PROCEEDINGS OF THE ROYAL SOCIETY

B | BIOLOGICAL SCIENCES

Animal vocal sequences: not the Markov chains we thought they were

Arik Kershenbaum, Ann E. Bowles, Todd M. Freeberg, Dezhe Z. Jin, Adriano R. Lameira and Kirsten Bohn

Supplementary data	"Data Supplement" http://rspb.royalsocietypublishing.org/content/suppl/2014/08/18/rspb.2014.1370.DC1.html
References	**This article cites 50 articles, 5 of which can be accessed free** http://rspb.royalsocietypublishing.org/content/281/1792/20141370.full.html#ref-list-1
Subject collections	Articles on similar topics can be found in the following collections behaviour (1249 articles)
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Proc. R. Soc. B* go to: **http://rspb.royalsocietypublishing.org/subscriptions**

Animal vocal sequences: not the Markov chains we thought they were

Arik Kershenbaum[1], Ann E. Bowles[2], Todd M. Freeberg[3], Dezhe Z. Jin[4], Adriano R. Lameira[5,6] and Kirsten Bohn[7]

[1]National Institute for Mathematical and Biological Synthesis, Knoxville, TN, USA
[2]Hubbs SeaWorld Research Institute, San Diego, CA 92109, USA
[3]Department of Psychology, University of Tennessee, Knoxville, TN, USA
[4]Department of Physics and the Center for Neural Engineering, Penn State University, University Park, PA, USA
[5]Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Sciencepark 904, 1098 XH, Amsterdam, The Netherlands
[6]Pongo Foundation, Papenhoeflaan 91, 3421 XN, Oudewater, The Netherlands
[7]Department of Biological Sciences, Florida International University, Miami, FL, USA

Many animals produce vocal sequences that appear complex. Most researchers assume that these sequences are well characterized as Markov chains (i.e. that the probability of a particular vocal element can be calculated from the history of only a finite number of preceding elements). However, this assumption has never been explicitly tested. Furthermore, it is unclear how language could evolve in a single step from a Markovian origin, as is frequently assumed, as no intermediate forms have been found between animal communication and human language. Here, we assess whether animal taxa produce vocal sequences that are better described by Markov chains, or by non-Markovian dynamics such as the 'renewal process' (RP), characterized by a strong tendency to repeat elements. We examined vocal sequences of seven taxa: Bengalese finches *Lonchura striata domestica*, Carolina chickadees *Poecile carolinensis*, free-tailed bats *Tadarida brasiliensis*, rock hyraxes *Procavia capensis*, pilot whales *Globicephala macrorhynchus*, killer whales *Orcinus orca* and orangutans *Pongo* spp. The vocal systems of most of these species are more consistent with a non-Markovian RP than with the Markovian models traditionally assumed. Our data suggest that non-Markovian vocal sequences may be more common than Markov sequences, which must be taken into account when evaluating alternative hypotheses for the evolution of signalling complexity, and perhaps human language origins.

1. Introduction

Many species of animals produce vocalizations comprising multiple element types, combined into complex sequences. Some species have vocal repertoires of tens or even hundreds of discrete elements; others have only a handful, but use them to generate a wide variety of combinations. For example, an individual mockingbird *Mimus polyglottos* can mimic over 100 distinct song types of different species, and combine them into diverse sequences [1]. Even the rock hyrax *Procavia capensis*, using no more than five discrete vocal elements, creates long vocal sequences that are rarely the same on repetition [2]. Thus, even species with few vocal elements can sometimes generate an apparently unbounded range of possible combinations. Such varied vocal behaviour raises the question of the role and origin of complexity in animal vocal communication, and the comparison of vocal complexity across taxa, including human speech.

Complexity seems easy to identify, but hard to define, and even harder to quantify [3]. Numerous metrics have been suggested to ascribe a value to the complexity of vocal repertoires. However, these metrics all rely, either explicitly or implicitly, on assumptions of the underlying process that generated sets of sequences. For instance, a frequently cited complexity measurement, Shannon entropy, is only appropriate when each element in a sequence is produced independently of all other elements (i.e. an independent production process),

2

although the assumption of independence is rarely tested [4–7]. If vocal sequences are generated by a non-independent random process, however, Shannon entropy is probably not suitable for quantifying complexity [8]. Whether vocal sequences are random independent processes or conform to some other non-independent stochastic model, identifying the process operating is an essential task for quantifying and comparing sequence properties.

Beyond the application to complexity metrics, uncovering the processes underlying vocal sequence generation in animals may prove crucial to our understanding of language origins. Vocal complexity naturally brings to mind human language; however, the comparison appears to be inappropriate. One of the main differences between language and non-human animal communication is the grammar used to produce sequences. Human language uses 'context-free grammars' (CFGs) that are capable of generating recursive sequences and unbounded correlations [9,10]. By contrast, animal vocal sequences are usually described as 'regular grammars', the simplest class of formal grammars [11], and many researchers have analysed animal vocalizations as such [12–15]. Regular grammars correspond to finite state automata (FSA), because they comprise a set of rules that could instruct a simple machine (automaton) to move between a (finite) number of well-defined states. In the case of vocal sequences, each state is an acoustic element. Finite state automata can be deterministic; for example, *syllable 'A' is always followed by syllable 'B'*. They can also be probabilistic (pFSA), in which case multiple possible transitions between states are governed by fixed probabilities; for example, *syllable 'A' is followed by syllable 'B' 90% of the time, and by syllable 'C' 10% of the time*. In contrast to deterministic finite state automata, different sequences can be generated each time a pFSA is used. pFSAs are an example of a Markov chain [16], the most common model used to examine animal vocal sequences [14]. The pFSA (or Markovian) paradigm assumes that future occurrences (or the probability of each future occurrence) are entirely determined by a finite number of past occurrences. This property of a stochastic sequence is known as the Markov property. For example, the probability of the next syllable in a sequence being of type '*A*' is determined by the types of the immediately preceding syllables—or at most some finite number of preceding syllables.

pFSAs remain popular for characterizing animal vocal sequences [11,14], as the mechanism for producing Markov chains is easily understood, and simple neural mechanisms for implementing them have been postulated, based on neuroanatomical observations [17,18]. However, Markov chains are insufficient for producing the complexity of any human language [9], and there exist grammatical structures that no pFSA can generate, in particular tree-like syntax such as 'the hyrax ate the grass that grew near the rock under the tree' [11]. Furthermore, no intermediate grammatical form exists between pFSA models and the CFG of human language [9]. It is not clear what adaptive force could drive the gradual evolution of CFGs in a species that uses only pFSA vocal communication. In computer science, the addition of register memory, which provides the ability to count the number of repetitions of a syllable, appears to be a simple transition from regular to context-free automata [19]. However, such models have not been described in animal communication.

