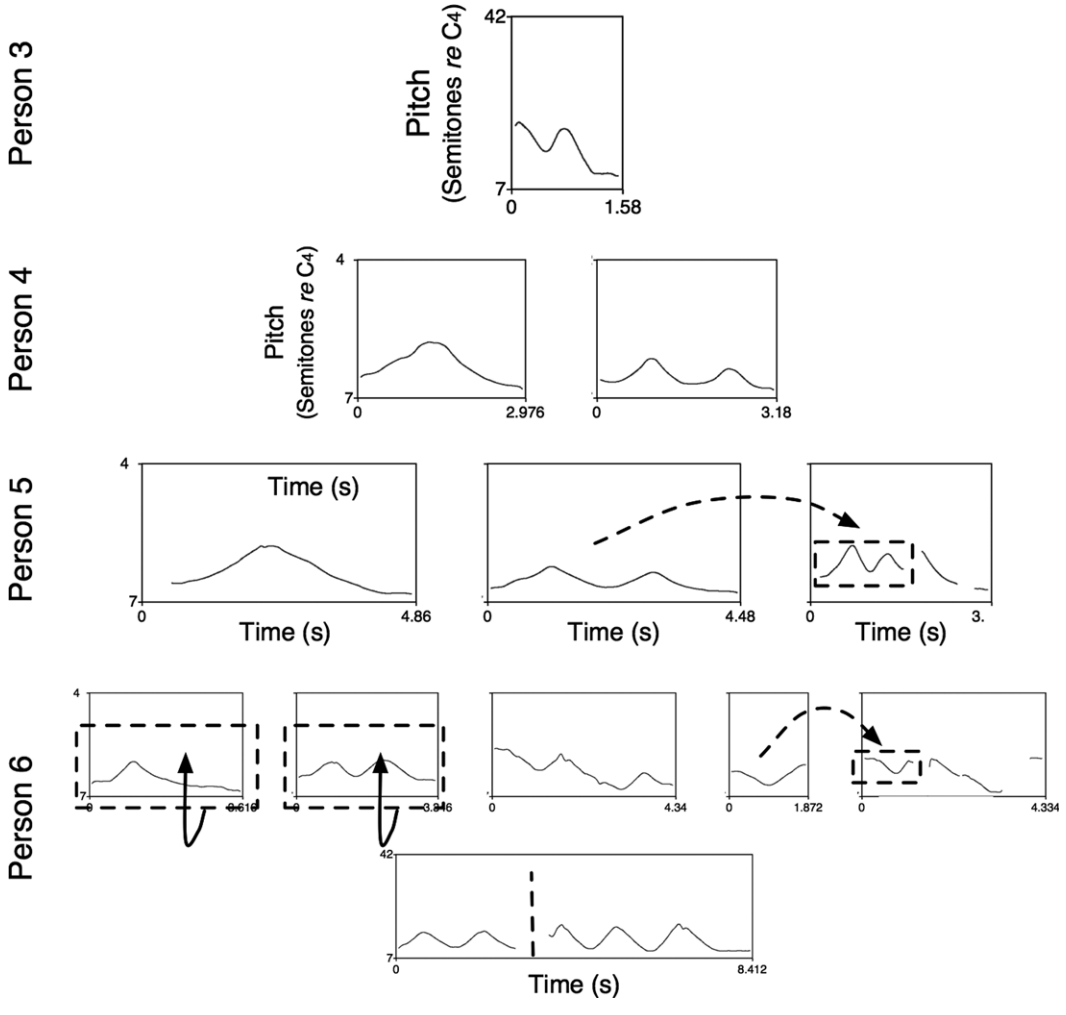

Fig. 7. Development of structure in a chain from the intact condition. The whistle on the first row seems to be an example for two new whistles in the next generation: one with one "bump" and another with two. The "two-bump" whistle is starting to be reused and combined with another pattern, and in generation six both the one-bump and two-bump whistles are being reused (appearing more than once in the set), mirrored (flipped vertically, indicated with reflexive arrows), and recombined (to occur in combination with other patterns) more widely.

generations, both in the intact ($L = 855$, $m = 4$, $n = 9$, $p = .86$) and mixed ($L = 921$, $m = 4$, $n = 9$, $p = .3$) condition. Instead, it seems to differ from chain to chain: In some chains, whistles tend to have many segments, in others less, but this seems to become more consistent within each chain.

