Supplementary information

Language-like efficiency and structure in house finch song

Mason Youngblood [[1]](#footnote-20) [[2]](#footnote-22)

# 1 Data

Recordings from 1975 were collected with a Nagra III reel-to-reel tape recorder and a Sennheiser 804 shotgun microphone and converted to digital files (32 bit, 96 kHz) by the Cornell Lab of Ornithology in 2013 ([1](#ref-Ju2019)). Recordings from 2012 (16 bit, 44.1 kHz) and 2019 (32 bit, 48 kHz) were collected with a Marantz PD661 solid-state recorder and a Sennheiser ME66 shotgun microphone. The recordings from 1975 were downsampled to 48 kHz prior to analysis (Luscinia processes both 44.1 kHz and 48 kHz files). In all three years special precautions were taken to avoid recording the same bird twice ([1](#ref-Ju2019)). Each site was visited only once. Within a site, only one individual was recorded within a 160 m radius until they stopped singing or flew away.

All songs were analyzed by ([3](#ref-youngbloodContentBiasCultural2022)) using Luscinia, a database and analysis program developed specifically for birdsong (<https://rflachlan.github.io/Luscinia/>). Songs were analyzed with a high-pass threshold of 2000 Hz, a maximum frequency of 9000 Hz and 5 dB of noise removal. 965 songs (26.2%) were excluded from the analysis due to high levels of noise. Continuous traces with more than 20 ms between them were classified as syllables ([2](#ref-Mundinger1975)).

Figure S1 shows an example of the first ten syllables of a house finch song recorded by ([2](#ref-Mundinger1975)) and analyzed in Luscinia, where the blue line is the mean frequency over time.

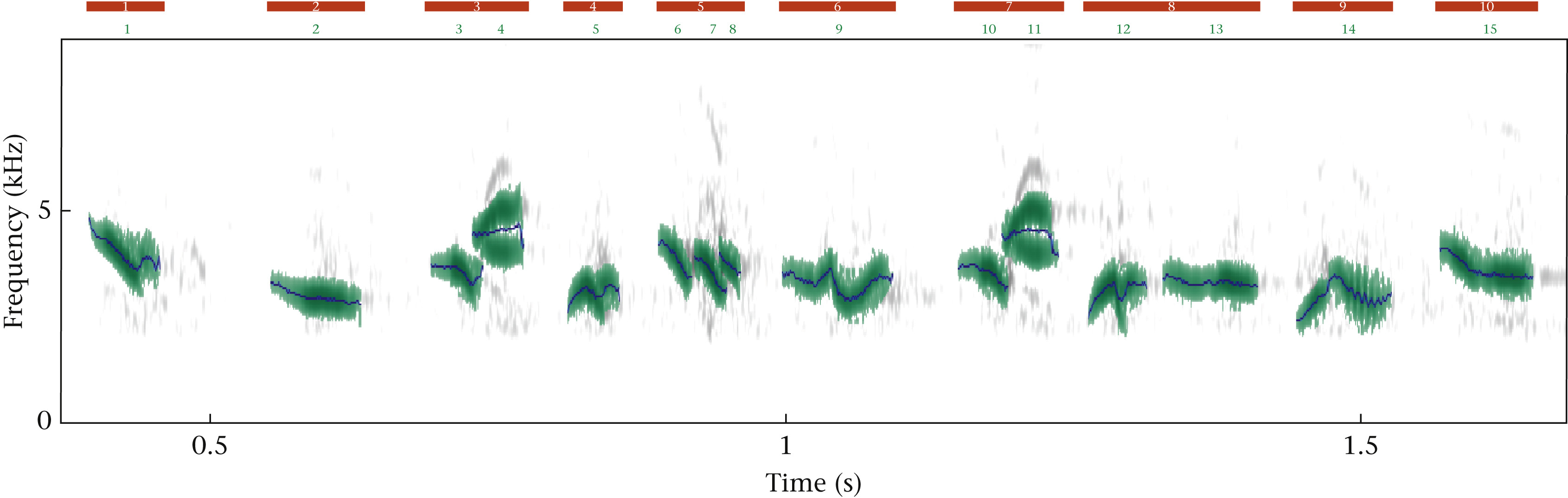


Figure S1: The first 10 syllables from a song recorded by ([2](#ref-Mundinger1975)) and analyzed in Luscinia. Each red bar corresponds to a syllable, and each green number corresponds to an element within that syllable. The blue traces represent mean frequency. In this song, syllable 3 and syllable 7 were classified as the same syllable type during dynamic time warping, and all other syllables are unique types. Reprinted from ([3](#ref-youngbloodContentBiasCultural2022)).

# 2 Clustering

The deep split parameter () determines the granularity of clustering by controlling the value of two other parameters: the maximum core scatter (), which controls the maximum within-group variation, and the minimum gap (), which controls the minimum between-group variation ([4](#ref-langfelder_etal08),[5](#ref-ju16)). correspond to , and . In their analyses of house finch song, ([1](#ref-Ju2019)) and ([6](#ref-Roginek2018)) manually set and while ([3](#ref-youngbloodContentBiasCultural2022)) used (corresponding to and ), all of which led to a similar number of syllable types. I will follow ([3](#ref-youngbloodContentBiasCultural2022)) in using the simpler deep split parameter to control granularity in clustering, as this approach is recommended by the creators of dynamic tree cut ([4](#ref-langfelder_etal08)) and has been widely used for a variety of applications ([7](#ref-liu_etal22),[8](#ref-zhao_etal20)) including vocal analysis ([9](#ref-burkett_etal15)).

## 2.1 Sampling Coverage

Unseen species models, fit using the iNEXT package in R ([10](#ref-hsiehINEXTPackageRarefaction2016)), were used to assess sampling coverage (see caption for details).

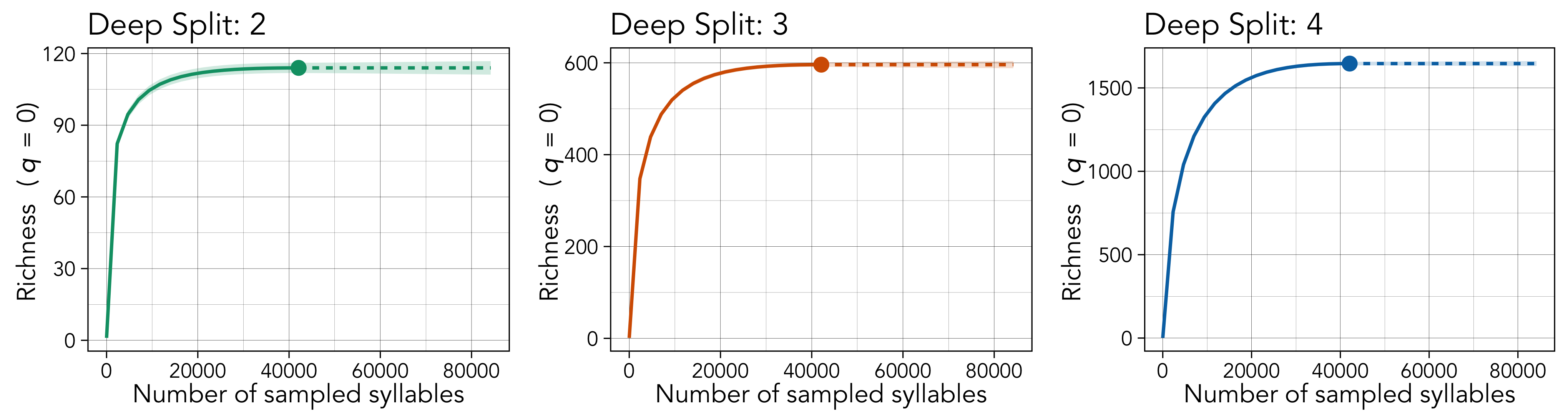


Figure S2: The results of unseen species models applied to the full frequency distribution of syllables classified at each level of deep split, run using the iNEXT package in R ([10](#ref-hsiehINEXTPackageRarefaction2016)). The point in each plot marks the observed richness (y-axis) for the total number syllables sampled (x-axis). The solid lines mark interpolated richness, the dashed lines mark extrapolated richness, and the shaded areas mark 95% confidence intervals. The interpolated richness values are the number of types you observe if you subsample that number of tokens from the real data. The extrapolated values are the projected number of types you would observe if you were able to continue sampling tokens from the same population. More details can be found in Hsieh et al. ([10](#ref-hsiehINEXTPackageRarefaction2016)): <https://doi.org/10.1111/2041-210X.12613>. We appear to have complete coverage of syllable types regardless of the level of deep split that is used.

## 2.2 Assessment of Duration Bias

Automated clustering methods may detect more variation in the features of longer syllables compared to shorter syllables, leading to a bias in classification where longer syllables are parsed into more categories. The risk of such bias should be minimized due to the use of dynamic time warping, which warps signals in the time domain (in this study up to 10% of the average signal length) before computing similarity ([11](#ref-Ratanamahatana2005)), and dynamic tree cut, which varies the granularity of classification to maximize the connectedness of items at the tips of branches ([4](#ref-langfelder_etal08)), but it is still worth highlighting.

If a hierarchical clustering algorithm is less sensitive to variation in short syllables, then it should tend to over-lump them into types with longer syllables at all cut heights. However, it is likely to over-lump them less at lower cut heights, where syllables are already being parsed into types with greater granularity. Based on this logic, I would predict two things from a hierarchical clustering algorithm that is biased with respect to duration:

1. First, I would expect the distribution of the median durations of syllable types to vary across different levels of granularity (tested in Figure S3A).
2. My second prediction requires more explanation. As granularity in clustering changes, the type that a syllable token is assigned to changes as well. This means that we can track the average duration of types that individual syllables are assigned to as granularity increases, which I will refer to as “trajectories”. I would expect a biased clustering algorithm to lead to biased trajectories, where the duration of types that syllables are assigned to changes systematically in one direction over time (tested in Figure S3B).

