

# Iconicity and Structure in the Emergence of Combinatoriality

Matthias Hofer (mhofer@mit.edu), Roger Levy

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology,  
43 Vassar St, Cambridge, MA 02139 USA

## Abstract

One design feature of human language is its combinatorial phonology, allowing it to form an unbounded set of meaningful utterances from a finite set of building blocks. Recent experiments suggest how this feature can evolve culturally when continuous signals are repeatedly transmitted between generations. Because the building blocks of a combinatorial system lack independent meaning, combinatorial structure appears to be in conflict with iconicity, another property salient in language evolution. To investigate the developmental trajectory of iconicity during the evolution of combinatoriality, we conducted an iterated learning experiment where participants learned auditory signals produced using a virtual slide whistle. We find that iconicity emerges rapidly but is gradually lost over generations as combinatorial structure continues to increase. This suggests that iconicity biases, whose presence was revealed in a signal guessing experiment, manifest in nuanced ways. We discuss implications of these findings for different ideas about how biases for iconicity and combinatoriality interact in language evolution.

**Keywords:** phonology; language evolution; combinatorial structure; iterated learning; iconicity

## Introduction

Combinatorial phonology is an important design feature of human language, allowing it to form an unbounded set of novel, meaning-bearing words from a small set of building blocks. How did it emerge in language? As part of a larger research program that attempts to explain linguistic properties through biases operating during language acquisition and use (Christiansen & Chater, 2016; Kirby, Cornish, & Smith, 2008), recent laboratory experiments have suggested how combinatorial structure could have arisen from continuous signals through a process called iterated learning (Verhoef, Kirby, & de Boer, 2014; Giudice, 2012). But while combinatorial structure might confer a range of advantages to language, it appears to be in conflict with another salient feature of communication systems: iconicity. In order to participate freely as primitives in larger composite forms that carry arbitrary meanings, the building blocks of a combinatorial system should be meaningless (Dingemans, Blasi, Lupyan, Christiansen, & Monaghan, 2015). Iconic signs, on the other hand, are motivated by properties of the meanings they refer to.

Evidence suggests that iconicity plays an important role in bootstrapping communication. In a study where subjects had to develop novel communication systems, Fay, Arbib, and Garrod (2013), found that gesture was preferentially adopted over speech, and explained their findings in terms of gesture's

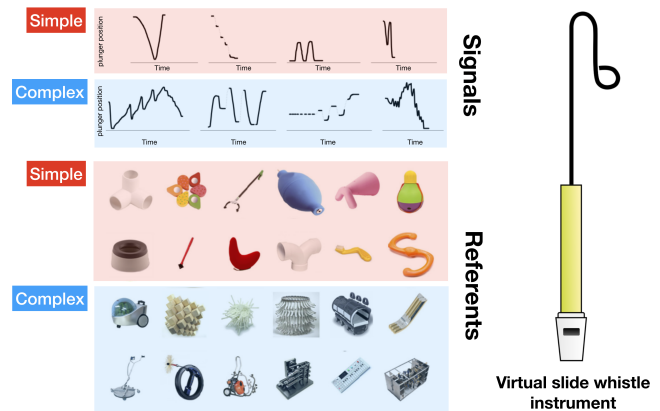


Figure 1: Depiction of stimuli and virtual slide whistle used in the iterated learning experiment to investigate the relation between iconicity and combinatoriality. Visual referent stimuli are from Lewis and Frank (2016).

stronger affinity for iconic representation. On the above account, signals tend to eventually lose these iconic origins as they develop into combinatorial systems. Goldin-Meadow and McNeill (1999) have similarly argued that iconicity is the default strategy and that combinatoriality is not adopted for the benefits it provides but to compensate when iconicity is not available. Consistent with this account, Verhoef, Kirby, and de Boer (2016) found that the onset of combinatorial structure in an iterated learning experiment was delayed when signal/referent mappings were scrambled between generations, making it harder for iconicity to develop, relative to a condition where mappings were kept intact.

