

# Can rarefaction be used to estimate song repertoire size in birds?

Kathleen R. PESHEK<sup>1</sup>, Daniel T. BLUMSTEIN<sup>1, 2\*</sup>

<sup>1</sup>The Rocky Mountain Biological Laboratory, Box 519, Crested Butte, CO 81224, USA

<sup>2</sup>Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095-1606, USA

**Abstract** Song repertoire size is the number of distinct syllables, phrases, or song types produced by an individual or population. Repertoire size estimation is particularly difficult for species that produce highly variable songs and those that produce many song types. Estimating repertoire size is important for ecological and evolutionary studies of speciation, studies of sexual selection, as well as studies of how species may adapt their songs to various acoustic environments. There are several methods to estimate repertoire size, however prior studies discovered that all but a full numerical count of song types might have substantial inaccuracies associated with them. We evaluated a somewhat novel approach to estimate repertoire size—rarefaction; a technique ecologists use to measure species diversity on individual and population levels. Using the syllables within American robins' *Turdus migratorius* repertoire, we compared the most commonly used techniques of estimating repertoires to the results of a rarefaction analysis. American robins have elaborate and unique songs with few syllables shared between individuals, and there is no evidence that robins mimic their neighbors. Thus, they are an ideal system in which to compare techniques. We found that the rarefaction technique results resembled that of the numerical count, and were better than two alternative methods (behavioral accumulation curves, and capture-recapture) to estimate syllable repertoire size. Future estimates of repertoire size, particularly in vocally complex species, may benefit from using rarefaction techniques when numerical counts are unable to be performed [*Current Zoology* 57 (3): 300–306, 2011].

**Keywords** Behavioral accumulation curve, Bioacoustics, Capture-recapture, Numerical count, Rarefaction, Song repertoire size

Bird songs vary in several aspects, including but not limited to their acoustics, composition, and repertoire size—the number of different notes, syllables, phrases, or songs produced (Williams, 2004). Repertoires differ between species and individuals (Catchpole and Slater, 2008). For example, thrushes, recognizable for their elaborate songs, may have a repertoire of 200 different song types, while other species have a single song type (Ince and Slater, 1985; Brumm et al., 2009; Catchpole and Slater, 2008). Components of avian repertoires include notes, syllables, phrases, which are assembled into songs. Brenowitz et al (1997). define a note as a single sound, while syllables are a sequence of notes. At least one syllable generates a phrase and certain phrase arrangements repeat to establish song types. Repertoire complexity relates to these components and their repetition or order (Williams, 2004).

Documenting this acoustic diversity within individuals and a species is a challenge (Krebs and Kroodsma, 1980; Kroodsma, 1982; Catchpole and Slater, 2008), but often is required for studies of geographic variation,

speciation, adaptation, and to document how animals respond to anthropogenic stimuli. Several methods have been used, and each has its own set of assumptions. A common assumption is that repertoire size is fixed and does not vary within a population (Botero et al., 2008). Arguably the best method is to fully count the syllables or song types. This method only assumes that the repertoire size is fixed and requires sufficient effort to fully enumerate each individual's repertoire (Botero et al., 2008). With large repertoires, it may be difficult or impractical to conduct a full count (Kroodsma and Parker, 1977; Botero et al., 2008). The remaining methods require sampling. The most common sampling method is to use behavioral accumulation curve equations (Dias et al. 2009). This technique graphs the number of songs recorded verses the number of song types (Wildenthal, 1965; Catchpole and Slater, 2008). For either an individual or for the entire population the maximum repertoire size is identified when the curve reaches an asymptote. The curve is asymptotic because as the number of songs recorded increases, the number of new song

---

Received July 10, 2010; accepted Jan 14, 2011.

\* Corresponding author. E-mail: marmots@ucla.edu

© 2011 *Current Zoology*

types detected should decrease. A third technique uses the ecological technique of capture-recapture. This technique relies on randomly choosing a sample of recorded syllables and then “recapturing” some of those syllables in successive samples (e.g. Garamszegi et al., 2002, 2005; Catchpole and Slater 2008). This approach assumes that the recordings are sufficiently long enough to estimate the repertoire size, and that song elements occur randomly at identical frequencies (Botero et al., 2008),

The above techniques are not perfect; they may over- and under-estimate the true repertoire size (Catchpole and Slater, 2008). Prior research has discovered that estimates using models produced incorrect results when compared to comprehensive numerical counts (Botero et al., 2008). This creates the need for another method that is both convenient and accurate in its estimates. Techniques using methods from estimating species diversity may be useful to estimate song repertoire size (Garamszegi et al., 2002).

