Citation:

Harris AJ, Wilson DR, Graham BA, Mennill DJ (2016) Estimating repertoire size in a songbird: a comparison of three techniques. *Bioacoustics*, 25: 211–224. doi: 10.1080/09524622.2016.1138416

Estimating repertoire size in a songbird: A comparison of three techniques

Alexander J. Harris[a], David R. Wilson[a,b], Brendan A. Graham[a], and Daniel J. Mennill[a,*]

[a]Department of Biological Sciences, University of Windsor, 401 Sunset Ave, Windsor, ON, Canada N9B 3P4

[b]Department of Psychology, Memorial University of Newfoundland, 232 Elizabeth Avenue, St. John's, NL, Canada A1B 3X9

*Corresponding author. Email: dmennill@uwindsor.ca

Running head: Comparing repertoire size estimation techniques

ABSTRACT

Many animals produce multiple types of breeding vocalizations that, together, constitute a vocal repertoire. In some species, the size of an individual's repertoire is important because it correlates with brain size, territory size, or social behaviour. Quantifying repertoire size is challenging because the long recordings needed to sample a repertoire comprehensively are difficult to obtain and analyze. The most basic quantification technique is simple enumeration, where one counts unique vocalization types until no new types are detected. Alternative techniques estimate repertoire size from subsamples, but these techniques are useful only if they are accurate. Using 12 years of acoustic data from a population of rufous-and-white wrens in Costa Rica, we used simple enumeration to measure the repertoire size for 40 males. We then compared these to the estimates generated by three estimation techniques: Curve Fitting, Capture-Recapture, and a new technique based on the Coupon Collector's Problem. To understand how sampling effort affects the accuracy and precision of estimates, we applied each technique to six different-sized subsets of data per male. When averaged across subset sizes, the Capture-Recapture and Coupon Collector techniques showed the highest accuracy, whereas the Curve Fitting technique underestimated repertoire size. Precision (the average absolute difference between the estimated and true repertoire size) was significantly better for the Capture-Recapture technique than the Coupon Collector and Curve Fitting techniques. Both accuracy and precision improved as subset size increased. We conclude that Capture-Recapture is the best technique for estimating the sizes of small repertoires.

Key words: Capture-Recapture, Coupon Collector's problem, Curve Fitting, repertoire size, simple enumeration, vocal repertoire

INTRODUCTION

Variation in vocal characteristics is associated with fitness in many species. For example, structural variation in vocalizations can signal fighting ability and aggression (Linhart, Slabbekoorn, and Fuchs, 2012), facilitate adaptive antipredator responses (Manser, 2013), and enable animals to communicate effectively in the presence of variable background noise (Slabbekoorn, 2013). Many animals have multiple types of breeding vocalizations that they can produce, and, together, these constitute an animal's vocal repertoire. Repertoire sizes vary considerably within species and populations (e.g. Peters et al., 2000), and this variation has been correlated with reproductive success (e.g. Reid et al., 2004), territory size (e.g. Aweida, 1995), and cognitive abilities (e.g. Sewall, Soha, Peters, and Nowicki, 2013). Our understanding of the adaptive significance of animal repertoires hinges on accurate and precise quantification of repertoire size.

Determining an animal's repertoire size can be a challenging task. The most basic technique is simple enumeration, which involves counting the number of unique types of vocalizations that an individual produces. Ideally, an individual would be followed for its entire lifetime to ensure that no vocalizations are missed. Because this is impractical, a rule must be established to limit sampling effort to a practical level. The sampling effort required for simple enumeration should reflect the effort typically needed to quantify an individual's entire repertoire, based on previous findings that involve thorough recordings. If no previous findings exist, then the sampling effort should be high enough that the researcher obtains many new recordings without detecting any new vocalization types. The amount of effort required to quantify repertoire size using simple enumeration is influenced by the size of the animal's repertoire, the pattern with which the animal selects its vocalizations, the frequency with which the animal vocalizes, and whether an animal is a closed-ended learner (i.e. all songs are learned early in life and adult repertoire size is fixed) or an open-ended learner (i.e. songs continue to be learned throughout life). Simple enumeration can work well for species with small repertoire sizes, species that cycle through their entire repertoire cyclically, and species that vocalize often (Botero et al., 2008). Simple enumeration requires much greater effort for species with larger repertoires, species that choose different types of vocalizations with different probabilities or a broader range of probabilities, and species that vocalize rarely.

Several estimation techniques have been developed to reduce the amount of effort required to obtain an accurate measure of repertoire size. Two common techniques are Curve Fitting and Capture-Recapture. The Curve Fitting technique uses the formula described by Wildenthal (1965) to fit a line of best fit to a small subset of data. The horizontal asymptote of the line then becomes the repertoire size estimate. Curve Fitting has been used for repertoire size estimation in several species (Derrickson, 1987, Botero et al., 2008).

The Capture-Recapture technique involves a different approach that is based on a comparison of the number of unique types of vocalizations recorded during two or more sampling occasions. The proportion of vocalization types from an initial sample that are observed again in subsequent samples is then used to estimate total repertoire size (Baillargeon & Rivest, 2007). Capture-Recapture has been popular for estimating the sizes of populations in ecological studies, but proves equally useful for estimating animals' repertoire sizes (Garamszegi et al., 2002, Garamszegi et al., 2005).

Previous studies on the accuracy of Curve Fitting and Capture-Recapture techniques have yielded inconsistent findings. Garamszegi et al. (2005) demonstrated that Capture-Recapture could accurately estimate a bird's syllable repertoire size using only 15 songs. The method was especially useful for species with large repertoires and heterogeneous selection probabilities (Garamszegi et al. 2005). In another study that focused on species with large repertoire sizes (≥ 160 element types), Botero et al. (2008) found that Capture-Recapture and Curve Fitting were both inaccurate when the sample size was small, and that they only became accurate when the sample size was so large that simple enumeration was also feasible (Botero et al. 2008).

