EVOLUTIONARY BIOLOGY

Language-like statistical structure arises in learned signaling: evidence from birdsong

Simon Kirby[1]*, Kazuo Okanoya[2]*, Ellen C. Garland[3,4], Miki Takahasi[5], Inbal Arnon[1,6]*

All languages have statistically coherent subsequences (e.g., words) whose frequency distribution follows a power law. These properties facilitate language learning in humans, making them good candidates for arising through cultural transmission as a way to help faithful transmission across generations. Recently, both properties were found in whale song, which is also culturally transmitted, leading to the strong prediction that they should be found wherever complex sequential signaling is culturally transmitted. Here, we use the same tools used to analyze human data and whale song to reveal that culturally transmitted Bengalese finch song also has statistically coherent subsequences whose distribution follows a power law. We additionally show that statistical coherence increases over development, but the power law is present throughout, suggesting that it reflects a fundamental principle of learned representations. Finding these parallels between evolutionarily distant species illustrates the importance of cultural transmission in shaping communication and suggests that core properties of language arise through convergent evolution.

INTRODUCTION

Language has two statistical properties that are strikingly rare in other species. First, it is made up of statistically coherent units (e.g., words) consisting of sequences of sounds, whose internal structure is relatively predictable. Second, the frequency distribution of these units follows a characteristic shape known as a Zipfian distribution where the most frequent unit appears approximately twice as often as the second most frequent one, three times as often as the third, and so on (*1*, *2*). While Zipfian distributions are found across the natural world, with different explanations applicable in different domains (*3*), recent work has shown that this distribution serves a learning function in human language: Having statistically coherent subsequences (*4*) and having their frequency follow a Zipfian distribution facilitate language learning in humans (*5*–*9*). For example, infants, children, and adults are better able to discover words in continuous speech when their frequency distribution follows this particular power law distribution (*5*, *7*, *9*). The fact that both statistical properties aid learning makes them good candidates for arising through cultural transmission: An extensive literature shows properties that aid learning may be present in language precisely because they help its faithful transmission over generations of learners (*10*–*12*). There is experimental evidence that cultural transmission can lead to the emergence of statistically coherent subsequences that follow a Zipfian distribution in humans, even in the absence of meaning or communication (*13*). Both statistical properties were thought to be uniquely human; however, the same statistical structure has been recently found in humpback whale (*Megaptera novaeangliae*) song, a culturally transmitted complex signaling system where individual sound elements are concatenated to form longer sequences (*14*). Using tools inspired by how human babies segment speech, the study found that whale song has language-like statistical structure (*15*): The song contains statistically coherent subsequences whose distribution is Zipfian. If the similarity between these two very different species is related to culture, then we can make the strong prediction that similar structure will be found wherever complex sequential signaling is culturally transmitted.

Passerine birds make an ideal test case for this prediction: Songs are made up of individual syllables used to form longer sequences. The song is complex and learned (*16*, *17*), and we have data on song development (currently lacking for whales). Previous work on birdsong has examined the distribution of the individual syllables, finding that it is highly skewed (*18*–*23*), with some studies reporting a good fit to a Zipfian distribution (*18*–*20*). However, these findings only pertain to the individual sound elements, whose learning and transmission may be affected by different learnability pressures than those affecting larger sequences. Existing findings cannot tell us whether birdsong, such as human language and whale song, contains statistically coherent subsequences whose distribution is Zipfian, as would be expected if these statistical properties arise to facilitate faithful transmission. We test the prediction that similar structure will be found in culturally transmitted birdsong when applying the same infant-inspired tools used to analyze human data and humpback whale song. The method uses dips in transitional probability between the basic acoustic building blocks to segment the song into subsequences. This is the same cue that is used by infants to segment speech (*4*): Because words are statistically coherent, transitional probabilities between syllables are higher (on average) within words than between words (*24*). Using developmental data, we can additionally ask whether this structure is also present in juvenile song or only in the adult target.

We focus here on Bengalese finches (*Lonchura striata* var. *domestica*) whose song has been extensively studied (*25*). Bengalese finches are a domesticated variant of a wild songbird, the white-rumped munia (*L. striata*) (*26*). Bengalese song is learned, culturally transmitted, and has complex structure (*27*). Only males sing (*28*), and songs differ between birds, with each bird's song typically consisting of around

[1]Centre for Language Evolution, University of Edinburgh, Edinburgh, UK. [2]Department of Life Sciences, Advanced Comprehensive Research Institute Teikyo University, Tokyo, Japan. [3]Sea Mammal Research Unit (SMRU), Scottish Oceans Institute, School of Biology, University of St Andrews, St Andrews, UK. [4]Centre for Social Learning and Cognitive Evolution, School of Biology, University of St Andrews, St Andrews, UK. [5]Language Development Subgroup, Laboratory for Molecular Mechanism of Brain Development, RIKEN Center for Brain Science, Saitama, Japan. [6]Psychology Department, Hebrew University, Jerusalem, Israel.
*Corresponding author. Email: simon.kirby@ed.ac.uk (S.K.); kazuookanoya@gmail.com (K.O.); Inbal.arnon@mail.huji.ac.il (I.A.)

eight distinct syllables that are combined to form longer sequences (*25*). Bengalese finches are sensitive to transitional probabilities (*29*), the cue used by our segmentation method. In a seminatural environment, Bengalese chicks can learn from multiple tutors (*29*). They splice sections of song from different adult males when the transition between syllables in the tutor's song is less predictable (*29*). This indicates that the song has statistical structure and that this information is relied on by chicks during learning: Bengalese finches do not learn songs as a stereotyped, fixed sequence but learn them as smaller chunks based on the transitional probabilities between syllables.

Here, we apply the same segmentation method used on human data and whale song to a corpus of Bengalese finch song (*30*). The corpus includes data from six male birds, each of which was recorded on five separate days throughout development, from the earliest analyzable song to fully crystallized song. If cultural transmission is a driving force in the emergence of statistically coherent parts and their Zipfian distribution, then we should find similar statistical structure in Bengalese finch song. We take advantage of the existence of developmental data to ask whether the statistical structure changes as the song approaches the adult target. One possibility is that a Zipfian distribution of statistically coherent subsequences will only be found in the adult song. Alternatively, these statistical properties could be found throughout development, suggesting that they reflect a more fundamental organizing principle of learned representations.