On the basis of the measures described so far, there does not seem to be a quantitative difference in overall trend between the two conditions. Both the intact and the scrambled condition lead to a gradual increase of structure and more learnable systems toward the

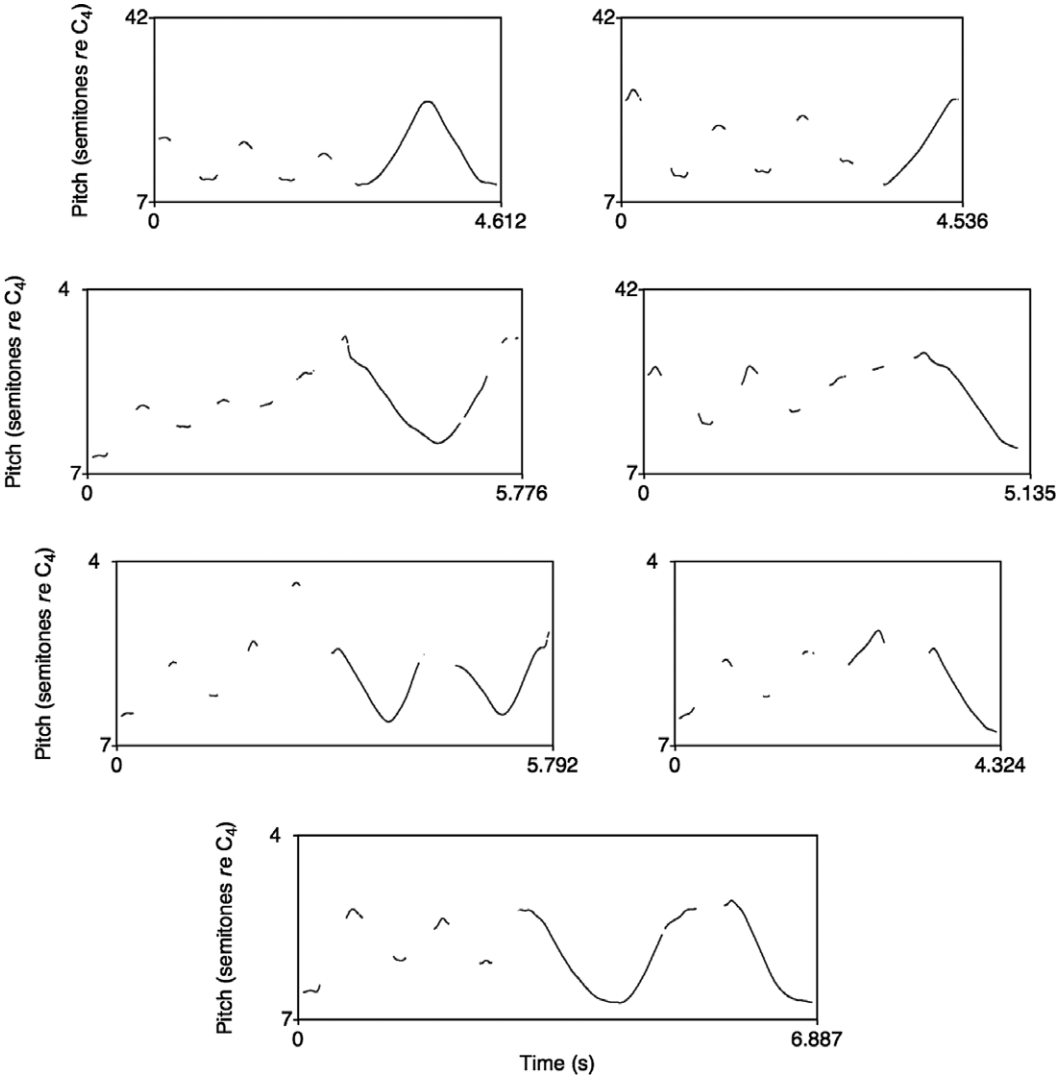

Fig. 8. Fragment from the whistle set produced by the last participant in a chain from the scrambled condition. Basic building blocks can be identified: short level notes, falling-rising slides, rising-falling slides, and falling or rising slides as well as a high-low alternating pattern.