A new estimation technique based on the Coupon Collector's Problem (Erdös and Rényi, 1961; Feller, 1968; Dawkins, 1991) was recently debuted by Kershenbaum, Freeberg, and Gammon (2015). The Coupon Collector's Problem describes a situation in which all items in a set must be collected, and where sampling occurs with replacement. Under this model, the initial items are collected rapidly, and the last few items take much more extensive sampling to acquire. This situation has obvious parallels to the sampling of an animal's vocal repertoire, particularly when the animals select their vocalization types at random (Kershenbaum et al., 2015). Observed repertoire size grows rapidly at the beginning of sampling, but then tapers off as more of the repertoire is sampled, until it plateaus when the entire repertoire has been sampled. Using a modification of the Coupon Collector's Problem that accounts for unequal probabilities of each song type (i.e. heterogeneous selection probability), Kershenbaum et al. (2015) showed that this technique is a more accurate predictor of repertoire size than other estimation techniques for species with heterogeneous selection probability. Their study estimated repertoire sizes at the population level, rather than the individual level, and so whether the Coupon Collector technique provides accurate estimates of the repertoire sizes of individual animals remains to be studied.

In this study, we compare repertoire size estimation techniques by analyzing historical repertoire size data from a population of rufous-and-white wrens (*Thryophilus rufalbus*). We compare the Curve Fitting, Capture-Recapture, and Coupon Collector techniques in terms of their ability to produce repertoire size estimates that match with the results from extensive simple enumeration. Rufous-and-white wrens are neotropical songbirds found in forests throughout western Central America and northwestern South America. Males of this species are closed-ended learners that sing one song type repeatedly before switching to a different song type, and may cycle through the same song types many times before singing their entire repertoire (i.e. an eventual variety, non-cyclic singing style; Mennill & Vehrencamp 2005, Hennin et al. 2009). When song type switches occur, certain song types are selected more often than others, giving a heterogeneous selection probability to each song type in their repertoire (unpublished data). Repertoire estimation techniques are thought to perform poorly when animals are undersampled (Derrickson 1987) or when they do not select song types with equal probability (Kroodsma 1982); this makes rufous-and-white wrens an interesting test case for studying these three estimation techniques.

Our first goal was to determine the repertoire sizes of male rufous-and-white wrens using twelve years of historical data collected in the field in Costa Rica. Many of our study animals have been recorded extensively, and we could quantify their repertoire size with confidence using simple enumeration. Our second goal was to compare the accuracy and precision of repertoire size estimations from the Curve Fitting, Capture-Recapture, and Coupon Collector techniques. We applied these techniques to different-sized subsets of our data and compared the repertoire size estimates to the repertoire size we determined through simple enumeration, which we used as a proxy for the animals' true repertoire sizes.

METHODS

Recording Vocal Repertoires

Data were collected at Sector Santa Rosa, Area de Conservación Guanacaste, Costa Rica (10°40' N, 85°30' W), where our research group has been conducting a long-term study of communication behaviour in a colour-banded population of rufous-and-white wrens since 2003. We analyzed data from 40 male wrens that we recorded during 1 to 7 successive breeding seasons (average ± SE: 3.7 ± 0.2) between 2003 and 2014. Birds were recorded between March and July of each year, coinciding with the onset of the breeding season of this species, when male vocal output reaches its peak (Topp & Mennill 2008). Birds were captured in their territories using mist nets and then banded with a unique combination of three coloured leg bands and a metal band to facilitate identification in the field. Rufous-and-white wrens are renowned for their vocal duets (Mennill & Vehrencamp 2008, Kovach et al. 2014), but we focused the current analyses on the vocalizations produced by males (both songs produced as solos and as contributions to duets), given their high song output and our extensive sampling of their songs (Mennill & Vehrencamp 2005; Topp & Mennill 2008).

Analysis of Field Recordings

We collected two types of field recordings: focal recordings and automated recordings. Focal recordings involved a recordist following a male through his territory at distances of 10 to 30m, dictating the bird's identity after each song. All focal recordings were collected between 0445 h and 1100 h. Focal recordings were collected with a shotgun microphone (Sennheiser MKH70 or ME67) and a solid-state digital recorder (Marantz PMD660 or PMD670; 22,050 Hz sampling rate, 16-bit encoding accuracy, WAVE format). Focal recordings were collected every year between 2003 and 2014, and they comprise the majority of recordings in this analysis (approximately 60%).

To complement focal recordings, and to sample birds' repertoires over longer periods than was possible with focal recordings, we collected automated recordings with three different types of equipment, all used to sample birds' songs at times when focal recordists were not present. (1) Microphone array recordings were collected in 2003 and 2004 by placing an array of eight stationary omni-directional microphones throughout birds' territories (sampling frequency: 22,050 Hz; full equipment details in Mennill et al., 2006). (2) Automated recorders consisting of elevated omni-directional microphones (Sennheiser ME62) and solid-state digital recorders (Marantz PMD670) were placed near the centre of the focal pair's territory in 2007 through 2010 (sampling frequency: 44,100 Hz; full equipment details in Mennill, 2014). (3) Automated Song Meter recorders (model: SM2-GPS, Wildlife Acoustics Inc., Concord, Massachusetts, USA) were placed in the centre of a pair's territory in 2011-2014, usually within 10m of the focal pair's nest (sampling frequency: 22,050 Hz; full equipment details in Mennill et al. 2012). We confirmed the identities of the birds in these unattended, automated recordings by ensuring that the song types matched between the focal recordings and the automated recordings; in all cases the songs recorded with the automated recorders unambiguously matched with the songs in the focal recordings of the known male from the same area. We distinguished between the voices of males versus females following previously established criteria (see Mennill and Vehrencamp, 2005). Our ongoing field studies involve re-sighting the birds throughout the field season to monitor their breeding behaviour, and we ensured that focal animals were located in the same territory before and after automated recordings were collected. Given that our study birds have large breeding territories (territory sizes range from 5678 ± 548 m^2 to 13497 ± 1043 m^2; Osmun and Mennill, 2011, Mennill and Vehrencamp, 2008), with substantial undefended spaces between adjacent territories (Osmun and Mennill 2011), our automated recorders placed centrally within birds' territories recorded only the target individuals. Any songs produced by rare territorial intruders were readily distinguished from the resident birds by cross-referencing repertoire data of neighbouring animals; even though song types are shared between individuals (Mennill & Vehrencamp 2005), shared song types have individually distinctive characteristics.