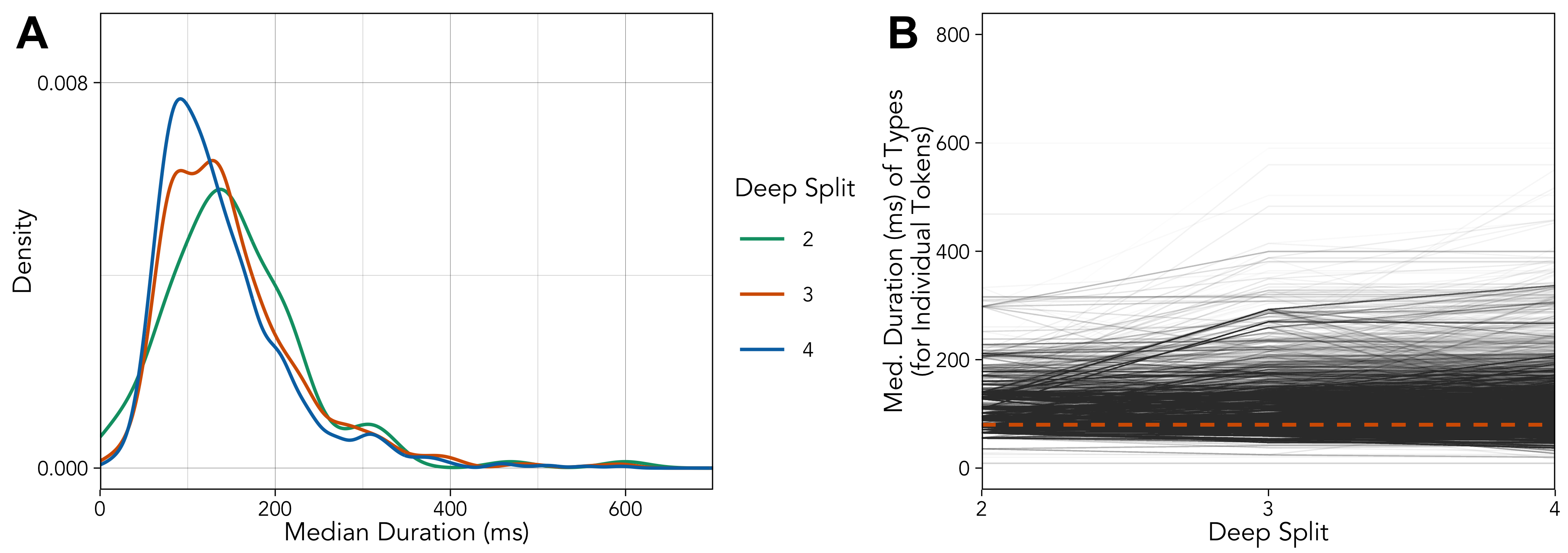


Figure S3: Panel A shows the median duration of types from each level of granularity in syllable clustering. The black lines in panel B are the “trajectories” of individual syllable tokens—the median duration of types that individual tokens are assigned to with increasing levels of granularity. The orange dotted line is the median value across all syllables.

There is a subtle decrease in the distribution of the median duration of types as the granularity of clustering increases (Figure S3A), but the effect is not significant in a simple linear model (duration ~ granularity) (p > 0.05). If we track the median duration of the types that individual tokens are assigned to at increasing levels of granularity (Figure S3B), there appears to be no clear trend (avg. is the orange line). A simple linear model with type as a random effect (duration ~ granularity + (1|type)) does detect a significant increase in the median duration of types that tokens are assigned to with increasing granularity (estimate: 0.0018; 95% CI: 0.00061, 0.0029), but this effect is extremely low and only comes out to an increase of about 1 ms per unit increase in deep split (for signals that are ~100 ms on average).

Based on this analysis, I think that duration-bias in clustering is unlikely to be an issue in the analysis, but just to be sure I have replicated the results for Zipf’s law of abbreviation with manually-classified syllables from an experimental cross-fostering study conducted in house finches ([12](#ref-Mann2020)). See [Replication with Data from Mann et al.](#replication-with-data-from-mann-et-al.) below for details.

## 2.3 Variance Captured by Types

The hierarchical clustering leads to syllable types that explain a large amount of the variation in the duration, bandwidth, and excursion of tokens, but only captures concavity at the higher levels of deep split.

Table S1: The proportion of the variance (R^2^) in the acoustic features of tokens that is accounted for by types identified at all three levels of granularity in syllable clustering, computed from frequentist models with the following specification: feature ~ (1|type).

| **Parameter** | **DS: 2** | **DS: 3** | **DS: 4** |
| --- | --- | --- | --- |
| Duration | 0.76 | 0.71 | 0.70 |
| Bandwidth | 0.72 | 0.84 | 0.90 |
| Concavity | 0.08 | 0.24 | 0.35 |
| Excursion | 0.36 | 0.49 | 0.59 |

# 3 Zipf’s Rank-Frequency Law

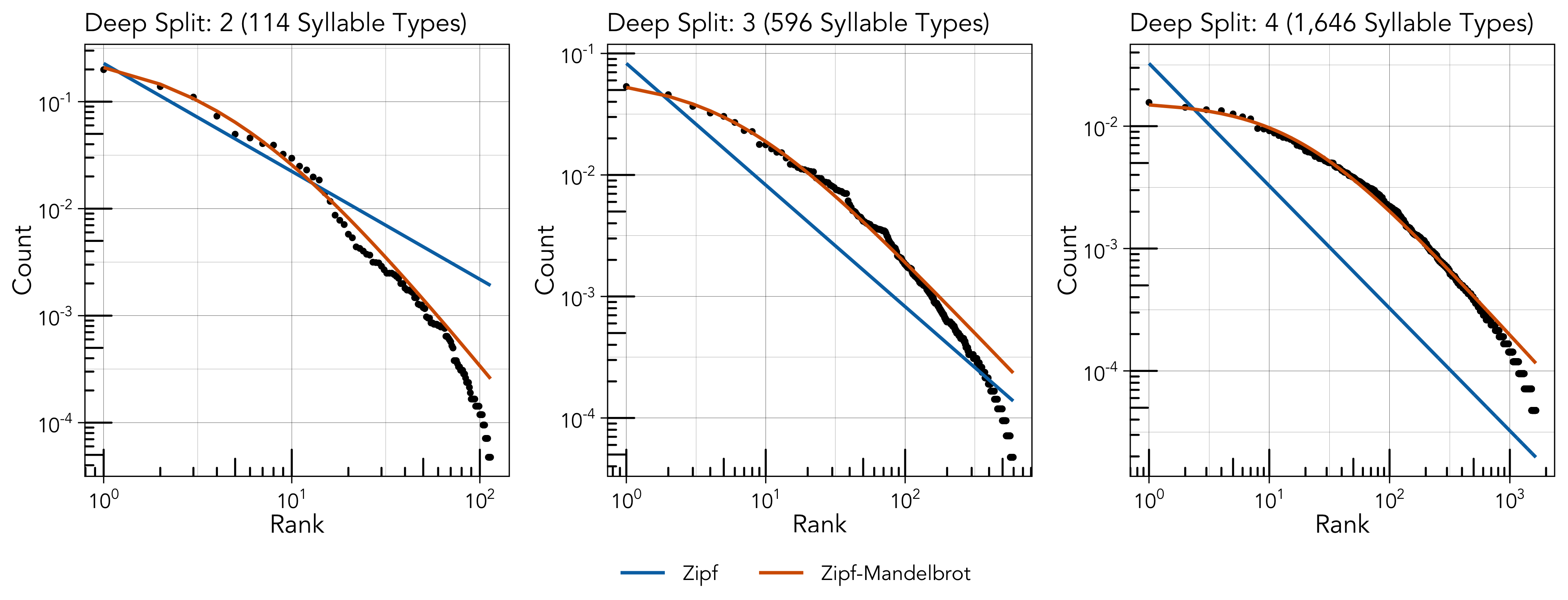


Figure S4: The relationship between rank (x-axis) and count (y-axis) at each level of deep split (left, center, and right). The blue and orange lines denote the expected distributions according to Zipf’s rank-frequency law (blue) and Mandelbrot’s extension of it (orange).

## 3.1 Expanded Description of the Model

Mandelbrot’s generalization of Zipf’s rank-frequency law takes the following form ([13](#ref-Mandelbrot1953),[14](#ref-Mandelbrot1962)):

where is the normalized frequency at each rank , is a normalization term, and and are parameters that control slope and convexity (respectively). According to ([15](#ref-izsak06)), the bounds of (1) are , , and . When , this function simplifies to the original form of Zipf’s rank-frequency law: .

is usually a normalization constant ([15](#ref-izsak06)–[17](#ref-macutek22)) defined as:

In practice, this form of Zipf’s rank-frequency law is notoriously difficult to fit to data due to strong correlations between and , which in turn determine ([15](#ref-izsak06)–[17](#ref-macutek22)). Here, I use a simplified version of (1) that treats as a third parameter that is estimated alongside and ([18](#ref-ausloos14)), as has been done in studies of chickadee calls ([19](#ref-hailman94)–[21](#ref-ficken_etal94)), which should be interpreted as an approximation of the Zipf-Mandelbrot distribution.

The model was fit as a non-linear model on the original scale, as opposed to the log-log scale.

## 3.2 Priors and Diagnostics

Table S2: Prior specification for the Zipf and Zipf-Mandelbrot models fit across the three levels of granularity.

| **Parameter** | **Class** | **Prior** | **Lower Bound** |
| --- | --- | --- | --- |
| a | b | normal(0, 10) | 1 |
| b | b | normal(0, 10) | -1 |
| c | b | normal(0, 10) | 0 |
|  | sigma | student\_t(3, 0, 2.5) | 0 |

Table S3: WAIC values from the Zipf and Zipf-Mandelbrot models fit across the three levels of granularity.

| **DS** | **Zipf** | **Zipf-Mandelbrot** |
| --- | --- | --- |
| 1 | -799 | -1,075 |
| 2 | -5,328 | -7,817 |
| 3 | -17,566 | -25,617 |

## 3.3 Deep Split: 2

Table S4: Estimates and diagnostics for the Zipf model.

| **Param.** | **Estimate** | **l-95% CI** | **u-95% CI** | **Rhat** | **Bulk\_ESS** | **Tail\_ESS** |
| --- | --- | --- | --- | --- | --- | --- |
| a | 1.01 | 1.00 | 1.03 | 1 | 5,510 | 3,580 |
| c | 0.23 | 0.22 | 0.24 | 1 | 7,223 | 6,440 |

Table S5: Estimates and diagnostics for the Zipf-Mandelbrot model.

| **Param.** | **Estimate** | **l-95% CI** | **u-95% CI** | **Rhat** | **Bulk\_ESS** | **Tail\_ESS** |
| --- | --- | --- | --- | --- | --- | --- |
| a | 2.20 | 2.03 | 2.39 | 1 | 2,129 | 2,702 |
| b | 4.66 | 4.00 | 5.41 | 1 | 2,132 | 2,653 |
| c | 9.43 | 5.22 | 16.68 | 1 | 2,102 | 2,515 |