Despite the possible loss of iconicity in the emergence of combinatorial phonology, many familiar forms of iconicity such as onomatopoeia or sound symbolism continue to play an important role in language (Dingemans et al., 2015). How do these various forms of iconicity develop as signals undergo their transition from holistic to combinatorial structure? To reconcile the existence of iconicity at different stages of language evolution, we focus on a more subtle form of iconicity, recently described by Lewis and Frank (2016), that exists between word length and conceptual complexity. In their analysis of monosyllabic words across 80 languages, the authors

found that longer words are systematically associated with more complex meanings. Whereas the languages considered by Lewis and Frank (2016) are already fully combinatorial, we examine whether this form of iconicity also arises in continuous signal spaces and use it to address questions about the relationship between iconicity and combinatoriality. To this end, we conducted an iterated learning experiment where subjects evolved a set of signals through iterated reproduction. Participants learned artificial languages consisting of whistled signals that were paired with referents taken from Lewis and Frank (2016). Both signals and referents varied in complexity (Figure 1) but were paired in such a way that there was no systematic relationship between simple and complex items in the beginning. Based on the literature presented above, we predicted that such a relationship, indicative of iconicity, would emerge but eventually disappear as the communication systems become more combinatorial.

Using the languages resulting from the iterated learning study, we present results from a series of experiments designed to answer the following questions:

1. Did the languages evolve combinatoriality? This was assessed by asking subjects to rate the amount of combinatorial structure that existed in the languages.
2. Did the signals evolve iconicity? Iconicity, defined in terms of congruent complexity associations between signals and referents, was measured by collecting complexity judgments for the evolved signals.
3. Which underlying cognitive structures support our inferences about iconicity? Previous studies suggest the existence of strong biases for the development of combinatoriality and iconicity (Lewis & Frank, 2016; Verhoef et al., 2016). To better understand the role of iconicity biases and how they manifest in our experiment, we devised a guessing game where naive listeners were asked to choose the most likely referent for each signal.

After presenting our results, we close by discussing how our findings relate to different ideas about the evolution of iconicity and combinatoriality.

## Experiments

To investigate how iconicity develops during the emergence of combinatorial structure, we conducted an iterated learning experiment. Miniature artificial languages were repeatedly acquired and subsequently transmitted by one ‘generation’ of subjects to the next. This took place across several independent transmission chains. We adopted the signal space used in Verhoef et al. (2014), in which subjects produced signals using a slide whistle instrument. Since we conducted the experiment online, we developed an on-screen, virtual version of the instrument, depicted in Figure 1. Pitch was controlled by moving the plunger up and down using the mouse. Sounds were produced by pressing down the space bar and continued until the space bar was released. Before the experiments

started, participants were given an opportunity to familiarize themselves with this interface. Using the languages that evolved during iterated learning, we subsequently conducted four additional experiments to address the aforementioned questions about the emergence of combinatorial structure and iconicity.

### Iterated learning experiment

The iterated learning experiment consisted of 15 independent chains, each consisting of 10 generations. Per chain and generation, a single subject learned and later reproduced an artificial language. The first subject in each chain was given a language constructed according to principles described below, while subsequent generations learned the language produced by the previous generation.

**Materials** Each language consisted of eight whistled sounds paired with different referents. Figure 1 shows which signals were used to initialize each experimental chain. The signals were obtained from whistles recorded and subsequently rated for their complexity in a pilot experiment. The signals were paired with unfamiliar visual objects selected from a stimulus set used in Lewis and Frank (2016), which was normed for complexity. Referents were categorized as either simple or complex. For each chain, four simple and four complex referents were selected at random from the stimuli depicted in 1 and assigned to signals with the constraint of counterbalancing between signal and referent complexity classes (half of the complex signals were paired with complex referents and while the other half was paired with simple ones and vice versa). This procedure ensured that the relation between signal and referent complexity was initially fully un-systematic.