Assigning Songs to Song Types

Rufous-and-white Wrens have vocal repertoires of songs, where each song type is readily classified into different song types based on the visual and aural characteristics of the three sections of their song: the introductory syllables, trill notes, and terminal syllables (as in Mennill & Vehrencamp 2005; Barker 2008). Following previous work by Barker (2008), songs were classified manually into types by comparing structural characteristics such as syllable length, minimum and maximum frequencies, frequency of maximum amplitude, bandwidth, and inter-syllable interval for the three song sections. In an analysis of song type categorization that relied on discriminant analysis with cross-validation, Barker (2008) showed that fine structural measurements are useful for accurately distinguishing different song types.

We annotated the audio files from all focal and automated recordings in SYRINX-PC sound-analysis software (J. Burt, Seattle, Washington, USA). We annotated each song and recorded its song type, manually comparing each song to a library of all previously recorded song types from that animal. When a bird produced a song that had a different song type from the previous song, we counted it as a song type switch. We determined the repertoire size of each bird from the total number of song types recorded throughout the entire study for that bird. Using these data, we constructed accumulation curves that showed the number of song type switches sampled on the x-axis versus the number of unique song types detected on the y-axis for each bird (Figure 1). Rather than using the total number of songs recorded, we used song type switches as the unit of interest when calculating repertoire size (as in other studies, for example, Valderrama et al. 2008; Sosa-López & Mennill 2014a, 2014b). We did this because rufous-and-white wrens sing with eventual variety, repeating a given song type, on average, eleven times before switching to a new song type (Mennill & Vehrencamp, 2005). Indeed, an animal may sing a specific song type more than 100 times in a row before switching to a new song type, leading to large plateaus in song type collection if sampling effort is measured relative to number of songs sung instead of number of song type switches. Within these long bouts of repeated songs, the song type of subsequent songs are not independent. For this reason, we treated song type switches as our unit of analysis.

We used simple enumeration to measure the actual repertoire size of each rufous-and-white wren because individuals used in this study had been recorded extensively (see Results).

This estimate was used as the benchmark to which the other three techniques were compared. Only individuals with 150 or more recorded song type switches were used in the analysis. We chose this number because 95% of the individuals had no new song types discovered after 150 song type switches using simple enumeration.

Repertoire Size Estimation

To determine the effect of sampling effort on the accuracy of each estimation technique, we created subsets of the data for each bird, using the first 25, 50, 75, 100, 125, and 150 song type switches recorded from each individual. This allowed us to examine the estimates produced by each technique from different amounts of sampling effort. We used R (R Core Team, 2014) to generate data subsets and to generate all repertoire size estimates. The raw data and the relevant R code are included in the online supplementary material.

For the Curve Fitting technique, we generated prediction curves for each possible repertoire size between 1 and 30 song types (i.e. a range that encompassed repertoire sizes we have encountered in our population in the last 12 years). We used the formula presented in Wildenthal (1965):

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

where n is the number of unique song types expected in a sample containing T song type switches; N is the assumed repertoire size. Thus, for each possible repertoire size between 1 and 30 song types, we generated a unique curve with an asymptote at that value. We applied an iterative process in which we generated a predictive model for each possible repertoire size, and then assessed the fit of each model by comparing it to the observed data using a least squares technique. Specifically, for each subset size and for each male, we selected the model that generated the smallest value when the absolute differences between the predicted and observed values were summed across all song type switches. The *N* from this model became the best estimate of repertoire size.

For the Capture-Recapture technique, we used *Rcapture* (R package; Rivest & Baillargeon, 2014) to estimate repertoire size. For each combination of male and subset size, we created a capture history that indicated which song types were captured during which capture occasions (0 = not captured; 1 = captured). Following Garamszegi et al. (2005) and Botero et al. (2008), we defined a capture occasion as 5 song type switches, which divided evenly into all of our subset sizes. Our Capture-Recapture models were based on a closed population, since our preliminary analyses suggest that repertoire size does not change throughout an adult's lifetime in this species (i.e. Rufous-and-white Wrens are closed-ended learners; Mennill & Vehrencamp 2015; DJM unpublished data). *Rcapture* can incorporate several different sources of variation that can each affect capture probabilities (Baillargeon & Rivest, 2007). We used Darroch's M_h model, which allows the probability of capture to vary among units (Darroch et al. 1993). This model thereby accounts for the possibility of common and rare song types when predicting repertoire size.