Despite the widespread use and simplicity of pFSA, there are other, non-Markovian stochastic processes, in particular models where future occurrences are determined by the (infinite) entirety of preceding events [20]. Non-Markovian processes have been used to describe (non-vocal) animal behaviour, for instance the renewal process (RP) model in the reproductive behaviour in sticklebacks, canaries and *Drosophila* [21], and the psychohydraulic model (PHM) of motivation proposed by Konrad Lorenz [22] for basic drives such as hunger. Although we are not aware of any prior work using non-Markovian processes to describe vocal behaviour, they seem likely candidates for vocal production. For example, non-Markovian mechanisms are able to describe both rapid shifts among vocal elements and long strings of repeated elements. Here, we test whether vocal sequences in several species are more consistent with a Markovian pFSA model, or a non-Markovian process, such as the RP or PHM. Non-Markovian stochastic processes like the RP have properties somewhat between the pFSA and the CFG, and the investigation of language evolution would not be complete without consideration of other, biologically realistic sequence-generating mechanisms.

Both RP and PHM models are considered non-Markovian because they do not rely on finite memory. In RP models, a particular behaviour (for instance, production of a particular vocal syllable) is repeated for some probabilistically determined time. Transitions between syllables of different types are still defined by a transition table as with a pFSA, but the number of repeats of each syllable in between transitions may be drawn from a distribution (e.g. Poisson). Although at first surprising, it can be shown that the sequence generated by such a process is non-Markovian [23] and cannot be well described by a pFSA. The RP does not fit the Markovian paradigm of finite memory, since the Poisson tail is unbounded. The PHM also relies on a nominally unbounded memory; in this case, the probability of a particular syllable occurring increases with the time since its last occurrence, and then falls to a minimum as soon as the syllable is used.

We gathered vocal sequences from seven taxa: the Bengalese finch *Lonchura striata domestica* [24], Carolina chickadee *Poecile carolinensis* [25,26], free-tailed bat *Tadarida brasiliensis* [13], rock hyrax *Procavia capensis* [2], short-finned pilot whale *Globicephala macrorhynchus* [27], killer whale *Orcinus orca* [28], and orangutan *Pongo abelii* and *Pongo pygmaeus wurmbii* [29]. For comparison with a human sequence corpus, we also analysed letter order in a sample of English (the text of the play *Hamlet* [30]), although the intention was not to imply that letter order in human language has any relevance to the evolution of vocal sequences in animals. These sequences were coded for distinct vocal elements (syllables) as described in the above-cited previous works. We aimed to match these sequences to the most likely generation model, from a range of possible models of varying complexity by testing each species's sequences against stochastic production from six different prospective processes: (i) a zero-order Markov process (ZOMP), (ii) a first-order Markov process (FOMP), (iii) a second-order Markov process (SOMP), (iv) a hidden Markov model (HMM), (v) an RP and (vi) a PHM process.

2. Material and methods

(a) Description of the stochastic processes

First, we describe each of these processes in detail. Consider a sequence S of n elements $[1 \ldots n]$, taken from a set of C different element types. The ZOMP defines a production process where each element is generated according to a fixed prior probability π,

3

independent of the preceding elements, so that the probability of the nth element S_n being of type $i = [1 \ldots C]$ is given by $\Pr(S_n = i) = \pi_i$. In the FOMP, the probability that the nth element will be of type i is determined only by the preceding element j, and the $C \times C$ transition matrix $\mathbf{T}$, which defines the probability that element i will occur after element j, so that $\Pr(S_n = i \mid S_{n-1} = j) = T_{j,i}$. Similarly, the SOMP defines the probability of the nth element in terms of the two preceding elements: $\Pr(S_n = i \mid S_{n-1} = j, S_{n-2} = k) = U_{(j,k),i}$. Note that the size of the second-order transition matrix $\mathbf{U}$ is of size $C^2 \times C$, which indicates the rapid increase in sample size required for accurate estimates of the transition probabilities, as the order of a Markov process increases [31].

The HMM [32] provides a more parsimonious and memory-efficient representation of higher-order Markov processes, and has been used successfully to capture the characteristics of vocal sequences from different species (e.g. [24,33]). As with the traditional Markov models, in generating an HMM sequence, successive elements are chosen probabilistically given the current state. Unlike traditional Markov models, though, in the HMM the states themselves are not explicitly defined in terms of preceding sequences of a fixed number of elements, but are constructed from the data by an expectation-maximization optimization known as the Baum–Welch algorithm [32]. This allows the HMM to represent a combination of low- and high-order Markov relationships within the same model.

The RP is defined by a first-order transition matrix, which determines the pFSA transitions between different elements. This matrix $\mathbf{R}$ is defined in a similar way to the FOMP transition matrix $\mathbf{T}$, but with zeroes along the main diagonal. Instead, those self-transitions are generated by a separate stochastic process. In this case, we define the number of repeated elements as being drawn from a Poisson distribution with mean $\boldsymbol{\mu}$, with a separate Poisson distribution for each element type i. A graphical description of the differences between the RP and Markovian processes can be found in [8], and is reproduced in the electronic supplementary material, figure S1.

Although the PHM has not previously been used to describe animal communication, it forms a useful counterpoint to the RP. Whereas in an RP model repeated elements occur more often than would be expected in a Markov model, in a PHM repeated elements are less common than expected. We implement a simplified PHM by defining for each element type i a function of the form $A_i(t_i) = 1 - e^{-\kappa_i t_i}$, where t_i is the time elapsed since element i last appeared, k_i is an element-specific rate constant and A_i is the equivalent of what Lorenz coined the 'action-specific energy' (i.e. the driving motivational force that builds up within an animal until a particular behaviour is precipitated). The probability of the next element being of type i is then given by $\Pr(i) = A_i / \sum_i A_i$.

(b) Generation of synthetic sequences

We determined the maximum-likelihood estimator parameters for each of the processes, given the empirical data. For the ZOMP, the parameter $\boldsymbol{\pi}$ is simply the observed prior probabilities of each of the element types i. For the FOMP and SOMP, the matrices $\mathbf{T}$ and $\mathbf{U}$ are estimated from the number of occurrences of the specific transitions between element types within the observed sequences.