**Procedure** Subjects were told that they had to learn an artificial language produced using a slide whistle with the goal of teaching the language to a computer program. After familiarizing themselves with the instrument, subjects engaged in five learning blocks, where they were shown each of the eight signal/referent pairs in random order. Each trial first displayed the visual referent, then the slide whistle playing back the corresponding signal. The whistle then stayed on screen and participants were instructed to repeat the signal. No feedback was given during learning. Subjects were admitted to the reproduction phase if they reached a learning criterion to assess how well they learned the language. The criterion test consisted of eight 2-AFC trials. Each of the eight signals was played to subjects once and they had to choose the correct referent from a set of two. The distractor item was sampled from among the remaining three items of the same complexity class, preventing participants to identify the correct referent based on referent complexity alone. No feedback was given during these trials. To advance to the final stage of the experiment, participants had to correctly identify at least six of the eight items. Participants that reached the learning criterion advanced to the reproduction phase, which

was framed as a computer teaching paradigm. Subjects were asked to record each signal for a computer program that will attempt to learn the language from them. All referent stimuli were presented simultaneously on screen and subjects could chose the order in which they recorded signals by clicking on the corresponding item.

While chosen to prevent subjects from producing the same signal multiple times, the framing of the task as a teaching paradigm did not fully prevent a loss of expressivity (see, e.g., Kirby et al., 2008). Throughout the experiment, 7% of signals were identified as duplicates using a dynamic time warping-based similarity measure, and replaced with signal versions produced during learning with the constraint of being sufficiently distinct from the remaining test phase signals. This approach is conservative since signals produced during learning typically very closely resemble the input the participant was given and thus limit the amount of change (relative to the input) experienced by the next participant.

**Subjects** A total of 382 subjects were recruited on Amazon's Mechanical Turk. 250 subjects passed a preliminary headphone check (Woods, Siegel, Traer, & McDermott, 2017), implemented to ensure consistent listening conditions, and were admitted to the main experiment. Of those, 164 subjects reached the learning criterion. Data from 14 subjects was due to technical reasons, leaving us with 150 subjects, one subject per chain and generation.

### Quantification of combinatorial structure

To answer whether signals evolved combinatoriality, we conducted a rating experiment. Naive participants were asked to rate the amount of structure present in the languages from the iterated learning experiment. Participants saw languages from either one of two conditions: In the intact condition, languages were randomly selected from across chains and generations in the iterated learning experiment. Participants were blind to which generation or chain a language came from. In the scrambled condition, participants were shown languages where signals from the different chains of each generation were randomly mixed together. Including this baseline condition allows us to assess to what extent combinatoriality judgments are about properties of the signals in the context of the language they evolved in, or simply about the structure of signals irrespective of their relation to the other signals in their language.

**Subjects and Procedure** Subjects were told that they had to rate the amount of structure of different newly discovered whistle languages. Structure was described as the existence of building blocks or principles that are shared among the signals in a language. Signals were presented visually in a presentation format similar to Figure 3 (but in a single row). This allowed subjects to make holistic judgments and facilitated comparisons between items in the language. Subjects were asked to report how structured a given language was, ranging from least to most structured, using a continuous slider. A to-

tal of 314 subjects took part in the rating experiment and each participant rated 24 items.

### Quantification of signal complexity and iconicity

To address our second question, it was necessary to quantify signal complexity in order to assess if signals developed to match the conceptual complexity of their referents. Two experiments were conducted, one where signals were presented visually and one where they were presented auditorily.

**Visual complexity ranking** Similar to the previous experiment, languages were presented in the form of a visual array. Subjects were instructed to sort the signals from least to most complex. Complexity was defined as signals that have many parts and that are difficult to memorize or reproduce. 374 subjects took part in this part of the experiment and each subject rated 16 items. After realizing that effect sizes of the iconicity measure are likely too small and that the noise introduced from using a perceptual modality different from the original, auditory modality could potentially mask important differences, we conducted a second rating experiment.

**Auditory complexity rating experiment** Focusing on just the first five generations of iterated learning, in the second complexity rating experiment, signals were presented similar to the main experiment, with the slide whistle playing back the recorded signals. Signals were randomly selected from across chains and generations, which enabled us to obtain absolute complexity judgments (compared to the rank-level judgments obtained in the visual experiment). Subjects judged the complexity of each signal from least to most complex using a slider. 175 subjects took part in the experiment and each subject rated 16 items.