## 3.4 Deep Split: 3

Table S6: Estimates and diagnostics for the Zipf model.

| **Param.** | **Estimate** | **l-95% CI** | **u-95% CI** | **Rhat** | **Bulk\_ESS** | **Tail\_ESS** |
| --- | --- | --- | --- | --- | --- | --- |
| a | 1.00 | 1.00 | 1.00 | 1 | 6,264 | 4,253 |
| c | 0.08 | 0.08 | 0.09 | 1 | 7,540 | 5,954 |

Table S7: Estimates and diagnostics for the Zipf-Mandelbrot model.

| **Param.** | **Estimate** | **l-95% CI** | **u-95% CI** | **Rhat** | **Bulk\_ESS** | **Tail\_ESS** |
| --- | --- | --- | --- | --- | --- | --- |
| a | 1.20 | 1.18 | 1.22 | 1 | 2,132 | 1,965 |
| b | 5.69 | 5.42 | 5.98 | 1 | 2,187 | 2,477 |
| c | 0.52 | 0.48 | 0.56 | 1 | 2,094 | 2,123 |

## 3.5 Deep Split: 4

Table S8: Estimates and diagnostics for the Zipf model.

| **Param.** | **Estimate** | **l-95% CI** | **u-95% CI** | **Rhat** | **Bulk\_ESS** | **Tail\_ESS** |
| --- | --- | --- | --- | --- | --- | --- |
| a | 1.00 | 1.00 | 1.00 | 1 | 6,266 | 4,117 |
| c | 0.03 | 0.03 | 0.03 | 1 | 8,434 | 7,188 |

Table S9: Estimates and diagnostics for the Zipf-Mandelbrot model.

| **Param.** | **Estimate** | **l-95% CI** | **u-95% CI** | **Rhat** | **Bulk\_ESS** | **Tail\_ESS** |
| --- | --- | --- | --- | --- | --- | --- |
| a | 1.08 | 1.07 | 1.09 | 1 | 2,079 | 2,879 |
| b | 17.29 | 16.79 | 17.79 | 1 | 2,076 | 3,021 |
| c | 0.34 | 0.32 | 0.36 | 1 | 2,032 | 2,774 |

## 3.6 Analysis by Year

Table S10: ΔWAIC comparing the fit of Zipf-Mandelbrot to Zipf's law separately to the data from each year at each level of deep split. Zipf-Mandelbrot provides a better fit in all conditions.

| **Year** | **DS: 2** | **DS: 3** | **DS: 4** |
| --- | --- | --- | --- |
| 1975 | -174 | -1,263 | -5,207 |
| 2012 | -156 | -1,958 | -5,518 |
| 2019 | -212 | -1,700 | -4,813 |

Table S11: The R² for the Zipf-Mandelbrot distribution fit separately to the data from each year at each level of deep split.

| **Year** | **DS: 2** | **DS: 3** | **DS: 4** |
| --- | --- | --- | --- |
| 1975 | 0.993 | 0.985 | 0.994 |
| 2012 | 0.987 | 0.991 | 0.986 |
| 2019 | 0.991 | 0.993 | 0.996 |

## 3.7 Departures at High Ranks

It appears that the Zipf-Mandelbrot distribution has a poorer fit to the rarer syllable types. Systematic departure from the Zipf-Mandelbrot distribution at high ranks is a well-known phemonenon in human language ([22](#ref-ferrericancho_sole01)–[24](#ref-piantadosi14)). It is thought to be caused by suboptimal sample sizes ([25](#ref-gorina_etal20)) and other data quality issues ([26](#ref-cristelli_etal12)), and can be reproduced by more sophisticated models of Zipf’s rank-frequency law ([27](#ref-montemurro_zanette02),[28](#ref-tunnicliffe_hunter22)).

Below are some simple simulations that demonstrate this principle in the house finch song data, conducted with the following steps at the intermediate level of granularity in syllable clustering (deep split of 3):

1. The fitted parameter values of the Zipf-Mandelbrot distribution (a = 1.20, b = 5.69, c = 0.52) were used to generate the expected frequencies of the 596 possible syllable types.
2. Tokens were pseudorandomly sampled with probability equal to the expected frequencies of the types.
3. The simulated frequency and rank of types was plotted alongside the theoretical distribution from the fitted parameter values.

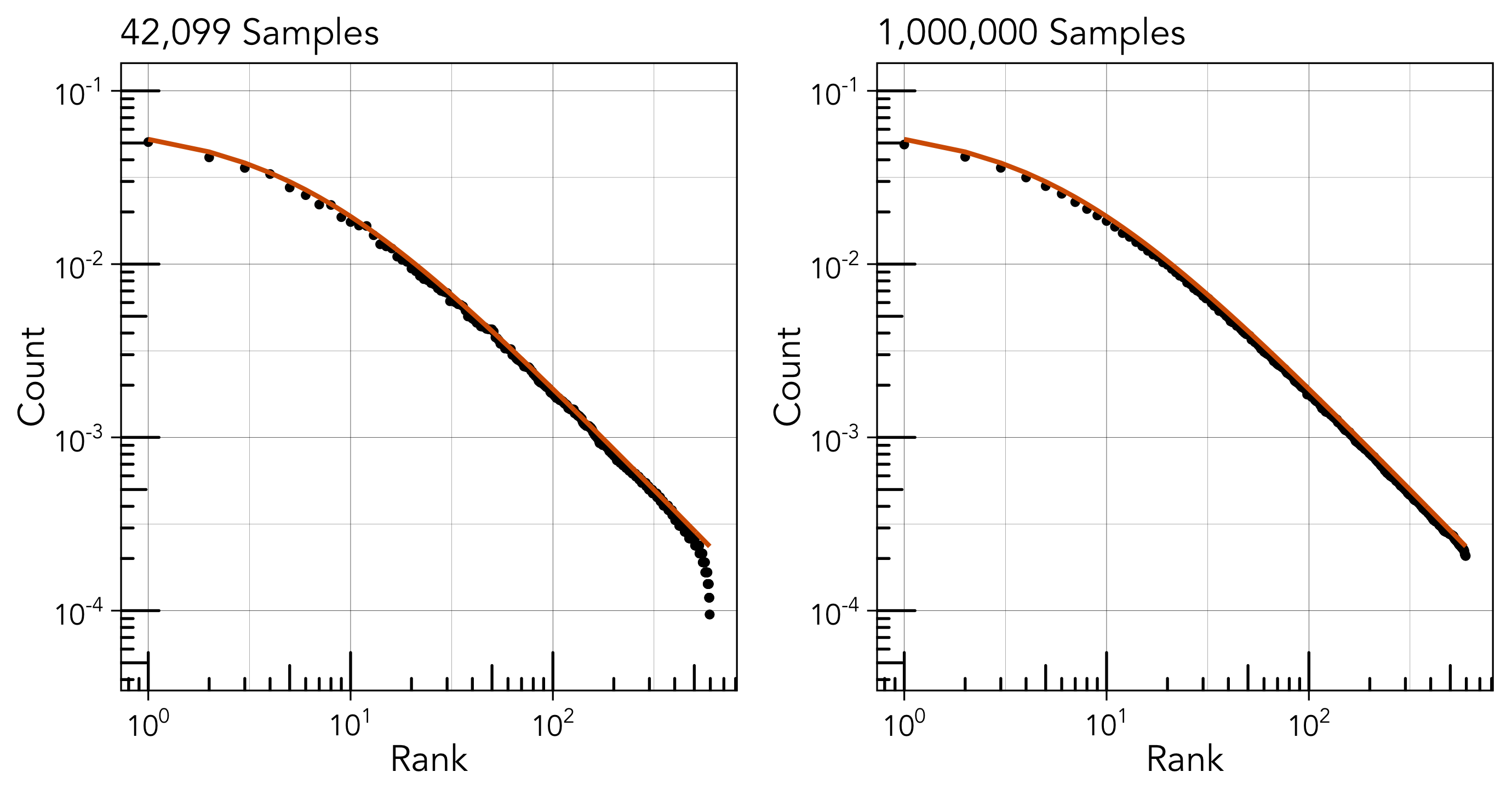


Figure S5: The relationship between rank (x-axis) and count (y-axis) at the intermediate granularity level (DS = 3, 596 syllable types). The orange lines denote the expected distributions according to Mandelbrot’s extension of Zipf’s rank-frequency law (a = 1.20, b = 5.69, c = 0.52). The points in the left and right panel are the expected counts and ranks when 42,099 and 1,000,000 tokens are sampled, respectively.

As you can see, there is a subtle but systematic departure from the Zipf-Mandelbrot distribution at high ranks when 42,099 tokens are sampled from the theoretical distribution (the same number observed in the real dataset). When 1,000,000 tokens are sampled this departure disappears. The logic here is that when sample sizes are limited rare types tend to be under-counted, leading to a “droop” ([28](#ref-tunnicliffe_hunter22)) in the curve at higher ranks.

