Ethology

How Reliable are the Methods for Estimating Repertoire Size?

Carlos A. Botero*,†,‡, Andrew E. Mudge†, Amanda M. Koltz§, Wesley M. Hochachka† &
Sandra L. Vehrencamp†,¶

* Center for Ecological and Evolutionary Studies, University of Groningen, Groningen, The Netherlands
† Cornell Laboratory of Ornithology, Ithaca, NY, USA
‡ Instituto de Investigación en Recursos Biológicos Alexander von Humboldt, Bogota, Colombia
§ Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA
¶ Department of Neurobiology and Behavior, Cornell University, Ithaca, NY, USA

Correspondence
Dr. Carlos A. Botero, Center for Ecological and Evolutionary Studies, University of Groningen, 9751 NN Haren, Groningen, The Netherlands. E-mail: C.A.Botero@rug.nl

Received: January 11, 2008
Initial acceptance: March 9, 2008
Final acceptance: August 11, 2008
(D. Zeh)

doi: 10.1111/j.1439-0310.2008.01576.x

Abstract

Quantifying signal repertoire size is a critical first step towards understanding the evolution of signal complexity. However, counting signal types can be so complicated and time consuming when repertoire size is large, that this trait is often estimated rather than measured directly. We studied how three common methods for repertoire size quantification (i.e., simple enumeration, curve-fitting and capture-recapture analysis) are affected by sample size and presentation style using simulated repertoires of known sizes. As expected, estimation error decreased with increasing sample size and varied among presentation styles. More surprisingly, for all but one of the presentation styles studied, curve-fitting and capture–recapture analysis yielded errors of similar or greater magnitude than the errors researchers would make by simply assuming that the number of types in an incomplete sample is the true repertoire size. Our results also indicate that studies based on incomplete samples are likely to yield incorrect ranking of individuals and spurious correlations with other parameters regardless of the technique of choice. Finally, we argue that biological receivers face similar difficulties in quantifying repertoire size than human observers and we explore some of the biological implications of this hypothesis.

Introduction

A proper knowledge of the repertoire size of different species is essential for understanding the evolution of complexity in animal communication. Estimates of repertoire size abound in the scientific literature (e.g., Ballard & Kovacs 1995; Boisseau 2005; Clark 1982; Cleveland & Snowdon 1982; Davidson & Wilkinson 2002; MacDougall-Shackleton 1997; McShane et al. 1995; e.g., Moynihan 1970; Read & Weary 1992; Saulitis et al. 2005; Searcy 1992; Smith 1977, 1986; Wong et al. 1999) and this trait has been correlated with male quality (Catchpole 1996; Nowicki et al. 1998; Kipper et al. 2006), parental provisioning (Buchanan & Catchpole 2000), parasite load (Buchanan et al. 1999), body condition (Lampe & Espmark 1994), age (Catchpole & Slater 1995), resource holding potential (Howard 1974), lifetime reproductive success (Hiebert et al. 1989), and brain morphology (DeVoogd et al. 1993). Although repertoire size is often estimated from incomplete samples, little is known about the robustness and reliability of the methods involved (but see Derrickson 1987; Garamszegi et al. 2005; Kroodsma 1982).

Three common methods for assessing repertoire size are simple enumeration, curve-fitting (Wildenthal 1965), and capture–recapture analysis (Catchpole & Slater 1995; Garamszegi et al. 2002). Simple enumeration is the act of counting the number of types present in a sample of signals and is ideal for species with small repertoires. However, when repertoire size is large, counting all types requires large samples and a large investment of time and effort (Kroodsma & Parker 1977). In those cases, repertoire size is often estimated, rather than measured directly, via curve-fitting (Wildenthal 1965). Curve-fitting predicts true repertoire size by fitting an exponential curve to the plot of accumulation of signal types as a function of sample size. It assumes that repertoire size is a fixed value and that all signal types have an equal and random probability of occurrence. Curve-fitting yields poor estimates when sample size is small (Derrickson 1987) or when singers do not present their song types at random (Kroodsma 1982). Another method for estimating repertoire size from incomplete samples is capture–recapture analysis, which is based on an adaptation of pre-existing ecological models (Catchpole & Slater 1995; Garamszegi et al. 2002). Capture–recapture analysis also assumes that repertoire size is a fixed quantity (i.e., it is based on 'close-population' models, see White et al. 1978). Contrary to curve-fitting, it does not assume that all signal types have the same probability of occurrence (Garamszegi et al. 2002). This method involves dividing a sample into groups of consecutive elements known as 'trapping occasions' and keeping track of which signal types are observed in each trapping occasion. Repertoire size is then estimated based on the total number of types observed and the estimated probability that a new type will be observed in a new trapping occasion. Given the recent introduction of this technique, there is little information on how it is affected by sample size or singing style (but see Garamszegi et al. 2005).

We studied the effects of presentation style and sample size on the techniques described above by applying them to simulated sequences of signals drawn from different repertoires of known sizes. Our sequences were modeled after syllable sequences in songbirds (one of the areas where these techniques are most commonly applied), but may also be translated to other signaling systems in which there is a fixed (and large) number of signal types. We explored the different ways in which animals can present their signal repertoires by varying the following parameters in our simulated sequences: probability of occurrence of different types, tendency to repeat each type before introducing a new one (i.e., immediate vs. eventual variety singing), tendency to deliver some elements always in the same combinations (e.g., in standard song types or bout types), and tendency to present all the repertoire in a single standard sequence. Because of the generality of these parameters, our results can be extrapolated to a wide array of species, repertoire presentation styles and signaling modalities.

Our analyses indicate that model-based estimation of repertoire size is not an ideal substitute for simple enumeration, especially when relative differences between individuals are of interest. To illustrate this and other points, we analyze real syllable sequences from the tropical mockingbird, *Mimus gilvus*, a species with large vocal repertoires.

