CHICAGO JOURNALS

American Society of Naturalists

The University of Chicago

Uneven Sampling and the Analysis of Vocal Performance Constraints

David R. Wilson,[1,*] Pierre-Paul Bitton,[1] Jeffrey Podos,[2] and Daniel J. Mennill[1]

1. Department of Biological Sciences, University of Windsor, Windsor, Ontario N9B 3P4, Canada;   2. Department of Biology, University of Massachusetts, Amherst, Massachusetts 01003

ABSTRACT: Studies of trilled vocalizations provide a premiere illustration of how performance constraints shape the evolution of mating displays. In trill production, vocal tract mechanics impose a trade-off between syllable repetition rate and frequency bandwidth, with the trade-off most pronounced at higher values of both parameters. Available evidence suggests that trills that simultaneously maximize both traits are more threatening to males or more attractive to females, consistent with a history of sexual selection favoring high-performance trills. Here, we identify a sampling limitation that confounds the detection and description of performance trade-offs. We reassess 70 data sets (from 26 published studies) and show that sampling limitations afflict 63 of these to some degree. Traditional upper-bound regression, which does not control for sampling limitations, detects performance trade-offs in 33 data sets; yet when sampling limitations are controlled, performance trade-offs are detected in only 15. Sampling limitations therefore confound more than half of all performance trade-offs reported using the traditional method. An alternative method that circumvents this sampling limitation, which we explore here, is quantile regression. Our goal is not to question the presence of mechanical trade-offs on trill production but rather to reconsider how these trade-offs can be detected and characterized from acoustic data.

Keywords: birdsong, correlated evolution, quantile regression, trade-off, trill, upper-bound regression.

Understanding the evolution of organismal traits requires attention to proximate mechanisms that shape phenotypic variation, since phenotypic variation constitutes the raw material on which selection can act (Darwin 1859). Many studies on organismal evolution seek to understand how phenotypic variation is generated and maintained, yet it is equally important to understand how the expression of phenotypic variation can be constrained (Futuyma 2009). Constraints on trait expression often manifest as trade-offs (Arnold 1992). This is because when multiple traits

* Corresponding author; e-mail: drwilson76@gmail.com.

are functionally coupled, either because of mechanical relationships, dependency on common underlying resources, or common developmental pathways, the expression of one trait should limit the expression of the other (Roff and Fairbairn 2007). A classic example is the trade-off between gametic and somatic growth (MacLeod and MacLeod 2009; see also Stearns 1992).

In animals, the expression of behavioral traits can be circumscribed by performance constraints, which are limits on animals' abilities to execute the motor patterns that constitute behavior. As an example, differential investment in fast, fatigue-sensitive muscles versus slow, fatigue-insensitive muscles can create a performance constraint on an animal's capacity for both speed and endurance (van Damme et al. 2002). Performance constraints should be particularly evident for mating displays subject to sexual selection, which we expect should evolve to high levels of performance to ensure their reliability (Searcy and Nowicki 2005; Byers et al. 2010). If some aspect of an individual's quality limits its ability to perform a mating display well, then the individual's maximum performance of that display should provide an accurate indication of their quality to rivals and prospective mates. Studies of trilled vocalizations in vertebrates provide some of the strongest evidence that performance constraints affect the evolution of animal displays (reviewed by Podos et al. 2004*a*, 2009). In brief, trills are structurally simple vocalizations that consist of rapid repetitions of frequency-modulated syllables. Descriptive analysis of the trills of songbirds across the family Emberizidae revealed a trade-off between the range of frequencies contained in each syllable (i.e., frequency bandwidth) and the rate at which syllables are uttered in succession (Podos 1997; fig. 1). The same trade-off has since been shown in a wide range of taxonomic groups (table 1) and has been attributed to vocal mechanics. Specifically, broad modulations of the vocal apparatus, especially the vocal tract (trachea, mouth, beak, and associated structures), are required to produce trills
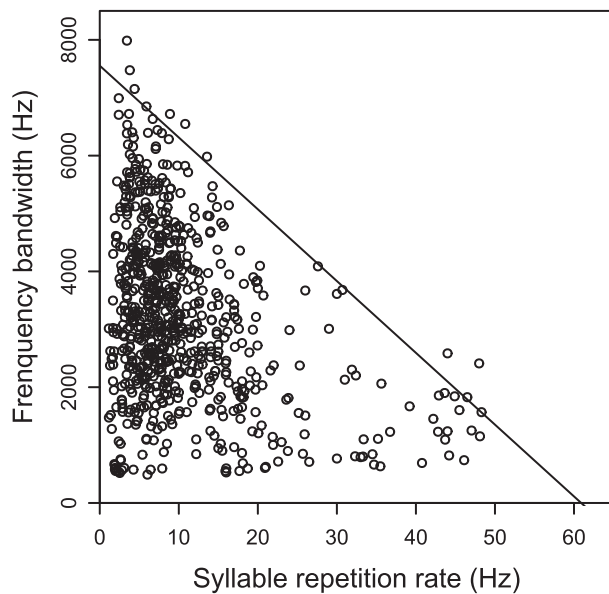
Figure 1: A performance trade-off on the trills of 34 songbird species in the family Emberizidae. The trade-off between syllable repetition rate and frequency bandwidth in trills results in a triangular distribution, with both narrow and wide frequency bandwidths possible at low syllable repetition rates yet only narrow frequency bandwidths possible at high syllable repetition rates. The inverse regression between syllable repetition rate and frequency bandwidth values in the upper tail of the frequency bandwidth distribution is a manifestation of the expected performance constraint on trill production. Data are from Podos (1997), and the regression line is calculated using the traditional upper-bound regression technique (see text for more details).

with wide frequency bandwidths that maintain the tonal purity of vocalizations across varying frequencies (West-neat et al. 1993; Hoese et al. 2000; Beckers et al. 2003; Podos et al. 2004*b*; Riede et al. 2006; Pasch et al. 2011). Because it is physically challenging to execute large vocal tract modulations both quickly and repetitively, the magnitude of frequency modulation necessarily limits the maximum rate of syllable repetition that an animal can produce.

The biomechanical trade-off in trill production yields a triangular distribution on the bivariate plot of syllable repetition rate versus frequency bandwidth, with both narrow and wide frequency bandwidths possible at low syllable repetition rates, yet only narrow frequency bandwidths possible at high syllable repetition rates (fig. 1). An inverse relationship between syllable repetition rate and frequency bandwidth values found in the upper tail of the frequency bandwidth distribution (i.e., the triangle's hypotenuse; Podos 1997) is an expected manifestation of the performance constraint; trills with relatively high syllable repetition rate and relatively wide frequency bandwidth are

thought to be more difficult to produce because they require greater neuromuscular performance from the vocal apparatus (Podos 2001). Accumulating evidence from many vertebrate taxa suggests that trill performance can reflect signalers' condition, life-history stage, and genetic quality (Janicke et al. 2008; Araya-Ajoy et al. 2009; Ballentine 2009; Juola and Searcy 2011; Vehrencamp et al. 2013; but see Cardoso et al. 2012), and that receivers indeed attend to variations in trill performance and modulate their responses accordingly (Draganoiu et al. 2002; Ballentine et al. 2004; Christensen et al. 2006; Illes et al. 2006; Cramer and Price 2007; Schmidt et al. 2008; deKort et al. 2009; Caro et al. 2010; DuBois et al. 2011; Pasch et al. 2011; Moseley et al. 2013). It is important to note, however, that many individuals and taxonomic groups produce trills that are well below the expected performance limit (note the prevalence of trills in the lower left corner of fig. 1), suggesting that selection for high-performance trills may not be universal.