end of the chains. However, there is a difference in the development of the structure in the two conditions. When looking at the development of entropy, we can see that the entropy in Fig. 6 in the intact condition tends to be higher than in the scrambled condition for almost all generations. A linear trend analysis of variance on the entropy with generation and condition as factors in a 2 × 9 mixed design ANOVA shows a main effect of condition, $F(1, 54) = 6.71$ ($p = .012$), as well as a main effect of generation, $F(8, 54) = 2.47$

(p = .023) (in accordance with the result of Page's trend test) and no interaction between generation and condition. This suggests that there is in fact a difference in the entropy between the two conditions. A post hoc Tukey's HSD test showed that entropy is significantly higher in the intact condition across generations as compared to the scrambled condition. Nevertheless, there is neither a significant difference in entropy between intact and mixed for the whistle sets produced by generation 1 (Mann–Whitney-U = 5, n = 4, p = .49), nor for generation 8 (Mann–Whitney-U = 9, n = 4, p = .89). Since both the starting entropy of the chains and the final result of overall decline in entropy are the same in both conditions, this higher entropy in the intact condition seems to indicate a delay in the drop of entropy as compared to the scrambled condition.

3.3. Iconicity

It is difficult to assess the role iconicity played in the two conditions based on an analysis of the signal-meaning mappings themselves, without human judgment. The results of the guessing game phases could indirectly reveal a potential influence. If the mappings were more transparent in the intact condition, we would expect participants in that condition to score higher on the identification task. However, the participants had been exposed to the data before the guessing game phases, since the guessing task only appeared after each learning round. It would therefore be impossible to know whether participants know the meaning because it is transparent, or because they remember it from learning before.

In order to deal with this issue, eight new participants were invited into the lab and asked to rate for each of the whistle-object pairs in all chains and for all generations in the intact condition how well they thought the sound fit with the object. This was expected to reveal whether a possible reduction of iconicity, measured as goodness-of-fit judgments, would coincide with the appearance of combinatorial structure in the condition where iconicity is possible. Overall, there did not seem to be any effect of generation on the degree of iconicity perceived by the participants on average, and when looking at each chain individually, only one out of the four chains showed a significant decrease of rated goodness-of-fit over generations with Page's (1963) trend test (L = 2,879, m = 12, n = 9, $p < .01$). Perhaps more important, intra-rater consistency between the eight raters was very low, as measured with intra-class correlation (Shrout & Fleiss, 1979) (ICC (2,1) = 0.0406). This suggests that what is transparent or iconic may be mostly subjective and experienced differently from person to person.

Given that iconicity may be subjective and depending on individual experience, it is difficult for an outside observer (such as the experimenter) to determine whether iconic structure is being used. However, some examples could be found in the form-meaning pairs in the current data and iconicity could take several different forms in these examples. Most often, the shape of the whistle (the pitch contour) would mimic certain features in the object. This could, for instance, be the overall shape of the object (round shape matched with curvy contour), the orientation of the object (long object placed on diagonal matched with one long falling contour), or the amount or direction of visually distinctive parts on the object (object with a certain number of distinctive parts on top

of each other matched with whistle consisting of a comparable number of sounding parts with rising contour). It should be noted though that these are subjective observations and that it is not necessarily the case that the participants would agree with or would be aware of the structural similarities between whistle and object as described. Fig. 9 shows a few examples of clear iconic form-meaning mappings that were encountered.

In some instances, a clear shift could be observed in the data from iconic holistic signals toward non-iconic signals that became part of the combinatorial system. Fig. 10 shows such an example. In this example, a signal emerges that clearly mimics the shape of the object. This signal is copied by following generations, although not perfectly. At some point a mirrored version of the signal is produced, which is equally iconic. Towards the end of the chain, however, we see that the signal gets altered in such a way that it loses its iconic relation and starts to fit better with the rest of the system that emerged.