RESULTS

Bengalese finch song has statistically coherent subsequences that follow a Zipfian distribution

If birdsong has statistically coherent parts, we should detect local dips in transitional probabilities between syllables. We used the syllable classification provided by Takahasi and Okanoya (*30*) (fig. S1). To identify dips between syllables, we used the same method used in the analysis of human data (*13*) and whale song (*15*). We estimated the transitional probability between each pair of consecutive syllables for each of the 30 recordings in our dataset (six birds, five recordings each). Each recording (but one) contains 10 song bouts [a song bout is defined as a sequence of singing in which syllables were not separated by a silent interval longer than 500 ms (*31*)]. For example, if a recording contained the sequence "AB." then we estimated the probability of B given a preceding A in the recording. We inferred segmentation boundaries when the ratio between two consecutive transitions was deemed low. We used ratios and not the transitional probabilities themselves to be able to detect relative dips: A transitional probability of 0.4 is not very low, but a drop from 0.9 to 0.4 may be meaningful. We used the same 0.5 cutting ratio used previously for segmenting whale song (*15*): A boundary was posited whenever the ratio was lower than 0.5 (meaning whenever the probability of the current transition was half that of the previous one). Prior work applying a similar segmentation pipeline to data from human participants used a threshold of 0.42 derived from a random baseline used in the experimental setup. However, we do not have such a random baseline for the real-world birdsong data (or for whale song). Instead, we chose to use the same 0.5 threshold used previously to segment whale song. Figure 1 illustrates the segmentation method for an extract of a sequence taken from one recording of one bird at day 100.

We ran the segmentation pipeline separately for each recording, preserving the boundaries between song bouts (since each song is a separate instance of the bird singing). Our method detected subsequences within the song: We conducted an average of 10.72 cuts for each song bout. We then calculated the frequency of the detected subsequences in each recording. To give an example, this is how our pipeline segmented a song bout produced by one bird on the final day of recording (each song syllable is represented by a unique letter):

cc | bbc | aaaaaa | bccddeefaaaaaa | ghccddeefaaa | ghcc | bccddeefaaa | ghccddeefaaa | gh | ghccddee | bccddeefaaaa | gh | ghccddeefaaa | ghc | bccddeefaaa | gh | bccddeefaaa | gh | bccddeefaaaaa | gh | bccdde

Figure 2A plots the frequency distribution of the segmented subsequences, collapsed over recordings. Figure 2B gives the same data on a log-log plot. The distribution shows an excellent fit to a power law [mean R^2 (coefficient of determination) = 0.89]. We estimated the slope of the distribution for each recording using maximum likelihood estimation. The mean slope (averaged over recordings) was 1.05, extremely similar to that found in human languages (*2*, *32*, *33*). Applying this infant-inspired pipeline to birdsong resulted in the detection of recurring subsequences that follow a Zipfian distribution. We repeated the analyses with two additional thresholds, 0.3 and 0.7, and find that the results are robust to changes in cutting sensitivity (see figs. S1 and S2). Changing the threshold affects how many cuts are made: Decreasing the threshold leads to fewer cuts (5.01 cuts on average per song bout when using a 0.3 threshold compared to 10.72 when using a 0.5 threshold), while increasing it leads to more cuts (16.75 cuts, on average, per song bout when using a 0.7 threshold). This reflects the fact that changing the threshold leads to detecting

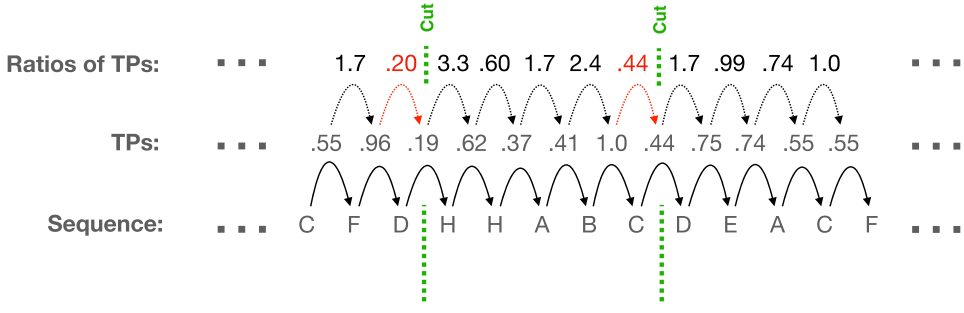

Fig. 1. Our segmentation pipeline as applied to an extract from a recording taken at day 100. The pipeline takes an unsegmented sequence of birdsong syllables, each of which is coded using a single unique letter. The transitional probabilities (TPs) are estimated by counting the frequency of pairs of syllables in each recording. The ratio of subsequent TPs is given in the top row of the diagram. Wherever this ratio drops below 0.5, shown here in red, a cut is made in the sequence. The subsequences found between cuts, e.g., HHABC here, are added to the inventory of subsequences for that bird on that day.