### Evaluation of signal iconicity

Finally, do people exhibit iconicity biases that explain their productions during iterated learning? To develop further insight into the nature of the biases that support iconicity, we conducted a guessing game where subjects were presented all eight signals from a language (in random order) and had to identify the most likely referent per signal. Subjects were instructed that they should always choose the referent that they thought most likely belonged to the signal, and that the same referent could be chosen more than once. This allowed us assess the existence, and to quantify the strength of iconicity biases that exist in signal interpretation. 218 subjects took part in the experiment and each subject rated 8 languages.

## Results

### Emergence of combinatorial structure

The first question we address in our analysis is whether languages developed combinatorial structure, as described in prior experiments (e.g., Verhoef et al., 2014; Giudice, 2012). In summary, we observe that signals in all chains develop combinatorial structure. Figure 3 depicts a representative example language from the final generation of chain 1, giving

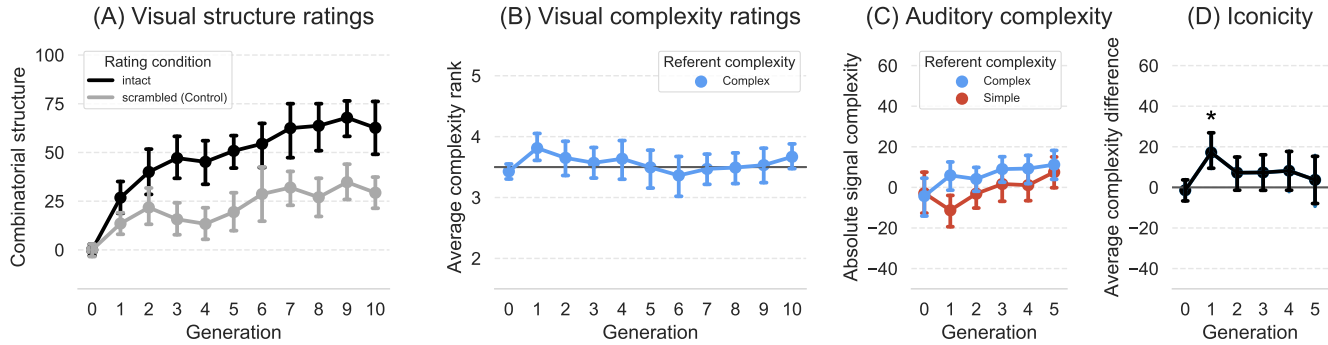


Figure 2: Results from the three experiments that we conducted to assess signal structure and complexity (visual and auditory), and the derived iconicity measure. Error bars are 95% confidence intervals.

a qualitative impression of the emergence of shared building blocks. Figure 2A shows the results from our combinatoriality measure. Across both the intact ( $t(14) = 8.06, p < 0.001$ ) as well as the scrambled condition ( $t(14) = 5.59, p < 0.001$ ), languages are judged to increase in structure over generations<sup>1</sup>, but languages in the intact condition are judged to increase more ( $t(14) = 2.5, p = 0.02$ ). This difference can only be explained by assuming that signals in the intact condition are structurally more similar to each other, which results in higher combinatoriality ratings compared to mixing languages across chains as in the scrambled condition.

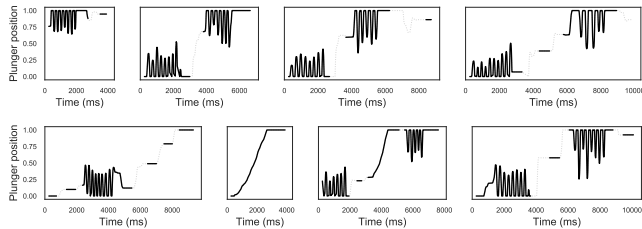


Figure 3: Set of stimuli from the last generation of chain 1. The language appears to consist of three building blocks: a short beep occurring at different pitch values, a long pitch sweep, and a wiggly segment of alternating pitches.

