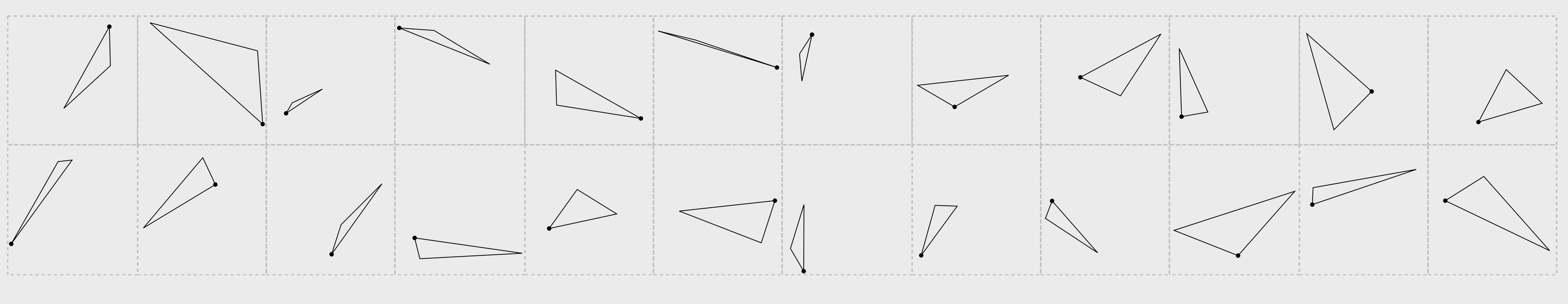
Communication increases the expressivity of emergent languages in an open-ended meaning space

INTRODUCTION

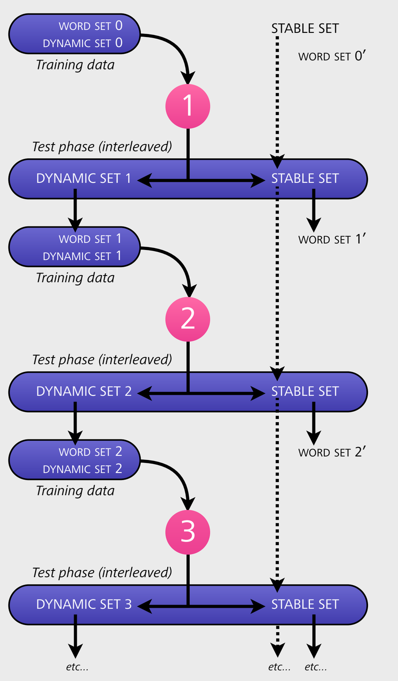
The meaning spaces typically used in iterated learning experiments (e.g. Kirby, Cornish, & Smith, 2008) are unlike natural language, which is characterized by open-ended structure. Some recent experiments have used continuous spaces (Perfors & Navarro, 2014; Silvey, Kirby, & Smith, 2013), but these do not fully address the open-ended nature of meaning. We have constructed a meaning space based on randomly generated triangles that is continuous, high-dimensional, open-ended, and not pre-determined by the experimenter. This experimental paradigm models discrete infinity (see e.g. Studdert-Kennedy, 2005 for some discussion), since a finite set of symbols is used to describe an infinite and ever-changing set of meanings.



Examples of the triangles that made up the meaning space. Stimuli were generated by randomly selecting three coordinates in a 500×500-pixel box. One vertex had a dot placed over it to give the triangle an orientation. This procedure allows for the generation of 6×10^{15} possible stimuli.

METHODS

EXPERIMENTAL PARADIGM Participants ($n = 40$) in our first experiment learned an artificial language describing 48 triangles. The first participant in a transmission chain learned words that were randomly generated from a finite set of syllables. Subsequent participants were trained on the output of the previous participant in the chain. Participants were then tested on a new set of stimuli that changed at every generation (see panel). Our second experiment was identical to the first, except each generation consisted of a pair of participants ($n = 80$) who used the language to communicate.