# 4 Zipf’s Law of Abbreviation

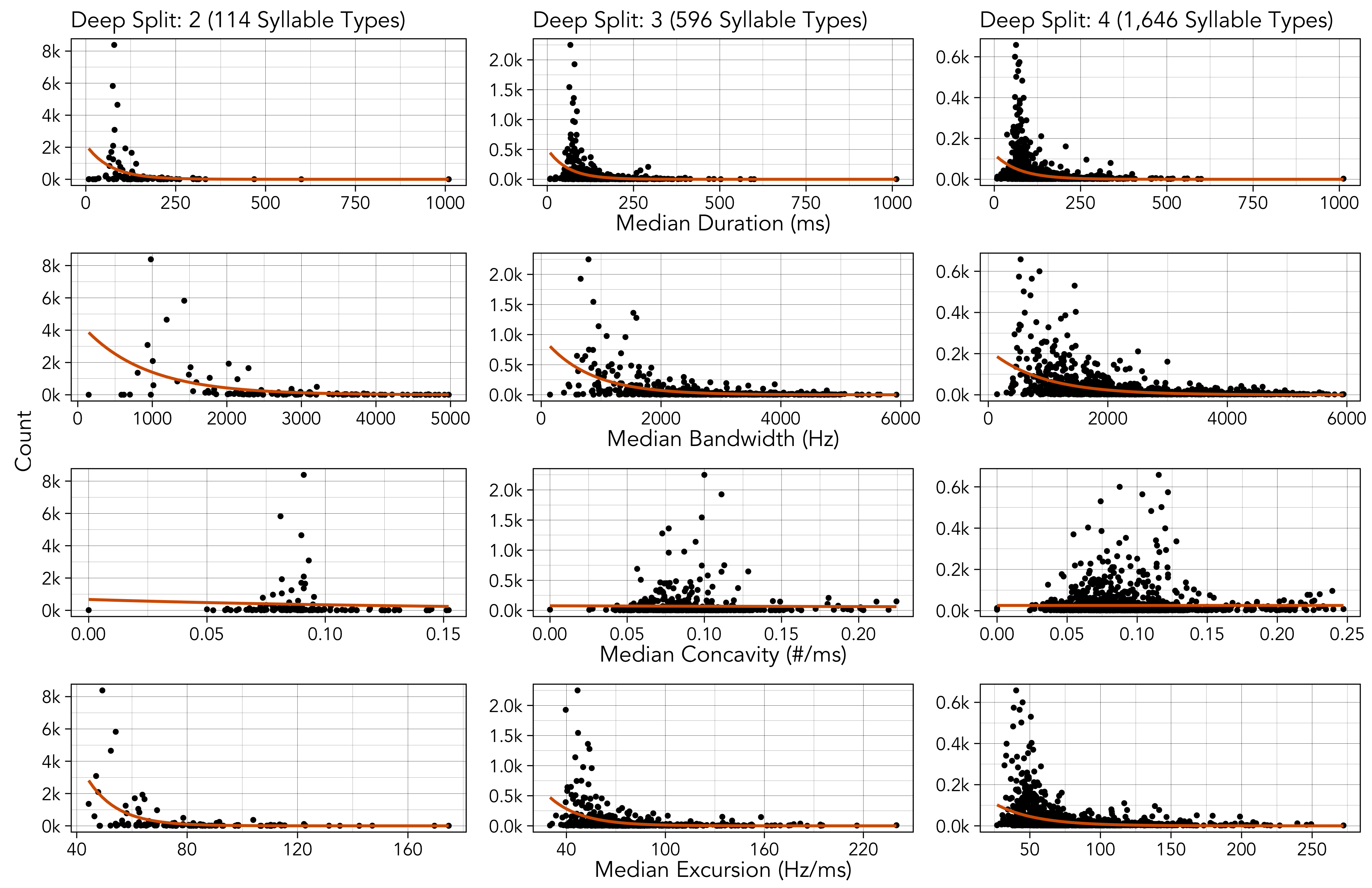


Figure S6: The relationship between four measures of production cost (x-axis) and count (y-axis) for each syllable type at each level of deep split (left, center, right). Each point shows the median value for a syllable type, so the orange best fit lines are from a simple Poisson model (count ~ cost) rather than the full log-normal model.

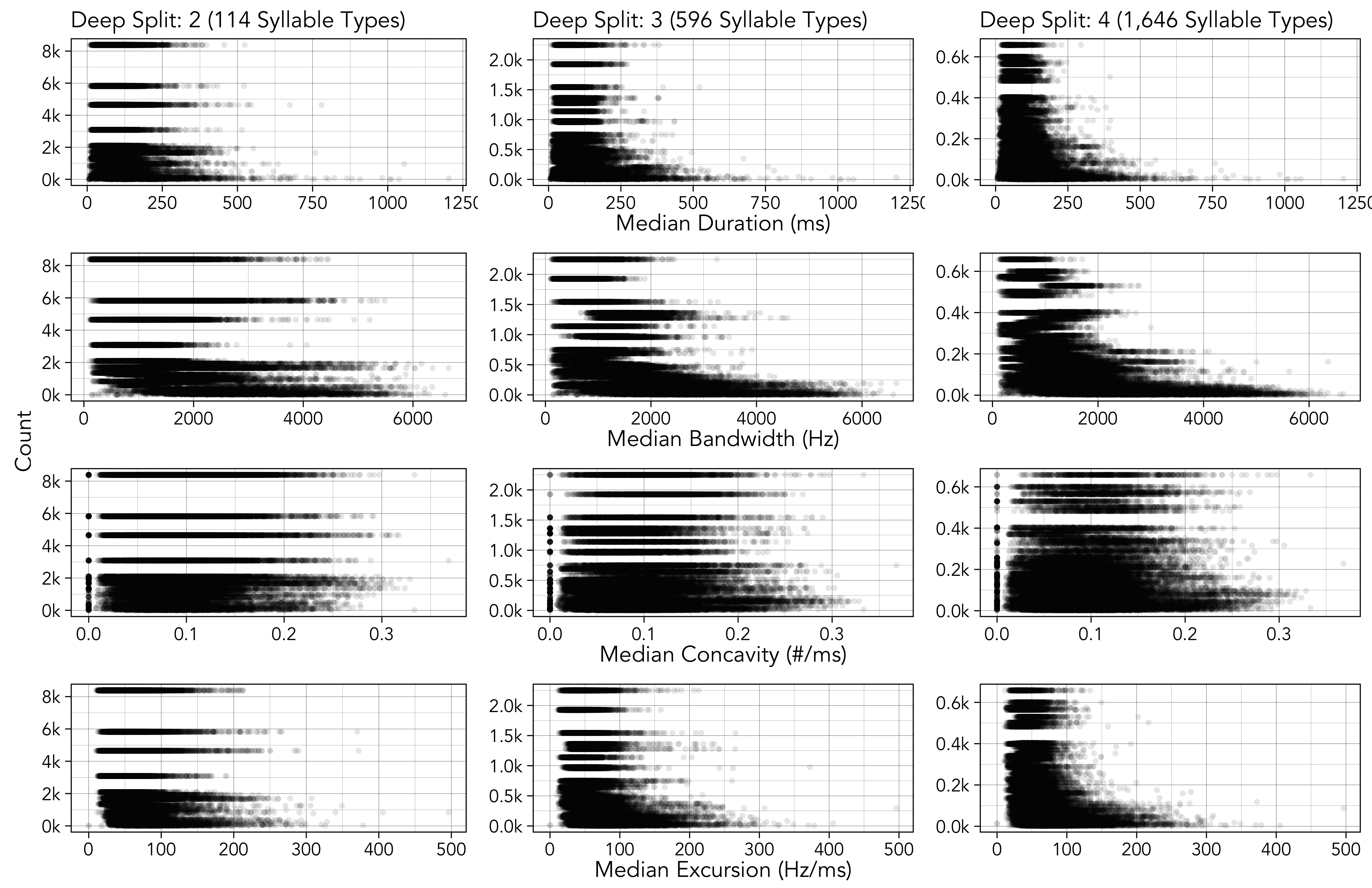


Figure S7: The relationship between four measures of production cost (x-axis) and count (y-axis) for each token (observation of each syllable type) at each level of deep split (left, center, right). Note that the y-axis is the count of each syllable type, which is why values are repeated across the individual tokens.

## 4.1 Priors and Diagnostics

Table S12: Prior specification for all four models of Zipf's law of abbreviation.

| **Class** | **Prior** | **Lower Bound** |
| --- | --- | --- |
| b | normal(0, 0.5) |  |
| Intercept | normal(0, 4) | 0 |
| sd | normal(0, 0.5) | 0 |
| sigma | normal(0, 0.5) | 0 |

Table S13: The model diagnostics for each model of Zipf's law of abbreviation.

| **Outcome** | **DS** | **Predictor** | **Rhat** | **Bulk ESS** | **Tail ESS** |
| --- | --- | --- | --- | --- | --- |
| Duration | 2 | Intercept | 1 | 712 | 913 |
| Count | 1 | 794 | 1,064 |
| 3 | Intercept | 1 | 451 | 915 |
| Count | 1 | 566 | 1,069 |
| 4 | Intercept | 1 | 268 | 580 |
| Count | 1 | 302 | 597 |
| Bandwidth | 2 | Intercept | 1 | 835 | 1,801 |
| Count | 1 | 1,047 | 2,088 |
| 3 | Intercept | 1 | 184 | 448 |
| Count | 1 | 230 | 586 |
| 4 | Intercept | 1 | 170 | 406 |
| Count | 1 | 230 | 525 |
| Concavity | 2 | Intercept | 1 | 1,059 | 2,059 |
| Count | 1 | 1,354 | 2,695 |
| 3 | Intercept | 1 | 1,307 | 2,561 |
| Count | 1 | 1,546 | 3,209 |
| 4 | Intercept | 1 | 723 | 1,469 |
| Count | 1 | 775 | 1,620 |
| Excursion | 2 | Intercept | 1 | 1,301 | 2,456 |
| Count | 1 | 1,508 | 2,661 |
| 3 | Intercept | 1 | 735 | 1,398 |
| Count | 1 | 855 | 1,592 |
| 4 | Intercept | 1 | 403 | 771 |
| Count | 1 | 479 | 956 |

## 4.2 Analysis by Year

Table S14: The estimated effect of count on each measure of production cost, in frequentist models that include year as a varying intercept, using the syllable classifications from each level of deep split. 95% confidence intervals that do not overlap with 0 are marked with an asterisk. The results are qualitatively identical to the main analysis.

| **Model** | **DS** | **Est.** | **2.5%** | **97.5%** |  |
| --- | --- | --- | --- | --- | --- |
| duration ~ count | 2 | -0.38 | -0.66 | -0.10 | \* |
| 3 | -0.47 | -0.59 | -0.35 | \* |
| 4 | -0.42 | -0.48 | -0.36 | \* |
| bandwidth ~ count | 2 | -0.56 | -0.77 | -0.35 | \* |
| 3 | -0.69 | -0.80 | -0.58 | \* |
| 4 | -0.64 | -0.70 | -0.58 | \* |
| concavity ~ count | 2 | -0.02 | -0.21 | 0.17 |  |
| 3 | -0.06 | -0.15 | 0.03 |  |
| 4 | -0.07 | -0.13 | -0.02 | \* |
| excursion ~ count | 2 | -0.27 | -0.42 | -0.12 | \* |
| 3 | -0.34 | -0.41 | -0.26 | \* |
| 4 | -0.32 | -0.36 | -0.28 | \* |

## 4.3 Replication using Kendall’s tau

I replicated the analysis using a method recently proposed by Lewis et al. ([29](#ref-lewis_etal23)) and released as the R package ZLAvian by Gilman et al. ([30](#ref-gilman_etal23)). This method computes Kendall’s tau, or the concordance between duration and frequency, where (tau + 1)/2 is the probability that if two random syllables are sampled (from either individuals or from the population) the longer note will be more common. The estimated tau is then compared against a null distribution that conservatively accounts for social learning ([30](#ref-gilman_etal23)).