For the HMM, the parameters of the model are calculated from the empirical data using the standard Viterbi algorithm [34]. In any HMM implementation, the number of states is a crucial factor in the model performance; therefore, we optimized the number of hidden states by minimizing the Akaike information criterion [35]. To do this, we calculated the log likelihood of generating the training sequence from the trained HMM and used the number of hidden states as the number of parameters in the information criterion calculation.

For the RP, the matrix of transitions between different elements $\mathbf{R}$ is calculated as for the FOMP, and the means $\boldsymbol{\mu}$ of the repeated-element Poisson distributions are estimated from the empirical distributions of number of repeats, separately for each element type i. For the PHM, the rate constants $\boldsymbol{\kappa}$ are also estimated from the empirical distributions of the intervals between elements of the same type.

Having extracted the maximum-likelihood estimator parameters for each model, we used these to generate artificial sequences based on each model, where the sequence lengths matched those of the original vocal datasets (figure 1). An overview of the dataset sizes and sequence lengths is given in table 1, and the data themselves are available in the electronic supplementary material (data.xls). For each species, we generated 200 artificial datasets using each of the model types.

(c) Comparison of artificial and recorded sequences

Determining whether a particular sequence, vocal or otherwise, is Markovian or not is a non-trivial problem. Rigorous tests for finite-state sequences exist [37], but are not easily applied to datasets of limited size. When limited data are available, transition probabilities are poorly estimated by the small number of transitions observed. In addition, rare states may be completely absent. Previous authors have used multiple measures of the statistical properties of the sequences, such as n-gram distribution [24]. However, these techniques measure aggregate similarity and do not directly address the similarity of the individual sequences. Aggregate comparisons may be an effective way of comparing very long sequences, where the probability distribution of n-grams would be expected to be limiting. However, they would be less accurate when comparing short strings such as those found in real recordings, and when the processes generating these strings may not be stationary (for instance, owing to shifting motivational state and responses to external events). We used a more direct method by comparing each simulated sequence with the corresponding original data, and calculating the edit (Levenshtein) distance [38] between the pair of sequences. Levenshtein distance measures the minimum number of insertions, deletions and replacements necessary to convert one sequence into another, and has been used for assessing vocal syntax in previous studies [2,39,40]. This distance gives a measure of dissimilarity between the simulated and original sequences, which we then averaged over the entire dataset. We calculated the Levenshtein distance between corresponding sequences, both in the simulated and original data, to generate a pairwise distance matrix. We then repeated this for 200 randomly generated sequence datasets for each model and each species.

Having measured the mean Levenshtein distance between a dataset and the maximum-likelihood estimator prospective models, we used multi-dimensional scaling [41] to convert the Levenshtein distance matrix to a series of Cartesian vectors, one for each sequence, which preserved to the greatest extent possible the pairwise Levenshtein distances between all of the sequences. The transformation of the distance matrix to a feature-space matrix allowed us to use classification algorithms for assigning the simulated data to the most likely model. For each dataset of N sequences, we used the MATLAB function $cmdscale$ to convert the $N \times N$ distance matrix to a matrix Y consisting of a series of N vectors of length p, where $p < N$ is the minimum dimensionality in which the N points can be embedded (i.e. where the pairwise distances between the points are conserved). We then reduced the dimensionality of each vector to length $q \leq p$, where q is the number of eigenvalues E of $Y \cdot Y'$ for which E is positive, and the change in successive eigenvalues $\Delta E = [E(r) - E(r+1)]/E(r)$, $r = [1 \ldots p - 1]$ is greater than 1%.

We used both a naive Bayesian classifier and a Z-test to determine from which of the six generation models the original
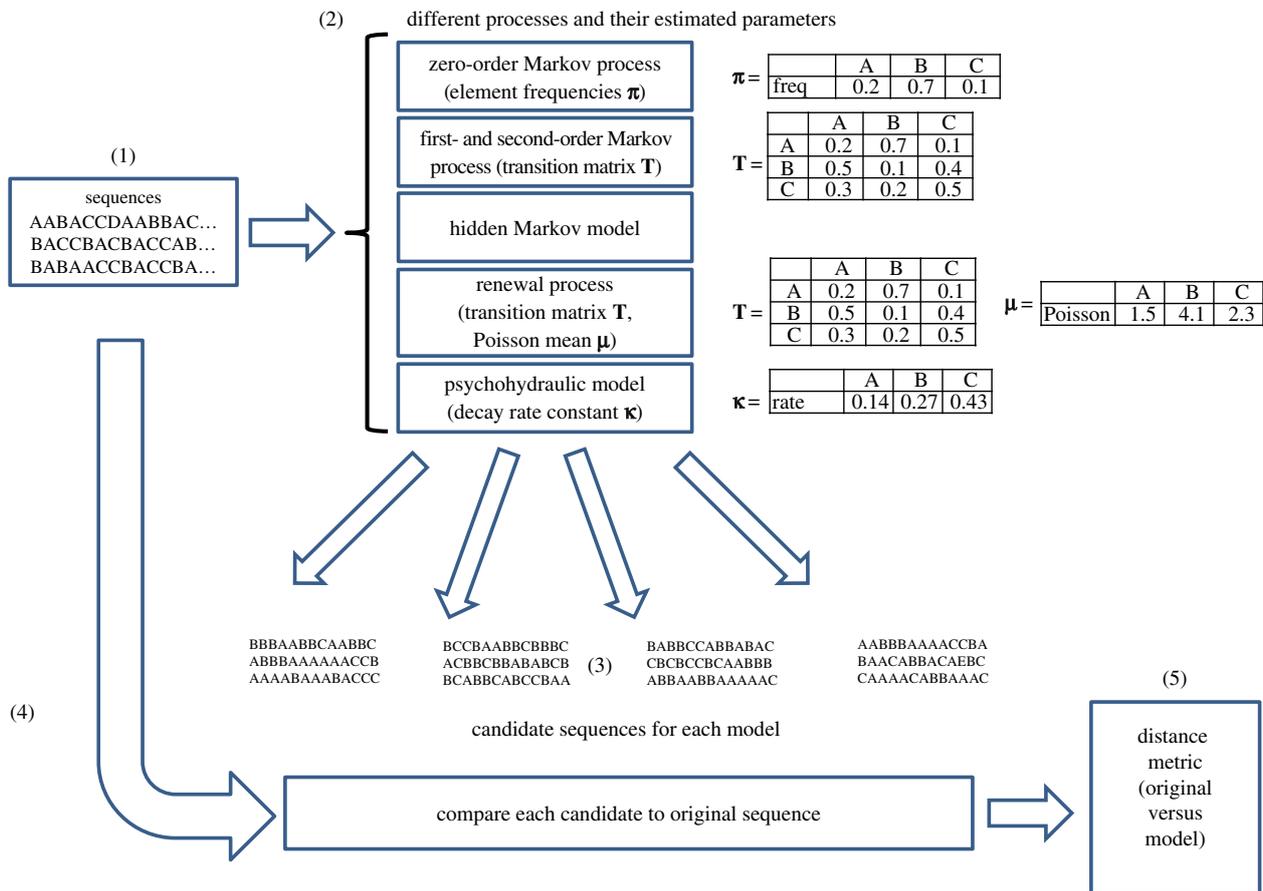
Figure 1. Flow diagram illustrating the calculation of the distance metric between model and empirical data. Empirical sequences (1) are used to derive maximum-likelihood estimator parameters for each of the models (2). Using these parameters, simulated sequences are generated (3) and compared to the corresponding original sequences (4). The average edit distance between these pairs of sequences is a measure of similarity between sequence and model (5).