The Coupon Collector's Problem is based on the idea of collecting a set of coupons that are hidden in cereal boxes (Dawkins 1991; Feller 1968). If there are *N* different coupons, it estimates the probability of collecting exactly *i* different coupons after purchasing *m* cereal boxes. The coupons are drawn at random and with replacement. For our study, we used the

Coupon Collector's Problem to estimate the probability of observing *i* of *N* different song types after sampling *m* song type switches. We implemented the Coupon Collector's Problem using a Monte Carlo simulation. For each possible repertoire size (*N*), and for each possible number of song type switches (*m*), we drew 100,000 independent samples. Each sample contained *m* songs and was drawn at random and with replacement from the repertoire of *N* song types. As in Kershenbaum et al. (2015), we modified the Coupon Collector's Problem to allow for unequal probabilities of song type selection. We set the probability of selecting each song type based on a Zipfian distribution, which has been used in previous studies to model the frequency of words in human languages, as well as the frequency of song types in avian vocal repertoires (Zipf 1949; Lemon & Chatfield 1973). Probabilities are calculated by the formula:

$$p(k; s, N) = \frac{1/k^s}{\sum_{n=1}^{n=N}(1/n^s)}$$

where $p(k; s, N)$ is the probability of selecting the k^{th} most common song type from a repertoire of *N* song types; *s* is the absolute value of the slope of the regression of the frequency of each song type on its corresponding rank, when plotted on a log-log scale. We used our raw data to calculate *s* for each subset size included in our analyses (25, 50, 75, 100, 125, or 150 song type switches). For each possible repertoire size (i.e. 1 to 30), and for each possible number of song type switches (i.e. 1 to 150), we calculated the expected number of song types as the average number of song types observed among the 100,000 samples. We used these values to create a prediction curve for each repertoire size. As in our analysis of the Curve Fitting technique, we assessed the fit of each prediction curve by comparing it to the observed data with a least squares technique. The *N* from the model that minimized the least squares was selected as the best estimate of repertoire size.

Statistical Analysis

We used a linear mixed-effects model in the R package *nlme* (Bates et al. 2015) to assess the effects of estimation technique and subset size on the accuracy of repertoire size estimates. We defined "accuracy" as the average difference between the repertoire size estimates generated with a particular technique and the true repertoire sizes determined through simple enumeration. In general, smaller deviations from zero indicated better accuracy; negative values indicated that a method was underestimating the true repertoire size, whereas positive values indicated that a method was overestimating the true repertoire size. We included the differences as a dependent variable in our analysis, and the estimation technique (i.e. Curve Fitting, Capture-Recapture, and the Coupon Collector technique), subset size (as a covariate), and 2-way interaction as independent variables with fixed effects. We did not include an intercept for the fixed effects because the hypothesized difference between the estimated and observed repertoire sizes was zero. To facilitate the interpretation of model coefficients, we centered subset size on zero. Bird identity was included as a subject variable with random intercepts to account for repeated measurements from the same individuals. We fit the model using restricted maximum likelihood estimation, and concluded that a particular estimation technique was accurate if the difference between its repertoire size estimates and the true repertoire sizes could not be distinguished statistically from zero.

We used a similar analysis to assess the effects of estimation technique and subset size on the precision of repertoire size estimates. In this study, we consider precision to be a measure of consistency in estimation. In estimating repertoire size, one might generate some overestimates and some underestimates of true repertoire size, but an average value that matches the true repertoire size; this is a situation with high accuracy, but low precision. We defined “precision” as the average absolute difference between the repertoire size estimates generated from a particular technique and the true repertoire sizes determined through simple enumeration. Smaller differences in these absolute values would indicate more consistency in the estimation of repertoire size, and therefore better precision. We again used a linear-mixed effects model as in our analysis of accuracy (above). We compared precision among the three estimation techniques using Tukey post-hoc comparisons, which we implemented in the R package *multcomp* (Hothorn et al. 2008).

All tests were two-tailed, and results were considered significant when $p \leq 0.05$. Both models complied with the parametric assumptions of linearity, homoscedasticity, and normality, as revealed by visual inspection of residual plots.

RESULTS

Enumerated Repertoire Size

Simple enumeration showed that the 40 male rufous-and-white wrens produced an average of 11.4 ± 0.3 song types each (mean ± SE; range: 8 – 15 song types), which is in accordance with a previous enumeration study of this species (Mennill and Vehrencamp, 2005). These results were based on extensive recordings of each individual (e.g. Figure 1), containing an average of 3619 ± 374 songs (mean ± SE; range: 744 – 11691) and 447 ± 43 song type switches (mean ± SE; range: 154 – 1882).

Accuracy of Repertoire Size Estimates

Estimation technique had a significant effect on the accuracy of repertoire size estimates (linear mixed-effects model: $F_{3,675} = 11.5$, $p < 0.001$; Fig. 2). The Capture-Recapture technique generated repertoire size estimates that were not statistically different from animals' true repertoire sizes ($t_{675} = -1.1$, $p = 0.283$; 95% CI for the difference: -0.8 – 0.2 songs types), underestimating the true repertoire size by only 0.3 ± 0.3 song types (mean ± SE; Table 1; Fig. 2). The Coupon Collector technique also generated repertoire size estimates that were not significantly different from animals' true repertoire sizes ($t_{675} = 0.8$, $p = 0.444$; 95% CI: -0.8 – 0.3 song types), underestimating the true repertoire size by only 0.2 ± 0.3 song types. In contrast, the Curve Fitting technique significantly underestimated repertoire size, with an average repertoire size estimate that was 1.0 ± 0.3 song types below the true repertoire size ($t_{675} = -3.4$, $p = 0.002$; 95% CI for the difference: -1.5 – -0.4 song types; Table 1; Fig. 2).

Subset size had a significant effect on the accuracy of repertoire size estimates, with larger subset sizes producing more accurate estimates for all estimation techniques. This effect was manifested through a significant interaction between estimation technique and subset size (estimation technique: $F_{3,675} = 11.5$, $p < 0.001$; subset size: $F_{1,675} = 2.0$, $p < 0.001$; interaction: $F_{2,675} = 14.8$, $p < 0.001$). Specifically, the Curve Fitting and Capture-Recapture techniques tended to underestimate repertoire size more at smaller subset sizes than at larger subset sizes. In contrast, the Coupon Collector technique tended to overestimate repertoire size more at smaller subset sizes than at larger subset sizes (Table 1; Fig. 2).

