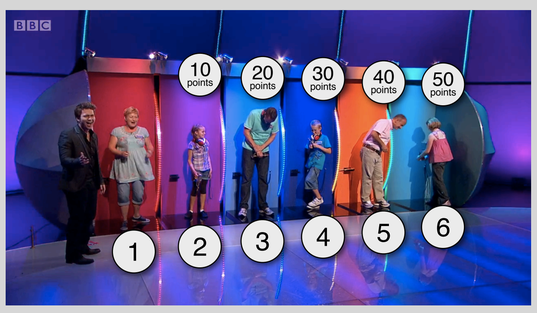
The emergence of combinatoriality in the cultural transmission of pop songs in a children's gameshow

INTRODUCTION

Duality of patterning says that the speech stream is composed of meaningful recombinable units (compositionality), which themselves are composed of meaningless recombinable units (combinatoriality). Kirby, Cornish, and Smith (2008) have shown that compositional structure can emerge under iterated learning. Verhoef (2012) has shown that iterated learning can also explain the emergence of combinatorial structure.

Copycats is a children's gameshow produced by the BBC in which two teams compete in games based on Chinese Whispers. In one game, each team must pass along a pop song from one team member to another by imitating it using a kazoo, a simple musical instrument. The team is awarded points based on how far they successfully transmit the song. The game has clear parallels with iterated learning and shares similarities with Verhoef's (2012) experimental setup.



During gameplay, barriers slide in and out of the set to separate the players. Player 1 listens to a pop song and reproduces it for player 2 using a kazoo. Player 2 subsequently repeats the tune to player 3 and so on down the line. Each player has 10 seconds to reproduce the song. Players wear headphones when not involved in playing or listening. Working backwards from player 6, the players are asked in turn to name the song. The number of points available decreases by 10 for each player who is unable to supply the correct answer.

METHODS

DATA COLLECTION Audio was collected from 102 episodes of the gameshow, yielding data for 204 teams. Only teams that scored 0 points were analysed (see panel opposite), which mitigates the effects caused by players who recognized their team's target song. Noise and other irrelevant material (e.g. laughter) were removed in Audacity, and the songs were converted into pitch tracks using Praat.

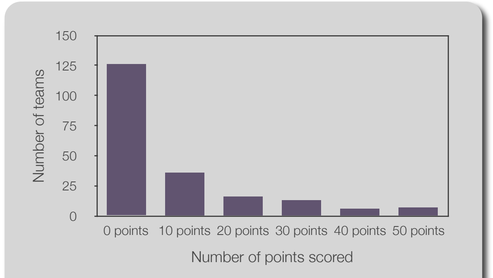
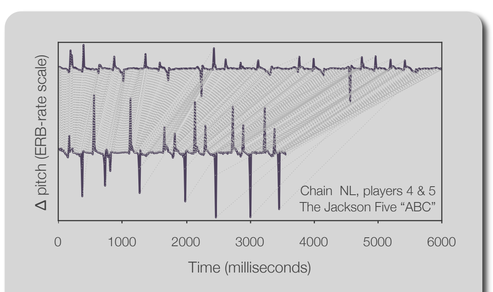


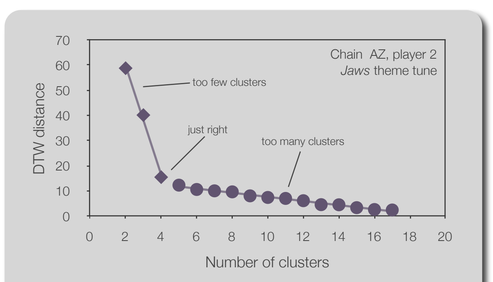
Chart showing how many teams scored each number of points. Only the 126 teams that scored 0 points were analysed in this study to mitigate the effects caused by players recalling songs from memory.

MEASURE OF LEARNABILITY The Derivative Dynamic Time Warping distance (Keogh & Pazzani, 2001) was used to measure the dissimilarity between consecutive players' songs. This algorithm warps the time axis to find the best fit between two time series and returns a measure of the distance between them (see panel opposite). This distance provides an estimate of how learnable a given player's song is by reference to how much error was made in reproducing it.