At the population level, Kendall’s tau is significantly lower than the null for all four measures of production cost at every level of granularity in syllable clustering. There is also an individual level effect for bandwidth and excursion at all three levels of granularity, but the results are more mixed for duration and concavity. I think this is likely because we only have a median of 6 songs recorded for individual birds. I also replicated the analysis using data from a cross-fostering experiment ([12](#ref-Mann2020)), where individual birds were much more heavily sampled (100, 64, and 62 songs recorded for each one), and found individual level effects of duration, bandwidth, and excursion (see [Replication with Data from Mann et al.](#replication-with-data-from-mann-et-al.) below).

Table S15: The estimate of Kendall's tau, computed using the method of Lewis et al. [@lewis\_etal23] and Gilman et al. [@gilman\_etal23] from the ZLAvian package in R. p-values values are marked with stars to denote their level of statistical significance: \* for p < 0.05, \*\* for p < 0.01, \*\*\* for p < 0.001, \*\*\*\* for p < 0.0001.

| **Parameter** | **DS** | **Level** | **tau** | **p-value** |  |
| --- | --- | --- | --- | --- | --- |
| duration | 2 | individual | -0.20 | 3.0e-03 | \*\* |
| population | -0.32 | 1.9e-07 | \*\*\*\* |
| 3 | individual | -0.06 | 1.2e-01 |  |
| population | -0.35 | 7.1e-36 | \*\*\*\* |
| 4 | individual | 0.02 | 4.8e-01 |  |
| population | -0.33 | 6.8e-88 | \*\*\*\* |
| bandwidth | 2 | individual | -0.32 | 1.0e-03 | \*\*\* |
| population | -0.48 | 1.8e-14 | \*\*\*\* |
| 3 | individual | -0.18 | 1.0e-03 | \*\*\* |
| population | -0.42 | 8.3e-53 | \*\*\*\* |
| 4 | individual | -0.09 | 1.0e-03 | \*\*\* |
| population | -0.38 | 1.3e-115 | \*\*\*\* |
| concavity | 2 | individual | -0.09 | 7.2e-02 |  |
| population | -0.24 | 1.0e-04 | \*\*\* |
| 3 | individual | -0.04 | 4.5e-02 | \* |
| population | -0.17 | 3.0e-10 | \*\*\*\* |
| 4 | individual | -0.02 | 8.6e-02 |  |
| population | -0.12 | 3.1e-12 | \*\*\*\* |
| excursion | 2 | individual | -0.24 | 1.0e-03 | \*\*\* |
| population | -0.36 | 5.4e-09 | \*\*\*\* |
| 3 | individual | -0.19 | 1.0e-03 | \*\*\* |
| population | -0.31 | 9.9e-29 | \*\*\*\* |
| 4 | individual | -0.14 | 1.0e-03 | \*\*\* |
| population | -0.27 | 2.2e-57 | \*\*\*\* |

## 4.4 Replication with Data from Mann et al.

I contacted two labs with large corpora of field-recorded house finches, and neither of them had manually-classified syllables with analyzed acoustic features. However, there is one cross-fostering study conducted by Paul Mundinger in 1972, and analyzed and published by Mann et al. in 2021 ([12](#ref-Mann2020)), that includes both manually-classified syllables and acoustic features. I replicated the analysis using these data to ensure that my results for Zipf’s law of abbreviation were not an artifact of the automated classification procedure.

Mann et al. ([12](#ref-Mann2020)) compared the acoustic features of songs from (1) house finches tutored by other house finches, (2) house finches tutored by canaries, (3) house finches reared without tutors, and (4) canaries. Mann et al. ([12](#ref-Mann2020)) classified syllable types separately for each individual bird, so I analyzed the repertoires of the most heavily sampled bird from the first three categories (B1-74 (house finch tutor): 62 songs; B1 (canary tutor): 100 songs; D4 (no tutor): 64 songs). The results of the model, run with the exact same specification as the in main analysis, can been seen below.

Table S16: The estimated effect of count on each measure of production cost, using the syllable classifications from individual birds from Mann et al. [@Mann2020] who were tutored by a house finch, a canary, or were reared without a tutor. 95% confidence intervals that do not overlap with 0 are marked with an asterisk.

| **Model** | **Tutor** | **Est.** | **Err.** | **2.5%** | **97.5%** |  |
| --- | --- | --- | --- | --- | --- | --- |
| duration ~ count | House Finch | -0.1679 | 0.0839 | -0.3329 | -0.0004 | \* |
| Canary | -0.3484 | 0.1303 | -0.6059 | -0.0866 | \* |
| None | -0.2329 | 0.2127 | -0.6402 | 0.2005 |  |
| bandwidth ~ count | House Finch | -0.1874 | 0.0832 | -0.3518 | -0.0275 | \* |
| Canary | -0.1933 | 0.1708 | -0.5372 | 0.1501 |  |
| None | -0.3981 | 0.1714 | -0.7191 | -0.0353 | \* |
| concavity ~ count | House Finch | -0.0980 | 0.2890 | -0.6572 | 0.4836 |  |
| Canary | -0.3833 | 0.3076 | -0.9716 | 0.2296 |  |
| None | -0.0978 | 0.3652 | -0.8137 | 0.6168 |  |
| excursion ~ count | House Finch | -0.2081 | 0.1051 | -0.4213 | -0.0025 | \* |
| Canary | -0.3189 | 0.1519 | -0.6124 | -0.0129 | \* |
| None | -0.1825 | 0.1639 | -0.5052 | 0.1580 |  |

Under typical circumstances, where a house finch learns song from another house finch, the results are qualitatively the same as the main analysis—duration, bandwidth, and excursion all have strong negative effects on count. Interestingly, only duration and excursion negatively predict count in the canary-tutored house finch, and only bandwidth negatively predicts count in the house finch reared without a tutor.

I also replicated the analysis using a method recently proposed by Lewis et al. ([29](#ref-lewis_etal23)) and released as the R package ZLAvian by Gilman et al. ([30](#ref-gilman_etal23)). This method computes Kendall’s tau, or the concordance between duration and frequency, where (tau + 1)/2 is the probability that if two random syllables are sampled the longer note will be more common. The estimated tau is then compared against a null distribution that conservatively accounts for social learning ([30](#ref-gilman_etal23)). Kendall’s tau can be computed at the population level (i.e. random syllables are sampled from the population rather than a single bird, see above), but here I only report the results at the individual level since I am using data from individual repertoires.

Table S17: The estimate of Kendall's tau, computed using the method of Lewis et al. [@lewis\_etal23] and Gilman et al. [@gilman\_etal23] from the ZLAvian package in R. In this case, were data represent individual repertoires, tau is only computed at the individual level. p-values less than 0.05 are marked with a single star, and values less than 0.01 are marked with two stars.

| **Parameter** | **Tutor** | **tau** | **p** |  |
| --- | --- | --- | --- | --- |
| duration | house finch | -0.301 | 0.006 | \*\* |
| canary | -0.515 | 0.009 | \*\* |
| none | -0.506 | 0.019 | \* |
| bandwidth | house finch | -0.264 | 0.012 | \* |
| canary | -0.116 | 0.287 |  |
| none | -0.460 | 0.030 | \* |
| concavity | house finch | -0.164 | 0.104 |  |
| canary | -0.077 | 0.321 |  |
| none | -0.214 | 0.186 |  |
| excursion | house finch | -0.315 | 0.003 | \*\* |
| canary | -0.429 | 0.012 | \* |
| none | -0.230 | 0.159 |  |

The results are mostly consistent with the Bayesian model above: Kendall’s tau is significantly negative for duration under all conditions, for bandwidth when birds learned from a house finch or were raised in isolation, and for excursion when birds learned from a house finch or a canary. Interestingly, we found that Kendall’s tau for duration is consistently significantly negative at the individual level for the Mann et al. ([12](#ref-Mann2020)) data, whereas it is only consistently significantly negative at the population level for the Youngblood and Lahti ([3](#ref-youngbloodContentBiasCultural2022)) data (see [Replication using Kendall’s tau](#replication-using-kendalls-tau) above). This is likely because there are many more songs recorded from the individual birds in the Mann et al. ([12](#ref-Mann2020)) data (N = 100, 64, and 62) compared to the Youngblood and Lahti ([3](#ref-youngbloodContentBiasCultural2022)) data (median of 6).

## 4.5 Robustness to Lognormality Assumption

Table S18: The estimated effect of count on each measure of production cost in frequentist models that drop the lognormality assumption and do not log-transform the y-axis, using the syllable classifications from each level of deep split. 95% confidence intervals that do not overlap with 0 are marked with an asterisk. The results are qualitatively identical to the main analysis.

| **Model** | **DS** | **Est.** | **2.5%** | **97.5%** |  |
| --- | --- | --- | --- | --- | --- |
| duration ~ count | 2 | -0.94 | -1.65 | -0.23 | \* |
| 3 | -0.97 | -1.27 | -0.66 | \* |
| 4 | -0.79 | -0.94 | -0.65 | \* |
| bandwidth ~ count | 2 | -1.23 | -1.73 | -0.73 | \* |
| 3 | -1.38 | -1.66 | -1.11 | \* |
| 4 | -1.19 | -1.33 | -1.04 | \* |
| concavity ~ count | 2 | 0.00 | -0.15 | 0.14 |  |
| 3 | 0.01 | -0.13 | 0.14 |  |
| 4 | 0.02 | -0.07 | 0.10 |  |
| excursion ~ count | 2 | -0.62 | -0.93 | -0.31 | \* |
| 3 | -0.73 | -0.93 | -0.53 | \* |
| 4 | -0.67 | -0.79 | -0.55 | \* |

## 4.6 Robustness to Long Tail

Table S19: The estimated effect of count on each measure of production cost in frequentist models that only include the middle 90% of the distribution, using the syllable classifications from each level of deep split. 95% confidence intervals that do not overlap with 0 are marked with an asterisk. The results are qualitatively identical to the main analysis.

| **Model** | **DS** | **Est.** | **2.5%** | **97.5%** |  |
| --- | --- | --- | --- | --- | --- |
| duration ~ count | 2 | -0.28 | -0.40 | -0.15 | \* |
| 3 | -0.32 | -0.39 | -0.24 | \* |
| 4 | -0.29 | -0.33 | -0.25 | \* |
| bandwidth ~ count | 2 | -0.40 | -0.54 | -0.26 | \* |
| 3 | -0.51 | -0.60 | -0.43 | \* |
| 4 | -0.49 | -0.53 | -0.44 | \* |
| concavity ~ count | 2 | -0.01 | -0.07 | 0.05 |  |
| 3 | 0.01 | -0.04 | 0.05 |  |
| 4 | 0.01 | -0.02 | 0.05 |  |
| excursion ~ count | 2 | -0.20 | -0.27 | -0.13 | \* |
| 3 | -0.24 | -0.28 | -0.19 | \* |
| 4 | -0.22 | -0.25 | -0.19 | \* |