### Development of iconic signal structure

Turning to our second question, Figure 2B depicts the results of the visual complexity ranking experiment where participants were asked to order signals from least to most complex (coded as 0 to 7). The depicted mean complexity rank represents the average rank of the four signals that were associated with complex referents. An at chance association between signals and referents corresponds to a mean rank of 3.5. While nearly all chains in the first generation have an average complexity rank of greater than 3.5, quantitatively this difference does not reach significance after correcting for multiple

<sup>1</sup>Analyses compare the regression coefficients fitted to the fifteen chains to a zero slope.

comparisons ( $t(14) = 2.60, p = 0.02$  before,  $p = 0.21$  after *Holm-Bonferroni* correction). Two features of the experimental measure could potentially mask this effect: the ranking score only captures ordinal differences and not differences in magnitude. Secondly, measurements may be noisy because the experiment was conducted in the visual instead of the auditory modality. While not posing a problem to the combinatoriality measure reported earlier (because of the larger effect sizes), this might hinder detection of iconicity in the languages.

To address these points, a second experiment collected complexity ratings in the auditory domain, restricting ourselves to the first five generations of iterated learning. Figure 2C shows the resulting ratings, grouped and averaged by associated referent complexity. These data were used to derive an iconicity measure, depicted in Figure 2D, by subtracting the average complexity of signals associated with simple referents from signals associated with complex referents. Positive values indicate the presence of iconicity in a congruent direction. As suggested earlier, we find that iconicity emerges immediately after initialization in generation one ( $t(14) = 3.71, p = 0.002$  before,  $p = 0.01$  after correction). While the return of the iconicity measure to chance is not significant within the first five generations of the auditory measure ( $t(14) = -1.56, p = 0.14$ ), the visual complexity measure from 2C strongly suggests that iconicity drops back to chance in subsequent generations and thus, taken together, licenses the inference that iconicity eventually disappears from the languages.

### Relationship between iconicity and structure

The previous analyses have demonstrated that languages develop both combinatoriality and iconicity over the course of the experiment. To develop insight into state-dependent trade-offs between iconicity and combinatoriality, we now look at the development of iconicity as a function of combinatorial structure instead of generation. Figure 4A shows the evolutionary trajectories of all fifteen languages in terms of their combinatorial structure and their iconicity (based on Figures 2A and D). The plots suggest that languages, while