The derivatives of two pitch tracks for a pair of consecutive players. The connecting lines show the best fit between the pitch tracks. Essentially, the sum of these lines is used as a measure of distance.

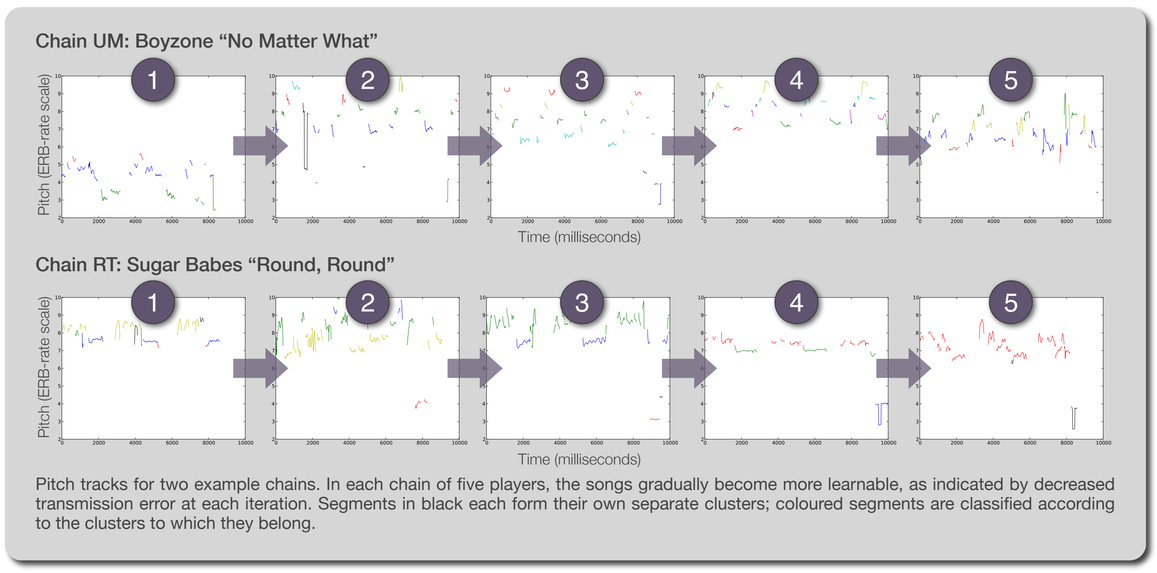
MEASURE OF COMBINATORIALITY The songs were first divided into their component segments. Boundaries between segments were indicated by (a) a period of silence or (b) a dramatic change in pitch. The segments were then clustered with agglomerative hierarchical clustering using Dynamic Time Warping (Sakoe & Chiba, 1978) as the distance metric. The number of clusters was determined using the L-method (Salvador & Chan, 2004; see panel opposite). The combinatoriality of each song was then estimated using Shannon entropy (Shannon, 1948).



Plot of number of clusters against average distance between members of the clusters. The "knee" of the curve, determined by finding the two best regression lines, indicates the number of clusters in the song.