Precision of Repertoire Size Estimates

The precision of repertoire size estimates was affected significantly by estimation technique (linear mixed-effects model: $F_{3,675} = 13.3$, $p < 0.001$), subset size ($F_{1,675} = 201.0$, $p < 0.001$), and the 2-way interaction between them ($F_{2,675} = 6.0$, $p = 0.003$). The precision of the Capture-Recapture technique was 0.9 ± 0.2 song types (mean ± SE; 95% CI: 0.5 – 1.4 song types), which was significantly better than the Coupon Collector technique (1.4 ± 0.2 song types; 95% CI: 0.9 – 1.9 song types; Tukey post-hoc comparison: $Z = 6.3$, $p < 0.001$; Table 1), but was statistically indistinguishable from the Curve Fitting technique (1.2 ± 0.2 song types; 95% CI: 0.7 – 1.7 song types; Tukey post-hoc comparison: $Z = 2.2$, p = 0.066). The precision of the Curve Fitting technique was statistically indistinguishable from the precision of the Coupon Collector technique (Tukey post-hoc comparison: $Z = -1.5$, $p = 0.314$). Precision improved with increasing subset size for all three techniques, although it improved more dramatically for the Capture-Recapture and Coupon Collector techniques than it did for the Curve Fitting or Capture-Recapture techniques (Table 1; Fig. 3).

DISCUSSION

Our comparison of three techniques for estimating song repertoire sizes of male rufous-and-white wrens revealed that the Capture-Recapture and Coupon Collector techniques produced more accurate estimates than the Curve Fitting technique, and that the Capture-Recapture technique produced more precise estimates than the Coupon Collector and Curve Fitting techniques. Both Capture-Recapture and Coupon Collector estimates were statistically indistinguishable from actual repertoire size values based on simple enumeration, whereas

Curve Fitting estimates consistently underestimated the birds' repertoire sizes. Therefore, we recommend using either Capture-Recapture or Coupon Collector estimation techniques for generating accurate estimations of repertoire size, particularly for species with small or medium sized repertoires, heterogeneous song type selection probability, and closed-ended learning, like the rufous-and-white wren.

The Capture-Recapture technique had the best performance of the three techniques, providing estimates that were statistically indistinguishable from our enumerated calculations of repertoire size, and doing so even with a small sampling effort. The Capture-Recapture technique estimated repertoire size to within 0.02 to 0.60 song types, and provided an exceptionally accurate estimate of repertoire size with 100 or more song type switches (Figure 2). With subsets of just 25 song type switches, the repertoire size estimates derived from the Capture-Recapture technique provided truer estimates than the Curve Fitting technique, as did the Coupon Collector technique (Figure 2). Furthermore, the Capture-Recapture technique had significantly better precision than the other two estimation techniques. Although precision was similar for the three estimation techniques with small subsets of data, the Capture-Recapture technique surpassed the precision of the other two techniques at higher sampling levels (Figure 3). Our conclusions are consistent with Garamszegi et al. (2005) who provided evidence that Capture-Recapture is a compelling technique for estimating repertoire size.

The Coupon Collector technique is a newer estimation technique than the other two we explore here. In the only other published study of the Coupon Collector technique, Kershenbaum et al. (2015) found that this technique provided better estimates than the Curve Fitting and Capture-Recapture techniques. Kershenbaum et al. (2015) generated estimates for the very large repertoire sizes that exist among a population of animals, instead of the relatively small repertoire sizes found within individuals. Our study is the first to assess the Coupon Collector technique for estimating the repertoire sizes of individual animals. This technique was the only technique that we explored here to over-estimate repertoire size, which occurred only at our smallest sampling level (25 song type changes). At all higher sampling levels, the Coupon Collector technique generated accurate estimates of repertoire size. Overall, the Coupon Collector technique generated estimates with similarly high accuracy to the Capture-Recapture technique, but with low accuracy at small sample sizes, and lower precision at all sample sizes.

The Curve Fitting technique produced estimates that underestimated repertoire size by an entire song type. This underestimation likely arose due to uncommon song types present in the repertoires of many rufous-and-white wrens. The Curve Fitting equation devised by Wildenthal (1965) cannot account for uncommon song types because it is strongly affected by the rapid presentation of common song types early in the sample. As sampling effort increased, the Curve Fitting technique produced estimates with better accuracy and precision. Botero et al. (2008) also explored the Curve Fitting technique for repertoire estimation and drew similar conclusions that this technique underestimates repertoire size, especially when sampling effort is small.

Many of our estimations resulted in under-estimates of repertoire size, including the Curve Fitting technique estimations at all sampling levels, and the other two estimation techniques at some sampling levels. Estimation techniques that under-estimate repertoire size may still be well-suited for determining an individual's biologically relevant repertoire size. For example, some birds in our study had song types that were only detected after thousands of songs and hundreds of song type switches had already been recorded. Additionally, some song types were very rare, and made up less than 0.1% of a bird's song production, occurring a few times across multiple field seasons. Songs that are sung so infrequently that they require extensive sampling to detect may have little impact on the bird's life history (Derrickson, 1987). For example, in sedge warblers, repertoire size affects mate attraction (Buchanan and Catchpole, 1997), so if females do not take the time to listen for rare song types, then rare song types will have little to no impact on mate choice. Additionally, individuals may modify their song type selection based on social contexts, and this could lead to a further decrease in an individual's effective repertoire size. For example, Trillo and Vehrencamp (2005) found that banded wrens modify their repertoire use in the presence of females to increase the production of song types with specific acoustic characteristics and to song type match with neighbouring males. Similarly, Hennin, *et al*. (2009) found that male rufous-and-white wrens use a subset of their total repertoire when they are trying to attract a mate. Techniques that consistently under-estimate repertoire size, as we have revealed for the Curve Fitting technique here, may offer realistic estimates of how other birds assess an individual's repertoire size by ignoring rare song types or song types that are not used in specific social contexts.