# 5 Menzerath’s Law

## 5.1 Priors and Diagnostics

Table S20: Prior specification for the model of Menzerath's law.

| **Class** | **Prior** | **Lower Bound** |
| --- | --- | --- |
| b | normal(0, 0.1) |  |
| Intercept | normal(0, 3) | 0 |
| sd | normal(0, 0.5) | 0 |
| sigma | normal(0, 0.5) | 0 |

Table S21: Estimates and diagnostics for the model of Menzerath's law applied to the real data.

| **Param.** | **Estimate** | **l-95% CI** | **u-95% CI** | **Rhat** | **Bulk\_ESS** | **Tail\_ESS** |
| --- | --- | --- | --- | --- | --- | --- |
| Intercept | 4.48 | 4.47 | 4.49 | 1 | 6,740 | 7,103 |
| Song Length | -0.05 | -0.07 | -0.04 | 1 | 4,794 | 6,044 |

Table S22: The R-hat values from the two null models applied to each of the 10 simulated datasets.

|  | **Simple Null Model** | | **Production Null Model** | |
| --- | --- | --- | --- | --- |
| **Dataset** | **Intercept** | **Song Length** | **Intercept** | **Song Length** |
| 1 | 0.9998875 | 0.9996983 | 0.9997456 | 1.000656 |
| 2 | 0.9997542 | 0.9997727 | 0.9999726 | 1.000328 |
| 3 | 1.0000627 | 0.9997787 | 0.9998281 | 1.001598 |
| 4 | 0.9997391 | 0.9997015 | 1.0005752 | 1.000895 |
| 5 | 0.9999400 | 1.0004971 | 0.9997567 | 1.000817 |
| 6 | 0.9999405 | 0.9998756 | 0.9997009 | 1.000332 |
| 7 | 0.9999780 | 0.9998101 | 0.9997015 | 1.001533 |
| 8 | 1.0005872 | 0.9998251 | 0.9998900 | 1.000225 |
| 9 | 0.9999938 | 0.9997752 | 0.9998704 | 1.001099 |
| 10 | 0.9997444 | 0.9998224 | 1.0001865 | 1.002680 |

## 5.2 Analysis by Year

Table S23: Estimates for a frequentist version of the model of Menzerath's law applied to the real data, with year added as a varying intercept. The results are qualitatively identical to the main analysis.

| **Param.** | **Estimate** | **2.5%** | **97.5%** |
| --- | --- | --- | --- |
| Intercept | 4.500 | 4.40 | 4.600 |
| Song Length | -0.047 | -0.06 | -0.033 |

# 6 Small-Worldness Index

## 6.1 Analysis by Year

Table S24: The small-worldness index computed separately from the data from each year at each level of deep split.

| **Year** | **DS: 2** | **DS: 3** | **DS: 4** |
| --- | --- | --- | --- |
| 1975 | 1.88 | 5.58 | 10.45 |
| 2012 | 1.76 | 5.71 | 13.34 |
| 2019 | 2.09 | 6.95 | 14.30 |

# 7 Mutual Information

Table S25: The WAIC and R-Squared value for each model at each level of deep split.

| **DS** | **Model** | **WAIC** | **R-Sq** |
| --- | --- | --- | --- |
| 2 | Exponential | -829 | 0.951 |
| Power-Law | -766 | 0.910 |
| Composite | -840 | 0.952 |
| 3 | Exponential | -617 | 0.982 |
| Power-Law | -487 | 0.927 |
| Composite | -668 | 0.986 |
| 4 | Exponential | -599 | 0.990 |
| Power-Law | -432 | 0.926 |
| Composite | -688 | 0.993 |

## 7.1 Priors and Diagnostics

Table S26: Prior specification for all three models of mutual information decay.

| **Parameter** | **Prior** | **Lower Bound** |
| --- | --- | --- |
| a | normal(0, 1) | 0 |
| b | normal(0, 1) | 0 |
| c | normal(0, 1) | 0 |
| d | normal(0, 1) | 0 |

Table S27: The WAIC values for each model, at each level of deep split, for increasing maximum distances between syllables.

|  | **2** | | | **3** | | | **4** | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Exp** | **PL** | **Comp** | **Exp** | **PL** | **Comp** | **Exp** | **PL** | **Comp** |
| 100 | -829 | -765 | -841 | -617 | -486 | -668 | -598 | -432 | -688 |
| 200 | -1,557 | -1,513 | -1,560 | -1,232 | -1,068 | -1,281 | -1,201 | -967 | -1,267 |
| 300 | -2,101 | -2,086 | -2,114 | -1,635 | -1,555 | -1,695 | -1,626 | -1,468 | -1,686 |
| 400 | -2,683 | -2,676 | -2,712 | -2,008 | -1,977 | -2,131 | -2,082 | -1,950 | -2,142 |
| 500 | -2,984 | -2,988 | -3,073 | -2,078 | -2,093 | -2,276 | -2,260 | -2,217 | -2,387 |
| 600 | -3,114 | -3,176 | -3,248 | -2,264 | -2,299 | -2,527 | -2,454 | -2,449 | -2,639 |
| 700 | -3,291 | -3,430 | -3,471 | -2,493 | -2,539 | -2,808 | -2,650 | -2,669 | -2,901 |
| 800 | -3,620 | -3,734 | -3,766 | -2,479 | -2,598 | -2,798 | -2,209 | -2,240 | -2,485 |
| 900 | -3,615 | -3,785 | -3,801 | -2,108 | -2,379 | -2,467 | -1,568 | -1,804 | -1,886 |
| 1000 | -3,712 | -3,931 | -3,942 | -1,958 | -2,350 | -2,408 | -1,235 | -1,608 | -1,654 |
| 1200 | -3,175 | -3,244 | -3,246 | -1,483 | -1,711 | -1,731 | -823 | -1,084 | -1,104 |

## 7.2 Deep Split: 2

Table S28: Estimates and diagnostics for exponential model.

| **Param.** | **Estimate** | **l-95% CI** | **u-95% CI** | **Rhat** | **Bulk\_ESS** | **Tail\_ESS** |
| --- | --- | --- | --- | --- | --- | --- |
| a | 0.17 | 0.16 | 0.19 | 1 | 4,336 | 4,595 |
| b | 0.39 | 0.35 | 0.44 | 1 | 4,511 | 4,996 |

Table S29: Estimates and diagnostics for power-law model.

| **Param.** | **Estimate** | **l-95% CI** | **u-95% CI** | **Rhat** | **Bulk\_ESS** | **Tail\_ESS** |
| --- | --- | --- | --- | --- | --- | --- |
| c | 0.13 | 0.12 | 0.14 | 1 | 6,738 | 7,059 |
| d | 1.08 | 1.01 | 1.15 | 1 | 6,437 | 6,285 |

Table S30: Estimates and diagnostics for composite model.

| **Param.** | **Estimate** | **l-95% CI** | **u-95% CI** | **Rhat** | **Bulk\_ESS** | **Tail\_ESS** |
| --- | --- | --- | --- | --- | --- | --- |
| a | 0.16 | 0.12 | 0.18 | 1 | 3,354 | 3,685 |
| b | 0.43 | 0.38 | 0.48 | 1 | 4,592 | 4,387 |
| c | 0.02 | 0.00 | 0.04 | 1 | 3,021 | 2,579 |
| d | 0.66 | 0.29 | 0.96 | 1 | 3,237 | 2,798 |

## 7.3 Deep Split: 3

Table S31: Estimates and diagnostics for exponential model.

| **Param.** | **Estimate** | **l-95% CI** | **u-95% CI** | **Rhat** | **Bulk\_ESS** | **Tail\_ESS** |
| --- | --- | --- | --- | --- | --- | --- |
| a | 0.60 | 0.57 | 0.64 | 1 | 4,478 | 5,210 |
| b | 0.27 | 0.26 | 0.29 | 1 | 4,453 | 5,343 |

Table S32: Estimates and diagnostics for power-law model.

| **Param.** | **Estimate** | **l-95% CI** | **u-95% CI** | **Rhat** | **Bulk\_ESS** | **Tail\_ESS** |
| --- | --- | --- | --- | --- | --- | --- |
| c | 0.52 | 0.49 | 0.56 | 1 | 6,895 | 5,998 |
| d | 0.95 | 0.90 | 1.00 | 1 | 7,133 | 7,055 |

Table S33: Estimates and diagnostics for composite model.

| **Param.** | **Estimate** | **l-95% CI** | **u-95% CI** | **Rhat** | **Bulk\_ESS** | **Tail\_ESS** |
| --- | --- | --- | --- | --- | --- | --- |
| a | 0.53 | 0.47 | 0.59 | 1 | 2,820 | 3,907 |
| b | 0.30 | 0.28 | 0.32 | 1 | 4,896 | 4,646 |
| c | 0.08 | 0.04 | 0.13 | 1 | 2,882 | 2,864 |
| d | 0.61 | 0.42 | 0.77 | 1 | 2,978 | 3,096 |

## 7.4 Deep Split: 4

Table S34: Estimates and diagnostics for exponential model.

| **Param.** | **Estimate** | **l-95% CI** | **u-95% CI** | **Rhat** | **Bulk\_ESS** | **Tail\_ESS** |
| --- | --- | --- | --- | --- | --- | --- |
| a | 0.74 | 0.71 | 0.77 | 1 | 3,787 | 4,572 |
| b | 0.24 | 0.23 | 0.25 | 1 | 4,287 | 5,362 |

Table S35: Estimates and diagnostics for power-law model.

| **Param.** | **Estimate** | **l-95% CI** | **u-95% CI** | **Rhat** | **Bulk\_ESS** | **Tail\_ESS** |
| --- | --- | --- | --- | --- | --- | --- |
| c | 0.69 | 0.64 | 0.73 | 1 | 6,835 | 6,785 |
| d | 0.92 | 0.87 | 0.97 | 1 | 6,638 | 6,848 |

Table S36: Estimates and diagnostics for composite model.

| **Param.** | **Estimate** | **l-95% CI** | **u-95% CI** | **Rhat** | **Bulk\_ESS** | **Tail\_ESS** |
| --- | --- | --- | --- | --- | --- | --- |
| a | 0.65 | 0.59 | 0.71 | 1 | 2,387 | 2,851 |
| b | 0.26 | 0.25 | 0.27 | 1 | 4,566 | 4,959 |
| c | 0.10 | 0.05 | 0.15 | 1 | 2,433 | 2,760 |
| d | 0.61 | 0.45 | 0.74 | 1 | 2,360 | 2,726 |