Our study evaluates the ecological technique of rarefaction (Gotelli and Colwell, 2001; Longeno et al., 2002; James and Wamer, 1982) to estimate repertoire sizes. Rarefaction is a technique for sampling distributions within certain categories, such as individuals in a population or species in a community (James and Wamer, 1982). This method is typically used for studying species diversity to estimate the predicted number of species from a sample of individuals (James and Wamer, 1982). A rarefaction curve begins with a steep slope and then becomes asymptotic as fewer new species are added in each successive sample (Gotelli and Colwell, 2001). Therefore rarefaction curves, like behavioral accumulation curves, are asymptotic. Unlike behavioral accumulation curves, rarefaction uses random resampling of a specific subset of the sample to calculate an average of each sampling unit in each sample (typically a species, but in our case we will use individuals). By plotting these averages, a rarefaction curve is created (Gotelli and Colwell, 2001). Rarefaction analyses standardize the sample size, allowing accurate species richness comparisons to be made (James and Wamer, 1982). Because of this, the repertoire estimate should be more accurate than the commonly used behavioral accumulation curve. Some prior research has used rarefaction to estimate repertoire size in bird songs (Forstmeier and Balsby, 2002), but it has not formally studied its efficacy.

To evaluate the use of rarefaction to estimate song diversity, we will treat the syllables within a song as

species, and the number of recorded individuals as samples. This permits us to account for the syllable diversity among individuals because individuals do not sing the same number and type of particular syllables, similar to the typical rarefaction’s management of species variation. Rarefaction, like the other methods, has a set of assumptions. These include; the sample size is standardized across samples, sufficiently long recordings are made to accurately portray the repertoire size, and that the results fit the sample size but cannot be extended to larger sample sizes (Gotelli and Entsminger, 2009). A potential shortcoming of rarefaction is that it may underestimate the repertoire size in species if syllables are sung non-randomly within songs (Forstmeier and Balsby, 2002).

We chose the American robin as a model species, and used field recordings, to compare rarefaction to other methods for several reasons. Robins are the most common thrush in North America (Sallabanks and James, 1999) and thrushes are notable for their elaborate songs (Ince and Slater, 1985; Brumm et al., 2009). Prior studies on song elements discovered at least 12 phrases that differed within an individual robin’s repertoire (Borror, 1965; Thomas, 1979). However, these prior studies did not include an upper phrase limit for individuals, and prior studies suggested little syllable sharing between individuals (Thomas, 1979; Sallabanks and James, 1999; Johnson, 2006). This large repertoire is somewhat remarkable since robins are not vocal mimics, which tend to have large repertoires because they learn songs from other species throughout their lives (Williams, 2004). Therefore diverse robin repertoires may result from two non-mutually-exclusive reasons. First, robins are one of the few species that invent their songs. Second, juveniles learn their songs prior to dispersal, and they disperse across the landscape (Konishi, 1965; Johnson, 2006). Because of this substantial diversity, we elected to estimate robin repertoires at the syllable level.

We acknowledge that creating a simulated data set (e.g., Botero et al., 2008) would have been an alternative method to using field recordings to compare repertoire size estimation algorithms. We chose, however, to use field estimates, even when potentially incomplete, to see how these methods performed with field data. Full repertoire estimates may be required for some, but not all studies. Female choice is hypothesized to drive the evolution of large song repertoires, yet a given female is unlikely to be able to hear the entire repertoire before mating (Byers and Kroodsma, 2009). And, while syllables can be re-arranged to create song variation,

estimating syllable variation should require less intensive sampling.

## 1 Materials and Methods

### 1.1 Recordings and syllable analysis

To compare the techniques for estimating repertoires, from May 30 until June 11 2009, we recorded the dawn and/or dusk songs of 38 American Robins in and around the Rocky Mountain Biological Laboratory (RMBL—38°57'46"N 106°59'34"W). A single observer walked around recording the robins and the GPS locations of each singing robin. We aimed to record a set of individuals on two separate occasions by returning to the same location and recording the territorial resident. We aimed to record individuals for five minutes at least twice. While we were unable to achieve this in all cases, we used the same recordings for all the estimation techniques; this allows us to accurately compare them. Songs were recorded using a Sennheiser ME67 directional microphone with a K6 power module (Sennheiser Electronic Corporation, Old Lyme, Connecticut, USA), a Rycote windscreen (Rycote Microphone Windshields Ltd, Gloucestershire, United Kingdom), and a Marantz Professional Solid State Recorder PMD660 (Marantz America, Inc., Mahwah, New Jersey, USA). We did not record songs during rainy or windy conditions because such environmental noise reduced the quality of recordings.