Table 1. Summary of the datasets used and their characteristics.

species	no. element types	no. sequences	total sequence length	source
free-tailed bat	3	71	514	[13]
Tadarida brasiliensis				
rock hyrax	5	263	3296	[2]
Procavia capensis				
Bengalese finch	7	2130	27 858	[24]
Lonchura striata domestica				
Carolina chickadee	7	4246	37 094	[25,26]
Poecile carolinensis				
short-finned pilot whale	20	18	246	[27]
Globicephala macrorhynchus				
orangutan	7	32	373	[29]
Pongo spp.				
killer whale	5	8	224	[28]
Orcinus orca				
English language	25	455	3816	[30]

sequences were most likely to have been drawn. The naive Bayesian classifier [42] calculated the posterior probabilities of belonging to each of the six model clusters, in q-dimensional space, given the distribution of the 200 sequences for each of the six models. The model with the highest posterior probability was chosen as the candidate model. We then performed an additional Z-test, using the MATLAB *normfit* function, to compare the mean distance of the original data to 200 simulated samples
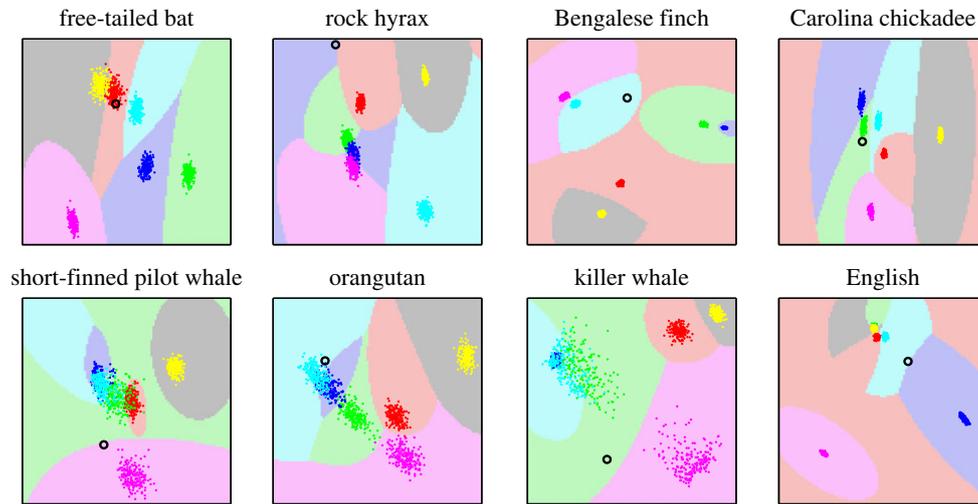
Figure 2. Location of the simulated sequences and original data (black circle) in two-dimensional Levenshtein distance space. Coloured points indicate only the first 30 randomly generated sequences from each model, for clarity: ZOMP (red), FOMP (green), SOMP (blue), HMM (cyan), RP (magenta) and PHM (yellow). Solid colours indicate the domains of the naive Bayesian classifier for each model type.
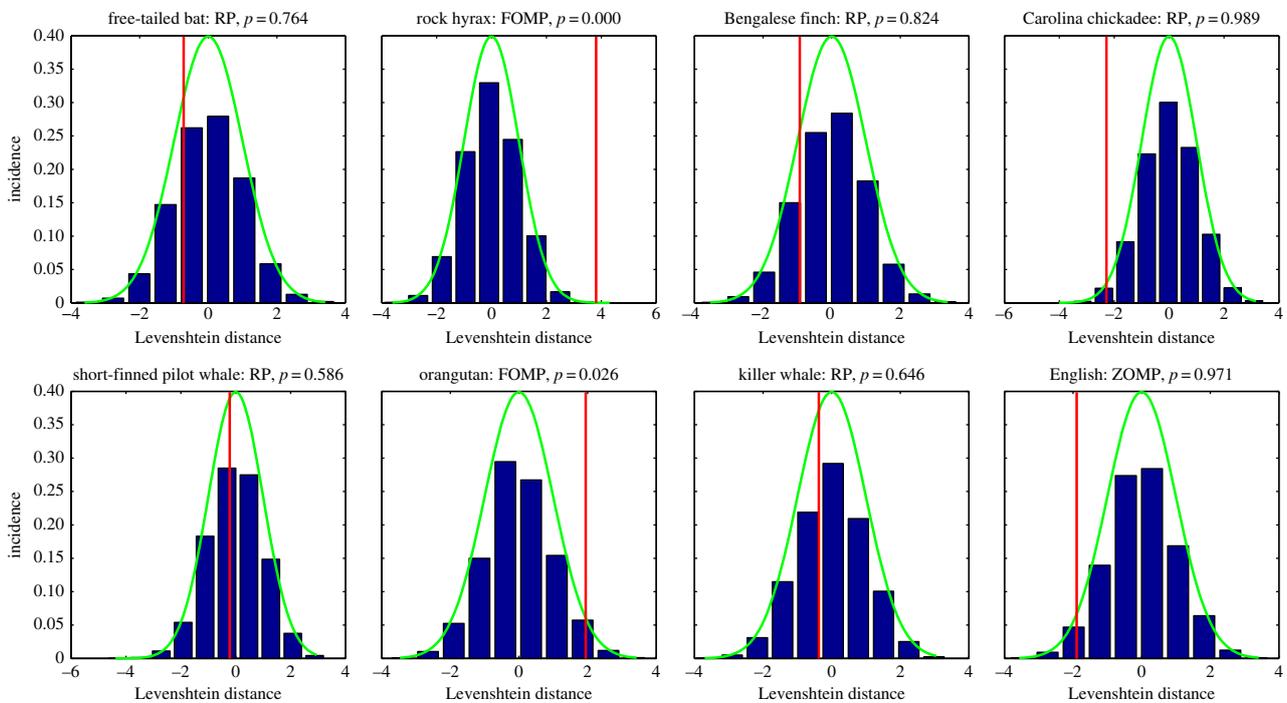