Methods

Simulation of Artificial Song Sequences

We created five imaginary individuals with repertoire sizes of 200, 190, 180, 170, and 160 element types. These values provide a realistic range of inter-individual differences in repertoire size (maximum difference = 20%, minimum difference = 5%) and can be used to study how well the different methods allow us to discriminate between pairs of individuals that have either large differences or small differences in repertoire size. The values used in our simulations reflect the approximate repertoire sizes reported for several species with large repertoires (e.g., mockingbirds, nightingales, wrens, and *Acrocephalus* warblers) without being extreme values for this parameter (Derrickson 1987; Read & Weary 1992). We used Matlab 7 (Mathworks Inc., http://www.mathworks.com) to simulate six song sequences of 2000 elements for each imaginary individual based on the following presentation styles: (1) completely random presentation of elements, RSQ (i.e., any type could occur at any place in the sequence); (2) cyclic presentation of the repertoire, CYC (i.e., types were presented one after the other and were only repeated after the rest of the repertoire had been exhausted); (3) types presented in standardized clusters with each cluster being a unique series of five different element-types always presented in the same order, SCR (the sequence of cluster types was randomly selected); (4) same as in (3) but simulating eventual variety by repeating each standardized cluster five times before introducing a new one, SCE; (5) types presented in completely random clusters of five elements repeated five times before switching to a new one, RCE (i.e., random clusters presented with eventual variety); and (6) types presented

Table 1: Summary of the five simulated singing styles and examples of the sequences they generate

Singing style	Example[a]
Random sequence (RSQ)	AKDFUGTRHNDLSOIRJFNCVAKFYHB…
Cyclic presentation (CYC)	ABCDEFGHIJKLMNOPQRSTUVWXYZA…
Random string of standardized cluster types (SCR)	(FGHIJ), (UVXYZ), (LMNOP)…
Standardized cluster types, eventual variety (SCE)	[(FGHIJ) × 5], [(UVXYZ) × 5]…
Random clusters, eventual variety (RCE)	[(HEOXP) × 5], [(FGLWM) × 5]…
Heterogeneous probability (HET)	BKDUUGTRSNDLSOTRJFNJVNTRYHR…

[a] Different letters represent different element types and parentheses mark the beginning and ending of an element-cluster (when applicable).

with heterogeneous probability of occurrence, HET (i.e., half of the types in each repertoire were defined as common and the other half as rare; common types were allowed to occur five times more often than rare ones). Examples of the sequences generated for each simulated style are found in Table 1.

From each simulated sequence we extracted eight subsequences including the first 250, 500, 750, 1000, 1250, 1500, 1750, and 2000 elements (i.e., a total of 240 datasets given the five simulated individuals, six singing styles, and eight sampling levels). These datasets allowed us to compare how the three methods performed at different sampling levels.

Curve-Fitting

Curve-fitting estimations were performed with CURVEXPERT v. 1.37 (http://curveexpert.webhop.biz). The curves typically used for curve-fitting are of the form

$$n = N(1 - e^{-T/N}),$$

where n is the number of distinct types observed in the sample, T is the number of elements sampled, and N is the estimated total number of types in the repertoire (Wildenthal 1965). To account for variation in the rate at which new types accumulate over time for different singing styles (see Kroodsma 1982), we used Davidson & Wilkinson's (2002) modified equation, which includes a curvature parameter, A, that predicts shallower curves at larger values:

$$n = N(1 - e^{-T/A*N}).$$

Capture–Recapture Analysis

We used the program CAPTURE (Rexstad & Burnham 1991) with the PC interface CAPTURE2 (J. Hines, USGS, http://www.mbr-pwrc.usgs.gov) to run all capture–recapture models. Capture–recapture analysis can account for different sources of variation in syllable detection probability, namely frequency of use (h), time (t), and behavior (b). The program CAPTURE includes a model selection algorithm that facilitates the identification of the proper source(s) of variation in a given dataset. Although models with all of the possible combinations of sources of variation are theoretically possible, CAPTURE cannot currently compute models that include simultaneously time, behavior, and heterogeneity effects, i.e., M(tbh). Thus, in cases in which M(tbh) was the most appropriate model for our data, we used the second most appropriate model instead.

Following Garamszegi et al. (2005), we began by defining a trapping occasion as a cluster of five consecutive elements, resembling the songs of our empirical example (five elements is the average number of syllables per song in the tropical mockingbird). This sampling scheme generated very large data matrices that exceeded the maximum number of trapping occasions that can be currently analyzed with the program CAPTURE (i.e., 80 trapping occasions, J. Hines, personal communication). Because of the software limitation, trapping occasions of five syllables could only produce estimates of repertoire size at the 250-elements sampling level and, thus, were clearly inappropriate for comparing the performance of the capture–recapture technique with that of the other methods. An exploration of other trapping occasion sizes indicated that small trapping occasions not only force users to analyze a smaller number of elements overall but also tend to produce more variable estimates and larger estimation errors at a given sample size than large trapping occasions (Fig. 1). Thus, we used a trapping occasion of 250 elements to evaluate the performance of the capture–recapture technique at its best. This sampling scheme generated a maximum of eight trapping occasions and thus allowed estimation of repertoire size at all sampling levels (except at 250 syllables because two or more trapping occasions are needed for capture–recapture analysis).

Estimation Errors and Statistical Analysis

We computed the mean relative error at each sampling level for each estimation technique as:

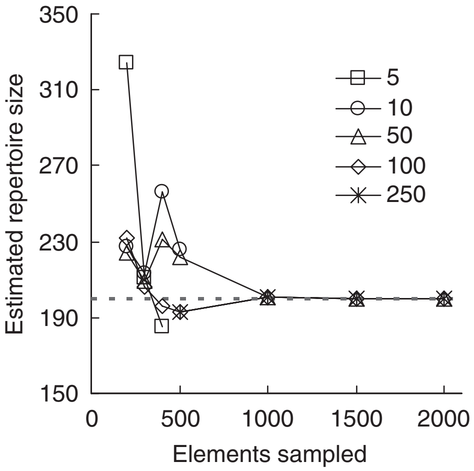

Fig. 1: Effect of trapping occasion size (i.e., number of elements included in each trapping occasion) on capture–recapture estimation. The estimates shown are based on a simulated sequence with completely random presentation of elements for an individual with 200 element types (true repertoire = dotted line). Data are not available for all trapping occasion sizes at all sampling levels because the maximum number of trapping occasions that can be currently analyzed with the program CAPTURE is 80.

$$\text{Relative Estimation Error} = |\text{ERS} - \text{TRS}|/\text{TRS},$$

where ERS is the estimated repertoire size and TRS is the true repertoire size for the corresponding individual. We used a General Linear Mixed Model (GLMM) to test for differences in relative errors as a function of estimation technique, presentation style, sampling level, and the two-way interactions between estimation technique and the remaining two factors. Individual identity was included as a random effect with a variance component covariance structure (SAS 9.3, PROC MIXED default) so as to account for non-independence in our data. The distribution of the residuals in this model was highly skewed and could not be approximated to normality with simple transformations. Thus, to test for differences in relative errors we ranked all errors from lowest to highest (rank for lowest error = 1), averaging ranks whenever there were ties, and ran the model again on the natural log of these ranks.