Studies of trill performance can provide insight into how animal signals are shaped not just by sexual selection, but also by natural selection. Given the central role of vocal tract modulations in trill production, if components of the vocal apparatus are also involved in nonsignaling functions, then natural selection could affect vocal modulation and performance indirectly (Nowicki et al. 1992; Podos 1997). In Darwin's finches, for example, beak morphology is known to evolve precisely in response to interspecific competition and local foraging conditions, and these species have evolved remarkable diversity in beak form and function (e.g., Schluter and Grant 1984; Gibbs and Grant 1987; Herrel et al. 2005). Producing trills with high vocal performance requires birds to open and close their beaks widely and quickly, yet this is physically incompatible with a large beak specialized for high-force application (Hoese et al. 2000; Podos and Nowicki 2004). Consequently, the divergence of beak morphology via natural selection has indirectly affected the performance of mating signals in these species and, as a result, provided an evolutionary mechanism for promoting speciation, preventing hybridization, and maintaining species diversity (Podos 2001, 2010; Podos and Nowicki 2004; Herrel et al. 2009). Evidence for parallel beak-related trade-offs on song evolution are now available for a number of other avian taxa (e.g., Seddon 2005; Ballentine 2006; Huber and Podos 2006; Derryberry 2009; Sockman 2009; Derryberry et al. 2012; but see Slabbekoorn and Smith 2000).

Our ability to detect and characterize the expected trade-off between trill rate and frequency bandwidth depends, in part, on the analytical methods used. Following Blackburn et al. (1992), Podos (1997) and many subsequent studies of performance constraints in trills have estimated performance trade-offs using upper-bound regression. The

Table 1: Assessment of performance trade-offs in 70 published data sets (55 within-species and 15 between-species) that contain measurements of syllable repetition rate and frequency bandwidth

Taxon	Common name	Sample size	No. bins	Sampling limitation	Slopes from four regression analyses				Consensus? Is there a performance trade-off?	Reference
					Traditional upper bound	Equal samples per bin	Quantile (90%)	Sample size as covariate		
Within-species data sets:										
Agelaius phoeniceus	Red-winged blackbird	113	6	−6	−140	−107	7	−55	No	Cramer and Price 2007
Aimophila ruficeps	Rufous-crowned sparrow	28	3	−11	−173	−229	−161	−868	No	Podos 1997
Ammodramus humeralis	Grassland sparrow	5	3	1	−3	−39	−55	−55	No	Podos 1997
Amphispiza bilineata	Black-throated sparrow	15	5	−1	−140	−46	−2	−71	No	Podos 1997
Amphispiza quinquestriata	Five-striped sparrow	12	2	−2	−414	−414	−399	0	No	Podos 1997
Artemisiospiza belli	Sage sparrow	7	2	3	455	−108	−35	0	No	Podos 1997
Calamospiza melanocorys	Lark bunting	45	5	−2	−123*	−256**	−109	−28	No	Podos 1997
Camarhynchus parvulus	Small tree finch	33	10	−1**	−93	−9	27	425	No	Christensen et al. 2006
Cardinalis cardinalis	Northern cardinal	19	3	−8	−698	660	223	−299	No	Podos 1997
C. cardinalis	Northern cardinal	66	8	−1	−1,024***	−1,163*	−804*	−1,005**	Yes	Narango 2012
Cardinalis sinuatus	Pyrrhuloxia	4	2	−2	905	753	753	0	No	Podos 1997
Catharacta antarctica	Brown skua	262	13	−3*	−4,308***	−2,384**	−2,237**	−3,132**	Yes	Janicke et al. 2008
Chondestes grammacus	Lark sparrow	88	8	−3*	−131***	−137	−60	−134*	No	Podos 1997
Emberiza calandra	Corn bunting	4	4	0	−82*	−82*	−93	−82	No	Podos 1997
Emberiza schoeniclus	Reed bunting	5	3	0	−18	−18	−67	−13	No	Podos 1997
Fregata minor	Great frigatebird	103	12	0	−17*	−14	−10**	−17***	Yes	Juola and Searcy 2011
Geospiza fortis	Medium ground finch	26	2	−2	−193	−193	−161	0	No	Podos 1997
G. fortis	Medium ground finch	37	9	−1**	−553	−211	−191	−202	No	Huber and Podos 2006
Geospiza scandens	Common cactus finch	5	2	−3	−161	−460	−161	0	No	Podos 1997
Junco hyemalis	Dark-eyed junco	17	4	−3*	−17	−10	6	−186	No	Podos 1997
J. hyemalis	Dark-eyed junco	188	15	−2***	−162***	−99**	−65	−212**	Yes	Cardoso et al. 2007
Junco phaeonotus	Yellow-eyed junco	9	5	0	−71	−43	−58	−54	No	Podos 1997
Lonchura oryzibora	Java sparrow (2 data points excluded)	69	12	−1	−408**	−198	−366	−163	No	Kagawa and Soma 2013
L. oryzibora	Java sparrow (all data included)	71	14	−1**	−36	56	(85**)	275	No	Kagawa and Soma 2013
Luscinia megarhynchos	Common nightingale	21	9	0	1,448	1,709	2,260	1,446	No	Schmidt et al. 2008
Melospiza georgiana	Swamp sparrow	33	4	−2	−183	−10	−17	−88	No	Podos 1997
M. georgiana	Swamp sparrow	58	10	0	−235	−156	−115	−125*	No	Ballentine 2006