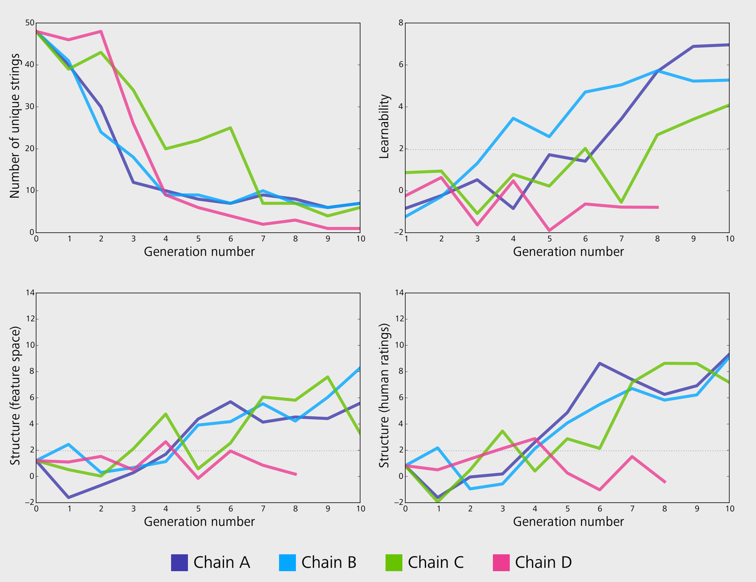


A participant learns from the previous participant's DYNAMIC SET and is tested on a new DYNAMIC SET plus the STABLE SET, which provides a consistent set of stimuli for the purpose of measuring structure and learnability.

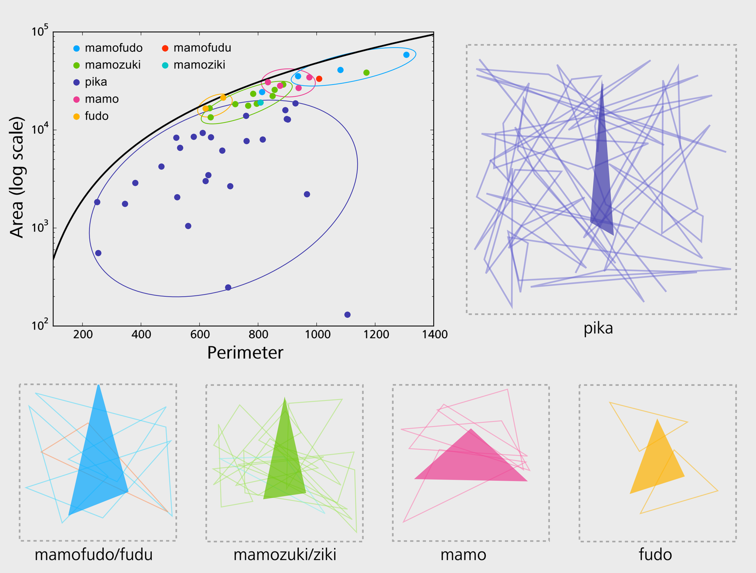
STRUCTURE MEASURE To measure structure in the languages, we correlate string dissimilarity with meaning dissimilarity for all pairs of stimuli in a given participant's output. The Levenshtein (1966) edit-distance was used to measure string dissimilarity. The dissimilarity between two triangles was calculated by taking the Euclidean distance in an 18-dimensional feature space. To verify the psychological reality of this space, we conducted an online dissimilarity ratings experiment with 103 participants and found that participants' ratings were highly correlated with the feature space estimates ($r = 0.502$, $z = 8.259$).

EXPERIMENT 1 RESULTS

The results for experiment 1 are shown in the panels below. The emergent languages arbitrarily divided the meaning space into a small number of categories based on the size and shape of the triangle stimuli.