Overall, we found that the Capture-Recapture and Coupon Collector techniques provide the most accurate estimates of repertoire size, and that the Capture-Recapture technique provides the most precise estimates of repertoire size. The Curve Fitting technique did not perform as well, tending to underestimate repertoire size to a statistically significant degree at smaller sample sizes. Future research should explore the use of Capture-Recapture for estimating actual repertoire size in other species with small repertoire sizes and heterogeneous song type probabilities. Curve Fitting and Coupon Collector techniques may be useful for estimating an individual's biologically relevant repertoire size in contexts where rare song types have little to no impact.

ACKNOWLEDGEMENTS

We thank the staff of Sector Santa Rosa of the Guanacaste Conservation area for logistical support. We thank A. Kirshebaum and an anonymous reviewer for ideas that improved the manuscript. This research was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Undergraduate Summer Research Award to AJH; by an Ontario Graduate Scholarship (OGS), a Queen Elizabeth II Graduate Scholarship in Science and Technology (QEII-GSST), a Chapman Grant from the American Museum of Natural History, a Student Research Grant from the Animal Behaviour Society, and an Alexander Wetmore Research Award from the American Ornithologist Union to BAG; by an NSERC Postdoctoral Fellowship to DRW; and by an NSERC Discovery Grant, an NSERC Accelerator Grant, and two NSERC Research Tools and Instruments Grants to DJM. Further support was provided by the Canadian Foundation for Innovation, the Government of Ontario, and the University of Windsor to DJM.

LITERATURE CITED

Aweida, M. K. (1995). Repertoires, territory size, and mate attraction in western meadowlarks. The Condor 97:1080-1083.

Baillargeon, S. and L.-P. Rivest (2007). Rcapture: loglinear models for Capture-Recapture in R. Journal of Statistical Software 19:1-31.

Barker, N. K. (2008). Effective communication in tropical forests: song transmission and the singing behaviour of rufous-and-white wrens (*Thryophilus rufalbus*) (Master's thesis). University of Windsor.

Botero, C. A., A. E. Mudge, A. M. Koltz, W. M. Hochachka, and S. L. Vehrencamp (2008). How reliable are the methods for estimating repertoire size? Ethology 114:1227-1238.

Buchanan, K. L., and C. K. Catchpole (1997). Female choice in the sedge warbler *Acrocephalus schoenobaenus*: multiple cues from song and territory quality. Proceedings of the Royal Society of London B 24:521-526.

Darroch, J. N., Fienberg, S. E., Glonek, G. F. V, and B. W. Junker (1993). A three-sample multiple-recapture approach to census population estimation with heterogeneous catchability. Journal of the American Statistical Association 88:1137-1148.

Dawkins, B. (1991) Siobhan's problem: the Coupon Collector revisited. The American Statistician 45:76-82.

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

Erdös, P. and A. Rényi (1961). On a classical problem of probability theory. Magyar Tudományos Akadémia Értesítöje 6:215-220.

Feller, W. (1968). An Introduction to Probability Theory and its Applications (vol. I, 3rd ed.), New York: John Wiley.

Garamszegi, L. Z., T. J. S. Balsby, B. D. Bell, M. Borowiec, B. E. Byers, T. Draganoiu M. Eens, W. Forstmeier, P. Galeotti, D. Gil, L. Gorissen, P. Hansen, H. M Lampe, S. Leitner, J. Lontkowski, L. Nagle, E. Nemeth, R. Pinxten, J.-M. Rossi, N. Saino, A. Tanvez, R. Titus, J. Török, E. Van Duyse, and A. P. Møller (2005). Estimating the complexity of bird song by using Capture-Recapture approaches from community ecology. Behavioral Ecology and Sociobiology 57:305-317.

Garamszegi, L. Z., T. Boulinier, A. P. Møller, J. Török, G. Michl, and J. D. Nichols (2002). The estimation of size and change in composition of avian repertoires. Animal Behavior 63:623-630.

Hennin, H. L., N. K. S. Barker, D. W. Bradley, and D. J. Mennill (2009). Bachelor and paired male rufous –and-white wrens use different singing strategies. Behavioral Ecology and Sociobiology 64:151-159.

Hothorn, T., Bretz, F. and P. Westfall (2008). Simultaneous inference in general parametric models. Biometrical Journal 50:346-363.

Kershenbaum, A., T. M. Freeberg, and D. E. Gammon (2015). Estimating vocal repertoire size is like collecting coupons: A theoretical framework with heterogeneity in signal abundance. Journal of Theoretical Biology 373:1-11.

Kroodsma, D.E. (1982). Song repertoires: problems in their definition and use. In: Kroodsma DE, Miller EH, editors. Acoustic Communication in Birds: Song Learning and its Consequences. Academic Press; New York, New York, USA.

Lemon, R. E., and C. Chatfield (1973). Organization of song of rose-breasted grosbeaks. Animal Behaviour 21: 28-44.

Linhart, P., H. Slabbekoorn, and R. Fuchs (2012). The communicative significance of song frequency and song length in territorial chiffchaffs. Behavioral Ecolgy 23:1338-1347.

Manser, M. B. (2013). Semantic communication in vervet monkey and other animals. Animal Behavior 86:491-496.

Mennill, D. J. (2014). Variation in the vocal behavior of common loons (*Gavia immer*): insights from landscape-level recordings. Waterbirds 37:26-36.

Mennill, D.J., M. Battiston, D. R. Wilson, J. R. Foote, and S. M. Doucet (2012). Field test of an affordable, portable, wireless microphone array for spatial monitoring of animal ecology and behaviour. Methods in Ecology and Evolution 3:704-712.

Mennill, D. J., J. M. Burt, K. M. Fristrup, and S. L. Vehrencamp (2006). Accuracy of an acoustic location system for monitoring the position of duetting tropical songbirds. Journal of the Acoustical Society of America 119:2832-2839.