216

M. georgiana	Swamp sparrow	88	9	−2*	−117*	−34	−61	−85	No	Liu et al. 2008
M. georgiana	Swamp sparrow	280	10	−6***	−100*	−1	26	49	No	Ballentine et al. 2004
Melospiza lincolnii	Lincoln's sparrow	44	7	−2	−55	35	−33	−13	No	Podos 1997
M. lincolnii	Lincoln's sparrow	1,680	12	−21*	−64**	−38	−46***	−58*	Yes	Sockman 2009
Melospiza melodia	Song sparrow	54	6	−2	−112**	−7	−90	−82	No	Podos 1997
M. melodia	Song sparrow (buzzes)	198	14	−3**	−18***	−14***	−10*	−11**	Yes	Maddison et al. 2012
M. melodia	Song sparrow (trills)	176	9	−6**	−262***	−51	34	−60	No	Maddison et al. 2012
Melozone aberti	Abert's towhee	4	2	2	1,089	1,135	1,089	0	No	Podos 1997
Peucaea aestivalis	Bachman's sparrow	49	6	−2	−103*	−115	−78	−98	No	Podos 1997
Peucaea botteri	Botteri's sparrow	12	4	0	1	−24	−21	−23	No	Podos 1997
Peucaea carpalis	Rufous-winged sparrow	19	5	−1	−30	259	112	269	No	Podos 1997
Peucaea cassinii	Cassin's sparrow	24	8	0	−20	−19	−28	18	No	Podos 1997
Peucaea sumichrasti	Cinnamon-tailed sparrow	6	3	0	−425	−425	−419	−425	No	Podos 1997
Pipilo erythrophthalmus	Eastern towhee	48	8	−1	−29	−37	−45	6	No	Podos 1997
Pooecetes gramineus	Vesper sparrow	66	4	−3	−97*	−17	−100	−65	No	Podos 1997
Rhynchospiza strigiceps	Stripe-capped sparrow	5	4	0	−55	−75	−62	42	No	Podos 1997
Scotinomys teguina	Alston's brown mouse	101	10	−1	−1,227**	−557*	−700**	−569*	Yes	Pasch et al. 2011
Serinus canaria	Island canary	40	6	−2	−61*	−60*	−48	−42	No	Podos 1997
S. canaria	Island canary	547	9	−16***	−160***	−158*	−100***	−185*	Yes	Draganoiu et al. 2002
S. canaria	Island canary	59	10	0	−27	−48	−88	23	No	Suthers et al. 2012
Setophaga petechia	Yellow warbler	228	10	−5*	−231*	−2	7	−41	No	Beebee 2004
Spizella passerina	Chipping sparrow	9	4	−1	−207	−202	−269	−124	No	Podos 1997
Thryophilus pleurostictus	Banded wren (bins combined)	130	14	−2***	−169***	−137***	−139***	−143	Yes	Illes et al. 2006
T. pleurostictus	Banded wren (bins not combined)	130	15	−1*	−105	−108	−139**	16	No	Illes et al. 2006
Troglodytes aedon	House wren (bins combined)	3,298	10	−110*	−169***	−150***	−95***	−168***	Yes	Cramer 2013
T. aedon	House wren (bins not combined)	3,298	17	−24	−92*	−138***	−95***	−69**	Yes	Cramer 2013
Zonotrichia leucophrys	White-crowned sparrow	16	7	0	−43*	−29	−37	−29	No	Podos 1997
Z. leucophrys	White-crowned sparrow	375	10	−13	−59**	−72**	−55**	−30*	Yes	Derryberry 2009
Between-species data sets:										
Aimophila	American sparrows	146	9	−5	−145***	−124**	−112	−47	No	Podos 1997
Amphispiza	American sparrows	34	5	−2	−347**	−331***	−332	−319*	Yes	Podos 1997
Dendrocolaptinae	Woodcreepers (clade 1)	89	11	−1	−53	−131	−106	−21	No	Derryberry et al. 2012
Dendrocolaptinae	Woodcreepers (clade 2)	136	12	−2	−56	67	(96**)	−30	No	Derryberry et al. 2012
Dendrocolaptinae	Woodcreepers (clade 3)	202	10	−4*	−51	10	28	−117	No	Derryberry et al. 2012
Dendrocolaptinae	Woodcreepers (clade 4)	123	15	−1**	−23	−4	−28	3	No	Derryberry et al. 2012
Dendrocolaptinae	Woodcreepers (all data)	550	15	−7***	−101**	−15	4	14	No	Derryberry et al. 2012
Emberizidae	Buntings and New World sparrows	740	10	−27**	−124***	−167**	−89***	−63	Yes	Podos 1997

217

Table 1 (*Continued*)

Taxon	Common name	Sample size	No. bins	Sampling limitation	Traditional upper bound	Equal samples per bin	Quantile (90%)	Sample size as covariate	Consensus? Is there a performance trade-off?	Reference
Emberizidae	Buntings and New World sparrows	74	9	−3**	−223*	−66	−53	97	No	Podos 2001
Emberizidae	Buntings and New World sparrows	153	10	−3***	−149	190	(158*)	161	No	Grant and Grant 2002
Geospiza	Darwin's finches	38	2	−12	−193	−193	23	0	No	Podos 1997
Melospiza	Song sparrows	131	8	−5*	−117*	−112	−77	−26	No	Podos 1997
Parulidae	Wood warblers (1 species excluded)	2,620	9	−128**	−239***	−144**	−69*	−134	Yes	Cardoso and Hu 2011
Parulidae	Wood warblers (all species included)	2,662	10	−108**	−131**	−256	−4	121	No	Cardoso and Hu 2011
Pipilo	Towhees	57	8	−2	−37	−54	−47	16	No	Podos 1997

Note: For each data set, we list the taxonomic group studied, sample size, the number of bins used in the upper-bound regression analyses, and the slope of the sampling limitation regression. We also show the slopes generated by four regression methods, as described in the text: (1) the traditional upper-bound regression (sensu Blackburn et al. 1992), (2) the equal samples per bin method, (3) the 90% quantile regression method, and (4) the sample size as covariate method. Brackets indicate that a slope is significantly positive; otherwise, all significant slopes were negative. We conclude that performance trade-offs are present when at least two of the final three regression slopes (consensus) were negative and statistically significant. Empty cells indicate that a particular method could not be conducted because of insufficient sample size.

* $P \leq .05$

** $P \leq .01$

*** $P \leq .001$

traditional upper-bound regression method plots the frequency bandwidth of trills versus syllable repetition rate, divides the syllable repetition rate axis into equal-width bins of arbitrary size, and regresses the maximum bandwidth value from each bin against its corresponding syllable repetition rate (detailed description in fig. 2). A significant negative upper-bound regression has been

regarded as evidence of a performance constraint (Podos 1997).