Once the songs were recorded we analyzed them with Raven Pro 1.3 (Krein et al., 2009, [www.birds.cornell.edu/raven](http://www.birds.cornell.edu/raven)). Spectrograms of the songs (256 point FFTs) were viewed in a standardized window of 0–5.5 kHz and 0–0.8 seconds, using Raven Pro 1.3. We used the 100-contrast level and a brightness level of 70 and created a syllable key to aid in syllable identification. We identified each syllable in every recording that was > 1 min. Because we were interested in estimating syllable diversity, we did not control the number of songs recorded for each individual.

Three data sets (a large, medium and small one) were created from the recordings of individuals that sang > 1 minute and that contained at least 50 syllables. The Large data set (hereafter, abbreviated L) contained the first 50 syllables from 16 individuals. The Medium data set (hereafter, abbreviated M) contained the first 100 syllables from 14 individuals. The Small data set (hereafter, abbreviated S) contained the first 150 syllables from five individuals. The syllable count included both new and repeated syllables until the total 50-, 100-, and 150- syllables were reached. We acknowledge that these

data sets are likely to not have captured each individual's *full* repertoire, however for the purposes of comparing methods, and for the purpose of comparing sampling efforts, they are appropriate since we used the same number of syllables from each individual in the data sets. Also, we recorded mainly at dawn when male robins have the most extreme songs and believe that while our recordings are not long, they should properly represent the typical songs females heard before selecting mates (Slagsvold, 1996; Byers and Kroodsmma, 2009).

### 1.2 Comparing methods to quantify repertoire size

With the set of syllables sung by each bird, we used four different methods to estimate an individual's and the population's syllable repertoire size. For comparison purposes we estimated the average repertoire size of both the population and individuals. Averages were estimated to account for individual variation in syllable repertoire size, and we believe that this technique may aid others in creating recognition algorithms that can discriminate among individuals because these estimates will set a limit on the number of syllables an individual sings.

First, we counted the syllables of all individuals in each data set and then found the average, median, and variance for an individual's repertoire size, and the 95% confidence interval for the individuals in all three data sets (L, M, and S). However, this method was biased because we did not have the same number of syllables recorded for each bird. We acknowledge that it also was likely incomplete, especially for those birds with shorter recordings. However, for our purposes, it was sufficient because we could compare this count to the estimates made from sampling techniques. The count of the population's complete repertoire did not have a confidence interval because it was neither an estimate nor an average.

Second, we used a behavioral accumulation curve (Dias et al., 2009) to calculate the robin repertoire size. The counts of the first 50-, 100-, 150- syllables sung were fitted to a behavioral accumulation curve,

$$P_{Sobs} = T_{Sobs} / (a/b)$$

where,  $P_{Sobs}$  is the portion of syllables sung,  $T_{Sobs}$  is the total number of syllables sung,  $a$  is the intercept, and  $b$  is the slope (Dias et al., 2009). A quadratic equation best explained the accumulation of robin syllables in previous studies (Dziadosz, 1977), as well as in our study. Thus, we estimated these parameters using a quadratic formula. We then averaged the results from the above

calculations for the L, M, and S data sets and calculated 95% confidence intervals.

Third, we used capture-recapture techniques to estimate the repertoire size by randomly selecting a set of syllables from all observed syllables 11 independent times. The first selection was the set of syllables that were “captured and marked”; the remaining 10 selections were the “recaptures”. We applied the capture-recapture formula,

$$N(t) = N*(N/n)$$

where,  $N$  is the sample size and  $n$  is the number of recaptured syllables (Catchpole and Slater 2008). This capture-recapture procedure was repeated three times with data sets containing accumulations of 50-, 100-, and 150- syllables from each individual. Therefore the data sets consisted of 800, 1400, and 750 accumulated syllables from which we randomly chose 50-, 100-, and 150- syllables. Once we had calculated the ten repertoire estimations for the population, we calculated an individual’s repertoire. To accomplish this we divided the population results by the number of individuals within the data set, repeating this for all data sets. To compare these results with the other methods we averaged the ten estimates, for the individual and population estimates of every data set, and calculated the 95% confidence intervals on these mean values.