Tropical Mockingbird Data

We also applied the techniques described above to the estimation of syllable repertoire size of six adult male tropical mockingbirds from Villa de Leyva, Colombia. To control for behavioral and social context in our samples, all focal birds were dominant, breeding males recorded during their corresponding periods of sustained song output prior to egg laying in 2004 (i.e., during courtship). Syllable sequences were obtained from recordings made at close proximity with a Marantz PMD690 recorder and a Sennheiser ME67 directional microphone. Following Garamszegi et al. (2005), we analyzed continuous samples of ca. 2000 syllables per male. The classification of all syllables for all birds was done jointly by AEM and CAB based on overall similarity in structure and duration (Fig. 2). As in the previous section, we used trapping occasions of 250 syllables and divided each syllable sequence into eight datasets including the first 250, 500, 750, 1000, 1250, 1500, 1750, and the maximum number of elements per sample (i.e., a total of 48 datasets given the six birds and eight sampling levels per bird). Given that we did not have 2000 consecutive syllables for every individual, the largest datasets used for capture–recapture analysis included seven trapping occasions with 250 syllables each (i.e., 1750 syllables) and the largest datasets used for curve-fitting estimation included 1992 syllables (i.e., the maximum sample size available for all six birds).

Results

Graphical summaries of our simulation results are presented in Figs 3 and 4. Figure 3 shows the estimated repertoire sizes for the different techniques and presentation styles and Fig. 4 shows how the simulated individuals ranked in terms of the total

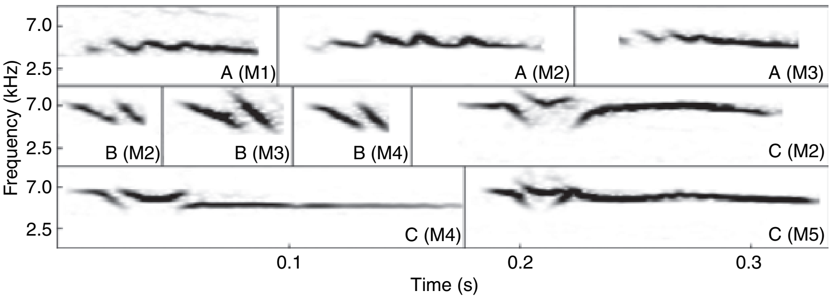

Fig. 2: Examples of inter-type and inter-individual variation in the syllable types of the tropical mockingbird. Types are identified by capital letters and singers are presented in parentheses.