This article is motivated by our observation that maximum bandwidth values in each bin of upper-bound regression analyses might be influenced not just by the expected mechanical trade-off but also by variation in the number of data points contained within each bin. Statis-
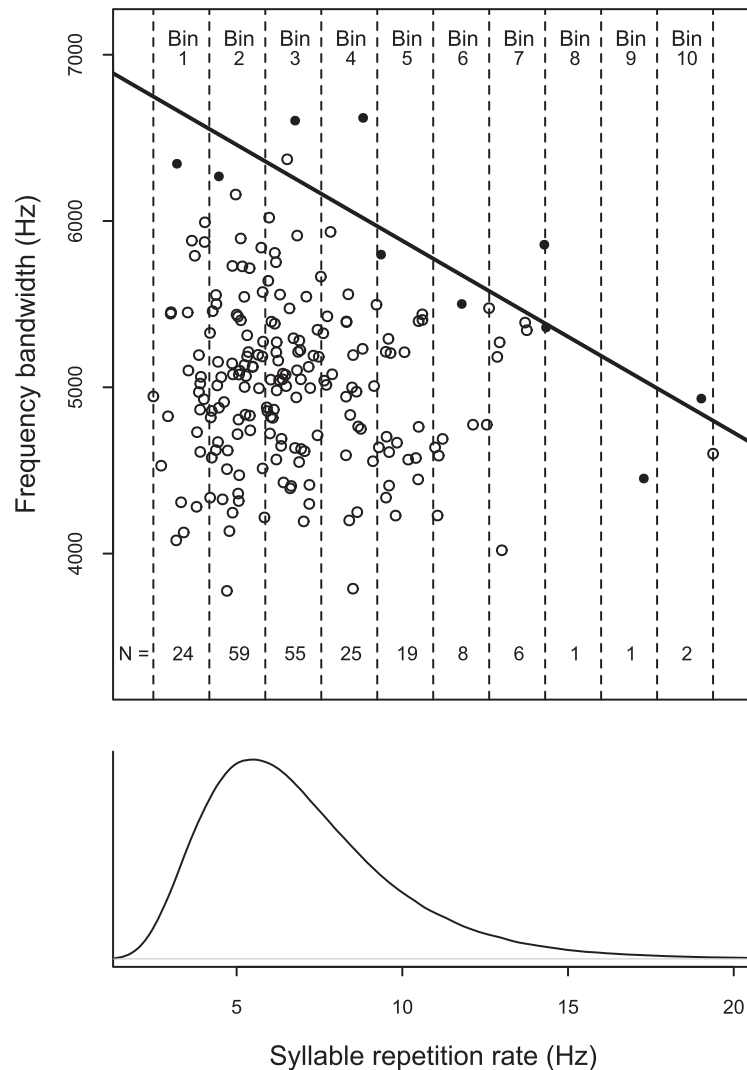
Figure 2: Upper-bound regression is the standard method for measuring performance trade-offs (Blackburn et al. 1992). In this simulation, we plotted frequency bandwidth by syllable repetition rate for 200 hypothetical trills with randomly generated x and y values. More specifically, bandwidth values were drawn at random from a normal distribution (mean = 5,000 Hz, SD = 500 Hz), and syllable repetition rates were drawn at random from the skewed distribution shown beneath the abscissa (mean = 7 Hz, SD = 3 Hz). We binned trills according to their syllable repetition rates using 10 bins of equal width and then regressed maximum frequency bandwidth values in each bin (filled circles) by their corresponding syllable repetition rates. The resulting upper-bound regression shows a significant negative relationship and therefore implies a significant performance trade-off (slope = -115.4, y-intercept = 7,035.6, $F_{1,8} = -4.6$, $P = .002$, $R^2 = 0.73$). However, given that these data were generated randomly, the performance trade-off is an artifact of the inherent sampling limitation, which is defined by the slope of the simple linear regression between bin number and the number of data points in each bin (slope = -5.7, y-intercept = 51.0, $F_{1,8} = -3.81$, $P = .005$, and $R^2 = 0.64$).

tical theory mandates that the maximum value of a random sample drawn from any frequency distribution should increase, on average, with sample size (fig. A1; figs. A1, A2 available online; Hartley and David 1954; Sokal and Rohlf 1995). Bins containing many data points should therefore have higher maximum bandwidth values than bins containing few data points, simply as a result of sampling effort and the natural variation inherent in the distribution of bandwidth values. If bins associated with high syllable repetition rates systematically contain fewer data points than bins associated with low syllable repetition rates, as would be expected when the distribution of syllable repetition rates is positively skewed, then negative slopes from upper-bound regression analyses could be, at least in part, a sampling artifact caused by the diminishing sample sizes among the bins (figs. 2, A2). Yet no study on performance trade-offs in trills has accounted for this potentially confounding effect of uneven sampling. This limits our understanding of the prevalence with which performance constraints have shaped the evolution of animal vocal signals.

The potentially confounding effects of sampling limitations on upper-bound regression analysis are likely not limited to the trade-off between syllable repetition rate and frequency bandwidth in trilled vocalizations, though this particular trade-off is the most widely documented in the literature (table 1; Blackburn et al. 1992). For example, upper-bound regression has been used to calculate performance trade-offs among other trait pairs that might be linked mechanically, such as maximum frequency and duration in the songs of ocellated antbirds (*Phaenostictus mcleannani*; Araya-Ajoy et al. 2009) and syllable repetition rate and call duration in the advertisement calls of gray tree frogs (*Hyla versicolor*; Reichert and Gerhardt 2012). Upper-bound regression has also been used outside of the context of performance trade-offs to evaluate duetting behavior (e.g., Logue et al. 2008), plumage divergence (e.g., Price and Whalen 2009), social networks (e.g., Roberts et al. 2009), and the relationship between animal abundance and body size (e.g., Blackburn et al. 1992; reviewed in Scharf et al. 1998).

In this study, we focus on the effects of uneven sampling on the detection and quantification of frequency bandwidth versus syllable repetition rate trade-offs in trills. Our objective is to better understand the prevalence with which this performance trade-off is expressed and to explore and recommend unbiased analytical techniques for detecting and quantifying performance trade-offs in this rapidly growing field of research. We review the literature and identify all data sets containing frequency bandwidth and syllable repetition rate data. For each data set, we measure the degree to which data are distributed asymmetrically among the bins used in a traditional upper-bound re-

gression; we refer to this as the "sampling limitation" of each data set. We then compare the measurement of the performance trade-off derived from traditional upper-bound regression analysis to those derived from three alternative analytical techniques that alleviate or control for sampling limitations. Taken together, these analyses help us produce more valid estimates of when performance trade-offs are being expressed. More generally, our study reveals the potentially confounding effects of sampling limitations on upper-bound regression analysis and highlights alternative methods for analyzing trade-offs and other biological phenomena that might be affected by uneven sampling.

Material and Methods

Data Collection

We used ISI Web of Science and Google Scholar to identify all published papers citing Podos's (1997) study of performance constraints (searches conducted in August 2013). We identified papers containing data sets with information on the syllable repetition rate and frequency bandwidth of trills. To be comprehensive, we included all published data sets, whether based on single or multiple species and irrespective of whether performance trade-offs had been assessed as part of the original study. We included multiple data sets from the same species if they were derived from different individuals and different published articles. We included multiple data sets derived from the same individuals if those data sets had been used in the original article to assess trade-offs at different taxonomic levels. For example, we included the species-level, genus-level, and family-level data sets presented in Podos (1997), even though the same individuals were sometimes represented in multiple data sets. When a data set was subdivided by population, subspecies, sex, morph, or recording year, we conducted analyses only on the pooled species-level data set. For one data set (Maddison et al. 2012), trills had been subdivided into slow "trills" and fast "buzzes," and each had been analyzed separately; we also analyzed these data separately because trills and buzzes are thought to involve different respiratory mechanisms and thus be subject to different performance trade-offs (Hartley and Suthers 1989; Podos 1997). Finally, there was one single-species data set and one multispecies data set where outliers were removed before analysis (Cardoso and Hu 2011; Kagawa and Soma 2013). To minimize assumptions about outliers, we analyzed these data sets twice, excluding and then including the outlying data points.