Participants filled out a post-participation questionnaire in which they were asked to describe their specific strategy (if any) for remembering the pairs and whether they

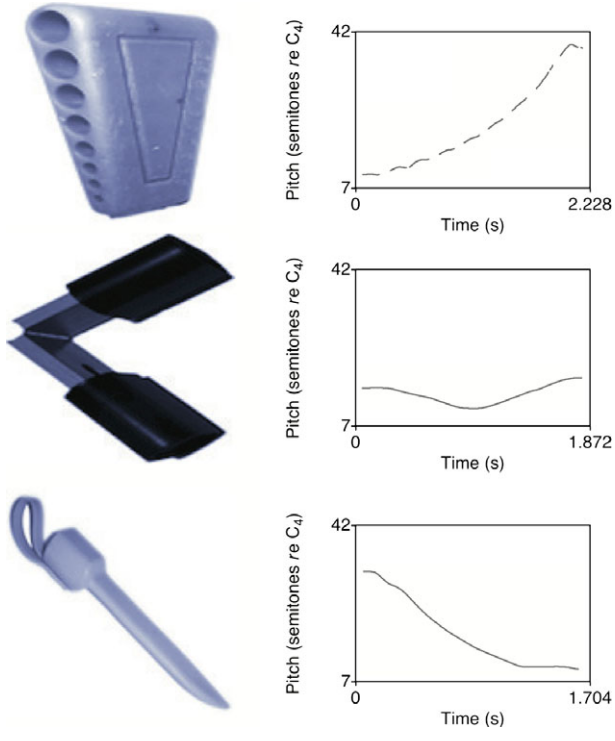

Fig. 9. Examples of iconic whistle-object pairs in the data. The first shows how the holes in the object that are arranged from the bottom to the top and become bigger are iconically depicted as a sequence of notes in a rising pattern. The second shows how the shape of the object is mimicked in the pitch contour. The third shows how the orientation of the object is imitated in the pitch contour.

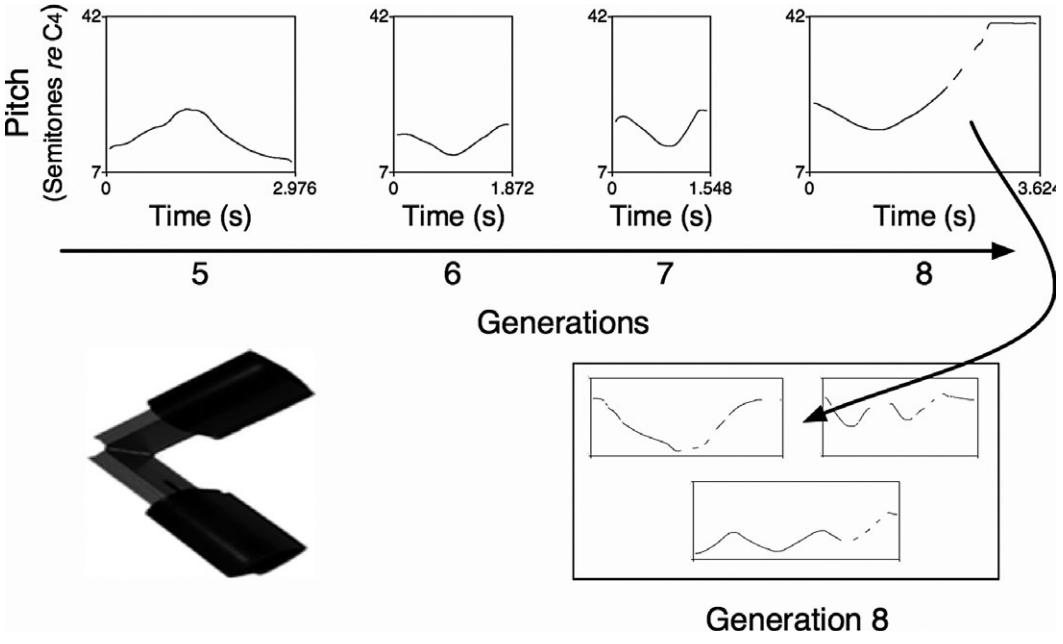

Fig. 10. An example of iconicity that is lost over generations. In generation 5, a signal emerges that clearly mimics the shape of the object (a V-shape). This signal is copied by following generations, although not perfectly: a mirrored version of the signal is produced, which is equally iconic. Toward the end of the chain, however, we see that the signal gets altered in such a way that it loses its iconic relation and starts to fit better with the rest of the system that emerged, in which most signals contain a staccato-like part, as shown in some of the other examples from generation 8.