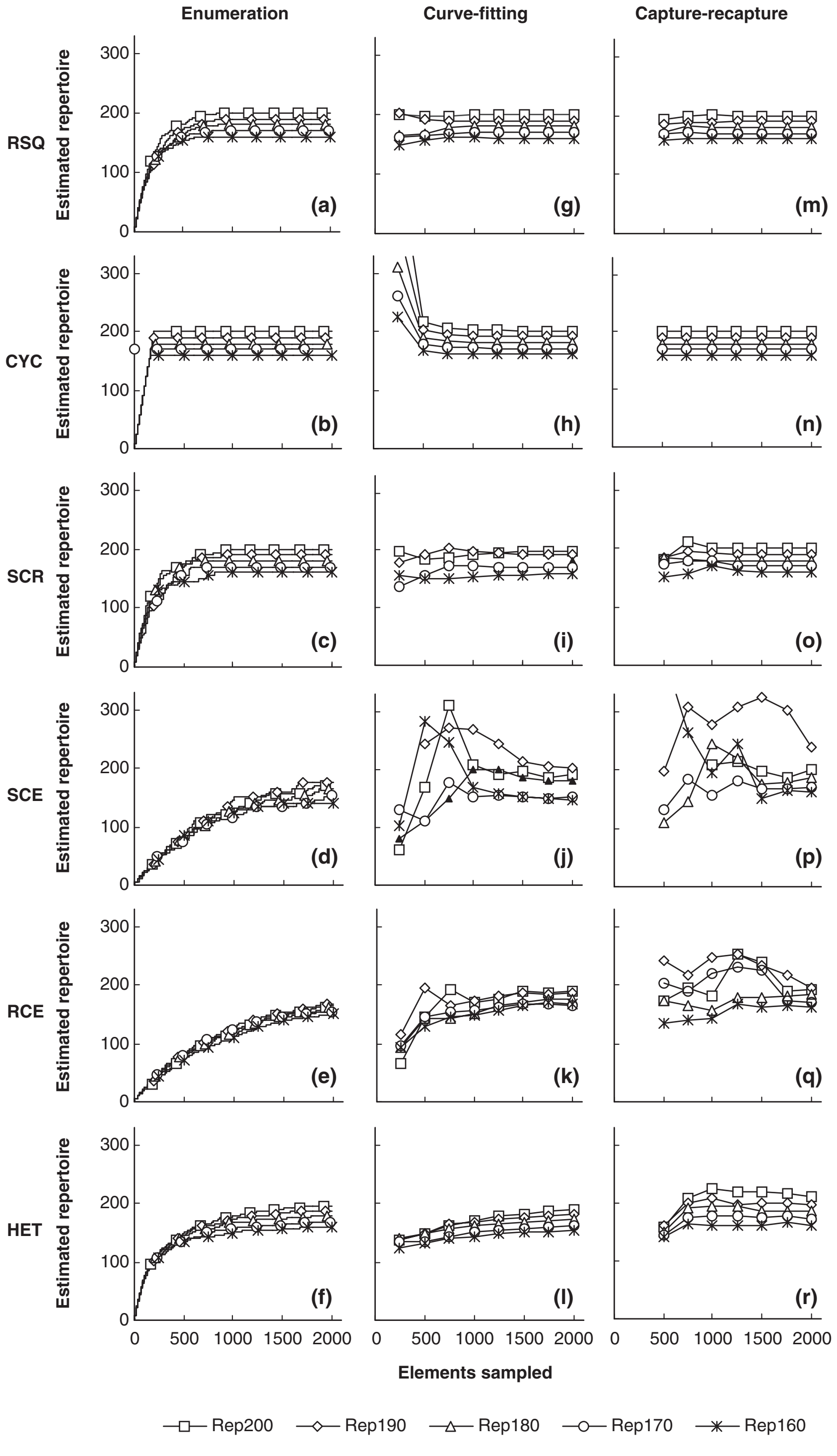

Fig. 3: Estimated repertoire size as a function of estimation technique, sample size, and singing style. Five simulated individuals are identified by the size of their repertoire (e.g., Rep200 = Individual with 200 types in its repertoire). The abbreviations for the different presentation styles follow Table 1.

number of types observed or estimated at the different sampling levels (rank = 1 is for the male with the largest repertoire, rank = 5 is for the male with the smallest repertoire). Estimation errors varied significantly between methods as a function of sample size and presentation style (Table 2, Fig. 5). All methods were more accurate when working with larger samples but this effect was slightly less pronounced in curve-fitting than in simple enumeration or capture–recapture analysis ($p < 0.001$).

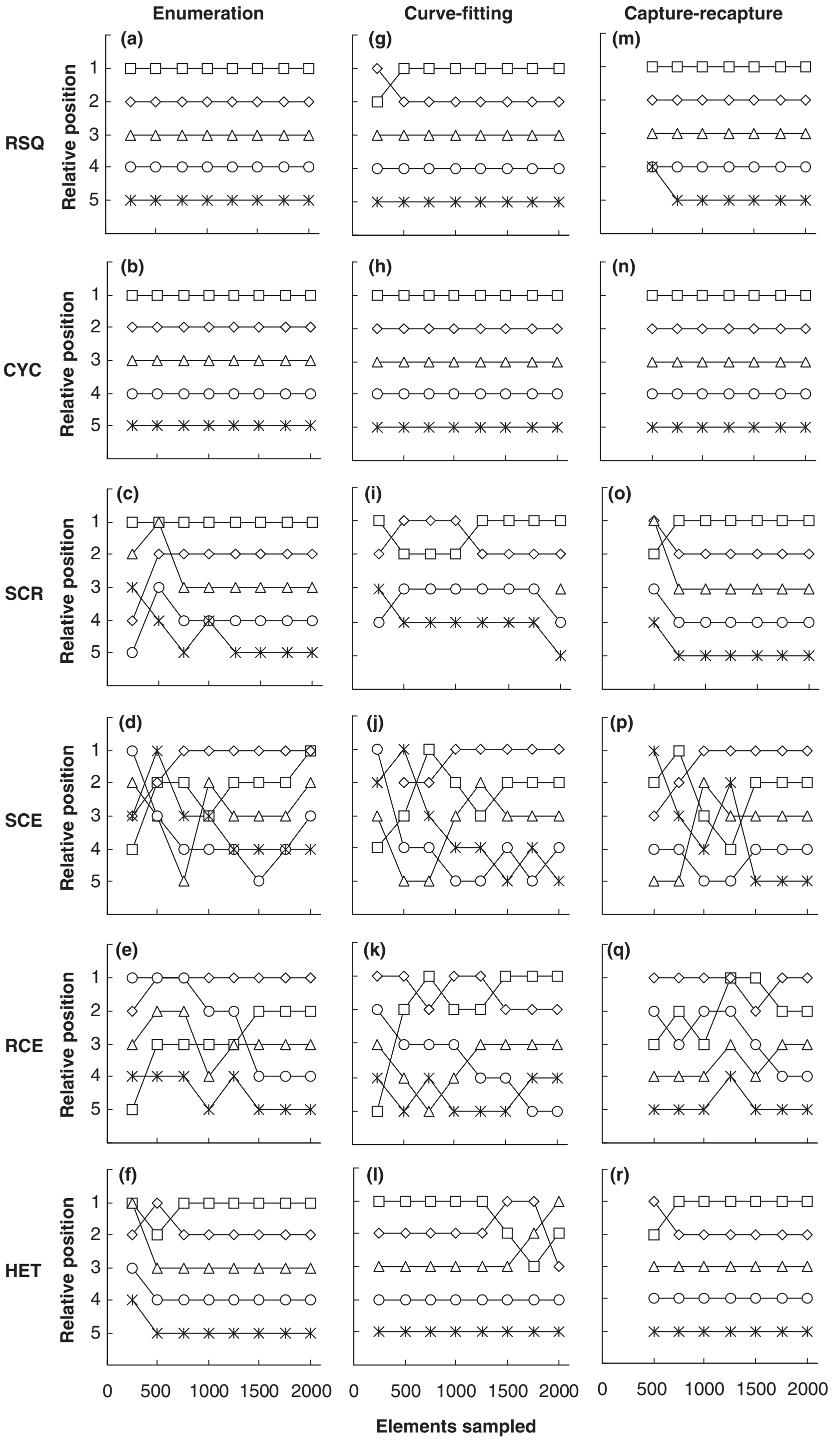

Fig. 4: Relative ranking of five individuals based on estimated repertoire size as a function of estimation technique, sample size, and singing style. The five simulated individuals are identified by the size of their repertoire (e.g., Rep200 = Individual with 200 types in its repertoire). The abbreviations for the different presentation styles follow Table 1.