Fourth, we used EcoSim 7.72 (Gotelli and Entsminger, 2009) to apply rarefaction methods for the L, M, and S data sets. For each data set we created a table containing individuals and the syllables they sang. Each row in the table contained one syllable. We then entered the times an individual sang that syllable in the corresponding cell. Once this table was completed we transferred the data into EcoSim 7.72 (Gotelli and Entsminger, 2009) by creating a matrix with 220 rows and 16-, 14-, and 5- columns (one for each robin in the data set). Using the species diversity function in EcoSim, we then calculated the richness within the data sets, using the formula,

$$E(S_n) = S - \left(\frac{N}{n}\right)^{-1} \sum_{i=1}^s \left(\frac{N - N_i}{n}\right)$$

where,  $E(S_n)$  is the estimated number of syllables,  $N$  is the total number of recorded individuals within a sample,  $n$  is a sub-sample of randomly selected recorded individuals, and  $S$  is the number of syllables (Heck et al., 1975; James and Wamer, 1982). In the species diversity window of EcoSim we used the all the default settings, besides sampling the whole sample (data set) and allowing the program to fit the rarefaction algorithm.

## 2 Results

### 2.1 A numerical count for repertoire size

Of the 38 individuals recorded on separate occasions, nine were recorded for at least five minutes and another 14 for at least three minutes; the remainder were recorded < 1 min. Individuals recorded for < 1 min were not included in this analysis. Using the set of recordings > 1 min, we identified 220 distinct syllables (unpublished data).

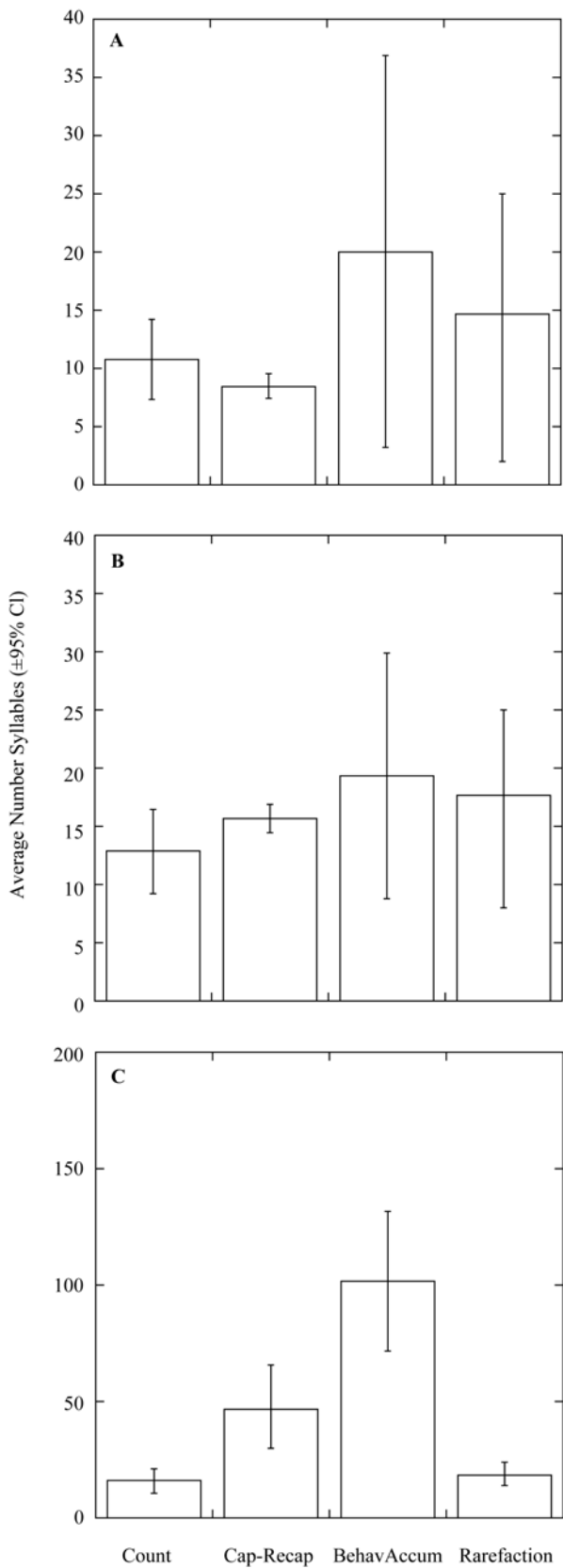
When we counted the first 50 syllables per individual, we estimated the average robin repertoire to be 10.79 syllables (95% CI: 7–14 syllables) for the L data set. The M syllable data set estimated the robin repertoire to be 12.86 syllables (95% CI: 9–16 syllables). The S data set sample estimated the robin repertoire to be 16 syllables (95% CI: 11–21 syllables). The population repertoires for the numerical count technique do not have confidence intervals. Our population’s average repertoire size was estimated as 166, 180, and 78 in the L, M, and S data sets respectively.