Figure 3. Histograms of the Levenshtein distances of simulated sequences from each other (blue bars) for the best-fit model (indicated in the title of each panel), and the fitted normal distribution (green line) using the MATLAB *normfit* function. The red line shows the mean Levenshtein distance of the original data from the simulated sequences, and the *p*-value indicates the probability of this mean distance (or greater) having been drawn from the distribution of simulated sequences (*Z*-test).

of the candidate model. We used a Monte Carlo method to take into account the variation within each model, and gave an estimate of the probability that the observed data were drawn from a distribution characterized by the 200 simulated samples of the candidate model. Simulated sequences that are very similar to each other (low variance) are clustered together in distance space, whereas simulated sequences with a high variance are spread out in distance space (figure 2). Therefore, any particular empirical dataset is more likely to fall within 95% confidence limits of a high-variance model than a low-variance one.

Higher-order Markov models are, by definition, generalizations of lower-order models, and in particular, the HMM is a generalization of any arbitrary-order Markov model. Therefore, it might appear that an HMM model must necessarily provide

a maximum-likelihood estimator of the model parameters that is at least as accurate as lower-order models, if less parsimonious (having a greater number of model parameters). However, we compared the original sequences directly to the corpus of generated sequences, so our similarity metric more broadly measured the appropriateness of each model, and often showed a better fit from the lower-order models (figure 3). We also performed an analysis of variance and a post-hoc Tukey test to assess whether the Levenshtein distances between the original sequences and their corresponding simulated sequences are significantly different among the different models.

Given that transition probability estimates are likely to be inaccurate for small sample sizes, we tested the robustness of our conclusions by repeating the analyses using smaller subsets

of the empirical data. We sub-sampled each dataset and determined the best-fit model for each sample size. If the dataset in its entirety is of sufficient size to estimate the best-fit model, we expect that the best-fit model would be consistent between the larger and full sample sizes.

3. Results

For the purpose of visualization, the naive Bayesian classifier is illustrated in figure 2 with a two-dimensional embedding, rather than a full q-dimensional embedding (although the two-dimensional embedding is in general insufficient to capture the distribution of the sequences in Levenshtein distance space). Figure 2 illustrates the location of the simulated sequences in distance space, the location of the original data and the domains of the classifier. Note that the spread of the simulated sequences varies substantially between models. For those models where the simulated sequences are tightly grouped (small Levenshtein distance between them), the Z-test is more likely to reject the hypothesis that the original data belong to this model, as the variance of the simulated sequences is small, and the original data are likely to fall several standard deviations from the mean of the simulated cluster. Figure 3 shows the results of the Z-test, comparing the distribution of distances within the simulated dataset of the candidate model and the distance of the original data from the simulated set. Where the original data distance is far from the intra-model distances, the data are unlikely to have been drawn from the model.

Table 2 shows the results of the Bayesian classification for each of the eight species (including English), along with the result of the Z-test for the most likely candidate model. The Shapiro–Wilk test for normality did not reject a normal distribution for any of the best-fit models, supporting the use of a Z-test. Electronic supplementary material, table S1 shows the results of the Z-tests for all models. Of the seven non-human species, none show clear Markovian behaviour. The Bengalese finch, Carolina chickadee, free-tailed bat, pilot whale and killer whale appear most similar to the non-Markovian RP, and the Z-test does not reject the RP model ($p = 0.824$, $p = 0.989$, $p = 0.764$, $p = 0.586$, $p = 0.646$, respectively). The orangutan and the hyrax are most similar to the Markovian FOMP, but for both of these species the FOMP is a poor fit to the data, and the vocal sequences are sufficiently different that we reject the null hypothesis of belonging to that model (orangutan $p = 0.026$, hyrax $p < 0.001$). Letter order in a sample of English writing appears to follow a Markovian ZOMP model ($p = 0.914$). The PHM did not appear to be a good model for any of the datasets tested.

The test of robustness by varying sample size showed that for all datasets used, except the pilot whale (which passed the Z-test) and the hyrax (which failed the Z-test), the conclusion of best-fit models was consistent at larger sub-sample sizes (see the electronic supplementary material, figure S2).

4. Discussion

Our results show that the vocal sequences of over half of the species studied—the Bengalese finch, the Carolina chickadee, the free-tailed bat, the pilot whale and the killer whale—can be better described as non-Markovian RPs, rather than traditional first-order, second-order or arbitrary-order HMMs.

Table 2. Results of the Bayesian classifier to find the best-fit model to the observed data. Embedding dimension shows the number of multidimensional scaling dimensions used for the classification, and p-values indicated by asterisk show that the Z-test rejects the hypothesis that the data belong to the best-fit model.

species	best-fit model	embedding dimension	p-value
free-tailed bat	RP	5	0.764
hyrax	FOMP	7	<0.001*
Bengalese finch	RP	7	0.824
chickadee	RP	8	0.989
pilot whale	RP	14	0.586
orangutan	FOMP	9	0.026*
killer whale	RP	14	0.646
English	ZOMP	6	0.914

*$p < 0.05$.

We cannot reliably identify a stochastic process generating the sequences of the hyrax or the orangutan, and it would be interesting to investigate why these vocalizations are qualitatively different from the others studied, whether because of phylogeny, functionality or other constraints.

This diversity of production models is quite unexpected, as previous works have overwhelmingly used the Markovian paradigm as a starting point for the analysis of animal vocal sequences [11,14]. Although putative Markovian generation processes are popular, partly because of their simplicity and partly because of the clear role that they fill in the Chomsky hierarchy [9,10], it is inappropriate to assume that they adequately describe the true generation process simply because of their utility. Indeed, it seems simplistic to assume that animals would primarily generate their vocal sequences based solely on a small number of preceding elements. RPs, in which a certain element is repeated until the animal is 'tired of it' (whether physically, cognitively or only figuratively), are alternative models, and indeed we have shown the RP to be a better approximation for the vocalizations of most of the species we examined.