A posteriori Tukey–Kramer tests revealed that in all but one presentation style, simple enumeration produced similar or smaller errors than the other two techniques. Specifically, mean relative errors for the different singing styles ranked as follows (inequalities imply a $p < 0.05$):

$$\text{Simple Enumeration} \approx \text{Capture-Recapture} < \text{Curve-Fitting (CYC, RSQ, SCR)}$$

$$\text{Simple Enumeration} \approx \text{Capture-Recapture} \approx \text{Curve-Fitting (HET, SCE)}$$

Table 2: Selected results from the General Linear Mixed Model testing the effects of estimation technique, presentation style, and sampling level on the mean relative estimation error

Effect	Numerator DF	Denominator DF[a]	F-value	p
Estimation technique	2	657.10	5.42	0.005
Presentation style	5	657.24	353.42	<0.001
Sampling level	1	657.01	413.82	<0.001
Technique × style	10	657.23	34.74	<0.001
Technique × sampling level	2	657.01	6.94	0.001

[a] Denominator degrees of freedom computed using Satterthwaite's (1946) approximation.

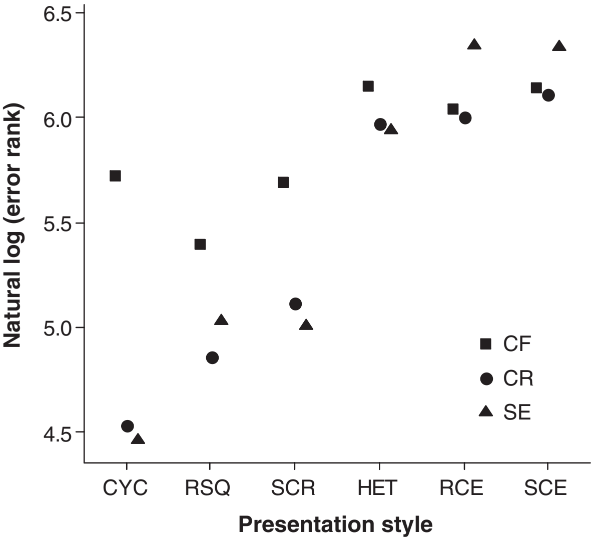

Fig. 5: Least squares means for the mean relative error of the different estimation techniques as a function of presentation style. Errors were ranked from lowest to highest and these ranks were natural log transformed to insure normally distributed residuals in our model. Larger Y-values imply larger errors in the estimation. SE = simple enumeration; CF = curve-fitting; CR = capture–recapture; Abbreviations for presentation styles follow Table 1.

$$\text{Capture-Recapture} \approx \text{Curve-Fitting} < \text{Simple Enumeration (RCE)}$$

Important details of the results obtained through each estimation technique are presented below.

Enumeration

With the exception of the cyclical presentation of types, all presentation styles yielded accumulation plots that resembled exponential curves (Fig. 3a–f). As expected, the more complex the presentation style, the larger the sample required for enumerating all types in a repertoire. The plots derived from simple enumeration also allowed us to observe some key features of singing behavior. For example, it was noteworthy that although cycling through the repertoire is the fastest way to show off all types (see Fig 3b), this presentation style precluded any distinction between singers unless the sample included more elements than the number of types in the smallest repertoire. In contrast, when types were presented with some randomness (e.g., Fig. 3a, c–f), differences between singers were apparent at lower sampling levels. In that case, however, mistakes in the ranking of males were more common (Fig. 4a, c–f).

Curve-Fitting

Davidson & Wilkinson's (2002) model produced curves that fitted our simulated data very well (Pearson correlation coefficient: $r \pm SE = 0.991 \pm 0.002$). However, the estimates produced with this method were proportional to the fraction of the repertoire that had been included in the sample and, as a consequence, curve-fitting tended to converge onto the true solution at around the same sampling level at which most of the types could already be counted through simple enumeration (Fig. 3). For sequences with cyclic presentation of types, curve-fitting led to drastic overestimation of repertoire size at the smallest sampling level (Fig. 3h). This overestimation was a product of fitting exponential curves to data that clearly do not accumulate in an exponential fashion (i.e., the exponential curves that have initial rates of accumulation of types as high as the ones observed in the cyclic sequences also have very large asymptotes). Curve-fitting also performed poorly at low sampling levels for singing styles with eventual variety (Figs 3j and 4j, k). In particular, it yielded clearly erroneous estimates of repertoire size $>10^4$ elements for seven out of the eight sample size datasets from the random string of standardized cluster types of individual Rep180 (these results were discarded and thus only one estimate is plotted for this individual in Fig. 3i).

As in Derrickson (1987), we found that the curves that best-fitted our data sometimes asymptoted below the total number of types observed in a sample (78 out of 240 datasets). This type of underestimation was most common when types were presented with heterogeneous probability of occurrence (31/40 datasets) or in standardized clusters with eventual variety (30/40 datasets).

Table 3: Estimates of repertoire size for the five simulated individuals using capture–recapture analysis based on 50 trapping occasions of five consecutive elements each

True repertoire	Singing style					
	RSQ	CYC	SCR	SCE	RCE	HET
200	214	490	218	41	60	177
190	186	388	184	45	65	166
180	178	215	198	44	60	139
170	171	258	135	47	65	170
160	158	185	173	43	70	123

RSQ, random sequences; CYC, cyclic presentation; SCR, random string of standardized cluster types; SCE, standardized cluster types, eventual variety; RCE, random clusters, eventual variety; HET, heterogeneous probability.

Capture–Recapture Analysis

The estimates generated through capture–recapture analysis were less strongly affected than in curve-fitting by the fraction of the repertoire that had been included in the sample. Nevertheless, this method also tended to converge onto the true solution at around the same sampling level at which most of the types could already be counted through simple enumeration.

For comparison, Table 3 shows the results of capture–recapture analyses using songs as trapping occasions (one song = five elements) and a total sample of 250 elements. These results also show an effect of singing style on relative estimation error as observed in the analyses with larger trapping occasions. Due to intrinsic differences in the rate of accumulation of types among singing styles (Fig. 3a–f), the 250-syllable samples used in this analysis contained a much smaller fraction of the total repertoire in RCE and SCE than in any other style, and this difference lead to conspicuously larger estimation errors.