## 7.5 Analysis by Year

Table S37: The WAIC values for the exponential, power-law, and composite models applied to the mutual information calculated separately from the data from each year at each level of deep split. The composite model outcompetes both alternatives in all conditions.

| **DS** | **Year** | **Exponential** | **Power-Law** | **Composite** |
| --- | --- | --- | --- | --- |
| 2 | 1975 | -653 | -607 | -656 |
| 2012 | -547 | -421 | -564 |
| 2019 | -521 | -427 | -571 |
| 3 | 1975 | -724 | -685 | -725 |
| 2012 | -598 | -446 | -607 |
| 2019 | -588 | -427 | -640 |
| 4 | 1975 | -687 | -671 | -723 |
| 2012 | -518 | -396 | -572 |
| 2019 | -511 | -370 | -587 |

## 7.6 Analysis with Song Sequences

Table S38: The WAIC values for the exponential, power-law, and composite models applied to the mutual information calculated from individual song sequences rather than from concatenated song bouts. The exponential model outcompetes both alternatives at deep split values of 3 and 4, while the power-law and composite models outcompete the exponential model at deep split of 2.

| **DS** | **Exponential** | **Power-Law** | **Composite** |
| --- | --- | --- | --- |
| 2 | -77 | -84 | -84 |
| 3 | -33 | -28 | -31 |
| 4 | -60 | -50 | -57 |

## 7.7 Analysis of Information in Bouts

This analysis was conducted to ensure that the ordering of songs within song bouts contributes mutual information to the decay curves.

Long-range dependencies in song sequences come from two places: the ordering of syllables within very long songs, and the ordering of songs within song bouts. To isolate the statistical signals of the latter, I created a “dummy” dataset where syllable sequences within songs were shuffled, but the ordering of bouts was the same. As a comparison, I created a second “dummy” dataset that has the same song sequences as the first, but with random song bouts (shuffled within individuals).

Wilcoxon signed-rank tests and t-tests show that the first dataset contains significantly more information than the second at all three levels of granularity in syllable clustering.

Table S39: The p-values from Wilcoxon signed-rank tests and t-tests comparing the amount of information contained in two "dummy" datasets: one that isolates the information contained in song bouts by shuffling syllable sequences within songs, and another that further shuffles song bout order within individuals. The first contains more information than the second (p < 0.05) across all three levels of granularity in syllable clustering.

|  | **DS: 2** | **DS: 3** | **DS: 4** |
| --- | --- | --- | --- |
| Wilcoxon signed-rank test | 3.2e-13 | 5.6e-26 | 3.2e-08 |
| t-test | 1.7e-02 | 6.2e-11 | 2.4e-02 |

# 8 Extended Discussion

A long-standing critique of Zipf’s laws is that they may be statistical artifacts of other processes ([31](#ref-caplan_etal20)), starting with Miller’s observation that randomly typing on keyboards can produce similar patterns ([32](#ref-miller57)). That being said, random typing accounts are not realistic causal descriptions of how communication systems emerge, and there are good empirical reasons to doubt that they undermine efficiency accounts ([24](#ref-piantadosi14)). Randomly-generated texts produce rank-frequency distributions that differ from those in real corpora ([33](#ref-ferrer-i-cancho_elvevag10)), random typing models are not truly neutral as they can be mathematically reframed as minimizing costs ([34](#ref-ferrer-i-cancho16),[35](#ref-ferrer-i-cancho_etal22)), and there is direct experimental evidence that Zipfian abbreviation emerges from pressure for efficient communication ([36](#ref-kanwal_etal17)). In my view, the most important contribution of the random typing account is to highlight that the problem of equifinality—different processes leading to similar outcomes ([37](#ref-Barrett2019))—means that patterns resembling Zipf’s laws are not sufficient to make conclusions about efficiency ([38](#ref-kanwal17),[39](#ref-semple_etal22)). Multiple lines of evidence should be presented alongside other work demonstrating that efficiency is shaping the system (e.g. physical ([12](#ref-Mann2020)) and environmental ([40](#ref-bermudez-cuamatzin_etal23)) constraints), as I have done here. See ([39](#ref-semple_etal22)), ([24](#ref-piantadosi14)), and ([38](#ref-kanwal17)) for more complete summaries of this debate.

Outside of linguistics, efficiency and complexity are often discussed in relation to cumulative cultural evolution (CCE). Definitions of CCE vary and a full review is outside of the scope of this study, but for convenience we will use the definition of ([41](#X4a3efed6eb38d2e9ab41c2a8023cf077128a4a5)): “the accumulation of sequential changes within a single socially learned behavior that results in improved function”. Discussions of CCE often focus on increasing complexity over time ([42](#ref-wilks_blakey18)), which was once thought to be a hallmark of human culture ([43](#ref-tomasello99)) but has now been observed in several non-human communication systems including humpback whale ([44](#ref-garland_etal22)) and Savannah sparrow song ([41](#X4a3efed6eb38d2e9ab41c2a8023cf077128a4a5)). ([45](#ref-gruber_etal22)) make a convincing argument that efficiency deserves more attention in CCE, as increases in complexity in one domain require increases in efficiency in another (see Equation 1 in the Introduction). House finch song may be a good research model for how the interplay between efficiency and complexity drives CCE, as male house finches have a social learning bias for more complex syllables ([3](#ref-youngbloodContentBiasCultural2022)), possibly as an adaptation to female preferences for more complex songs ([46](#ref-Nolan2004)–[48](#ref-Ciaburri2019)), and there appears to be pressure for efficiency at the level of both syllables and songs. That being said, CCE may not be the best framework for understanding the interaction between efficiency and complexity in birdsong, as its logic is more difficult to apply to “aesthetic” behavior ([49](#ref-sinclair_etal22)) especially when it is optimized for female preferences that evolve to maximize inclusive fitness rather than the specific properties of songs that males sing ([50](#ref-geller_lahti23)).

House finch song exhibits language-like efficiency and structure, but music-like structure has not been similarly studied in this species. In the last two decades researchers have identified aspects of birdsong, such as rhythm and pitch intervals in thrush nightingales ([51](#ref-roeskeCategoricalRhythmsAre2020)–[53](#ref-rothenberg_etal14)), that closely resemble aspects of human music. Future studies should explore language- and music-like properties of birdsong in parallel across multiple levels of granularity to inform the ongoing debate about whether birdsong is more akin to music or language ([54](#ref-fitch06)–[56](#ref-rohrmeier_etal15)).

# References

1. Ju C, Geller FC, Mundinger PC, Lahti DC. [Four decades of cultural evolution in house finch songs](https://doi.org/10.1093/auk/uky012). The Auk: Ornithological Advances. 2019;136:1–18.

2. Mundinger PC. Song dialects and colonization in the house finch, Carpodacus mexicanus, on the east coast. The Condor [Internet]. 1975;77(4):407–22. Available from: <https://doi.org/10.2307/1366088>

3. Youngblood M, Lahti D. [Content bias in the cultural evolution of house finch song](https://doi.org/10.1016/j.anbehav.2021.12.012). Animal Behaviour. 2022;185:37–48.

4. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics [Internet]. 2008 Mar 1 [cited 2023 Jun 2];24(5):719–20. Available from: <https://academic.oup.com/bioinformatics/article/24/5/719/200751>

5. Ju C. FinchCatcher [Internet]. Department of Biology, Queens College; 2016. Available from: <http://finchcatcher.net>

6. Roginek EW. Spatial variation of house finch (Haemorhous mexicanus) song along the American Southwest coast. Queens College; 2018.

7. Liu S, Lu Y, Geng D. Molecular Subgroup Classification in Alzheimer’s Disease by Transcriptomic Profiles. J Mol Neurosci [Internet]. 2022 Apr 1 [cited 2023 Jun 2];72(4):866–79. Available from: <https://doi.org/10.1007/s12031-021-01957-w>

8. Zhao H, Zhang S, Shao S, Fang H. Identification of a Prognostic 3-Gene Risk Prediction Model for Thyroid Cancer. Frontiers in Endocrinology [Internet]. 2020 [cited 2023 Jun 2];11. Available from: <https://www.frontiersin.org/articles/10.3389/fendo.2020.00510>

9. Burkett ZD, Day NF, Peñagarikano O, Geschwind DH, White SA. VoICE: A semi-automated pipeline for standardizing vocal analysis across models. Sci Rep [Internet]. 2015 May 28 [cited 2023 Jun 2];5(1):10237. Available from: <https://www.nature.com/articles/srep10237>

10. Hsieh TC, Ma KH, Chao A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). McInerny G, editor. Methods Ecol Evol [Internet]. 2016 Dec [cited 2022 Nov 29];7(12):1451–6. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12613>

11. Ratanamahatana CA, Keogh E. [Three myths about dynamic time warping data mining](https://doi.org/10.1137/1.9781611972757.50). In: Proceedings of the 2005 SIAM International Conference on Data Mining. 2005. p. 506–10.

12. Mann DC, Lahti DC, Waddick L, Mundinger PC. [House finches learn canary trills](https://doi.org/10.1080/09524622.2020.1718551). Bioacoustics. 2020;1–17.

13. Mandelbrot B. An informational theory of the statistical structure of language. Communication Theory. 1953;486–502.

14. Mandelbrot B. On the theory of word frequencies and on related Markovian models of discourse. 1962;190–219.

15. Izsák J. Some practical aspects of fitting and testing the Zipf-Mandelbrot model: a short essay. Scientometrics [Internet]. 2006 Apr [cited 2023 Sep 6];67(1):107–20. Available from: <http://link.springer.com/10.1007/s11192-006-0052-x>

16. Mouillot D, Lepretre A. [Introduction of relative abundance distribution (RAD) indices, estimated from the rank-frequency diagrams (RFD), to assess changes in community diversity](https://doi.org/10.1023/A:1006297211561). Environmental Monitoring and Assessment. 2000;63:279–95.