Repetition has long been recognized as a feature of animal behaviour (e.g. eventual variety in birdsong [43] and non-vocal behavioural repetition [21]), although the mechanisms responsible may be diverse [44,45]. Mating and aggressive displays make use of repetition to augment the magnitude of the display signal [44], and repeated displays appear more effective in attracting a mate or deterring a rival in species including songbirds [46] and fallow deer *Dama dama* [47]. However, a trade-off must exist between the benefit of signal repetition and energetic costs or physiological constraints [46,48]. Such a trade-off may be a proximal cause of a non-monotonic distribution of the number of repeats, such as the Poisson distribution of the proposed RP, which appears to be consistent with our empirical data.

Characterizing vocal sequences as Markov chains places animal vocal sequences in the category of regular grammars and distinguishes them from the more complex context-free structure of human language. However, we have shown that the oft-cited conclusion that all animal communication

7

rspb.royalsocietypublishing.org Proc. R. Soc. B 281: 20141370

conforms to regular grammars [11,18] is misleading. Indeed, little mention has been made in the literature of non-Markovian alternatives to the pFSA grammar. No attempt has been made until now to test whether animal vocal sequences are indeed most likely to be generated by pFSAs, or instead by some other, non-Markovian stochastic process. It has been pointed out [49,50] that insufficient attention has been given to the different levels of complexity in pFSAs of different types (i.e. different orders versus HMMs), and we extend this observation to non-Markovian processes. Claims that certain species such as European starlings *Sturnus vulgaris* perceive vocal sequences with a grammar more complex than regular grammar have met with scepticism [36,51,52]. However, our findings do not point to *greater* grammatical complexity, but to *different* grammatical processes, something so far barely examined in the literature.

For this comparison, we have used a small but diverse set of data. Some of the datasets, such as the sequences obtained from orangutans, were necessarily rather small because of the difficulty of working in the wild with an inaccessible, endangered and semi-solitary species. The sequences from the pilot whales potentially contained biasing information, since the audio recorders attached to the animals could also detect the calls of other individuals. However, such a bias would tend to produce a more independent (ZOMP-like) sequence, whereas our findings for the pilot whale indicated a low probability of independent generation. The pilot whale data were also unusual in that they consisted primarily of stereotyped calls, with few non-stereotyped calls; the occurrence of such sequences is probably highly context-dependent [27]. This could indicate either atypical behaviour or, possibly, unusually communicative behaviour. We believe that, despite these limitations, inclusion of these species helps to broaden the scope of our comparison, since primates and cetaceans are mammalian orders recognized as having particularly sophisticated acoustic communication.

Estimating the parameters of probabilistic models from small sample sizes is necessarily problematic [53,54]. For example, an FOMP transition table for the English alphabet is a matrix of size $26 \times 26 = 676$ cells, and assuming that at least 10 observations are required to provide a reasonable estimate of each transition probability, then at least 6760 transitions must be made. In practice, the required number of observations is much more, as some transitions may be rarely observed. For an SOMP, more than 10^5 observations are required. In most cases, our data fall far short of the desirable number of observations. However, reasonable estimates of model parameters can often be made with surprisingly small sample sizes [8,54]. The results of our test for robustness show that, with the exception of the pilot whale data, wherever a clear best-fit model is indicated, the chosen model is consistent between larger sub-sample sizes, and we believe that this is an indication of reliability of our conclusions.

We found that the letter order in the English language is best modelled by a ZOMP, whereas previous work has indicated that an SOMP is a more appropriate model [55]. However, English is clearly neither a ZOMP nor an FOMP, and so the resemblance of a corpus of English letters to one or the other may be more dependent on the metric used to assess similarity, rather than the underlying stochastic processes. Information-theoretic approaches [37] naturally lean towards the second-order Markovian paradigm; for example, because the letter 't' is so often followed by the letter 'h'.

However, we believe that our approach of comparing the string similarity of sequences generated by putative models provides a more useful comparison in the field of animal communication research, although possibly less useful for analysing human texts.

To the best of our knowledge, no extant species other than humans have a true language, with an unlimited ability to communicate abstract concepts [56]. Although many non-human animal species have essential precursor abilities, such as vocal production learning [57], contextual reference [58–61] and non-semantic syntax [2,27,62], only humans have a grammatical structure that is sufficiently complex for true linguistic potential [56]. Since no non-human species demonstrate proto-linguistic grammars, proposed mechanisms for the evolution of language in humans remain speculative (e.g. [63]). Among theories of language origin that posit language evolving from systems like extant non-human animal communication, it is debated whether language arose as a gradual adaptation of simpler vocal communication systems [64] or gestural systems [65], or whether essential linguistic abilities arose suddenly (or at least very rapidly) [11]. Although a conceptual path between regular and supra-regular grammars is well accepted in the computer science literature [19], an important question is whether an incremental *evolutionary* path exists between the pFSA regular grammars that heretofore were considered common in animals, and the CFG linguistic structures that exist in humans. The incremental hypothesis must explain the lack of 'proto-languages' in the animal kingdom, representing a link between animal and human linguistic capabilities [66]. Conversely, saltationary or rapid-evolution hypotheses must provide a convincing and evolutionarily plausible mechanism that could explain the qualitative gap between the regular grammar of animal communication and the CFG of human language. Examples would include metric (timing) features [11], or a synthesis of multiple regular grammars [63] as a 'bridge' between the two capabilities. Recent work has indicated that complex syntax can develop as the result of simple neurological changes; for example, in Bengalese finches, which have syntax qualitatively more complex than their wild ancestors [67].

Our findings appear to indicate that pFSA is *not* the ubiquitous nature of animal vocal sequences, and this requires re-evaluation of both gradual and saltational hypotheses. Application of our analysis to more species, and the use of more putative non-Markovian stochastic models, may reveal intermediate steps between known Markovian animal grammars and human CFGs, narrowing the gap between human and non-human animal communicative abilities.

References

1. Gammon DE, Altizer CE. 2011 Northern mockingbirds produce syntactical patterns of vocal mimicry that reflect taxonomy of imitated species. *J. Field Ornithol.* **82**, 158–164. (doi:10.1111/j.1557-9263.2011.00318.x)

2. Kershenbaum A, Ilany A, Blaustein L, Geffen E. 2012 Syntactic structure and geographical dialects in the songs of male rock hyraxes. *Proc. R. Soc. B* **279**, 2974–2981. (doi:10.1098/rspb.2012.0322)

3. Edmonds B. 1999 What is complexity?—the philosophy of complexity per se with application to some examples in evolution. In *The evolution of complexity* (eds F Heylighen, D Aerts), pp. 1–16. Dordrecht, The Netherlands: Kluwer.