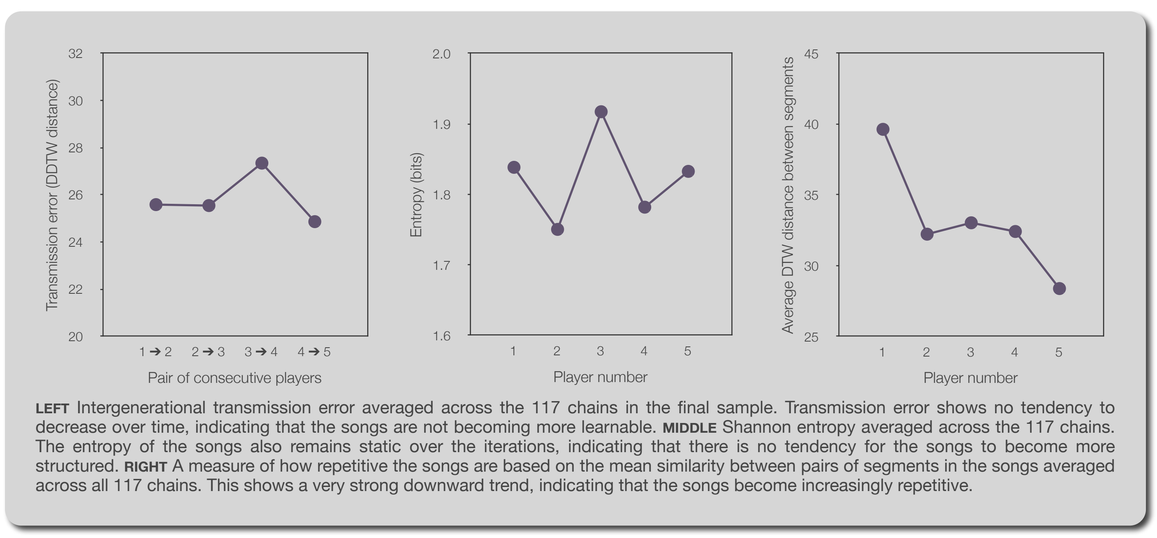
RESULTS

Nine teams were removed from the sample because a player had difficulty using the kazoo to produce any sound, bringing the sample down to 117 chains. Two example chains are given in the panel below.



Pitch tracks for two example chains. In each chain of five players, the songs gradually become more learnable, as indicated by decreased transmission error at each iteration. Segments in black each form their own separate clusters; coloured segments are classified according to the clusters to which they belong.

The panel below provides the aggregate results for transmission error (left) and entropy (middle). Neither of these measures showed the significant downward trend that was expected. The third graph (right) gives a measure of how repetitive the songs are, which is calculated by taking the average DTW distance between pairs of segments in the songs. This shows a downward trend that is highly significant according to Page's (1963) trend test ($L = 5513$, $m = 117$, $n = 5$, $p < 0.001$).



LEFT Intergenerational transmission error averaged across the 117 chains in the final sample. Transmission error shows no tendency to decrease over time, indicating that the songs are not becoming more learnable. **MIDDLE** Shannon entropy averaged across the 117 chains. The entropy of the songs also remains static over the iterations, indicating that there is no tendency for the songs to become more structured. **RIGHT** A measure of how repetitive the songs are based on the mean similarity between pairs of segments in the songs averaged across all 117 chains. This shows a very strong downward trend, indicating that the songs become increasingly repetitive.

DISCUSSION

LEARNABILITY One explanation for the lack of an interesting trend is that, as the songs are passed on, they tend to become shorter, so players tend to repeat the same short tune over and over. In many cases, consecutive players differed widely in how many reps they performed which would dramatically increase transmission error despite the tunes being very similar. Learnability for this dataset might better be explored by conducting a similarity judgement task with naive participants.

COMBINATORIALITY This is a noisy dataset, and although there are a large number of chains, there is very little data per player. For this reason, my approach to segmentation and clustering do not always perform very well, which may explain why there was no interesting trend in terms of entropy. However, a measure of the repetitiveness of the songs showed a highly significant trend as the songs increasing come to rely on a small set of similar sounds. In addition, my qualitative impression of the dataset hinted at the emergence of combinatorial structure.

REFERENCES

Keogh, E. J., & Pazzani, M. J. (2001). Derivative dynamic time warping. In V. Kumar & R. Grossman (Eds.), *Proceedings of the 1st SIAM international conference on data mining*.

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.

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

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

Salvador, S., & Chan, P. (2004). Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. In *Proceedings of the 16th international conference on tools with artificial intelligence* (pp. 576–584). Los Alamitos, CA: IEEE Computer Society.

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

Verhoef, T. (2012). The origins of duality of patterning in artificial whistled languages. *Language and Cognition*, *4*, 357–380.

Jon W. Carr
Language Evolution and Computation Research Unit
School of Philosophy, Psychology and Language Sciences
University of Edinburgh
j.w.carr@sms.ed.ac.uk