thought the whistles and objects fit well together. Often participants reported strategies in line with the observations described in the previous paragraph. Other strategies that were reported involved:

1. imagining how the object would sound and linking this with the whistle.
2. imagining how the object would move and linking the pitch contour with that.
3. linking the object with some real object they know and linking the whistle with the sound that object would make.

These reports further illustrate the subjectivity of form-meaning resemblance, at least in the context of the meanings and signals used in this study. The fact that people seemed to be using lots of different strategies for mapping form to meaning does not mean iconicity did not at all play a role in the intact condition. As we saw before, there is a significant difference in the development of both the recall error and combinatorial structure between the two conditions. The only difference between the two conditions was in the opportunity for the iconic structure to remain; therefore, it is unlikely that any other factor caused the different results.

4. Discussion

In the work presented here, whistled signals evolve from holistic, unstructured signals to structured signals that reuse a limited set of building blocks. These building blocks are used in different combinations in the 12 whistles in the set. Each chain appears to use different building blocks and different rules for combining the building blocks. Given that 12 signals is a very small number that does not exhaust the possibilities of the signaling space, this reuse of building blocks appears not to be driven by a pressure to keep signals distinct, as proposed by Hockett (1960). Instead, the emergence of structure appears to be a reflection of a human tendency to find and create structure in sets of signals in order to make them easier to learn. This finding is entirely analogous to the one described in Verhoef et al. (2014), and it therefore appears to be independent of the presence of meaning associated to the signals.

The meanings did influence the emergence of structure: Sets of signals in the intact condition (that is the condition in which the iconic structure could be transmitted from participant to participant) had significantly higher entropy, indicating that there was less combinatorial structure in these sets (although the amount of structure did increase over the generations). In this respect the results are in line with the findings described by Roberts et al. (2015), who found a negative influence of the amount of iconicity on the degree of combinatorial structure in early emerging communication systems. They did not investigate what happened when these systems were transmitted over generations, which is what we did in the current study.

Interestingly, it was also found that recall error was worse in the intact condition than in the scrambled condition, indicating that iconic signal-meaning associations may be learned less well than non-iconic ones. This may at first appear puzzling, as iconic signals are supposedly more transparent and easier to map, but a similar effect has been described by Ortega (2013), in an experiment where second language learners of a sign language were less precise in imitating iconic signs as opposed to arbitrary signs. The recall error we measured in the current experiment only looks at the precise shape of the signals. When the iconic structure is used, participants express a more or less abstract property (shape, size, texture etc.) of the meaning with their signal, and as long as the correct association between this property and the meaning is preserved, the precise realization of the signal is of less importance. This may result in signals that are more different according to the distance measure, but that are perceived as more similar. An example that illustrates this can be found in Fig. 10, where an iconic signal is realized in two different ways. Both signals mimic the shape of the object in the exact same way and both are equally iconic, but one of the signals goes up first and then down while the other goes down first and then up, resulting in two dissimilar signals. In addition, for iconic signals, language users need to not only agree and find alignment on what features of the signals are important and which variations are relevant, but also on which properties of the meaning are in focus when mapping form to meaning. In Al-Sayyid Bedouin Sign Language (Sandler et al., 2011), for instance, this seems to still be in progress, where different signers may use

different iconic signs for the same meaning. These signs are clearly iconic and focus on different properties of the expressed meaning—for example, in the sign for lemon either the act of squeezing it or the experience of sourness is expressed. This is related to our observation that in the experiment the perception and expression of the iconic structure is highly dependent upon the individual who uses it, especially with our set of meanings where there are no clear, culturally established conventions about what the salient properties of an object are. Although iconicity does appear to delay the emergence of structure, this appears to be only a transient effect: Combinatorial structure still emerges over only a very limited number of transmission/learning events.