### 2.2 A comparison of methods to a numerical count in estimating repertoire size

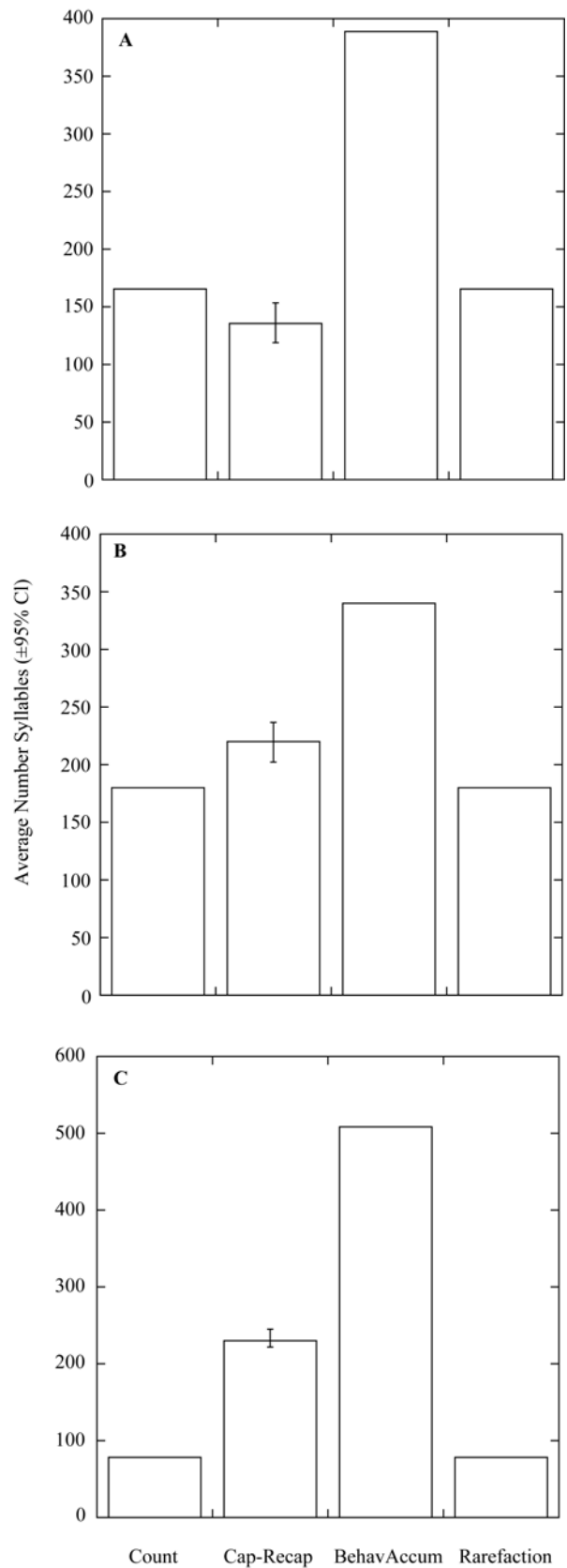
Fig. 1 and 2 compare the methods for both individual and population repertoires. The behavioral accumulation analysis also estimated larger repertoires than those seen in the count analyses. The L data set analysis estimated 20.03 syllables per individual (95% CI: 3–37 syllables), the M data set analysis estimated 19.3 syllables (95% CI: 9–30 syllables), and the S data set analysis estimated 101.73 syllables (95% CI: 72–132). The results for the behavioral accumulation curve do not overlap the other confidence intervals for the S data set estimates. Rather, the behavioral accumulation estimates for an individual’s repertoire are very large, ranging from 72 to 132. The population repertoire estimates are 389.1, 339.8, and 508.6 syllables, respectively.