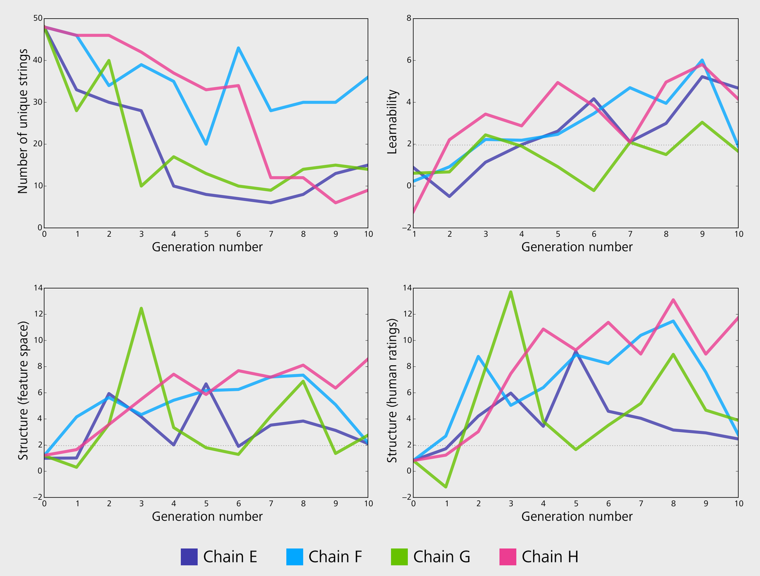
Results for experiment 1. Over time the languages become smaller and more learnable (adjusted for chance), and structure increases, suggesting that a systematic mapping has evolved between form and meaning.



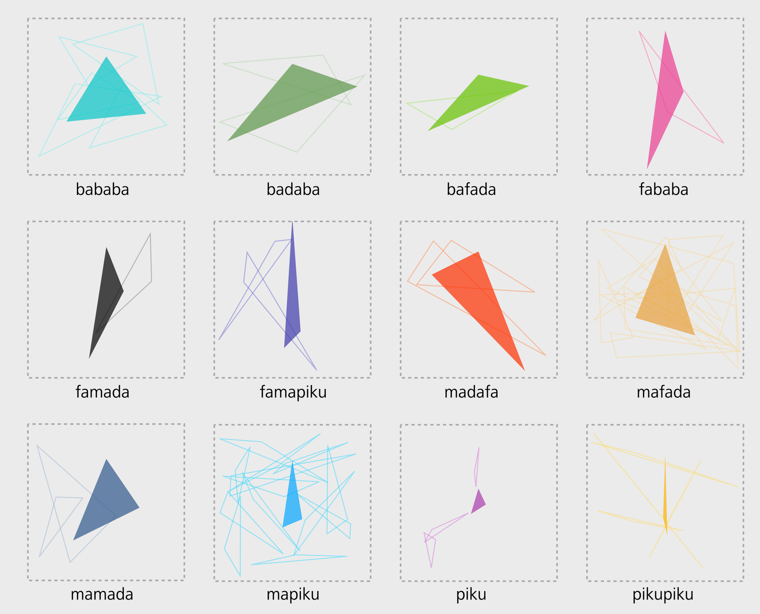
Example language from Chain B Generation 8. The plot approximates a shape-size space. The space of possible triangles was divided into five categories based on shape and size. Filled triangles show prototypes.

EXPERIMENT 2 RESULTS

Our second experiment added dyadic communication to the paradigm which increased the expressivity of the languages. These more expressive languages appear to make more nuanced distinctions by making use of compositional structure.



Results for experiment 2. Over time the languages become more learnable (adjusted for chance) and more structured. Compared to experiment 1, the number of unique strings remains high, suggesting that the languages are capable of making more distinctions.



Example language from Chain H Generation 8, which distinguishes 12 words and uses a small set of recombinable units (*ba*, *da*, *fa*, *ma*, & *piku*).