Many data sets identified in our literature search were presented not as raw numerical data but as scatterplots. To quantify these data, we converted scatterplots to digital

image files using Preview software (ver. 4.2; Apple, Cupertino, CA). We opened image files in ImageJ software (ver. 1.47d; W. Rasband, National Institutes of Health), calibrated the software using the scatterplot axes, and measured values of syllable repetition rate and frequency bandwidth for every data point. Analyses confirmed that our data extraction approach was accurate (see appendix, available online). For several scatterplots, data points were too clumped to be resolved with this approach, and we instead requested the data from the authors. In total, we obtained 70 data sets for our analysis (68 data sets were unique, but two were analyzed twice for a total of 70; see "Analysis" below), including 19 that we extracted from scatterplots, six that were published as raw numerical data, and 45 that were provided to us by the authors. To the best of our knowledge, our data set includes all published data sets with information on the frequency bandwidth and syllable repetition rate of trilled vocalizations (data for each trill are available in the appendix table in Dryad: http://dx.doi.org/10.5061/dryad.36847; Wilson et al. 2013). Original studies varied in whether they coded frequency bandwidth in hertz or kilohertz; here we coded all data sets in hertz.

In addition to values of frequency bandwidth and syllable repetition rate, we noted whether trills were recorded in the laboratory or the field (see appendix table in Dryad: http://dx.doi.org/10.5061/dryad.36847). High-frequency sound attenuates more quickly than low-frequency sound, so it is possible that the more variable recording distances associated with field recordings affected measures of maximum frequency and caused the distributions of frequency bandwidth to become positively skewed (Naguib et al. 2008). Only four data sets were derived from the laboratory, so it was not possible to compare recording environments statistically. However, the average skew of frequency bandwidth distributions (as estimated by g_1, a conventional measure of skew; Sokal and Rohlf 1995) was greater for data sets derived from the laboratory ($g_1 = 0.50 \pm 0.49$, mean $\pm$ SD) than for data sets derived from the field ($g_1 = 0.32 \pm 0.58$), suggesting that recording in the field did not increase the skew of frequency bandwidth distributions. In addition to the recording environment, we noted whether frequency bandwidth was measured relative to a standardized amplitude or energy threshold. Previous research shows that not measuring frequency bandwidth relative to standard thresholds can exaggerate the range of frequencies reported, thereby causing the distribution of frequency bandwidth values to become positively skewed (Zollinger et al. 2012). In total, 11 of 70 data sets did not measure frequency bandwidth relative to a specified quantitative energy or amplitude threshold. As predicted, the average skew of these distributions was greater ($g_1 = 1.01 \pm 0.85$, mean $\pm$ SD) than for data sets where

a threshold was used ($g_1 = 0.35 \pm 0.59$; independent samples t-test: $t_{68} = 3.21$, $P = .002$; data used in all meta-analyses are available in the appendix table in Dryad: http://dx.doi.org/10.5061/dryad.36847).

Analysis

For each data set, we applied the traditional upper-bound regression analysis (sensu Blackburn et al. 1992), which aims to detect the presence of performance trade-offs between syllable repetition rate and the frequency bandwidth values found in the upper tail of the frequency bandwidth distribution (Podos 1997). We applied the binning schemes used in the original articles when that information was available. For two data sets (Illes et al. 2006; Cramer 2013), the authors had combined certain bins because of a paucity of data points; however, since we are specifically testing for the effect of uneven sample sizes among bins, we analyzed each of these two data sets twice, once with the bins combined and once without. When the upper-bound regression had not been calculated in the original article ($N = 15$ of the 70 data sets), we divided syllable repetition rate into 10 bins of equal width along the X-axis, with the first bin beginning with the lowest value of syllable repetition rate and the last bin ending with the highest value. We selected 10 bins because this was similar to the number of bins used in previous studies, and generally ensured adequate statistical power in the upper-bound regression calculations.

We quantified each data set's "sampling limitation" as the slope of a simple linear regression between bin order number (independent variable) and the number of data points in each bin (dependent variable). Bin number 1 was always associated with the lowest syllable repetition rate and the last bin with the highest syllable repetition rate. We expect the sampling limitation to become more negative as the distribution of syllable repetition rates becomes more positively skewed (Sokal and Rohlf 1995; see figs. 2, A1).

As we report below, sampling limitations were detected in the majority of data sets. We thus applied three alternative, complementary analytical approaches that allowed us to control for sampling limitations. These three alternative approaches have all been used in other contexts to calculate regressions near the upper boundaries of bivariate scatterplots (e.g., Scharf et al. 1998), but have not been used previously to analyze performance trade-offs in trilled vocalizations.

1. *Equal samples per bin method.* For our first alternative analytical approach, we calculated new bin sizes for the original data and then reran the upper-bound regression analysis. Within each data set, we varied the width of individual bins so that each bin contained the same

number of data points, thereby removing a priori the sampling limitation. Whereas the traditional upper-bound regression technique applies bins of equal width yet variable sample size, this first alternative technique applies bins of variable width yet equal sample size. We divided data points into the same number of bins as in each original analysis to ensure similar statistical power between the equal samples per bin method and the traditional upper-bound regression.

2. Quantile regression method. For our second alternative analytical approach, we estimated performance trade-offs using quantile regression (Koenker and Bassett 1978). Unlike simple linear regression, which describes how variation in an independent variable affects the mean of a dependent variable, quantile regression describes how variation in an independent variable affects other aspects of the dependent variable's distribution. It is a powerful approach because, unlike upper-bound regression, it uses all of the original data and does not require that data be assigned to arbitrarily delimited bins. Here, we used quantile regression to estimate how changes in syllable repetition rate affected the ninetieth percentile of the frequency bandwidth distribution. We could have chosen a higher percentile, but given the mechanics of quantile regression, a percentile beyond 90% begins to be less resistant to outliers and to suffer reduced precision in estimating rates of change in the upper tail of dependent variable distributions (Cade et al. 1999). We computed confidence intervals for the ninetieth percentile slope coefficient using the rank method (Gutenbrunner and Jureckova 1992).