Mennill, D. J. and S. L. Vehrencamp (2005). Sex differences in the singing and duetting behavior of neotropical rufous-and-white wrens (*Thryothorus rufalbus*). Auk 122:175-186.

Mennill, D. J. and S. L. Vehrencamp (2008). Context-dependent functions of avian duets revealed through microphone array recordings and multi-speaker playback. Current Biology 18:1314-1319.

Osmun, A. E., and D. J. Mennill (2011). Acoustic monitoring reveals congruent patterns of territorial singing behaviour in male and female tropical wrens. Ethology 117:385-394.

Peters, S., W. A. Searcy, Beecher M. D., and S. Nowicki (2000). Geographic variation in the organization of song sparrow repertoires. Auk 117:936-942.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. and R Core Team (2015). nlme: linear and nonlinear mixed effects models. R package version 3.1-122.

R Core Team (2014). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reid, J. M., P. Arcese, A. L. E. V. Cassidy, S. M. Hiebert, J. N. M. Smith, P. K. Stoddard, A. B. Marr, and L. F. Keller (2004). Song repertoire size predicts initial mating success in male song sparrows, *Melospiza melodia*. Animal Behavior 68:1055-1063

Rivest, L.-P. and S. Baillargeon (2014). Rcapture: loglinear models for Capture-Recapture experiments. R package version 1.4-2.

Sewall, K. B., J. A. Soha, S. Peters, and S. Nowicki (2013). Potential trade-off between vocal ornamentation and spatial ability in a songbird. Biology Letters 9.

Slabbekoorn, H. (2013). Songs of the city: noise-dependent spectral plasticity in the acoustic phenotype of urban birds. Animal Behavior 85:1089-1099.

Sosa-Lopez, J.R. and D. J. Mennill (2014a). The vocal behaviour of the Brown-throated Wren (*Troglodytes brunneicollis*): Song structure, repertoires, sharing, syntax, and diel variation. Journal of Ornithology 155:435-446.

Sosa-Lopez, J.R. and D. J. Mennill (2014b). Vocal behaviour of the island-endemic Cozumel Wren (*Troglodytes aedon beani*): Song structure, repertoires, and song sharing. Journal of Ornithology 155:337-346.

Trillo, P. A. and S. L. Vehrencamp (2005). Song types and their structural features are associated with specific contexts in the banded wren. Animal Behavior 70:921-935.

Valderrama, S., Parra, J., Davila, N. and D.J. Mennill (2008). The vocal behavior of the critically-endangered Niceforo's Wren (*Thryothorus nicefori*). Auk 125:395-401.

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

Zipf, G. K. (1949) Human behavior and the principle of least effort. Addison-Wsley Press, Cambridge, MA.

Figures

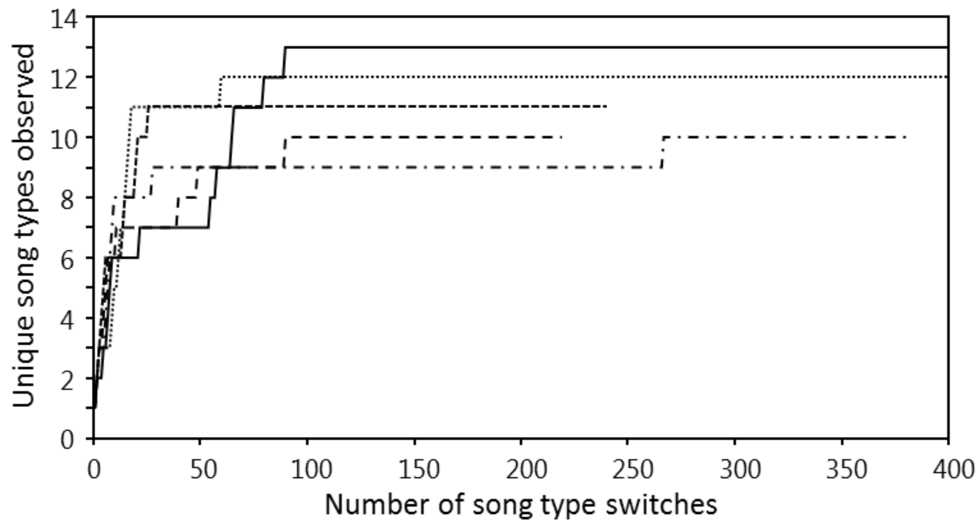

Figure 1. Simple enumeration data showing repertoire size estimates for five example male rufous-and-white wrens. Sampling effort (number of song type switches recorded) is on the x-axis and number of unique song types detected is on the y-axis. The large plateaus in the graph, where the number of unique song types does not increase despite large increases in sampling effort, suggest that a bird's repertoire has been sampled in its entirety. Note that in rare cases, such as the lowest curve, unique songs are detected even after extensive sampling.