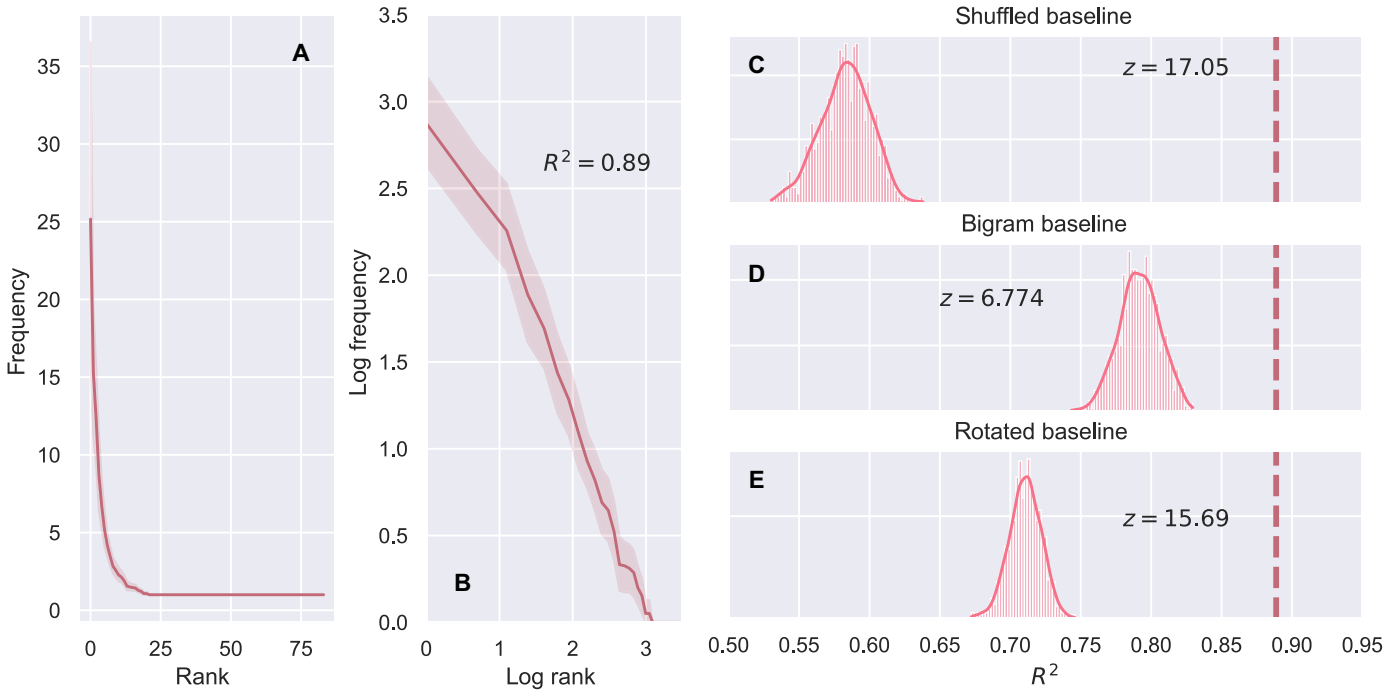

Fig. 2. Subsequences detected in birdsong by our infant-inspired segmentation method follow a Zipfian frequency distribution. We calculate the frequency-by-rank distribution for each bird and day (i.e., recording) separately and graph the mean and 95% confidence intervals around the mean. (**A**) Frequency of the detected subsequences plotted by rank. (**B**) Same data plotted on a log-log scale. The data show a good fit to a Zipfian distribution (as measured by the mean R^2 value across all 30 recordings). (**C**) Mean R^2 of the real data as a dashed line compared to the mean R^2s found by running our entire pipeline on 1000 randomly shuffled datasets. (**D**) Comparison to 1000 datasets generated using the bigram statistics of the real data (but disrupting larger sequential information). (**E**) Comparison to 1000 randomly rotated datasets. The *z*-scores demonstrate that the real data are a far better fit to a Zipfian distribution than any of the baseline datasets.

units of varying sizes: smaller units when the cutting ratio is higher and larger ones when it is lower. Using these two additional thresholds also leads to the detection of statistically coherent subsequences whose distribution shows a good fit to a Zipfian distribution ($R^2 = 0.75$ for the 0.3 threshold and $R^2 = 0.90$ for the 0.7 threshold). That is, the Zipfian distribution seems to be present at different levels of granularity, as has been found for human language (*34*).

Zipfian distributions reflect sequential structure in birdsong

It is possible that the Zipfian distribution of subsequences is simply an artifact of the distribution of individual syllables. To address this, we created 1000 pseudo datasets by shuffling the syllables within each song bout [following (*15*)]. This preserves the distribution of syllables but modifies the sequential structure. We reran our segmentation pipeline on each of the shuffled datasets and calculated the mean R^2 for each of these datasets to assess the fit to a power law. If the statistical structure we found in birdsong reflects its sequential structure, then we should find a lower fit in the shuffled datasets. Figure 2C shows that the mean R^2 derived from the real data has a much better fit to a power law than any of the shuffled datasets ($z = 17.05$, $P < 0.00001$; $z = 16.32$, $P < 0.00001$ for the 0.3 cutting threshold; $z = 9.04$, $P < 0.00001$ for the 0.7 cutting threshold). Since syllables in Bengalese finch song are not entirely independent of one another (there are predictive relations between subsequent syllables, (*35*)], we were interested to see if using bigram statistics alone would lead to the presence of statistically coherent subsequences which show as good a fit to a Zipfian distribution. If not, then this would suggest that Bengalese finches are sensitive to sequential information larger than the bigram. To implement this baseline, we generated a further 1000 pseudo datasets, which preserved the bigram statistics of the real song but disrupted larger sequential information. We ran our pipeline on each of these bigram-based datasets, calculated the distribution of the detected subsequences, and examined the fit to a Zipfian distribution. Figure 2D shows the fit to a Zipfian distribution in the bigram baseline and in the real data, indicating a better fit in the real data ($z = 6.77$, $P < 0.00001$; $z = 9.22$, $P < 0.00001$ for the 0.3 cutting threshold; $z = 1.87$, $P = 0.0307$ for the 0.7 cutting threshold). This demonstrates that the Zipfian distribution of subsequences in the song of Bengalese finches arises due to sequential structure above the bigram.

To further ensure that the results we found are related to our specific segmentation cue, we performed a more stringent validation that keeps the sequential order and the length of the detected subsequences, but where the cuts do not correspond to dips in transitional probability (*15*). We did this by generating "rotated" datasets, where the cutting points of the song bouts in the original data were shifted by a random number of steps (e.g., if a sequence was cut at position 5 in the original dataset and the random number was 6, then the cut would now be placed at position 11 in the sequence). We calculated the frequency distribution for 1000 "rotated" datasets. The fit to a Zipfian distribution in the real data, as assessed by mean R^2, was much better than in any of the rotated datasets as shown in Fig. 2E ($z = 15.69$, $P < 0.00001$; $z = 11.9$, $P < 0.00001$ for the 0.3 cutting threshold; $z = 9.661$, $P < 0.00001$ for the 0.7 cutting threshold).