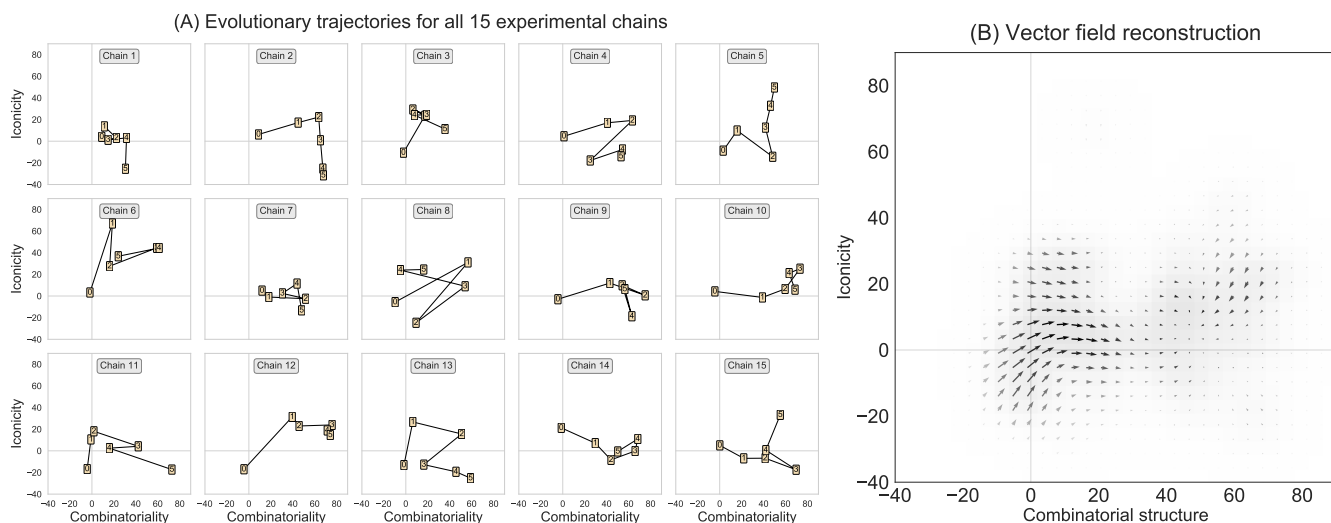


Figure 4: Visual representation of the evolutionary trajectories of all fifteen experimental chains, on the basis of which we constructed a vector field representation that shows inter-generational parameter changes for observed and hypothetical languages.

displaying common patterns, can vary substantially in terms of their developmental time constant. We summarize these data by constructing a vector field that shows extrapolated average magnitude and direction of inter-generational changes on a grid (based on the behavior of nearby languages). The model is constructed by considering all 75 vectors in Figure 4A that represent transitions from one generation to the next. For each grid point, an average magnitude and direction is estimated by computing the weighted linear combination of vectors using their distance, obtained with a multivariate Gaussian kernel centered around the grid point, as weights. Figure 4B shows the resulting vector space model. The total sum of Gaussian weights per grid point, superimposed in grey, corresponds to the number of vectors nearby that were used to construct the estimate. The model summarizes in which direction, and how much, hypothetical languages would change in terms of combinatoriality and iconicity, based on the observed data. Adding to the results reported above, gains in iconicity or maintenance of already existing iconicity is only observed when languages are still relatively unstructured. Languages lose their iconic structure as combinatoriality increases further. More sporadically observed ‘extreme’ levels of iconicity and combinatoriality appear to be unsustainable and eventually revert to lower levels.

### Inductive biases for iconicity

Which underlying cognitive structures support our inferences about iconicity? We asked naive subjects in a guessing game to pick the most likely referent for each signal. Figure 5 shows the probability of listeners choosing a complex referent as a function of signal complexity, indicating a strong tendency for choosing referents that match the perceived complexity of the signal. (Note that the ground truth referent information in Figure 5 is displayed as additional information and not part of the reported analysis.) Crucially, this

bias allowed subjects to reliably identify the correct referent complexity class for signals that exhibited the most iconicity ( $t(9) = 2.77, p = 0.024$ ). We compared the probability of choosing the correct referent class with chance performance for the ten languages that scored highest in iconicity (measure taken from Figure 2D).

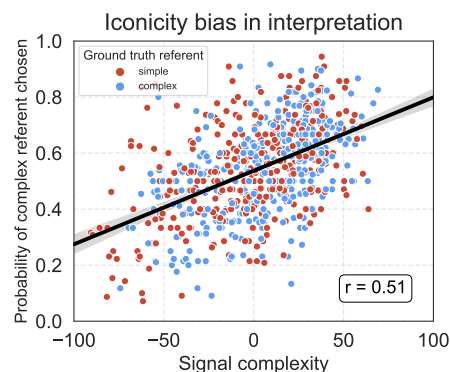


Figure 5: Results from the guessing game, conducted to quantify the strength of iconicity biases for the evolved signals.

## Discussion and Conclusion

We conducted an iterated learning experiment to investigate how iconicity develops when combinatoriality emerges in initially unstructured, continuous signals. Signals gradually became more combinatorial over the course of the experiment. The emergence of iconicity, measured in terms of signals matching the conceptual complexity of their referents, was shown to develop immediately, but iconicity eventually disappeared while combinatorial structure continued to increase. This result is particularly strong because languages were initialized at a point of complete arbitrariness.

The largest increases in combinatoriality were seen in generation one. This is consistent with the idea that signal structure is a consequence of cognitive biases for combinatoriality. Because of having selected a diverse signal set for initialization, memory demands during learning and reproduction were arguably the highest in the first generation. Since this affords the greatest potential for prior biases to manifest, we would expect the largest increase in structure here.