3. Sample size as covariate method. For our third alternative analytical approach, we reran the original upper-bound regression analysis with bins of equal width, as in traditional upper-bound regression, but incorporated into the analysis the number of data points in each bin as a covariate. We $\log_{10}$ transformed the number of data points in each bin before analysis because the relationship between the expected maximum value in a random sample and sample size is logarithmic (see fig. A1). The main advantage of this method is that it uses the same values of syllable repetition rate and frequency bandwidth that had been used in the traditional upper-bound regression. It also calculates the independent effects of syllable repetition rate and the sampling limitation on the observed maximum frequency bandwidth distribution.

We conducted an additional set of analyses to explore how the number of bins affects the slopes derived from traditional upper-bound regression, the equal samples per bin method, and the sample size as covariate method. For each method, we applied three standard binning schemes to each of the 70 data sets, including 8-bin, 10-bin, and 12-bin schemes. We report the slopes of each method and each binning scheme in the appendix table

in Dryad: http://dx.doi.org/10.5061/dryad.36847. For each method, we compare the slopes among the three binning schemes using repeated-measures ANOVA. Given that the three binning schemes produce similar results within each method (see results below), we consider only those results derived from the original binning schemes (i.e., the same binning schemes that were used in the original articles, or 10 bins when that information was unavailable) when assessing the presence and strength of potential performance trade-offs.

All analyses were conducted using R (R Development Core Team 2013). The 90% quantile regression analyses were conducted using the quantreg package (Koenker 2012).

Results

Analysis of Performance Trade-Offs

The past 16 years have witnessed significant interest in the study of acoustic trade-offs in trill structure. Overall, we found 70 data sets on acoustic trade-offs reporting values for syllable repetition rate and frequency bandwidth, including 55 within-species data sets and 15 between-species data sets (table 1). The data sets were derived from 26 unique published studies and represented a diverse range of animal taxa. Those that focused on a single species represented two classes, four orders, 11 families, 29 genera, and 40 species (39 avian, one mammalian). Those data sets that focused on multiple species represented six families, 48 genera, and a total of 146 avian species. Our search also revealed that interest in performance constraints has increased over time; publication year and the number of published studies showed a significant positive correlation (Pearson correlation analysis: $N = 16$ years (i.e., 1997–2012), $r = 0.60$, $P = .013$).

We detected sampling limitations (i.e., negative relationships between bin order number and the number of data points per bin; see "Material and Methods") in 63 of the 70 data sets analyzed (90%). The median slope of the regression between bin order number and the number of data points per bin was −2.0 (interquartile range: −4.5 to −0.7; range: −127.7 to 3.0). A total of 26 data sets (37% of all data sets) had sampling limitations with statistically significant negative slopes (simple linear regression: all $P \leq .05$; table 1).

Before controlling for sampling limitations, the average slope ($\pm$ SD) of performance trade-offs among the 70 data sets was −164.4 $\pm$ 612.2, as determined by the upper-bound regression method (table 1). When we removed or controlled for sampling limitations, the average slope of estimated trade-offs was −55.7 $\pm$ 448.5, as calculated using the quantile regression method. The number of data

sets that exhibited statistically significant regression slopes also decreased when we controlled for sampling limitations. Using the traditional upper-bound regression method, 33 of the 70 published data sets showed statistically significant regression slopes. When sampling limitations were controlled, the consensus of our three alternative analytical techniques showed that only 15 of those 33 data sets, and no additional data sets, exhibited statistically significant regression slopes (table 1). Each of our three alternative analytical approaches produced results similar to the overall consensus: 17 data sets retained statistically significant regression slopes with the equal samples per bin method; 14 data sets retained statistically significant regression slopes with the quantile regression method; and 14 data sets retained statistically significant regression slopes with the sample size as covariate method. We conclude that sampling limitations therefore accounted for approximately one-half of all statistically significant estimates of performance trade-offs in the published literature (i.e., 48%–58%, or 16 to 19 of 33 data sets; table 1).

It is noteworthy that the number of data sets that showed a significant performance trade-off after controlling for sampling limitations (i.e., 15 of 70, based on a consensus of the three alternative analytical techniques; table 1) still exceeded the number that could be attributed to Type I error (probability of Type I error in each data set $= 0.05$; binomial test: $P < .001$). Similarly, the number of data sets that showed a negative regression slope (i.e., 53 of 70, based on quantile regression method) was significantly greater than expected by chance (probability of slope being negative due to chance $= 0.5$; binomial test: $P < .001$).

For methods involving bins, slope estimates did not differ among three alternative binning schemes (data available in the appendix table in Dryad: http://dx.doi.org/10.5061/dryad.36847). For the traditional upper-bound regression method, the average slope was -213 ± 589 (mean $\pm$ SD) with 8 bins, -211 ± 615 with 10 bins, and -211 ± 629 with 12 bins (repeated-measures ANOVA: $F_{2,57} = 0.19$, $P = .825$). For the equal samples per bin method, the average slope was -119 ± 451 with 8 bins, -114 ± 442 with 10 bins, and -101 ± 445 with 12 bins ($F_{2,57} = 2.37$, $P = .102$). For the sample size as covariate method, the average slope was -77 ± 406 with 8 bins, -99 ± 509 with 10 bins, and -98 ± 478 with 12 bins ($F_{2,57} = 0.31$, $P = .739$). In addition to the slope estimates, the number of data sets in which we detected a statistically significant negative slope was similar among the three binning schemes for each method; it ranged between 30 and 35 data sets for the traditional upper-bound regression method, between 18 and 21 data sets for the

equal samples per bin method, and between 11 and 14 data sets for the sample size as covariate method.

Predictors of Performance Trade-Offs

Taxonomic affiliation did not predict whether our alternative analytical approaches detected performance trade-offs (based on the consensus of the three alternative analytical techniques; table 1). After controlling for sampling limitations, we detected at least one statistically significant performance trade-off in a variety of families, including Cardinalidae, Cricetidae, Emberizidae, Fregatidae, Fringillidae, Parulidae, Stercorariidae, and Troglodytidae. Within most families, however, our ability to detect performance trade-offs varied by species (i.e., for Cardinalidae, Emberizidae, Parulidae, and Troglodytidae). Furthermore, within many species, our ability to detect performance trade-offs varied among data sets (table 1).

The inclusion of a single species versus multiple species in a data set did not affect our ability to detect performance trade-offs. We detected consensus evidence for performance trade-offs in 12 of 55 data sets from single species (22%), and in three of 15 data sets from multiple species (20%; Fisher's exact test for a 2×2 contingency table: $N = 70$, $P > .999$; table 1).