Statistical coherence of detected subsequences increases during development

Bengalese finches typically start singing around day 35 at the earliest, but in some individuals, singing could start as late as day 50 (*36*). Our earliest recordings start around day 60, not far into song development, and end around day 120 when the song is already stable (*36*). Each bird was exposed to a single tutor, older than 120 days, meaning that they are learning only one target song throughout development. Figure 3A shows changes in the average transitional probability within subsequences and between subsequences during development: If statistical coherence increases over time, we should see a greater difference in the later recordings. We ran a mixed-effects model on the full set of transitional probabilities in the data with fixed effects of day and transition-type (within versus between) and their interaction, as well as a random effect of bird, and a by-bird random slope for transition-type. Here, as with all the models reported below, this was the maximal random effects structure that we were able to fit [following (*37*)]. We found a main effect of transition type and a significant interaction: transitional probabilities within subsequences were higher than between subsequences ($\beta = 0.39$, $SE = 0.037$, $P < 0.00001$) as would be expected since this was the cue we were using to segment sequences, but this difference increased with development ($\beta = 0.0005$, $SE = 0.0002$, $P = 0.033$). This could occur if the juvenile song is a noisy version of the adult target, reducing the statistical coherence of the detected subsequences in the juvenile song. To examine this, we compared the song bouts produced by each bird to song bouts produced by that bird's adult tutor. We used the inverse of normalized Levenshtein string distance to quantify this (Fig. 3B). We then averaged over all the comparisons for a particular recording (since each recording contains 9 to 10 song bouts, this is based on 90 to 100 comparisons). We ran a mixed-effects model on average similarity with a fixed effect of day and a random effect of bird. Similarity increased by day, meaning that the detected subsequences became more like the adult target over development ($\beta = 0.0021$, $SE = 0.0067$, $P = 0.0044$).

Detected subsequences follow a Zipfian distribution throughout development

The developmental data allows us to investigate whether the fit to a Zipfian distribution changes during learning. To examine this, we looked at the effect of day of recording on R^2 (we also ran this analysis using recording number rather than day, and found the same pattern of results, see supplementary materials). Figure 3C shows the fit to a Zipfian distribution (assessed using R^2) during development. We ran a mixed-effect model with bird as a random effect and found no effect of day on the fit to a power law ($\beta = 0.0003$, $SE = 0.0005$, $P = 0.56$). That is, the fit remained similarly high throughout development.

The slope of the Zipfian distribution increases during development

The presence of this distribution throughout song development is intriguing. Although the detected subsequences initially differ from

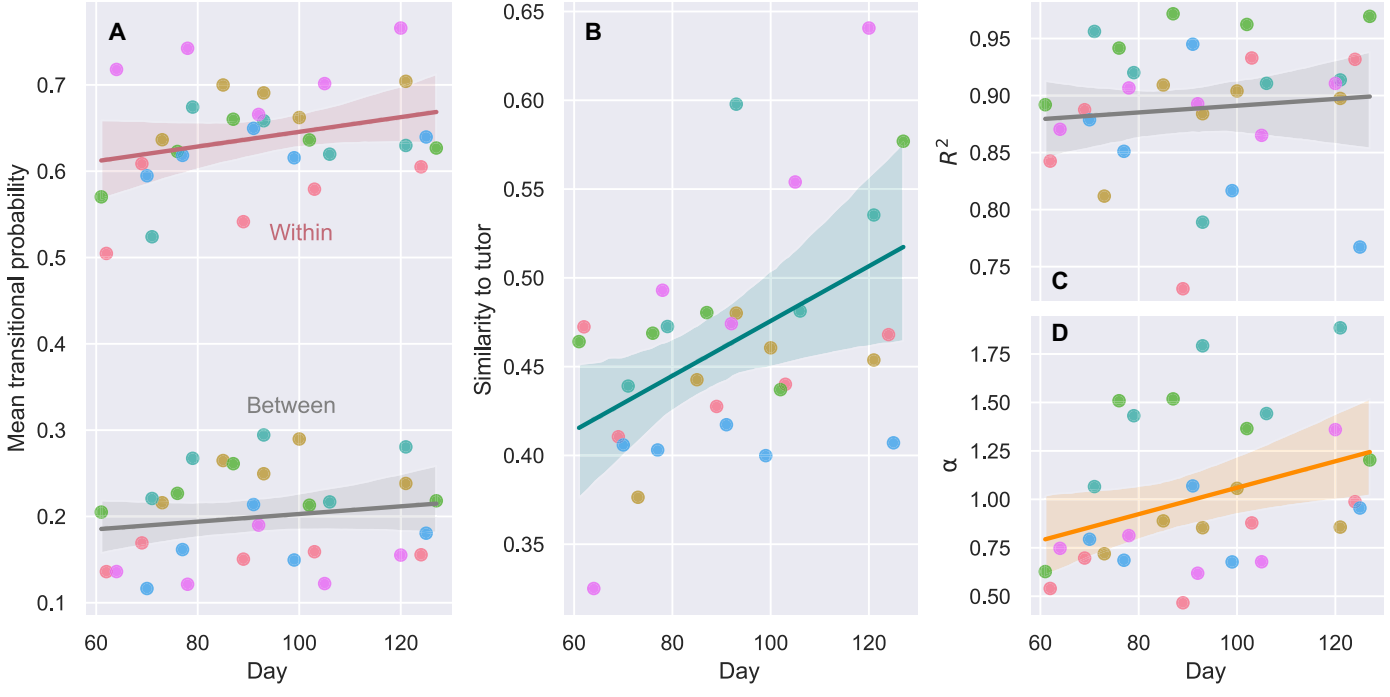

Fig. 3. Detected subsequences follow a Zipfian distribution throughout development, but their statistical coherence and slope increase as they approach the adult target. Our dataset consists of recordings from six Bengalese finches, each recorded on five days during development, from the earliest analyzable song to fully crystalized song. In these graphs, each bird is shown by dots with a particular color. (**A**) Statistical coherence increases over development. The cue for segmentation becomes increasingly clear over development, reflected in a growing difference between the transitional probabilities within subsequences and between subsequences. (**B**) The songs gradually approach the adult target, demonstrated by their growing similarity to song bouts of the adult tutor for that bird (measured using the inverse of normalized Levenshtein string distance between all pairs of song bouts, where a score of 1 indicates identical sequences). (**C**) Detected subsequences fit a Zipfian distribution throughout development, and the R^2 value does not change by day of recording. (**D**) Last, the slope of the Zipfian distribution—the exponent alpha of the power law—increases over development (see Materials and Method for details), meaning the distribution becomes more skewed.

the target, their frequency distribution nevertheless follows a power law. However, the slope of that distribution does increase with development (see Fig. 3D). We ran a mixed-effects model with a random effect of bird to confirm this ($\beta = 0.006$, SE = 0.002, $P = 0.0056$). This means that although the distribution is Zipfian from the start, it becomes more skewed as the song approaches the adult target. As an additional confirmation of this effect, we ran a combined model predicting log frequency from log count directly, with an interaction term for day and a random effect of bird. If the distribution becomes more skewed during development, then we would expect the interaction with day to be significant, which is what we found (full details are in the Supplementary Materials) ($\beta = 0.006$, SE = 0.0007, $P < 0.00001$).