The research presented here suggests that structured sets of signals will appear when a signaling system is repeatedly learned and transmitted, even when there is a possibility for iconic structure. Apparently in some domains, modern humans' tendency to find structure and to generalize in difficult learning situations sometimes trumps the advantage of using iconic structure. Although the iconic structure was used in this experiment, it did not appear to be a stable signaling strategy, perhaps because the perception and expression of the iconic structure is subjective and depends on the individual. Even though there are many shared (iconic) biases that may guide the emergence of words and structures, there are also strong individual differences in perception and expression of iconic mappings, and therefore it should not be assumed as a given that iconic signals emerge and persist more easily than arbitrary ones. Note that this point of discussion applies mainly to iconicity of the holistic type. Iconicity, of course, can also be part of a systematic and predictable system, for instance when a certain type of iconicity is used consistently for a semantic category or when there is a good mapping between the topologies of the form and meaning spaces (De Boer & Verhoef, 2012). Many stable uses of iconicity in languages show such systematicity between meaning structure and form structure, for instance, in patterned iconicity (Padden et al., 2013) or diagrammatic iconicity (Fischer & Nänny, 1999). In these cases there is not necessarily a tension between iconicity and structure. In the types of iconicity that have been found in iconicity-rich spoken languages, systematicity and regularities are indeed important. As Dingemanse pointed out: "It is the diagrammatic types of (. . .) iconicity that enable ideophones to move beyond the imitation of singular events toward cross-modal associations, perceptual analogies and generalizations of event structure" (Dingemanse, 2012, p. 659). The importance of patterns in the use of iconicity has been recognized for sign languages in particular (Meir, Padden, Aronoff, & Sandler, 2013; Padden et al., 2013). For each referent, there are often many different possible resembling forms, using different types of iconicity. Languages differ in which types they use or prefer and within a language the use of iconic types may be organized beyond simple resemblance. However, due to the lack of structure in the meaning space, the design of the experiment described in this article made the appearance of this kind of iconicity unlikely.

The design of the meaning space is therefore a first example of a design choice that may have influenced the specific results we found in ways that would perhaps make it different from the real origins of combinatorial structure in languages. The images were chosen as not to exhibit clear patterns of meaning (reoccurring features or systematic

differences in size or shape, for instance). This prevented the interference with compositional structure (combining meaningful building blocks into larger meaningful ensembles), and this choice was made for the purely methodological reason to isolate the phenomenon of combinatorial structure (combining meaningless elements of signal into larger meaningful signals) as much as possible. But excluding compositional structure at the same time excludes an important class of iconic structures, as described in the previous paragraph. In the emergence of structure in real language, however, it is likely that combinatorial structure and compositional structure emerge simultaneously and may influence each other.

A second design choice involves the images that were chosen in such a way that the participants did not have (culturally) shared associations with the meanings. This prevented the use of existing words in the language of the participants to be used as common ground for creating signals, for instance by mimicking syllable structure or other characteristics of the words in the whistles. It was thought to be a more realistic model of early language emergence than the use of easily recognizable images that tend to have rich (culturally) shared associations. However, it could be argued that such shared associations may predate or emerge together with the emergence of language.

Third, we used a one-way learning method. Participants were only exposed to a recorded set of output from the previous participant in their diffusion chain. This precluded any interaction between users of the signaling system. However, as mentioned in section 1, interaction has been shown to lead to rapid conventionalization (Garrod et al., 2010) while it is also likely that in interaction, distinctive properties of the signals could become exaggerated (Fay et al., 2008). This could in principle lead to either more structured or to more iconic systems of signals. Here, the choice was made to avoid interaction in order to focus solely on the influence of transmission and to exclude the possibility of explaining the emergence of structure on the basis of conscious creation or invention by single individuals. Structure emerges gradually over multiple generations in diffusion chains.

A last point involves the modality for signal production. Even though the slide whistles provided an easy-to-use tool for creating continuous auditory signals without the interference of previous experience with spoken language, the pairing with the meanings used in this study may not in all cases have elicited very intuitive ways to map iconically. Although there were some examples of clear iconic mappings found in the data (as shown in Fig. 9), most types of iconicity that could be expected to evolve here would need to rely on a high level of abstraction. From previous research, we know that modality and mappability play an important role in the development of iconicity and communicative success (Fay, Arbib, & Garrod, 2013; Verhoef, Roberts, & Dingemanse, 2015), which makes this an important design decision.