The loss of iconicity is consistent with the hypothesis that iconicity is in complementary distribution with combinatorial structure (Verhoef et al., 2016; Goldin-Meadow & McNeill, 1999; Roberts, Lewandowski, & Galantucci, 2015), since the building blocks of such a system must be stripped of their iconicity when they participate in larger meaning-bearing units. It is important to note, however, that it is not clear why the particular kind of iconicity we investigated here must be lost in order for combinatorial structure to arise. In the transition from holistic to combinatorial structure, complexity in the signal domain could be expressed equally well in terms of number of building blocks (Lewis & Frank, 2016). The observation that iconicity is nevertheless lost could, however, provide important insight into the nature of the transition process. Zuidema and de Boer (2018) recently distinguished between analytic and synthetic routes to combinatoriality. In the synthetic route, preexisting signals are combined to form larger combinatorial signals, while in the analytic route, potentially overlapping parts of preexisting signals are used to form new signal. The present findings are consistent with the holistic account, which predicts that productive recombination leads to new signals that are composite, therefore complex, irrespective of the complexity of their referent but simply due to the mechanics of recombination.

While our guessing game suggests that people have strong biases for iconicity, our results indicate that these biases manifest in subtle ways. Smith et al. (2017) presented evidence that the strength with which cognitive biases manifest in cultural evolution depend on a number of factors. The authors focused on properties of the transmission paradigm, such as how many different agents subjects learn from, which shapes the input to learning. In the present study, we found evidence that the manifestation of otherwise strong cognitive biases, such as a bias for iconicity, can also depend on properties of the input more directly, for instance, on how much structure signals exhibit. Understanding how properties of the input can modify the expression of biases more broadly is an avenue for future research. In addition, the novel vector field analysis we present suggests the possibility of testing specific combinations of iconicity and combinatoriality to develop a more complete picture of trade-offs in parameter space.

One further aspect that is not addressed in our study is the role of modality on the form of iconicity studied here. In work that explored the structure of signals that subjects created when more or less signal dimensions were available, Little, Eryılmaz, and de Boer (2017) found strong modality effects mediating the relationship between iconicity and combinatoriality.

Future work is needed to explore how our findings generalize to other signal modalities.

Finally, we note that the combinatoriality measure obtained via subject ratings is only an approximation to signal structure that emerged in the experiment. To better understand the patterns that exist in the evolved signals and how they are used productively, it is necessary to develop computational models. In ongoing work, we are developing statistical models of signal structure and learners' underlying inductive biases that will allow us to test more specific hypotheses about the evolution of iconicity and combinatoriality in language.

## References

- Christiansen, M. H., & Chater, N. (2016). *Creating language: integrating evolution, acquisition, and processing*. Cambridge, MA: The MIT Press.
- Dingemans, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences*, 19(10), 603–615.
- Fay, N., Arbib, M., & Garrod, S. (2013). How to Bootstrap a Human Communication System. *Cognitive Science*, 37(7), 1356–1367.
- Giudice, A. D. (2012). The emergence of duality of patterning through iterated learning: Precursors to phonology in a visual lexicon. *Language and Cognition*, 4(04), 381–418.
- Goldin-Meadow, S., & McNeill, D. (1999). The Role of Gesture and Mimetic Representation in Making Language the Province of Speech..
- 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.
- Lewis, M. L., & Frank, M. C. (2016). The length of words reflects their conceptual complexity. *Cognition*, 153, 182–195.
- Little, H., Eryılmaz, K., & de Boer, B. (2017). Signal dimensionality and the emergence of combinatorial structure. *Cognition*, 168, 1–15.
- 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.
- Smith, K., Perfors, A., Fehér, O., Samara, A., Swoboda, K., & Wonnacott, E. (2017). Language learning, language use and the evolution of linguistic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711), 256–278.
- 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., & de Boer, B. (2016). Iconicity and the Emergence of Combinatorial Structure in Language. *Cog-*

- nitive Science*, 40(8), 1969–1994.
- Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, 79(7), 2064–2072.
- Zuidema, W., & de Boer, B. (2018). The evolution of combinatorial structure in language. *Current Opinion in Behavioral Sciences*, 21, 138–144.