Capture-recapture generated similar individual repertoire size estimates as the direct count when we used the L and M data sets consisting of 50 (8.5 syllables, 95% CI: 7–14 syllables) and 100 syllables (15.69 syllables, 95% CI: 14–17 syllables). When we used the S data set, we estimated that each robin sang 45.98 syllables, (95% CI: 12–49 syllables). The population estimate for the L data set was lower (136.04, 95%, 118–154 syllables) than that calculated from the direct count, while the results from M (219.6 syllables, 95% CI: 202–237 syllables) and S (229.9, 95% CI: 221–246 syllables) data sets were larger.

The 95% confidence intervals from the rarefaction



**Fig. 1** Comparison of repertoire estimation techniques for individual robin repertoires in the Large data set (A), the Medium data set (B), and the Small data set (C)



**Fig. 2** Comparison of repertoire estimation techniques for the population of sampled robin repertoires in the Large data set (A), the Medium data set (B), and the Small data set (C)

analysis overlap those from the count analysis, although these intervals are larger for the rarefaction analysis. This overlap suggests that there are no significant differences between the estimate of the rarefaction and count methods. Rarefaction analysis estimated an average individual repertoire of 14.69 syllables (95% CI: 2–25) for the L data set, 17.64 syllables (95% CI: 8–25) for the M data set analysis, and 18.73 syllables (95% CI: 14–24) for the S data set analysis. The population repertoire estimates for this technique were 166, 180, and 78.

### 3 Discussion

We acknowledge that all methods for estimating repertoire sizes are imperfect; a full numerical count is the best method (Botero et al., 2008). However, for large repertoires, a full count may be quite difficult since the amount of effort required to fully quantify the repertoire is substantial (Garamszegi et al., 2002). We used count data (that we recognize may not be a comprehensive count but instead closely represents the repertoire that other birds react to (e.g., Byers and Kroodsma, 2009)) to compare two traditional methods (mark-recapture and behavioral accumulation curves) of estimating repertoire size with a method rarely used by ornithologists—rarefaction.

Confidence intervals from the rarefaction analysis estimating individual repertoires using all three data sets (S, M, and L) were not significantly different than those obtained by the numerical count. This suggests that rarefaction is as accurate as a numerical count and thus it is the best alternative to a full count that we examined. Importantly, rarefaction may decrease effort at estimating song (or syllable) repertoire size because it is effective with relatively short recordings. By contrast, other methods, e.g., capture-recapture (Forstmeier and Balsby, 2002), had significant shortcomings (Botero et al., 2008). Behavioral accumulation curves and capture-recapture are biased with small and/or incomplete samples (Botero et al., 2008). Most repertoire estimations in fact use incomplete samples to estimate the total repertoire, suggesting that future studies should use the most appropriate method, rarefaction, as the alternative to a numerical count, when resources are limited, rather than the more biased techniques.

Rarefaction permits us to sample each individual less because it retains accuracy with incomplete samples. Thus, it should be possible to sample more individuals, and at more locations, thereby facilitating studies of geographic variation.

Beyond studies of syllable repertoire size, rarefaction techniques can be used to estimate the repertoire size of any behavior. Behavioral biologists, like ornithologists, have searched for appropriate methods with which to accurately estimate behavioral repertoires in ways that decreased the amount of observation time (Wilson and Fagen, 1974). We suggest that rarefaction may be a useful solution to this problem as well.

**Acknowledgements** We thank Gary Entsminger for help with ECOSim, the Rocky Mountain Biological Laboratory for facilities, and NSF IDBR-0754247 for support. Research was conducted under UCLA ARC Protocol 2000-147, and under permits issued by the Rocky Mountain Biological Laboratory. Comments by four anonymous reviewers and the editor helped us focus and improve our presentation.