5. Conclusion

In the iterated learning experiments presented above, structure emerges (as shown by decrease in entropy) and learnability improves (as shown by decrease in recall error) over

the eight generations of participants. The results are therefore in line with the findings of a similar experiment without meaning (Verhoef, 2012; Verhoef et al., 2011). The meanings associated with the signals in the present experiment were expected to influence the emergence of structure through the possible use of iconicity. Indeed, some influence was found, although it seemed to last only for a short time in the transmission process. In both conditions the end result is the same, combinatorial structure emerges, but the significant difference in measured structure suggests that the route toward structure may differ and seemed to have been delayed in the intact condition where iconicity was possible.

The points discussed in the discussion section on design choices can all be used to define variations on the experiment presented here, and these are therefore paths for future work that we are currently pursuing. This involves, for instance, the use of different modalities, such as gestures, different degrees of interaction between participants, and different sets of form-meaning pairs with varying initial degrees of iconicity and structure.

In summary, we presented a method for studying the role of meaning in the cultural evolution of combinatorial structure in acoustic signals. This method has the potential to be extended in many different ways to shed more light on the emergence and evolution of combinatorial structure, iconic patterns, and the cognitive mechanisms that enable us to use it.

Acknowledgments

We thank Gisela Govaart for helping with data collection and Alex del Giudice, Kenny Smith, Wendy Sandler, and Carol Padden for discussions and suggestions. We are also grateful for the helpful comments we received from the editor and anonymous reviewers. This research was funded by a NWO Rubicon grant to T.V. and ERC project ABACUS (grant number 283435) to B.d.B.

Notes

1 In a similar fashion, signed languages combine basic elements such as handshapes, movements, and locations. In addition to such combinatorial structure, in both modalities meaningful elements (words) are combined into larger complexes (sentences). This second level of recombination (compositional structure) is, however, not the topic of this paper.
2 Preliminary findings of the experiment have been presented in Verhoef, Kirby, and de Boer (2013).

References

Bergen, B. (2004). The psychological reality of phonaesthemes. *Language*, *80*(2), 290–311.

Berwick, R. C., Okanoya, K., Beckers, G. J. L., & Bolhuis, J. J. (2011). Songs to syntax: The linguistics of birdsong. *Trends in Cognitive Sciences*, *15*(3), 113–121.

Boersma, P., & Weenink, D. (2013). *PRAAT: Doing phonetics by computer*. Amsterdam: Universiteit van Amsterdam.

Christiansen, M. H., & Kirby, S. (2003). Language evolution: Consensus and controversies. *Trends in cognitive sciences*, *7*(7), 300–307.

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

Cuskley, C., & Kirby, S. (2013). Synaesthesia, cross-modality, and language evolution. *Oxford Handbook of Synaesthesia*, *20*, 869–907.

De Boer, B., & Verhoef, T. (2012). Language dynamics in structured form and meaning spaces. *Advances in Complex Systems*, *15*(3), 1150021-1–1150021-20.

De Boer, B., & Zuidema, W. (2010). An agent model of combinatorial phonology. *Adaptive Behavior*, *18*(2), 141–154.

Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass*, *6*(10), 654–672.

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

Fay, N., Arbib, M. A., & Garrod, S. (2013). How to bootstrap a human communication system. *Cognitive Science*, *37*(7), 1356–1367.

Fay, N., Garrod, S., & Roberts, L. (2008). The fitness and functionality of culturally evolved communication systems. *Philosophical Transactions of the Royal Society of London B*, *363*, 3553–3561.

Fischer, O., & Nänny, M. (1999). *Introduction: Iconicity as a creative force in language use* (pp. 15–36). Amsterdam: Benjamins.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive science*, *29*(5), 737–767.

Garrod, S., Fay, N., Rogers, S., Walker, B., & Swoboda, N. (2010). Can iterated learning explain the emergence of graphical symbols? *Interaction Studies*, *11*(1), 33–50.