4. Da Silva ML, Piqueira JRC, Vielliard JME. 2000 Using Shannon entropy on measuring the individual variability in the rufous-bellied thrush *Turdus rufiventris* vocal communication. *J. Theor. Biol.* **207**, 57–64. (doi:10.1006/jtbi.2000.2155)

5. McCowan B, Doyle LR, Hanser SF. 2002 Using information theory to assess the diversity, complexity, and development of communicative repertoires. *J. Comp. Psychol.* **116**, 166–172. (doi:10.1037/0735-7036.116.2.166)

6. Suzuki R, Buck JR, Tyack PL. 2006 Information entropy of humpback whale songs. *J. Acoust. Soc. Am.* **119**, 1849–1866. (doi:10.1121/1.2161827)

7. Doyle LR, McCowan B, Hanser SF, Chyba C, Bucci T, Blue JE. 2008 Applicability of information theory to the quantification of responses to anthropogenic noise by southeast Alaskan humpback whales. *Entropy* **10**, 33–46. (doi:10.3390/entropy-e10020033)

8. Kershenbaum A. 2013 Entropy rate as a measure of animal vocal complexity. *Bioacoustics* **23**, 195–208. (doi:10.1080/09524622.2013.850040)

9. Chomsky N. 2002 *Syntactic structures*, 9th edn. The Hague, The Netherlands: de Gruyter Mouton.

10. Jäger G, Rogers J. 2012 Formal language theory: refining the Chomsky hierarchy. *Phil. Trans. R. Soc. B* **367**, 1956–1970. (doi:10.1098/rstb.2012.0077)

11. Berwick RC, Beckers GJ, Okanoya K, Bolhuis JJ. 2012 A bird's eye view of human language evolution. *Front. Evol. Neurosci.* **4**, 5. (doi:10.3389/fnevo.2012.00005)

12. Robinson JG. 1979 An analysis of the organization of vocal communication in the titi monkey *Callicebus moloch*. *Z. Tierpsychol.* **49**, 381–405. (doi:10.1111/j.1439-0310.1979.tb00300.x)

13. Bohn KM, Schmidt-French B, Schwartz C, Smotherman M, Pollak GD. 2009 Versatility and stereotypy of free-tailed bat songs. *PLoS ONE* **4**, e6746. (doi:10.1371/journal.pone.0006746)

14. ten Cate C, Okanoya K. 2012 Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Phil. Trans. R. Soc. B* **367**, 1984–1994. (doi:10.1098/rstb.2012.0055)

15. Briefer E, Osiejuk TS, Rybak F, Aubin T. 2010 Are bird song complexity and song sharing shaped by habitat structure? An information theory and statistical approach. *J. Theor. Biol.* **262**, 151–164. (doi:10.1016/j.jtbi.2009.09.020)

16. Grinstead CM, Snell JL. 1997 Markov chains. In *Introduction to probability* (eds CM Grinstead, JL Snell), pp. 405–470. Providence, RI: American Mathematical Society.

17. Jin DZ. 2009 Generating variable birdsong syllable sequences with branching chain networks in avian premotor nucleus HVC. *Phys. Rev. E* **80**, 051902. (doi:10.1103/PhysRevE.80.051902)

18. Katahira K, Suzuki K, Okanoya K, Okada M. 2011 Complex sequencing rules of birdsong can be explained by simple hidden Markov processes. *PLoS ONE* **6**, e24516. (doi:10.1371/journal.pone.0024516)

19. Minsky ML. 1967 *Computation: finite and infinite machines*. Englewood Cliffs, NJ: Prentice-Hall.

20. Van Kampen N. 1998 Remarks on non-Markov processes. *Braz. J. Phys.* **28**, 90–96. (doi:10.1590/S0103-97331998000200003)

21. Cane VR. 1959 Behaviour sequences as semi-Markov chains. *J. R. Stat. Soc. Series B Stat. Methodol.* **21**, 36–58.

22. Lorenz KZ. 1950 *Physiological mechanisms in animal behavior*, pp. 221–268. Oxford, UK: Academic Press.

23. Nelson R. 1995 *Probability, stochastic processes, and queueing theory: the mathematics of computer performance modeling*. New York, NY: Springer.

24. Jin DZ, Kozhevnikov AA. 2011 A compact statistical model of the song syntax in Bengalese finch. *PLoS Comput. Biol.* **7**, e1001108. (doi:10.1371/journal.pcbi.1001108)

25. Freeberg TM. 2012 Geographic variation in note composition and use of chick-a-dee calls of Carolina chickadees (*Poecile carolinensis*). *Ethology* **118**, 555–565. (doi:10.1111/j.1439-0310.2012.02042.x)

26. Freeberg TM. 2008 Complexity in the chick-a-dee call of Carolina chickadees (*Poecile carolinensis*): associations of context and signaler behavior to call structure. *Auk* **125**, 896–907. (doi:10.1525/auk.2008.07025)

27. Sayigh L, Quick N, Hastie G, Tyack P. 2012 Repeated call types in short-finned pilot whales, *Globicephala macrorhynchus*. *Mar. Mamm. Sci.* **29**, 312–324. (doi:10.1111/j.1748-7692.2012.00577.x)

28. Crance JL, Bowles AE, Garver A. 2014 Evidence for vocal learning in juvenile male killer whales, *Orcinus orca*, from an adventitious cross-socializing experiment. *J. Exp. Biol.* **217**, 1229–1237. (doi:10.1242/jeb.094300)

29. Lameira AR *et al*. 2013 Predator guild does not influence orangutan alarm call rates and combinations. *Behav. Ecol. Sociobiol.* **67**, 519–528. (doi:10.1007/s00265-012-1471-8)

30. Shakespeare W. 2014 *The tragedy of Hamlet, Prince of Denmark*. See http://shakespeare.mit.edu/hamlet/full.html (accessed 2014).

31. Doyle LR, McCowan B, Johnston S, Hanser SF. 2011 Information theory, animal communication, and the search for extraterrestrial intelligence. *Acta Astronaut.* **68**, 406–417. (doi:10.1016/j.actaastro.2009.11.018)

32. Cappé O, Moulines E, Rydén T. 2005 *Inference in hidden Markov models*. New York, NY: Springer Science Business Media.

33. Reby D, André-Obrecht R, Galinier A, Farinas J, Cargnelutti B. 2006 Cepstral coefficients and hidden Markov models reveal idiosyncratic voice characteristics in red deer (*Cervus elaphus*) stags. *J. Acoust. Soc. Am.* **120**, 4080–4089. (doi:10.1121/1.2358006)

34. Garivier A. 2012 The Baum–Welch algorithm for hidden Markov Models: speed comparison between octave / python / R / scilab / matlab / C / C++. See http://www.math.univ-toulouse.fr/~agarivie/Telecom/code/index.php (accessed 2013).