In contrast to taxonomic affiliation, certain statistical characteristics of the data sets predicted whether performance trade-offs were detected by the consensus of the three alternative approaches. Performance trade-offs were more likely to be detected for data sets with a large sample size (logistic regression analysis: $N = 70$, Wald statistic $= 5.81$, $P = .016$, odds ratio $= 1.002$), when bin numbers were large ($N = 70$, Wald statistic $= 10.56$, $P = .001$, odds ratio $= 1.382$), and when the slope of the sampling limitation was steeply negative ($N = 70$, Wald statistic $= 4.05$, $P = .044$, odds ratio $= 0.971$). We note, however, that sample size was positively correlated with the number of bins (Pearson correlation analysis: $N = 70$, $r = 0.35$, $P = .003$) and negatively correlated with the slope of the sampling limitation ($N = 70$, $r = -0.85$, $P < .001$). When the independent effects of each variable were assessed with multiple logistic regression, only the number of bins predicted whether performance trade-offs were detected (sample size: Wald statistic $= 1.05$, $P = .305$, odds ratio $= 1.002$, variance inflation factor $= 4.6$; number of bins: Wald statistic $= 4.95$, $P = .026$, odds ratio $= 1.298$, variance inflation factor $= 1.3$; slope of sampling limitation: Wald statistic $= 0.329$, $P = .566$, odds ratio $= 1.031$, variance inflation factor $= 4.1$). The probability of detecting performance trade-offs did not depend on mean syllable repetition rate, mean frequency bandwidth, skew of the syllable repetition rate distribution, or skew of the frequency bandwidth distribution (all

$N = 70$, all Wald statistics ≤ 1.37, all $P > .242$). Slopes calculated using the quantile regression method were not affected by sample size, mean syllable repetition rate, mean frequency bandwidth, skew of the syllable repetition rate distribution, skew of the frequency bandwidth distribution, slope of the sampling limitation, or the number of bins used in traditional upper-bound regression (all $F_{1,69} \leq 1.26$, all $P \geq .265$).

Discussion

The vast majority (90%) of published data sets analyzed demonstrate an inherent sampling limitation to some degree, with more data points expressed at low than at high syllable repetition rates. This pattern is consistent with our expectation for a performance trade-off that limits the range of frequency bandwidths that can be achieved at high syllable repetition rates. Nevertheless, as we argue here, this pattern could also emerge as a spurious result of the sampling algorithm (fig. 2). Recognition of this possibility motivated application of the three alternative analytical methods featured in this article. Of the 70 published data sets analyzed, 47% feature upper-boundary regressions with statistically significant negative slopes, when analyzed using traditional upper-bound regression. This implies that about half of all data sets analyzed show statistical evidence for performance trade-offs. Yet when we controlled for the sampling limitation, the number of data sets analyzed that retained statistically significant performance trade-offs decreased to only 21% of the 70 published data sets (based on the consensus of the three alternative approaches; range among the three alternative approaches was 20%–24%). We can therefore state that sampling limitations have confounded the analysis of performance trade-offs in this rapidly expanding body of literature.

Consequences of Uncontrolled Sampling Limitations

Once sampling limitations are controlled, 15 of the 70 published data sets (21%; table 1; based on the consensus of the three alternative approaches; range among the three alternative approaches was 20%–24%) show statistically significant negative relationships between syllable repetition rate and frequency bandwidth. For these taxa, we are in a strong position to argue that performance constraints limit animals' ability to produce broadband syllables in rapid succession, as originally proposed by Podos (1997). The animals that show clear evidence for performance constraints are taxonomically diverse, including one family of rodent, two families of seabird, and five families of songbird (table 1). Performance constraints on the production of trilled vocalizations are therefore widespread

among animals. This is not surprising; our understanding of vocal tract mechanics provides a strong a priori assumption that this performance constraint should occur for all animals that produce trilled vocalizations (Westneat et al. 1993; Hoese et al. 2000; Beckers et al. 2003; Podos et al. 2004b; Riede et al. 2006; Pasch et al. 2011).

There are several possible reasons for why many taxa did not show significant performance trade-offs after controlling for sampling limitations. First, it is possible that these taxa do not actually experience performance trade-offs between the two focal song parameters. Although we expect all trills to be subject to performance constraints at some level (Podos 1997), the syllable repetition rates and frequency bandwidths of some taxa may not be sufficiently high for trills to experience the expected trade-off. In other words, although all trilling species face potential trade-offs on trill structure, realized trill parameters may fall short of the regions in acoustic space where trade-offs occur (see fig. 2 in Podos et al. 2004a). Indeed, some taxonomic groups have probably never experienced selection for high-performance trills, and consequently, their trills fall well below the frequency bandwidths and syllable repetition rates for which trade-offs are expected. In other cases, selection may favor songs that achieve high performance levels, yet not in the two acoustic parameters addressed here. The original study on this topic (Podos 1997) included many species precisely because of the expectation that most individual species would not express enough acoustic variation to reveal the expected trade-off. Second, failure to detect performance trade-offs could reflect low statistical power of some of the data sets, especially those with small sample sizes. Thirty-one of the 55 data sets that did not show evidence of performance trade-offs with the consensus method had fewer than 50 data points each, whereas most data sets that showed evidence of a performance trade-off (14 of 15) had more than 50 data points. Yet, we also failed to detect performance trade-offs in several studies that had large sample sizes of more than 100 data points each (table 1: American sparrows, banded wrens, buntings and New World sparrows, red-winged blackbirds, song sparrows, swamp sparrows, yellow warblers, woodcreepers, and wood warblers; median $N = 150$ data points in these data sets). Third, some studies may have not recorded the focal species' highest performance trills in their sample. Since many species produce multiple types of trill, it is possible that recordings in some data sets did not contain the trill types that are under the strongest selection for performance, particularly if those trill types are rarely used by the species. Fourth, vocal performance is known to vary according to context, season, and motivational state (DuBois et al. 2009; Vehrencamp et al. 2013), suggesting that some studies might have missed high-performance trills because of when their re-

cordings were acquired. Fifth, the vocal tract mechanics responsible for generating rapid broadband trills vary among species, and not all configurations may be subject to equal performance trade-offs. For example, canaries produce broadband notes by producing sequential notes with the two sides of their syrinx (Suthers et al. 2012), and as such this species might be able to achieve greater trill rate and frequency bandwidth values, although they should still be subject to trade-offs at some level (Podos 1997; Draganoiu et al. 2002; Suthers et al. 2012).

Although our analyses support the hypothesis that performance trade-offs affect trill structure in a wide range of taxonomic groups, they also reveal limits on our ability to draw inferences about performance trade-offs based on comparative analyses of trill structure alone (table 1). These limits seem particularly stark in the case of swamp sparrows, a species that fails to demonstrate a performance trade-off after controlling for sampling limitations, yet for which experimental studies of song production and learning have provided strong evidence for the expression of performance trade-offs (e.g., Podos 1996; Hoese et al. 2000; Lahti et al. 2011).