Empirical Results

Male tropical mockingbirds sing songs composed of 1–31 syllables (mean ± SE = 4.63 ± 0.05 syllables per song) and 1–12 different syllable types (mean ± SE = 2.96 ± 0.03 types per song; n = 2569 songs). Syllable types vary in their frequency of occurrence (Fig. 6). Individuals sometimes sing the same syllable combinations in different days suggesting that syllables could be associated in standard ‘song types’ in this species (or ‘standardized clusters’ as referred to above). Different song types may share a few syllable types and tend to be presented with eventual variety. Figure 7 shows the estimated syllable repertoire sizes and relative rankings for our six focal males as a function of sampling size and estimation technique. As expected, these plots closely resemble the plots for the two simulated singing styles with eventual variety (Fig. 3d, e, j, k, p, q).

Davidson & Wilkinson’s (2002) model also produced curves that fitted the tropical mockingbird data very well (Pearson correlation coefficient: r ± SE = 0.988 ± 0.001). In this case, curve-fitting predicted repertoire sizes that were equal to or above the number of types observed in the sample 35 out of 48 times. The estimates of total syllable repertoire size derived from curve-fitting had not reached a point of stability by our maximum sample sizes suggesting that more syllables are required to enumerate the total repertoire size in this species.

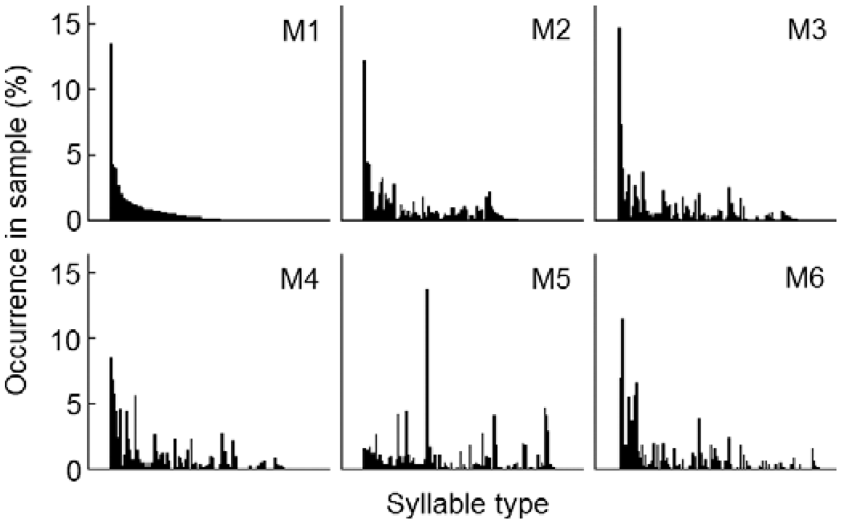

Fig. 6: Syllable type use in the tropical mockingbird. All the types present in the population are listed in the same order for every bird.

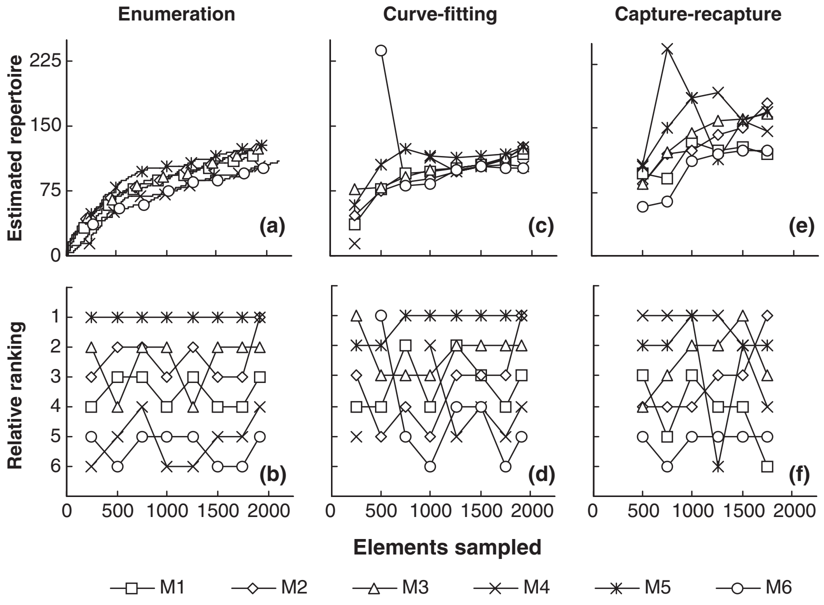

Fig. 7: Estimated repertoire size and relative ranking for six male tropical mockingbirds as a function of sample size and estimation technique.

Defining natural songs as trapping occasions for capture–recapture analysis often generated encounter histories that failed the test of population closure. Nevertheless, the assumption of population closure was sometimes met when only a few songs were analyzed (e.g., 15 songs as in Garamszegi et al. 2005) because of the eventual variety of the tropical mockingbird. The repertoire sizes estimated from such small samples were, however, deceptively small because they were based on a subsample of types that were being repeated over and over. For those interested in comparing the results of this analysis to Garamszegi et al. (2005), our estimates of total repertoire (and number of types present in each sample) based on 15 songs were: M1 = 46 (21), M2 = 23 (21), M3 = 24 (22), M4 = 13 (10), M5 = 40 (23), and M6 = 8 (6) (the data for M3 and M5 did not meet the assumption of closure under this sampling scheme).

Trapping occasions of 250 syllables generated encounter histories that met the assumption of stationarity in 31 out of 36 datasets (capture–recapture analysis cannot be performed with a single trapping occasion so 36 datasets come from six birds at six sampling levels each). The effect of sample size on the magnitude of the estimates produced through capture–recapture analysis was more pronounced in this case than in the simulations (see Fig. 7).

Discussion

We conclude that curve-fitting and capture–recapture analysis do not necessarily provide better estimates of the total repertoire size than simple enumeration when dealing with incomplete samples of signals from species with large repertoires. Although estimation techniques may seem to save time and effort, our results show that they will often yield errors of similar or greater magnitude than the errors researchers would make by simply assuming that the number of types present in an incomplete sample is the true repertoire size. The exception to this rule is when animals present their types in random clusters with eventual variety, in which case both curve-fitting and capture–recapture analysis may provide better estimates than simple enumeration at low sample sizes.