35. Burnham KP, Anderson DR. 2002 *Model selection and multimodel inference: a practical information-theoretic approach*. New York, NY: Springer.

36. Corballis MC. 2007 Recursion, language, and starlings. *Cogn. Sci.* **31**, 697–704. (doi:10.1080/15326900701399947)

37. DeDeo S. 2013 Evidence for non-finite-state computation in a human social system. *PLoS ONE* **8**, e75818. (http://arxiv.org/abs/1212.0018)

38. Ristad ES, Yianilos PN. 1998 Learning string-edit distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 522–532. (doi:10.1109/34.682181)

39. Gil D, Slater PJ. 2000 Song organisation and singing patterns of the willow warbler, *Phylloscopus trochilus*. *Behaviour* **137**, 759–782. (doi:10.1163/156853900502330)

40. Garland EC, Lilley MS, Goldizen AW, Rekdahl ML, Garrigue C, Noad MJ. 2012 Improved versions of the Levenshtein distance method for comparing sequence information in animals' vocalisations: tests using humpback whale song. *Behaviour* **149**, 1413–1441. (doi:10.1163/1568539X-00003032)

41. Cox TF, Cox MAA. 2000 *Multidimensional scaling*. Berlin, Germany: Springer.

42. Duda RO, Hart PE, Stork DG. 2012 *Pattern classification*. New York, NY: John Wiley & Sons.

43. Catchpole CK, Slater PJB. 2003 *Bird song: biological themes and variations*. Cambridge, UK: Cambridge University Press.

44. Payne RJH, Pagel M. 1997 Why do animals repeat displays? *Anim. Behav.* **54**, 109–119. (doi:10.1006/anbe.1996.0391)

45. Mowles SL, Ord TJ. 2012 Repetitive signals and mate choice: insights from contest theory. *Anim. Behav.* **84**, 295–304. (doi:10.1016/j.anbehav.2012.05.015)

46. Podos J. 1997 A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae). *Evolution* **51**, 537–551. (doi:10.2307/2411126)

47. Jennings DJ, Gammell MP, Carlin CM, Hayden TJ. 2005 Win, lose or draw: a comparison of fight structure based on fight conclusion in the fallow deer. *Behaviour* **142**, 423–439. (doi:10.1163/1568539054012001)

9

rspb.royalsocietypublishing.org   Proc. R. Soc. B **281**: 20141370

48. Drăgănoiu TI, Nagle L, Kreutzer M. 2002 Directional female preference for an exaggerated male trait in canary (*Serinuscanaria*) song. *Proc. R. Soc. Lond. B* **269**, 2525–2531. (doi:10.1098/rspb.2002.2192)

49. Hurford JR. 2011 *The origins of grammar: language in the light of evolution II*. Oxford, UK: Oxford University Press.

50. Petkov CI, Wilson B. 2012 On the pursuit of the brain network for proto-syntactic learning in non-human primates: conceptual issues and neurobiological hypotheses. *Phil. Trans. R. Soc. B* **367**, 2077–2088. (doi:10.1098/rstb.2012.0073)

51. Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC. 2006 Recursive syntactic pattern learning by songbirds. *Nature* **440**, 1204–1207. (doi:10.1038/nature04675)

52. Beckers GJ, Bolhuis JJ, Okanoya K, Berwick RC. 2012 Birdsong neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport* **23**, 139–145. (doi:10.1097/WNR.0b013e32834f1765)

53. Cover TM, Thomas JA. 1991 *Elements of information theory*, pp. 18–21. New York, NY: John Wiley & Sons.

54. Hausser J, Strimmer K. 2009 Entropy inference and the James–Stein estimator, with application to nonlinear gene association networks. *J. Mach. Learn. Res.* **10**, 1469–1484.

55. Kundu A, He Y. 1991 On optimal order in modeling sequence of letters in words of common language as a Markov chain. *Pattern Recognit.* **24**, 603–608. (doi:10.1016/0031-3203(91)90027-3)

56. Hauser MD, Chomsky N, Fitch W. 2002 The faculty of language: what is it, who has it, and how did it evolve? *Science* **298**, 1569–1579. (doi:10.1126/science.298.5598.1569)

57. Janik VM, Slater PJ. 1997 Vocal learning in mammals. *Adv. Study Behav.* **26**, 59–99. (doi:10.1016/S0065-3454(08)60377-0)

58. Slobodchikoff C, Kiriazis J, Fischer C, Creef E. 1991 Semantic information distinguishing individual predators in the alarm calls of Gunnison's prairie dogs. *Anim. Behav.* **42**, 713–719. (doi:10.1016/S0003-3472(05)80117-4)

59. Cheney DL, Seyfarth RM. 2005 Constraints and preadaptations in the earliest stages of language evolution. *Linguist. Rev.* **22**, 135–159. (doi:10.1515/tlir.2005.22.2-4.135)

60. Clay Z, Zuberbühler K. 2011 Bonobos extract meaning from call sequences. *PLoS ONE* **6**, e18786. (doi:10.1371/journal.pone.0018786)

61. Arnold K, Zuberbühler K. 2012 Call combinations in monkeys: compositional or idiomatic expressions. *Brain Lang.* **120**, 303–309. (doi:10.1016/j.bandl.2011.10.001)

62. Shapiro AD, Tyack PL, Seneff S. 2010 Comparing call-based versus subunit-based methods for categorizing Norwegian killer whale, *Orcinus orca*, vocalizations. *Anim. Behav.* **81**, 377–386. (doi:10.1016/j.anbehav.2010.09.020)

63. Miyagawa S, Berwick RC, Okanoya K. 2013 The emergence of hierarchical structure in human language. *Front Psychol.* **4**, 71. (doi:10.3389/fpsyg.2013.00071)

64. Jackendoff R. 2011 What is the human language faculty? Two views. *Language* **87**, 586–624. (doi:10.1353/lan.2011.0063)

65. Tomasello M. 2008 *Origins of human communication*. Cambridge, MA: MIT press.

66. Clark B. 2013 Syntactic theory and the evolution of syntax. *Biolinguistics* **7**, 169–197.

67. Katahira K, Suzuki K, Kagawa H, Okanoya K. 2013 A simple explanation for the evolution of complex song syntax in Bengalese finches. *Biol. Lett.* **9**, 20130842. (doi:10.1098/rsbl.2013.0842)