Subsequences in Bengalese song follow Zipf's law of brevity

Zipf's law of brevity states that more frequent units are typically shorter. This pattern is found across many species, including songbirds, and is taken as evidence for optimal coding (*18*, *38*–*43*). If the detected units are relevant to the birds, then we may expect them to follow Zipf's law of brevity. As Fig. 4 shows, there is a strong relationship between frequency and length: Longer subsequences tend to be less frequent than shorter ones [$\beta = -0.04$, SE = 0.0045, $P < 0.00001$ using Poisson mixed effects regression, with bird as a random effect, following Youngblood (*18*)].

DISCUSSION

In this paper, we test the hypothesis that cultural transmission underlies recently found similarities between human language and whale song. This hypothesis makes the strong prediction that we should find similar structure whenever complex sequential signaling is culturally transmitted. Using the same segmentation pipeline previously applied to humans and humpback whales, we find that Bengalese finch song does indeed exhibit this structure: The song is made up of statistically coherent subsequences whose distribution follows a power law, with a slope similar to that found in human language. This distribution is not commonly found in animal communication: While Zipf's law of brevity (where more frequent elements tend to be shorter) has been documented across species (*38*, *40*, *41*, *43*), there is much less evidence for the presence of Zipfian frequency distributions. Relatively few studies evaluate the law, and the ones that do report mixed findings. Zipfian distributions, or the less constrained Zipf-Mandelbrot distribution, have been reported for individual sound elements in some species [songbirds (*18*), dolphins (*44*), and humpback whales (*45*)], but not in others [marmoset vocalizations (*46*), skylark calls (*21*), and California Chickadee calls (*47*)]. Prior work has assessed the distribution of individual elements, not larger sequences, which may be prone to different learning pressures. Finding statistically coherent subsequences that follow a Zipfian distribution in humans (*1*, *2*, *7*, *13*), humpback whales (*15*), and now Bengalese finches demonstrates that cultural evolution can lead to convergent properties in unrelated species.

Bengalese finches provide an ideal species to study because we have access to recordings of birds at different stages of song development, and we know that they are sensitive to transitional probabilities during song learning (*29*), the same cue used by infants to segment speech (*4*), and by our segmentation method. Using a corpus of song development, we found that the statistical coherence of subsequences increases over development, as does the slope of the distribution, which gradually approaches the adult target, but, nevertheless, the subsequences follow a Zipfian frequency distribution throughout. The developmental change in slope that we found has also been suggested for the distribution of individual whistles in dolphins [but see (*48*)], where the slope of the distribution of whistles produced by infant dolphins was lower than that of ones produced by adult dolphins (*44*). That is, the juvenile song of Bengalese finches seems like a noisy version of the adult target in some respects, but not others. This raises the intriguing possibility that Zipfian distributions reflect a more general organizational principle underpinning the representation of complex behavior, even when those representations differ

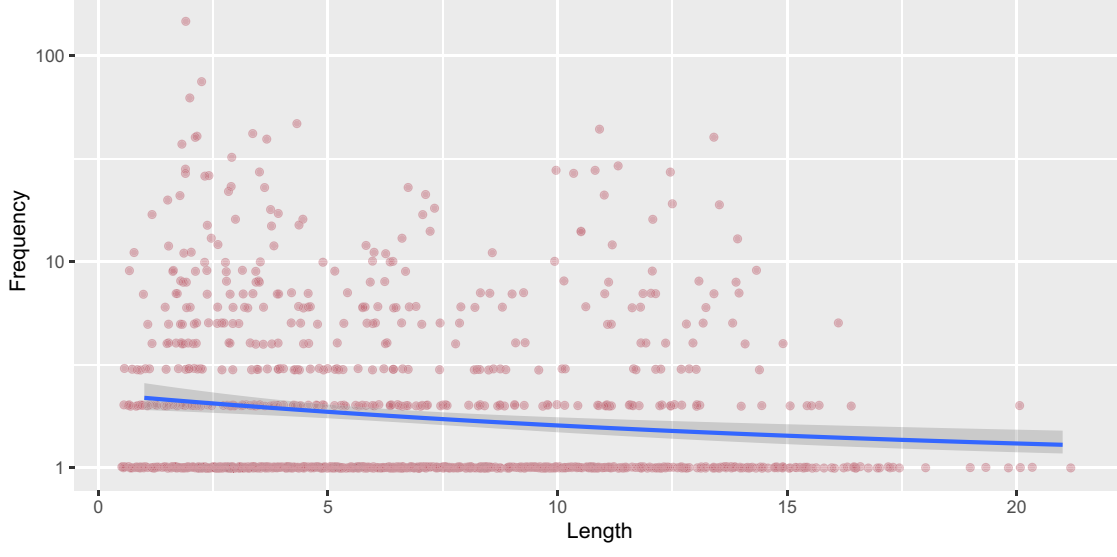

Fig. 4. Detected subsequences follow Zipf's law of brevity, frequent sequences are shorter. In this graph, we plot the frequency of the detected subsequences in each recording by their length (in syllables). The data show a strong relation between frequency and length, assessed using a Poisson mixed effects regression ($P < 0.00001$). The blue line shows a Poisson regression line collapsing over all data.

from the adult target. In our case, this means that although the juvenile song subsequences differ from the adult ones, they nevertheless follow a Zipfian frequency distribution. The idea that Zipfian distributions are a fundamental organizational principle receives some support from analyses of their presence in human language data. A Zipfian distribution is found not only when looking at the lexicon as a whole but also within different parts of speech [e.g., nouns, verbs, and prepositions (*32*)] and semantic categories [e.g., planet names and numbers (*2*)]. That is, the same organizational structure found in the whole domain (the lexicon) is also reflected in specific categories within it.