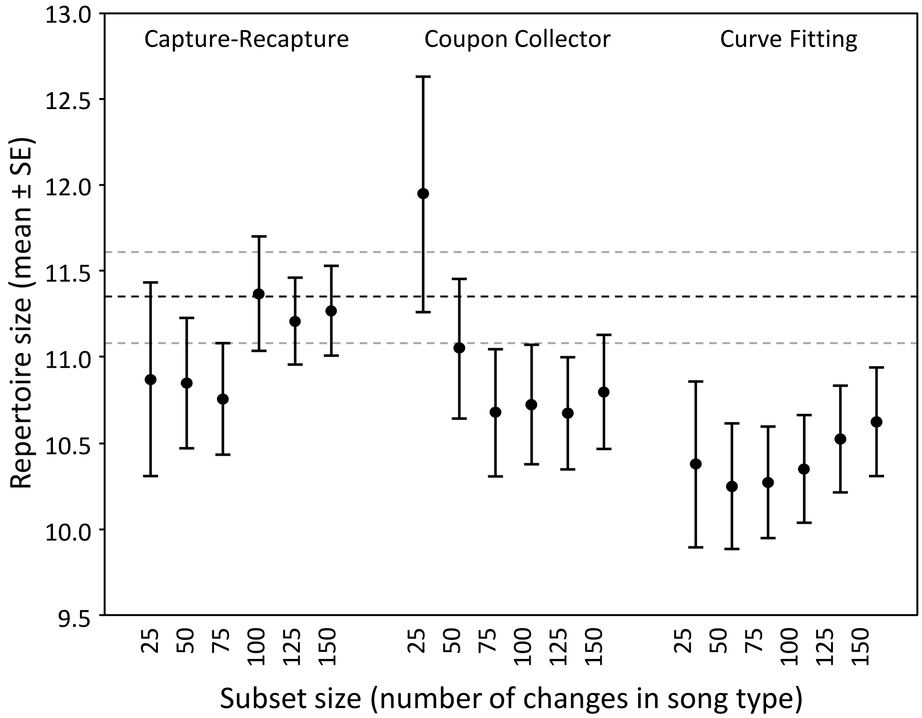

Figure 2. Estimated repertoire sizes from three different estimation techniques (Capture-recapture, Coupon Collector, and Curve Fitting techniques) for 40 male rufous-and-white wrens. For each of the three estimation techniques, the error bars show estimated repertoire sizes for each of six subset sizes: 25 (left), 50, 75, 100, 125, and 150 (right) song type switches. True repertoire sizes were measured through simple enumeration and are depicted by the hatched lines (mean = black hatched line; mean ± SE = gray hatched lines). Accuracy is defined as the average difference between the repertoire size estimated with a particular technique and subset size and the true repertoire size determined through simple enumeration. Smaller differences indicate better accuracy.

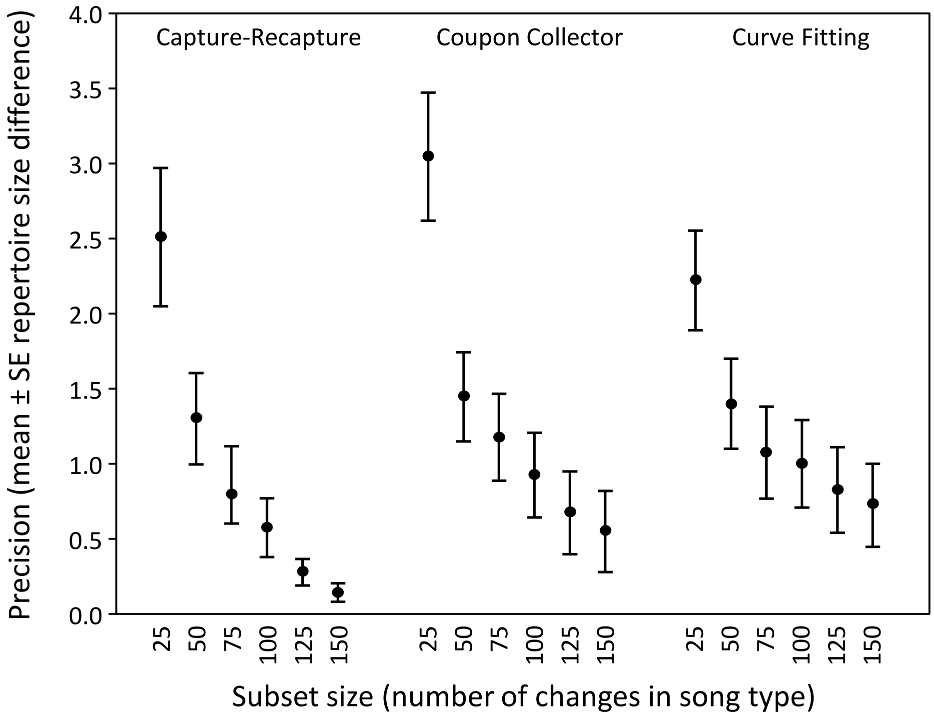

Figure 3. Effects of estimation technique and subset size on the precision of repertoire size estimates for 40 male rufous-and-white wrens. For each estimation technique, we show the precision of repertoire size estimates derived from subsets of 25 (left), 50, 75, 100, 125, and 150 (right) song type switches. Precision is defined as the average absolute difference between true repertoire size, as determined through simple enumeration, and the repertoire size estimated with a given technique and subset size; smaller absolute differences indicate better precision.

Table 1. Model coefficients from the analyses of the accuracy and precision of repertoire size estimates.

Dependent Variable	Parameter	Model Coefficient	SE
Accuracy	Capture-Recapture[1]	-0.298	0.277
	Curve Fitting[1]	-0.950	0.277
	Coupon Collector[1]	-0.213	0.277
	Subset size[2]	0.004	0.002
	Curve fitting x subset size[3]	-0.002	0.003
	Coupon collector x subset size[3]	-0.017	0.003
Precision	Capture-Recapture[1]	0.946	0.240
	Curve Fitting[1]	1.208	0.240
	Coupon Collector[1]	1.379	0.240
	Subset size[2]	-0.017	0.002
	Curve fitting x subset size[3]	0.007	0.003
	Coupon collector x subset size[3]	-0.002	0.003

[1]Model coefficients for the three estimation techniques indicate the average accuracy or precision of the technique (in song types), relative to zero, when all other variables are held constant.

[2]Model coefficients for subset size indicates how much the dependent variable changes (in terms of song types) with each 1-unit change in subset size, when averaged across all techniques.

[3]Model coefficients for the interaction terms indicate how much the dependent variable changes with each 1-unit change in subset size for a particular technique, relative to the amount of change observed for a 1-unit change in subset size for the reference category (i.e. Capture-Recapture).