Our results also indicate that correlations between repertoire size and variables such as reproductive success, male quality, etc., can be very misleading in species with very large repertoires. Given that individual ranking based on repertoire size is strongly dependent on the number of syllables classified (see Fig. 5), researchers are likely to observe spurious correlations (regardless of the technique of choice) if individual repertoires are not sampled extensively. When information on the true repertoire size of a species is not available to determine an appropriate sample size, or when more extensive sampling is simply not possible, efforts must at least be made to check the stability of the measurements/estimates before attempting correlation. We suggest that one possible way to do so is to subsample each sequence of songs/syllables and check whether the highest sampling levels arrive at similar conclusions. If there is still strong variation among the higher sampling levels, then researchers should be skeptical of any observed patterns. Another possible way to avoid spurious correlation is to compute confidence intervals for each estimate and to rank individuals differently only if these intervals do not overlap. Unfortunately, this procedure is likely to prevent ranking individuals altogether especially if the differences between them are not extreme or if the samples available are relatively small. For example, this method would preclude the ranking of individuals at all sampling levels in our simulated sequences of standardized clusters with eventual variety.

Our simulations confirm Kroodsma's (1982) intuition that presentation style is a particularly important factor to consider when attempting to quantify repertoire size in any species. At a very basic level, presentation style determines the rate at which new types accumulate over time and, thus, is the ultimate determinant of the fraction of the total repertoire that is included in a sample of any given size (see Fig. 3a–f). This means that more repetitive presentation styles will require larger samples than others so as to achieve comparable levels of accuracy. Additionally, some estimation techniques perform particularly poorly for certain presentation styles. For example, curve-fitting produces misleadingly high estimates for cyclical singers at low sampling levels because the initially steep rates of accumulation of types in this presentation style can only be approximated by exponential curves with extremely high thresholds (Table 3 suggests that capture–recapture based on songs as trapping occasions might also suffer from similar problems). In the same way, standardized clustering and eventual variety tend to increase the variability of the estimates derived from estimation techniques and thus to produce highly inaccurate rankings of individuals (see Figs 4 and 7). Because of all of these reasons, we believe that applying a standard methodology for repertoire size estimation in comparative analyses (e.g., Garamszegi et al. 2005), is probably not a good idea.

It has been suggested that model-based estimation may produce more biologically realistic measures of repertoire size than simple enumeration. For example, Derrickson (1987) noted that northern mockingbirds sing some syllable types only rarely and suggested that these rare types should not be considered part of the effective repertoire size of this species. He also suggested that curve-fitting is less affected by rare types than simple enumeration and leads to more realistic estimates of the biologically relevant repertoire size because it often predicts repertoire sizes that are below the total number of types observed in a sample (Derrickson 1987). We disagree with this interpretation because the probability of detection of the different types is not part of the curve-fitting algorithm and because, as a consequence, rare types are not preferentially discounted over more common types. Furthermore, it is not clear that biological receivers discount rare types at all, or that they do so using similar algorithms.

Another interesting point that emerges from our analyses is that the enumeration of types in a repertoire could be as complicated for biological receivers as it is for human observers. For example, if a female tropical mockingbird were to choose between the six neighboring males used as focal subjects in our study on the basis of repertoire size alone, she would have to invest at least 6 h to sample 2000 syllables from each male (assuming an average rate of seven songs per minute (Botero & Vehrencamp 2007) and five syllables per song). It is quite disconcerting to realize that even if she makes no mistakes when classifying syllables in real time, at the end of 6 h of very hard work and undivided attention she will still have a high degree of uncertainty as to which male is her best option. Given that the time, effort, and neuronal resources needed for this type of comparison will increase with repertoire size, it is possible that biological receivers in species with extremely large repertoires also estimate repertoire size from incomplete samples (Garamszegi et al. 2005) and thus, that they deal with similar problems to the ones discussed above. Alternatively, it is also possible that very large repertoire sizes are an indirect product of selection on other traits. For example, if females care about song matching rates during male–male countersinging interactions (Logue & Forstmeier 2008), then selection could be expected to favor males that can learn more songs and from more tutors. The hypothesis that large repertoire sizes may be a product of indirect selection contradicts prevailing views (see Buchanan & Catchpole 2000; Buchanan et al. 1999; Catchpole 1996; Catchpole & Slater 1995; Hiebert et al. 1989; see Howard 1974; Kipper et al. 2006; Lampe & Espmark 1994; Nowicki et al. 1998; Searcy 1992; Searcy & Yasukawa 1996) and must be tested with more data on species with very large repertoires. For those determined in undertaking such a challenge, we recommend a good dose of patience as well as the tried and true method of extensive sampling and (not so) ''simple'' enumeration.

Acknowledgements

Many thanks to G. White and J. Hines for advice and help using the programs Capture and MARK.

We thank the many undergraduates at Universidad de los Andes and Cornell University that were involved in the development of our classification methods and the initial stages of data collection. J.W. Bradbury, M.L. Hall, and W.A. Searcy provided helpful comments on earlier drafts of this manuscript. Our sincere thanks go as well to two anonymous reviewers, for thoughtful and constructive criticism that helped us improve substantially the quality of this manuscript. Funding for this study was provided by the National Institutes of Health, grant R01MH60461 (SLV).

Literature Cited

Ballard, K. A. & Kovacs, K. M. 1995: The acoustic repertoire of hooded seals (*Cystophora cristata*). Can. J. Zool. **73**, 1362—1374.

Boisseau, O. 2005: Quantifying the acoustic repertoire of a population: The vocalizations of free-ranging bottlenose dolphins in Fiordland, New Zealand. J. Acoust. Soc. Am. **117**, 2318—2329.

Botero, C. A. & Vehrencamp, S. L. 2007: Responses of male Tropical Mockingbirds (*Mimus gilvus*) to variation in within-song and between-song versatility. Auk **124**, 185—196.

Buchanan, K. L. & Catchpole, C. K. 2000: Song as an indicator of male parental effort in the sedge warbler. Proc. R. Soc. Lond., B, Biol. Sci. **267**, 321—326.

Buchanan, K. L., Catchpole, C. K., Lewis, J. W. & Lodge, A. 1999: Song as an indicator of parasitism in the sedge warbler. Anim. Behav. **57**, 307—314.

Catchpole, C. K. 1996: Song and female choice: Good genes and big brains? Trends Ecol. Evol. **11**, 358—360.