Hinton, L., Nichols, J., & Ohala, J. J. (1994). Introduction: Sound-symbolic processes. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 1–14). Cambridge, England: Cambridge University Press.

Hockett, C. (1960). The origin of speech. *Scientific American*, *203*, 88–111.

Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *The American Journal of Psychology*, *109*(2), 219–238.

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, *109*(1), 54–65.

Keogh, E., & Pazzani, M. (2001). Derivative dynamic time warping. Presented at the The 1st SIAM International Conference on Data Mining (SDM-2001), Chicago (Il).

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., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, *104*(12), 5241–5245.

Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi, & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 121–147). London: Springer.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Martinet, A. (1949). La double articulation linguistique. *Travaux du Cercle Linguistique de Copenhague*, *5*, 30–37.

Meir, I., Padden, C., Aronoff, M., & Sandler, W. (2013). Competing iconicities in the structure of languages. *Cognitive Linguistics*, *24*(2), 309–343.

Mitani, J. C., & Marler, P. (1989). A phonological analysis of male gibbon singing behavior. *Behaviour*, *109*, 20–45.

Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology*, *140*(3), 325.

Nielsen, A., & Rendall, D. (2012). The source and magnitude of sound-symbolic biases in processing artificial word material and their implications for language learning and transmission. *Language and Cognition*, *4*(2), 75–140.

Nowak, M. A., Krakauer, D., & Dress, A. (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. In Eli Fischer-Jørgensen and Nina Thorsen (Eds.), *Proceedings of ICPhS IX* (vol. 3, pp. 181–185). Copenhagen: Institute of Phonetcs.

Ortega, G. (2013). Acquisition of a signed phonological system by hearing adults: The role of sign structure and iconcity. PhD Thesis, University College London, London.

Padden, C., Meir, I., Hwang, S.-O., Lepic, R., Seegers, S., & Sampson, T. (2013). Patterned iconicity in sign language lexicons. *Gesture*, *13*(3), 287–308.

Page, E. (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., & McVay, S. (1971). Songs of humpback whales. *Science*, *173*, 585–597.

Perniss, P., Thompson, R., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, *1*, 227.

Ramachandran, V., & Hubbard, E. (2001). Synaesthesia–a window into perception, thought and language. *Journal of Consciousness Studies*, *8*(12), 3–34.

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.

Sakoe, H., & Chiba, S. (1978). Dynamic programming optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, *26*(1), 43–49.

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

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

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428.

Simner, J., Cuskley, C., & Kirby, S. (2010). What sound does that taste? Cross-modal mappings across gustation and audition. *Perception*, *39*(4), 553.

Smith, K., Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2008). Introduction: Cultural transmission and the evolution of human behaviour. *Philosophical Transactions of the Royal Society of London*, *363*(1509), 3469–3476.

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

Thompson, R., Emmorey, K., & Gollan, T. H. (2005). "Tip of the fingers" experiences by deaf signers: Insights into the organization of a sign-based lexicon. *Psychological Science*, *16*(11), 856–860.

Tolar, T. D., Lederberg, A. R., Gokhale, S., & Tomasello, M. (2008). The development of the ability to recognize the meaning of iconic signs. *Journal of Deaf Studies and Deaf Education*, *13*(2), 225–240.

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. (2013). Combinatorial structure and iconicity in artificial whistled languages. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3669–3674). Austin, TX: Cognitive Science Society.

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

Verhoef, T., Kirby, S., & Padden, C. (2011). Cultural emergence of combinatorial structure in an artificial whistled language. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 483–488). Austin, TX: Cognitive Science Society.

Verhoef, T., Roberts, S. G., & Dingemanse, M. (2015). Emergence of systematic iconicity: Transmission, interaction and analogy. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 2481–2486). Austin, TX: Cognitive Science Society.

Ward, J., Huckstep, B., & Tsakanikos, E. (2006). Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex*, *42*(2), 264–280.

Zuidema, W. (2003). How the poverty of the stimulus solves the poverty of the stimulus. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems 15* (pp. 51–58). Cambridge, MA: MIT Press.

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

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. More information on the experimental set-up, complete transmission chains, and details on the implementation of analyzed measures and signal preprocessing steps