There are, of course, substantial differences between human language on the one hand and whale song and birdsong on the other. A primary function of language is to communicate and convey meaning, whereas whales and birds use song primarily for sexual display (*49*). In other words, the detected subsequences in Bengalese finch songs are not expected to convey semantic meaning. This illustrates that key statistical properties of language can be found in culturally transmitted systems even in the absence of semantic content. This begs the question of why Zipfian distributions have not been found more widely in other species with learned communication systems. There is evidence that the distribution of individual sound elements in such systems is highly skewed, with some studies reporting a fit to a Zipfian distribution (*44*, *50*, *51*) and more finding a fit to the more convex Zipf-Mandelbrot extension (*18*–*23*, *45*). However, it is very hard to evaluate what counts as a "good fit" to a particular distribution (*20*, *48*, *50*). Our approach addresses this challenge by moving away from simply reporting the fit to a power law and instead comparing that fit to baselines. We can do this because we segment the song into statistically coherent parts and analyze their distribution, rather than look at the distribution of the individual sound elements. This allows us to generate baseline distributions and show that the fit to a power law that we are getting is driven by the sequential structure of the song, and the particular cues we are using to segment it. Applying this approach more broadly may reveal similar statistical structure in the communication systems of a wider range of species. Note, however, the directionality of our prediction: We expect culturally transmitted communication systems to be Zipfian, but we do not expect all Zipfian distributions to necessarily have a cultural origin (*3*).

These findings illustrate a parallel between three species separated by hundreds of millions of years of evolution. Humans, humpback whales, and Bengalese finches all have signaling systems made up of statistically coherent parts whose frequency distribution follows a power law. Our claim is that these properties arise due to their facilitative effect on learning and are found in these three species because their signaling systems are culturally transmitted, a process that leads to properties that maximize learnability. Looking across very different species using the same analytic techniques has uncovered previously undetected structure in their communication systems while also revealing the origins of properties of language previously thought to be uniquely human.

MATERIALS AND METHODS

Song recordings

The dataset is from (*30*). Bengalese finches typically start singing around day 35 at the earliest, but in some individuals, singing could start as late as day 50. Syllable classification is not possible until productions are stable enough, around days 60 to 70; therefore, the annotation of recorded data started at around days 60 to 70. The earliest developmental points are from around day 60, not far into song development (*36*). Birds were recorded every 2 weeks (on average) until the bird was around 120 days old. At this point, Bengalese finch songs are crystallized and do not develop further (*36*). Each song was recorded in a sound-proof chamber (640 mm by 1295 mm by 1720 mm, Science cabin) using a DAT recorder (Sony ZA5ES, Tokyo, Japan) or a personal computer (PC) with a microphone (Sony ECM- MS957). A song bout was defined as serial sound production separated from the next bout by an interval of longer than 500 msec (*31*). The recordings were carried out in compliance with Japanese ethical procedures at the time of the original study. SASLab Pro sound analysis (Avisoft, Germany) software was used to make sonograms for each bird. Song syllables were identified by eye by the lead researchers (see fig. S1). Each syllable was assigned a unique letter, and songs were transformed into a sequence of letters. Reliability was assessed by having two additional experienced coders identify syllables. Agreement between all coders was very high (93.3%) (*30*).

Segmentation pipeline

Our analysis was carried out on each recording separately. In other words, we ran the pipeline 30 times (six birds × five recordings per bird) for each cutting ratio (0.3, 0.5, and 0.7). The recordings contain previously coded sequences of syllables. Each recording contained 9 or 10 song bouts. We first estimated the transitional probabilities between song syllables in each recording by counting the frequency of each distinct syllable given each distinct previous syllable (we did not calculate frequencies for syllables that crossed song bout boundaries since these mark the end one of song bout and the start of another). We then scanned through each song bout, looking at pairs of adjacent transitional probabilities and posited a cut whenever the ratio between the previous transitional probability and the current one was smaller than our cutting ratio threshold (0.3, 0.5, and 0.7). The detected subsequences (across all song bouts) were added to the inventory of subsequences for that recording. We then calculated the frequency and length distributions of all the subsequences in the recording. The segmentation pipeline was implemented in Python and is available as code S1.

Statistical analyses

We calculated R^2 for the relation between the log frequency and log rank of the detected subsequences using Pearson's product moment correlation. The shuffled baseline involved randomizing the order of the syllables within each song bout in each recording, and running the pipeline again. We did this 1000 times. We then compare the mean R^2 across all the real song bouts (i.e., a single number, calculated over the 299 song bouts) to a distribution of the mean R^2 (calculated more than 299 shuffled song bouts) for each of the 1000 shuffled datasets.

To implement the bigram baseline, we generated multiple pseudo datasets whose bigram distribution was sampled from the distribution of the real song bouts. Each of these pseudo datasets contains one pseudo song bout for each real song bout in our data. To generate a pseudo song bout for a particular target song bout, we go through the following steps: (i) Start with an empty pseudo song bout; (ii) pick a random point in the target song bout and add the syllable found at that point to the pseudo song bout; (iii) find all of the occurrences of the current last syllable in the pseudo song bout and randomly select one of those, adding the syllable following it to the

pseudo song bout (if the syllable only occurs at the end of the target song bout, go to step (ii). Repeat step (iii) until the pseudo song bout is the same length as the target song bout. We ran our pipeline on each of these bigram-based datasets, calculated the distribution of the detected subsequences, and examined the fit to a Zipfian distribution.

The rotated baseline involved taking the cutting points from each song bout in the original data and shifting them by a random number of steps (e.g., if the random number was 5 for a particular song bout, a cut at position 10 of the sequence would now be placed in position 15). We rotated any steps that fell off the end of the song bout back to the start again. We then compare the mean R^2 across all the real recordings to that of the distribution of mean R^2 for each of 1000 rotated datasets.

To estimate the slope of the power law, we used the scipy.optimize library to fit the data for each recording to the power law using a maximum likelihood approach with the Nelder-Mead algorithm and report the exponent α in $f \approx \frac{1}{r^{-\alpha}}$, where f is frequency and r is rank frequency.