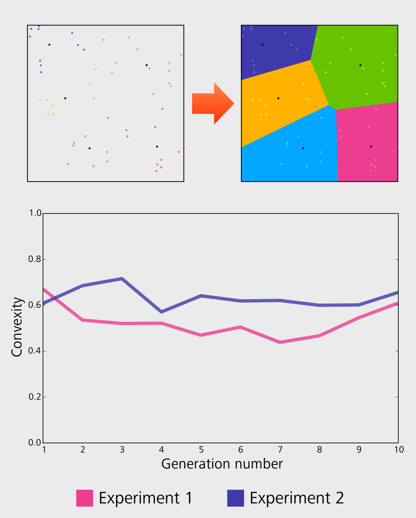
We are currently working on methods to determine how the sub-lexical structure corresponds to the dimensions of the meaning space (any ideas would be greatly appreciated). However, the differences between these two experiments suggest that communicative pressures are required for compositionality to arise in higher-dimensional meaning spaces.

DISCUSSION POINTS

CATEGORY BIASES Humans are biased towards encoding certain categories over others (e.g. Landau, Smith, & Jones, 1988 show a bias for shape). Iterated learning amplifies these biasses, allowing us to observe them in a laboratory setting. In experiment 1, participants primarily encoded shape and size (as opposed to e.g. location or rotation), but in experiment 2, participants were forced to make other distinctions to disambiguate the triangles.

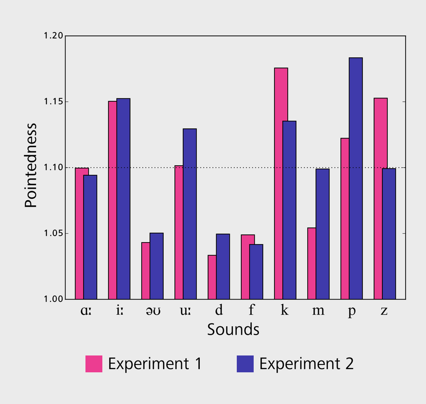
LINEAGE SPECIFICITY Although participants tended to encode the same properties, there were subtle differences between chains in how the space was divided, just as there are in natural languages.

CONVEXITY Contrary to the predictions of Gärdenfors (2000), category convexity in the emergent languages tended to be suboptimal and did not increase over time. We measured convexity by counting the number of triangles with a given name that fell into the corresponding cell of a Voronoi tessellation of the feature space (see panel to the right). While it is possible that our feature space does not fully capture the underlying conceptual space, this could suggest that humans do not discretize meaning spaces into perfectly convex regions.



ABOVE: Example of a Voronoi tessellation (right), which represents a perfectly convex ideal to which we compare the actual data (left). BELOW: Mean convexity in the experiments suggesting that the languages do not evolve increasingly convex meaning structure.

SOUND SYMBOLISM It has been suggested that sound symbolism facilitates word learning (see e.g. Monaghan, Christiansen, & Fitneva, 2011). We identified sound symbolic patterning that follows documented sound symbolic tendencies for more rounded vs. more angular shapes (see e.g. Ahlner & Zlatev, 2010). The results are shown to the right.



Mean pointedness for ten sounds that occur in generations 6–10. Dotted line shows chance level. The sounds [iː], [k], [p], and [z] are more likely to occur with more pointed triangles, while [əʊ], [d], [f], and [m] are more likely to occur with less pointed triangles.

REFERENCES

Ahlner, F., & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies*, *38*, 298–348.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, MA: MIT Press.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the USA*, *105*, 10681–10686.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*, 299–321.

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

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: General*, *140*, 325–347.

Perfors, A., & Navarro, D. J. (2014). Language evolution can be shaped by the structure of the world. *Cognitive Science*, *38*, 775–793.

Silvey, C., Kirby, S., & Smith, K. (2013). Communication leads to the emergence of sub-optimal category structures. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth, *Proceedings of the 35th annual conference of the Cognitive Science Society* (pp. 1312–1317). Austin, TX: Cognitive Science Society.

Studdert-Kennedy, M. (2005). How did language go discrete? In M. Tallerman, *Language origins: Perspectives on evolution* (pp. 48–67). Oxford, UK: Oxford University Press.

Jon W. Carr[1], Hannah Cornish[2], Kenny Smith[1], Simon Kirby[1]

1 School of Philosophy, Psychology and Language Sciences, University of Edinburgh
2 Psychology, School of Natural Sciences, University of Stirling

j.w.carr@ed.ac.uk