Catchpole, C. K. & Slater, P. J. B. 1995: Bird Song: Biological Themes and Variations. Cambridge University Press, Cambridge.

Clark, C. W. 1982: The acoustic repertoire of the southern right whale, a quantitative analysis. Anim. Behav. **30**, 1060—1071.

Cleveland, J. & Snowdon, C. T. 1982: The complex vocal repertoire of the adult cotton-top tamarin (*Saguinus oedipus oedipus*). J. Comp. Ethol. **58**, 231—270.

Davidson, S. M. & Wilkinson, G. S. 2002: Geographic and individual variation in vocalizations by male *Saccopteryx bilineata* (Chiroptera: Emballonuridae). J. Mammal. **83**, 526—535.

Derrickson, K. C. 1987: Yearly and situational changes in the estimate of repertoire size in northern mockingbirds (*Mimus polyglottos*). Auk **104**, 198—207.

DeVoogd, T. J., Krebs, J. R., Healy, S. D. & Purvis, A. 1993: Relations between song repertoire size and the volume of brain nuclei related to song: comparative evolutionary analyses amongst oscine birds. Proc. R. Soc. Lond., B, Biol. Sci. **254**, 75—82.

Garamszegi, L. Z., Boulinier, T., Moller, A. P., Torok, J., Michl, G. & Nichols, J. D. 2002: The estimation of size and change in composition of avian song repertoires. Anim. Behav. **63**, 623—630.

Garamszegi, L. Z., Balsby, T. J. S., Bell, B. D., Borowiec, M., Byers, B. E., Draganoiu, T., Eens, M., Forstmeier, W., Galeotti, P., Gil, D., Gorissen, L., Hansen, P., Lampe, H. M., Leitner, S., Lontkowski, J., Nagle, L., Nemeth, E., Pinxten, R., Rossi, J. M., Saino, N., Tanvez, A., Titus, R., Torok, J., Van Duyse, E. & Muller, A. P. 2005: Estimating the complexity of bird song by using capture-recapture approaches from community ecology. Behav. Ecol. Sociobiol. **57**, 305—317.

Hiebert, S. M., Stoddard, P. K. & Arcese, P. 1989: Repertoire size, territory acquisition and reproductive success in the Song Sparrow. Anim. Behav. **37**, 266—273.

Howard, R. D. 1974: Influence of sexual selection and interspecific competition on mockingbird song (*Mimus polyglottos*). Evolution **28**, 428—438.

Kipper, S., Mundry, R., Sommer, C., Hultsch, H. & Todt, D. 2006: Song repertoire size is correlated with body measures and arrival date in common nightingales, *Luscinia megarhynchos*. Anim. Behav. **71**, 211—217.

Kroodsma, D. E. 1982: Song repertoires: problems in their definition and use. In: Acoustic Communication in Birds: Song Learning and its Consequences (Kroodsma, D. E. & Miller, E. H., eds). Academic Press, New York, NY.

Kroodsma, D. E. & Parker, L. D. 1977: Vocal virtuosity in the brown thrasher. Auk **94**, 783—785.

Lampe, H. M. & Espmark, Y. O. 1994: Song structure reflects male quality in pied flycatchers, *Ficedula hypoleuca*. Anim. Behav. **47**, 869—876.

Logue, D. M. & Forstmeier, W. 2008: Constrained performance in a communication network: implications for the function of song-type matching and for the evolution of multiple ornaments. Am. Nat. **172**, 34—41.

MacDougall-Shackleton, S. A. 1997: Sexual selection and the evolution of song repertoires. In: Current Ornithology (Nolan, V. Jr, Ketterson, E. D. & Thompson, C. F., eds). Plenum Press, New York, NY.

McShane, L. J., Estes, J. A., Riedman, M. L. & Staedler, M. M. 1995: Repertoire, structure, and individual variation of vocalizations in the sea otter. J. Mammal. **76**, 414—427.

Moynihan, M. 1970: The control, supression, decay, disappearance and replacement of displays. J. Theor. Biol. **29**, 85—112.

Nowicki, S., Peters, S. & Podos, J. 1998: Song learning, early nutrition and sexual selection in songbirds. Am. Zool. **38**, 179—190.

Read, A. F. & Weary, D. M. 1992: The evolution of bird song: comparative analyses. Philos. Trans. R. Soc. Lond., B, Biol. Sci. **338**, 165—187.

Rexstad, E. & Burnham, K. P. 1991: User's Guide for Interactive Program CAPTURE. Abundance Estimation of Closed Animal Population. Colorado State University, Fort Collins, CO.

Satterthwaite, F. E. 1946: An approximate distribution of estimates of variance components. Biometrics Bull. **2**, 110—114.

Saulitis, E. L., Matkin, C. O. & Fay, F. H. 2005: Vocal repertoire and acoustic behavior of the isolated AT1 killer whale subpopulation in southern Alaska. Can. J. Zool. **83**, 1015—1029.

Searcy, W. A. 1992: Song repertoire and mate choice in birds. Am. Zool. **32**, 71—80.

Searcy, W. A. & Yasukawa, K. 1996: Song and female choice. In: Ecology and Evolution of Acoustic Communication in Birds (Kroodsma, D. E. & Miller, E. H., eds). Cornell University, Ithaca, NY.

Smith, W. J. 1977: The Behavior of Communicating: An Ethological Approach. Harvard University Press, Cambridge, MA.

Smith, W. J. 1986: Signaling behavior: contribution of different repertoires. In: Dolphin Cognition and Behavior: A Comparative Approach (Schusterman, R. J., Thomas, J. A. & Wood, F. G., eds). Erlbaum, Hillsdale, NJ.

White, G. C., Burnham, K. P., Otis, D. L. & Anderson, D. R. 1978: User's Manual for Program Capture. Utah State Univ., Logan, UT.

Wildenthal, J. L. 1965: Structure in primary song of the mockingbird (*Mimus polyglottos*). Auk **82**, 161—189.

Wong, J., Stewart, P. D. & Macdonald, D. W. 1999: Vocal repertoire in the European badger (*Meles meles*): Structure, context, and function. J. Mammal. **80**, 570—588.