To compare the song bouts in each recording to the adult target, we used Levenshtein string edit distance, with insertion, deletion, and substitution costs all set to 1. We normalized the edit distance by dividing by the length of the longest string and subtracting the result from 1. This creates a string similarity measure, where 1 indicates identical strings. For a given recording, we computed the average similarity of all of the song bouts in that recording to all of the song bouts in the song bouts of the adult tutor for that particular bird (so that for all recordings but one, this would involve 100 comparisons). We used linear mixed effects regressions to determine the effect of day of recording on our various measures (reported in the main text). We used the maximal random effects structure that led to convergence (*37*). We repeated these analyses after recoding day of recording from a continuous variable indicating the age of the bird into a consecutive number, such that the first recording of a particular bird is given the number 1 and the final recording of that bird is given the number 5. These analyses give an identical pattern of results and are reported in code S1. To evaluate Zipf's law of brevity (the relationship between frequency and length), we use a Poisson mixed effects regression with frequency (in a recording) as the outcome variable and a fixed effect for length in syllables, with random intercepts for bird, following Youngblood (*17*).

Supplementary Materials

This PDF file includes:

Figs. S1 to S3
Code S1
Dataset

REFERENCES

1. G. Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* (Addison-Wesley Press, 1949).
2. S. T. Piantadosi, Zipf's word frequency law in natural language: A critical review and future directions. *Psychon. Bull. Rev.* **21**, 1112–1130 (2014).
3. M. Newman, Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **46**, 323–351 (2005).
4. J. R. Saffran, R. N. Aslin, E. L. Newport, Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928 (1996).
5. C. Kurumada, S. C. Meylan, M. C. Frank, Zipfian frequency distributions facilitate word segmentation in context. *Cognition* **127**, 439–453 (2013).
6. A. T. Hendrickson, A. Perfors, Cross-situational learning in a Zipfian environment. *Cognition* **189**, 11–22 (2019).
7. O. Lavi-Rotbain, I. Arnon, The learnability consequences of Zipfian distributions in language. *Cognition* **223**, 105038 (2022).
8. L. Wolters, O. Lavi-Rotbain, I. Arnon, Zipfian distributions facilitate children's learning of novel word-referent mappings. *Cognition* **253**, 105932 (2024).
9. L. Wolters, M. Ota, I. Arnon, Skewed distributions facilitate infants' word segmentation. *Cognition* **263**, 106221 (2025).
10. S. Kirby, Culture and biology in the origins of linguistic structure. *Psychon. Bull. Rev.* **24**, 118–137 (2017).
11. S. Kirby, H. Cornish, K. Smith, Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10681–10686 (2008).
12. S. Kirby, M. Dowman, T. L. Griffiths, Innateness and culture in the evolution of language. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5241–5245 (2007).
13. I. Arnon, S. Kirby, Cultural evolution creates the statistical structure of language. *Sci. Rep.* **14**, 5255 (2024).
14. E. C. Garland, A. W. Goldizen, M. L. Rekdahl, R. Constantine, C. Garrigue, N. D. Hauser, M. M. Poole, J. Robbins, M. J. Noad, Dynamic horizontal cultural transmission of humpback whale song at the ocean basin scale. *Curr. Biol.* **21**, 687–691 (2011).
15. I. Arnon, S. Kirby, J. A. Allen, C. Garrigue, E. L. Carroll, E. C. Garland, Whale song shows language-like statistical structure. *Science* **387**, 649–653 (2025).
16. P. Marler, Three models of song learning: Evidence from behavior. *J. Neurobiol.* **33**, 501–516 (1997).
17. F. Nottebohm, Reassessing the mechanisms and origins of vocal learning in birds. *Trends Neurosci.* **14**, 206–211 (1991).
18. M. Youngblood, Language-like efficiency and structure in house finch song. *Proc. Biol. Sci.* **291**, 20240250 (2024).
19. V. Deslandes, L. R. Faria, M. E. Borges, M. R. Pie, The structure of an avian syllable syntax network. *Behav. Processes* **106**, 53–59 (2014).
20. A. Kershenbaum, V. Demartsev, D. E. Gammon, E. Geffen, M. L. Gustison, A. Ilany, A. R. Lameira, Shannon entropy as a robust estimator of Zipf's Law in animal vocal communication repertoires. *Methods Ecol. Evol.* **12**, 553–564 (2021).
21. E. Briefer, T. S. Osiejuk, F. Rybak, T. Aubin, Are bird song complexity and song sharing shaped by habitat structure? An information theory and statistical approach. *J. Theor. Biol.* **262**, 151–164 (2010).
22. J. P. Hailman, M. S. Ficken, R. W. Ficken, The 'chick-a-dee'calls of Parus atricapillus: A recombinant system of animal communication compared with written English. *Semiotica* **56**, 191–224 (1985).
23. M. L. Cody, E. Stabler, H. M. Sánchez Castellanos, C. E. Taylor, Structure, syntax and "small-world" organization in the complex songs of California Thrashers (Toxostoma redivivum). *Bioacoustics* **25**, 41–54 (2016).
24. K. Stärk, E. Kidd, R. L. A. Frost, Word segmentation cues in german child-directed speech: A corpus analysis. *Lang. Speech* **65**, 3–27 (2022).
25. K. Okanoya, Song syntax in bengalese finches: Proximate and ultimate analyses. *Adv. Study Behav.* **34**, 297–346 (2004).
26. E. Honda, K. Okanoya, Acoustical and syntactical comparisons between songs of the white-backed munia (*Lonchura striata*) and its domesticated strain, the bengalese finch (*Lonchura striata var. domestica*). *Zool. Sci.* **16**, 319–326 (1999).
27. K. Okanoya, Behavioural factors governing song complexity in bengalese finches. *Int. J. Comp. Psychol.* **25**, 44–59 (2012).
28. N. S. Clayton, Song learning in bengalese finches: A comparison with zebra finches. *Ethology* **76**, 247–255 (1987).
29. M. Takahasi, H. Yamada, K. Okanoya, Statistical and prosodic cues for song segmentation learning by bengalese finches (*Lonchura striata var. domestica*). *Ethology* **116**, 481–489 (2010).
30. M. Takahasi, K. Okanoya, Song learning in wild and domesticated strains of white-rumped munia, *Lonchura striata*, compared by cross-fostering procedures: Domestication increases song variability by decreasing strain-specific bias. *Ethology* **116**, 396–405 (2010).
31. T. Hosino, K. Okanoya, Lesion of a higher-order song nucleus disrupts phrase level complexity in Bengalese finches. *Neuroreport* **11**, 2091–2095 (2000).
32. O. Lavi-Rotbain, I. Arnon, Zipfian distributions in child-directed speech. *Open Mind* **7**, 1–30 (2023).
33. A. Mehri, M. Jamaati, Variation of Zipf's exponent in one hundred live languages: A study of the Holy Bible translations. *Phys. Lett. A* **381**, 2470–2477 (2017).
34. G. M. Linders, M. M. Louwerse, Zipf's law revisited: Spoken dialog, linguistic units, parameters, and the principle of least effort. *Psychon. Bull. Rev.* **30**, 77–101 (2023).
35. K. Katahira, K. Suzuki, K. Okanoya, M. Okada, Complex sequencing rules of birdsong can be explained by simple hidden Markov processes. *PLOS ONE* **6**, e24516 (2011).
36. K. Sasahara, O. Tchernichovski, M. Takahasi, K. Suzuki, K. Okanoya, A rhythm landscape approach to the developmental dynamics of birdsong. *J. R. Soc. Interface* **12**, 20150802 (2015).
37. D. J. Barr, R. Levy, C. Scheepers, H. J. Tily, Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* **68**, 255–278 (2013).