### References

- Borror DJ, 1965. *Birdsong: The anatomy of a miracle*. Audubon 67: 159–163.
- Botero CA, Mudge AE, Koltz AM, Hochach WM, Vehrencamp SL, 2008. How reliable are methods for estimating repertoire size? *Ethology* 114: 1227–1238.
- Brenowitz EA, Margoliash D, Nordeen KW, 1997. An introduction to bird song and the avian song system. *J. Neurobiol.* 33: 495–500.
- Brumm H, Lachlan RF, Reibel K, Slater PJB, 2009. On the function of song repertoires: Testing the ‘antiexhaustion hypothesis’ in chaffinches. *Anim. Behav.* 77: 37–42.
- Byers BE, Kroodsma DE, 2009. Female mate choice and songbird song repertoires. *Anim. Behav.* 77: 13–22.
- Catchpole CK, Slater PJB, 2008. *Bird Song: Biological Themes and Variations*. 2nd edn. Cambridge: Cambridge University Press.
- Dias PAD, Rangel-Negrín A, Coyohua-Fuentes A, Canales-Espinosa D, 2009. Behaviour accumulation curves: A method to study the completeness of behavioural repertoires. *Anim. Behav.* 77: 1551–1553.
- Dziadosz V, 1977. *The vocalizations of the American robin*. Doctoral Dissertation, Ohio State University, Columbus, Ohio.
- Forstmeier W, Balsby TJS, 2002. Why mated dusky warblers sing so much: Territory guarding and male quality announcement. *Behaviour*. 139: 89–111.
- Garamszegi LZ, Boulinier T, Møller, AP, Torok J, Michl G et al., 2002. The estimation of size and composition of avian song repertoires. *Anim. Behav.* 63: 623–30.
- Gotelli NJ, Colwell RK, 2001. Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology* 4: 379–391.
- Gotelli NJ, Entsminger GL, 2009. *EcoSim: Null Models Software for Ecology*. Version 7. Jericho, VT 05465: Acquired Intelligence Inc. & Kesey-Bear. <http://garyentsminger.com/ecosim.htm>.
- Heck KL, Van Belle G Jr, Simberloff, 1975. Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. *Ecology* 56: 1459–1461.
- Ince SA, Slater PJB, 1985. Versatility and continuity in the songs

- of thrushes *Turdus* spp. *Ibis* 127: 355–364.
- James FC, Wamer NO, 1982. Relationships between temperate forest bird communities and vegetation structure. *Ecology* 63: 159–171.
- Johnson SL, 2006. Do American robins acquire songs by both imitating and inventing? *Wilson J. Ornithol.* 118: 341–352.
- Krebs JR, Kroodsma DE, 1980. Repertoires and geographical variation in bird song. *Adv. Study Behav.* 11: 143–77.
- Krein T, Hawthorne D, Hu D, Strickman L, Nelson S, 2009. Raven Pro sound software. Bioacoustics Research Program, Cornell Lab of Ornithology. [www.birds.cornell.edu/raven](http://www.birds.cornell.edu/raven)
- Kroodsma DE, Parker LD, 1977. Vocal virtuosity in the brown thrasher. *Auk* 94: 783–785.
- Kroodsma DE, 1982. Song repertoires: Problems in their definition and use. In: Kroodsma DE, Miller EH ed. *Acoustic Communication in Birds*. New York: Academic Press, 125–146.
- Konishi M, 1965. Effects of deafening on song development in American robins and black-headed grosbeaks. *Z. Tierpsychol.* 22: 584–599.
- Longeno JT, Coddington J, Colwell RK, 2002. The ant fauna of tropical rain forest: Estimating species richness three different ways. *Ecology* 83: 689–702.
- Sallabanks R, James FC, 1999. American robin *Turdus migratorius*. In: Poole A ed. *The Birds of North America Online*. Ithaca: Cornell Lab of Ornithology; Retrieved from the Birds of North America Online: <http://bna.birds.cornell.edu/bna/species/46> doi:10.2173/bna.462.
- Slagsvold T, 1996. Dawn and dusk singing of male American robins in relation to female behavior. *Wilson Bull.* 108: 507–515
- Thomas DK, 1979. An analysis of the morning song of the American robin *Turdus migratorius* throughout the breeding season. M.Sc. Thesis, Bloomsburg State College, Bloomsburg, Pennsylvania.
- Wildenthal JL, 1965. Structure in primary song of the mockingbird *Mimus polyglottos*. *Auk* 82: 161–89.
- Williams H, 2004. Birdsong and singing behavior. *Annal. N.Y. Acad. Sci.* 1016: 1–30.
- Wilson EO, Fagen RM, 1974. On the estimation of total behavioral repertoires in ants. *N.Y. Entomol. Soc.* 82: 106–112.