17. Mačutek J. Why do parameter values in the Zipf-Mandelbrot distribution sometimes explode? Journal of Quantitative Linguistics [Internet]. 2022 Oct 2 [cited 2023 Sep 7];29(4):413–24. Available from: <https://www.tandfonline.com/doi/full/10.1080/09296174.2021.1887613>

18. Ausloos M. Zipf–Mandelbrot–Pareto model for co-authorship popularity. Scientometrics [Internet]. 2014 Dec [cited 2023 Sep 7];101(3):1565–86. Available from: <http://link.springer.com/10.1007/s11192-014-1302-y>

19. Hailman JP. Constrained permutation in “chick-a-dee”-like calls of a black-lored tit Parus xanthogenys. Bioacoustics [Internet]. 1994 Jan [cited 2023 Sep 19];6(1):33–50. Available from: <http://www.tandfonline.com/doi/abs/10.1080/09524622.1994.9753270>

20. Freeberg TM, Lucas JR. Information theoretical approaches to chick-a-dee calls of Carolina chickadees (Poecile carolinensis). Journal of Comparative Psychology [Internet]. 2012 [cited 2023 Sep 19];126(1):68–81. Available from: <http://doi.apa.org/getdoi.cfm?doi=10.1037/a0024906>

21. Ficken MS, Hailman ED, Hailman JP. The chick-a-dee call system of the Mexican chickadee. The Condor [Internet]. 1994 Feb [cited 2023 Sep 19];96(1):70–82. Available from: <https://academic.oup.com/condor/article/96/1/70-82/5124075>

22. Ferrer I Cancho R, Solé RV. Two Regimes in the Frequency of Words and the Origins of Complex Lexicons: Zipf’s Law Revisited. Journal of Quantitative Linguistics [Internet]. 2001 Dec [cited 2024 Feb 21];8(3):165–73. Available from: <https://www.tandfonline.com/doi/full/10.1076/jqul.8.3.165.4101>

23. Montemurro MA. Beyond the Zipf–Mandelbrot law in quantitative linguistics. Physica A: Statistical Mechanics and its Applications [Internet]. 2001 Nov [cited 2024 Feb 21];300(3-4):567–78. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0378437101003557>

24. Piantadosi ST. Zipf’s word frequency law in natural language: A critical review and future directions. Psychon Bull Rev [Internet]. 2014 Oct [cited 2023 Aug 3];21(5):1112–30. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4176592/>

25. Gorina OG, Tsarakova NS, Tsarakov SK. Study of Optimal Text Size Phenomenon in Zipf–Mandelbrot’s Distribution on the Bases of Full and Distorted Texts. Author’s Frequency Characteristics and derivation of Hapax Legomena. Journal of Quantitative Linguistics [Internet]. 2020 Apr 2 [cited 2024 Feb 21];27(2):134–58. Available from: <https://www.tandfonline.com/doi/full/10.1080/09296174.2018.1559460>

26. Cristelli M, Batty M, Pietronero L. There is More than a Power Law in Zipf. Sci Rep [Internet]. 2012 Nov 8 [cited 2024 Feb 21];2(1):812. Available from: <https://www.nature.com/articles/srep00812>

27. Montemurro MA, Zanette DH. New perspectives on Zipf’s law in linguistics: from single texts to large corpora. Glottometrics. 2002;4:87–99.

28. Tunnicliffe M, Hunter G. Random sampling of the Zipf–Mandelbrot distribution as a representation of vocabulary growth. Physica A: Statistical Mechanics and its Applications [Internet]. 2022 Dec [cited 2024 Feb 21];608:128259. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0378437122008172>

29. Lewis RN, Kwong A, Soma M, De Kort SR, Gilman RT. Java sparrow song conforms to Menzerath’s Law but not Zipf’s Law of Abbreviation [Internet]. Animal Behavior and Cognition; 2023 Dec [cited 2024 Jan 25]. Available from: <http://biorxiv.org/lookup/doi/10.1101/2023.12.13.571437>

30. Gilman RT, Durrant C, Malpas L, Lewis RN. Does Zipf’s law of abbreviation shape birdsong? [Internet]. Animal Behavior and Cognition; 2023 Dec [cited 2024 Jan 25]. Available from: <http://biorxiv.org/lookup/doi/10.1101/2023.12.06.569773>

31. Caplan S, Kodner J, Yang C. Miller’s monkey updated: communicative efficiency and the statistics of words in natural language. Cognition [Internet]. 2020 Dec [cited 2023 Sep 22];205:104466. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0010027720302857>

32. Miller GA. Some effects of intermittent silence. The American Journal of Psychology [Internet]. 1957 [cited 2023 Sep 21];70(2):311–4. Available from: <http://www.jstor.org/stable/1419346>

33. Ferrer-i-Cancho R, Elvevåg B. Random texts do not exhibit the real Zipf’s law-like rank distribution. Scalas E, editor. PLoS ONE [Internet]. 2010 Mar 9 [cited 2023 Sep 21];5(3):e9411. Available from: <https://dx.plos.org/10.1371/journal.pone.0009411>

34. Ferrer-i-Cancho R. Compression and the origins of Zipf’s law for word frequencies. Complexity [Internet]. 2016 Nov [cited 2023 Aug 28];21(S2):409–11. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/cplx.21820>

35. Ferrer-i-Cancho R, Bentz C, Seguin C. Optimal coding and the origins of Zipfian laws. Journal of Quantitative Linguistics [Internet]. 2022 Apr 3 [cited 2023 Sep 21];29(2):165–94. Available from: <https://www.tandfonline.com/doi/full/10.1080/09296174.2020.1778387>

36. Kanwal J, Smith K, Culbertson J, Kirby S. Zipf’s law of abbreviation and the principle of least effort: language users optimise a miniature lexicon for efficient communication. Cognition [Internet]. 2017 Aug [cited 2023 Aug 9];165:45–52. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0010027717301166>

37. Barrett BJ. [Equifinality in empirical studies of cultural transmission](https://doi.org/10.1016/j.beproc.2018.01.011). Behavioural Processes. 2019;161:129–38.

38. Kanwal J. Word length and the principle of least eﬀort: language as an evolving, efficient code for information transfer [Doctor of Philosophy]. School of Philosophy, Psychology, and Language Sciences, University of Edinburgh; 2017.

39. Semple S, Ferrer-i-Cancho R, Gustison ML. Linguistic laws in biology. Trends in Ecology & Evolution [Internet]. 2022 Jan [cited 2023 Aug 3];37(1):53–66. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0169534721002305>

40. Bermúdez‐Cuamatzin E, Slabbekoorn H, Macías Garcia C. Spectral and temporal call flexibility of House Finches (Haemorhous mexicanus) from urban areas during experimental noise exposure. Ibis [Internet]. 2023 Apr [cited 2023 Apr 12];165(2):571–86. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/ibi.13161>

41. Williams H, Scharf A, Ryba AR, Ryan Norris D, Mennill DJ, Newman AEM, et al. Cumulative cultural evolution and mechanisms for cultural selection in wild bird songs. Nat Commun [Internet]. 2022 Jul 11 [cited 2023 Apr 12];13(1):4001. Available from: <https://www.nature.com/articles/s41467-022-31621-9>

42. Wilks CEH, Blakey KH. In the jungle of cultural complexity. Evol Anthropol [Internet]. 2018 Sep [cited 2023 Aug 21];27(5):180–3. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/evan.21724>

43. Tomasello M. The Cultural Origins of Human Cognition. Harvard University Press; 1999.

44. Garland EC, Garrigue C, Noad MJ. When does cultural evolution become cumulative culture? A case study of humpback whale song. Phil Trans R Soc B [Internet]. 2022 Jan 31 [cited 2023 Aug 21];377(1843):20200313. Available from: <https://royalsocietypublishing.org/doi/10.1098/rstb.2020.0313>

45. Gruber T, Chimento M, Aplin LM, Biro D. Efficiency fosters cumulative culture across species. Phil Trans R Soc B [Internet]. 2022 Jan 31 [cited 2023 Apr 12];377(1843):20200308. Available from: <https://royalsocietypublishing.org/doi/10.1098/rstb.2020.0308>

46. Nolan PM, Hill GE. [Female choice for song characteristics in the house finch](https://doi.org/10.1016/j.anbehav.2003.03.018). Animal Behaviour. 2004;67(3):403–10.

47. Mennill DJ, Badyaev AV, Jonart LM, Hill GE. [Male house finches with elaborate songs have higher reproductive performance](https://doi.org/10.1111/j.1439-0310.2006.01145.x). Ethology. 2006;112(2):174–80.

48. Ciaburri I, Williams H. [Context-dependent variation of house finch song syntax](https://doi.org/10.1016/j.anbehav.2018.11.001). Animal Behaviour. 2019;147:33–42.

49. Sinclair NC, Ursell J, South A, Rendell L. From Beethoven to Beyoncé: do changing aesthetic cultures amount to “cumulative cultural evolution?” Front Psychol [Internet]. 2022 Feb 9 [cited 2023 Aug 21];12:663397. Available from: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.663397/full>

50. Geller FC, Lahti DC. [Is sexiness cumulative? Arguments from birdsong culture](https://doi.org/10.1016/j.anbehav.2023.09.006). Animal Behaviour. 2023;205:131–7.

51. Roeske TC, Tchernichovski O, Poeppel D, Jacoby N. Categorical Rhythms Are Shared between Songbirds and Humans. Current Biology [Internet]. 2020;30(18):3544–3555.e6. Available from: <https://doi.org/10.1016/j.cub.2020.06.072>

52. Roeske TC, Kelty-Stephen D, Wallot S. Multifractal analysis reveals music-like dynamic structure in songbird rhythms. Scientific Reports [Internet]. 2018;8(1):1–15. Available from: <http://dx.doi.org/10.1038/s41598-018-22933-2>

53. Rothenberg D, Roeske TC, Voss HU, Naguib M, Tchernichovski O. Investigation of musicality in birdsong. Hearing Research [Internet]. 2014;308(September):71–83. Available from: <http://dx.doi.org/10.1016/j.heares.2013.08.016>

54. Fitch WT. [The biology and evolution of music: A comparative perspective](https://doi.org/10.1016/j.cognition.2005.11.009). Cognition. 2006;100(1):173–215.

55. Shannon RV. Is birdsong more like speech or music? Trends in Cognitive Sciences [Internet]. 2016;20(4):245–7. Available from: <http://dx.doi.org/10.1016/j.tics.2016.02.004>

56. Rohrmeier M, Zuidema W, Wiggins GA, Scharff C. [Principles of structure building in music, language and animal song](https://doi.org/10.1098/rstb.2014.0097). Philos Trans R Soc B. 2015;370:20140097.

1. Minds and Traditions Research Group, Max Planck Institute for Geoanthropology, [masonyoungblood@gmail.com](mailto:masonyoungblood@gmail.com) [↑](#footnote-ref-20)
2. Institute for Advanced Computational Science, Stony Brook University [↑](#footnote-ref-22)