38. S. Semple, M. J. Hsu, G. Agoramoorthy, Efficiency of coding in macaque vocal communication. *Biol. Lett.* **6**, 469–471 (2010).
39. R. Heesen, C. Hobaiter, R. Ferrer-i-Cancho, S. Semple, Linguistic laws in chimpanzee gestural communication. *Proc. R. Soc. B Biol. Sci.* **286**, 20182900 (2019).
40. M. Huang, H. Ma, C. Ma, P. A. Garber, P. Fan, Male gibbon loud morning calls conform to Zipf's law of brevity and Menzerath's law: Insights into the origin of human language. *Anim. Behav.* **160**, 145–155 (2020).
41. R. Ferrer-i-Cancho, A. Hernández-Fernández, D. Lusseau, G. Agoramoorthy, M. J. Hsu, S. Semple, Compression as a universal principle of animal behavior. *Cognit. Sci.* **37**, 1565–1578 (2013).
42. R. Ferrer-i-Cancho, C. Bentz, C. Seguin, Optimal coding and the origins of Zipfian laws. *J. Quant. Linguist.* **29**, 165–194 (2022).
43. L. Favaro, M. Gamba, E. Cresta, E. Fumagalli, F. Bandoli, C. Pilenga, V. Isaja, N. Mathevon, D. Reby, Do penguins' vocal sequences conform to linguistic laws? *Biol. Lett.* **16**, 20190589 (2020).
44. B. McCowan, S. F. Hanser, L. R. Doyle, Quantitative tools for comparing animal communication systems: Information theory applied to bottlenose dolphin whistle repertoires. *Anim. Behav.* **57**, 409–419 (1999).
45. J. A. Allen, E. C. Garland, R. A. Dunlop, M. J. Noad, Network analysis reveals underlying syntactic features in a vocally learnt mammalian display, humpback whale song. *Proc. R. Soc. B Biol. Sci.* **286**, 20192014 (2019).
46. Y. B. Gultekin, D. G. Hildebrand, K. Hammerschmidt, S. R. Hage, High plasticity in marmoset monkey vocal development from infancy to adulthood. *Sci. Adv.* **7**, eabf2938 (2021).
47. T. M. Freeberg, J. R. Lucas, Information theoretical approaches to chick-a-dee calls of Carolina chickadees (*Poecile carolinensis*). *J. Comp. Psychol.* **126**, 68–81 (2012).
48. R. Suzuki, J. R. Buck, P. L. Tyack, The use of Zipf's law in animal communication analysis. *Anim. Behav.* **69**, 9–17 (2005).
49. E. C. Garland, P. K. McGregor, Cultural transmission, evolution, and revolution in vocal displays: Insights from bird and whale song. *Front. Psychol.* **11**, 544929 (2020).
50. B. McCowan, L. R. Doyle, J. M. Jenkins, S. F. Hanser, The appropriate use of Zipf's law in animal communication studies. *Anim. Behav.* **69**, 1–7 (2005).
51. R. Ferrer-i-Cancho, D. Lusseau, B. McCowan, Parallels of human language in the behavior of bottlenose dolphins. *Linguist. Front.* **5**, 5–11 (2022).

Acknowledgments

Funding: This work is supported by the Israel Science Foundation grant 445/20 (to I.A.), MEXT KAENHI Grant-in-Aid for Scientific Research (S) JP23H05428 and Grant-in-Aid for Transformative Research Areas (A) JP24H01560 and JP25H01949 (to K.O.), and the Royal Society University Research Fellowship UF160081 and URF/R/221020 (to E.C.G.) **Author contributions:** S.K., I.A., E.C.G., and K.O. conceived of the study. K.O. and M.T. collected and curated the data. K.O. and M.T. transcribed and analyzed the bird song data. I.A. and S.K. extended the human infant–inspired segmentation pipeline to bird song, with input from K.O. S.K. conducted the statistical analyses and generated all the visualizations except S3, created by K.O. S.K. and I.A. wrote the first drafts of the manuscript, and S.K., K.O., E.C.G., and I.A. contributed to the final draft. **Competing interests:** The authors declare that they have no competing interests. **Data, code, and materials availability:** All data and code needed to evaluate and reproduce the results in the paper are present in the paper and/or the Supplementary Materials. All materials needed to replicate the results are available in the Supplementary Materials.

Submitted 3 July 2025
Accepted 30 June 2026
Published 7 August 2026
10.1126/sciadv.aea3015

ScienceAdvances

Language-like statistical structure arises in learned signaling: evidence from birdsong

Simon Kirby, Kazuo Okanoya, Ellen C. Garland, Miki Takahasi, and Inbal Arnon

Sci. Adv. **12** (32), eaea3015. DOI: 10.1126/sciadv.aea3015

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.