Many studies on the functional salience of trills have relied on measures of the orthogonal deviation of points from the regression line that defines the performance trade-off (often referred to as "vocal deviation" or "vocal performance"; Podos 2001; Ballentine et al. 2004). If the regression line does not have a statistically significant negative slope after controlling for sampling limitations (i.e., if it does not provide statistical evidence of a performance trade-off), then vocal deviation from that line might not provide a reliable measure of vocal and neuromuscular performance—although biomechanical models of sound production still predict that trills with higher values in either parameter will be harder to produce. If the slope of the regression line remains statistically significant yet changes as a result of controlling for sampling limitations, then measures of vocal deviation may also need to be viewed with caution. As an illustration of this problem, consider two trills with slow and fast syllable repetition rates and equal deviation below an improperly defined regression line. If the slope of the regression line becomes shallower after controlling for a sampling limitation, then the vocal performance of the slow trill, as determined by its signed orthogonal distance from the regression line, will rise relative to that of the fast trill. Consequently, measures of vocal deviation derived from inaccurately defined performance trade-offs will likewise be inaccurate and, more problematically, confounded by syllable repetition rate.

Sources of Sampling Limitations

We have presented evidence that sampling limitations affect the analysis of performance trade-offs. The sampling limitations we identify here emerge when the distribution of syllable repetition rates is positively skewed. One possible reason for the observed positive skew is that syllable repetition rate, like many biological variables, has a minimum value of zero, which truncates the left side of its distribution. In contrast, there is no theoretical maximum value for syllable repetition rate, so trills with very high values will positively skew the distribution.

Another possible reason for positive skew is that performance trade-offs restrict the range of bandwidths available for trills with high syllable repetition rates. One way to illustrate this concept is to consider that the area available in the left half of the triangular frequency bandwidth by syllable repetition rate distribution (i.e., the area associated with low trill rates) is three times larger than the area available in the right half of the distribution (i.e., the area associated with high trill rates; e.g., fig. 1). If animals are more likely to produce trills in the larger acoustic space associated with low syllable repetition rates, then the distribution of trills along the syllable repetition rate axis should show positive skew. However, the same logic suggests that performance trade-offs should yield an equally skewed distribution of frequency bandwidth values, since the area available in the bottom half of the triangular distribution (i.e., the area associated with low-frequency bandwidth) is also three times larger than the area available in the top half of the distribution (i.e., the area associated with high-frequency bandwidth; e.g., fig. 1). Yet, among the 70 published data sets we analyzed here, the average skew was significantly greater for syllable repetition rate ($g_1 = 1.15 \pm 0.93$, mean $\pm$ SD) than for frequency bandwidth ($g_1 = 0.45 \pm 0.67$; paired *t*-test: $t_{69} = 4.75$, $P < .001$), suggesting that performance trade-offs alone cannot account for the sampling limitation. Regardless of the underlying mechanism, it is clear that syllable repetition rate is commonly skewed, which thus confounds the analysis of performance trade-offs.

Recommendations for Controlling Sampling Limitations

Our objective was to illustrate the potentially confounding effects of sampling limitations on the analysis of performance trade-offs and to recommend alternative analytical techniques for future studies. The three alternative methods that we employed—equal samples per bin, 90% quantile regression, and sample size as covariate—all controlled for potential sampling limitations. Of these three, we recommend quantile regression for future analyses of per-

formance trade-offs, for four reasons. First, quantile regression does not require assumptions about how to apportion data among bins, as is necessary for upper-bound regression. Second, quantile regression uses all of the original data to estimate the slope, so its statistical power increases with increasing sample size (Koenker and Hallock 2001). In contrast, the statistical power of upper-bound regression techniques is limited by the number of bins used in the analysis, since only a single data point from each bin is included in the analysis (Blackburn et al. 1992). Third, of the three alternative analytical techniques that we explored, quantile regression matched the consensus view most often, with the same conclusion reached in all but three of the 70 data sets (table 1). Fourth, slopes derived from quantile regression of the 90th percentile are resistant to outliers. In upper-bound regression, by contrast, outliers have a heavy influence on estimating slopes (Scharf et al. 1998; Koenker and Hallock 2001).

In using quantile regression, the choice to measure the slope of the nineteeth percentile is somewhat arbitrary; however, previous studies suggest that the nineteeth percentile is a reasonable choice because its slope can be estimated precisely, its slope is resistant to outliers, and its slope accurately estimates the expected trade-off found near the upper boundary of a triangular distribution (Scharf et al. 1998; Cade et al. 1999). Percentiles between the nineteeth and the hundredth percentiles may also reflect the expected trade-off, but ultimately, the maximum percentile that can be estimated precisely will be limited by sample size and the characteristics of the distribution being analyzed (Cade et al. 1999). Researchers might also consider measuring the slope of the fiftieth percentile and using it to approximate slopes found in the upper quantiles. This approach might seem particularly appealing because quantile regression of the median will tend to have greater statistical power than quantile regression of the upper quantiles. However, we caution researchers that such an approach will underestimate the expected upper boundary slopes, since slopes derived from the fiftieth percentile are predicted to be only half as steep as slopes derived from the upper quantiles of triangular distributions. Our data substantiate this concern. Across our 70 data sets, the average slope of 90% quantile regressions was -56, which was approximately twice as steep as the average slope derived from 50% quantile regressions (-26; Wilcoxon signed rank test: $\chi^2 = 3.30$, $P = .001$; data available in appendix table in Dryad: http://dx.doi.org/10.5061/dryad .36847). We therefore recommend quantile regression of the nineteeth percentile for estimating trill performance trade-offs.

In conclusion, we have shown that sampling limitations can confound the detection and quantification of performance trade-offs in the trilled vocalizations of a wide range of taxonomic groups. In particular, sampling limitations have artificially inflated our estimate of the number of taxonomic groups whose trills show evidence for being affected by performance trade-offs. Using analytical techniques that control for sampling limitations, we show that performance trade-offs are manifest across a broad range of animal taxa but that evidence for performance trade-offs is less prevalent than indicated by the literature. We propose quantile regression as an alternative analytical approach that is unaffected by the sampling limitation and which has broad applicability to all situations where sampling limitations have confounded upper-bound regression analysis. Ultimately, comparative analyses of signal diversity, such as those explored here, are best considered together with experimental or developmental tests of the role of constraints and trade-offs in signal expression.

Acknowledgments

Literature Cited

Araya-Ajoy, Y., J. Chaves-Campos, E. K. V. Kalko, and J. A. DeWoody. 2009. High-pitched notes during vocal contests signal genetic diversity in ocellated antbirds. PLoS ONE 4:e8137–e8138.

Arnold, S. J. 1992. Constraints on phenotypic evolution. American Naturalist 140(suppl.):S85–S107.

Ballentine, B. 2006. Morphological adaptation influences the evolution of a mating signal. Evolution 60:1936–1944.

———. 2009. The ability to perform physically challenging songs predicts age and size in male swamp sparrows, *Melospiza georgiana*. Animal Behaviour 77:973–978.

Ballentine, B., J. Hyman, and S. Nowicki. 2004. Vocal performance influences female response to male bird song: an experimental test. Behavioral Ecology 15:163–168.

Beckers, G. L., R. A. Suthers, and C. ten Cate. 2003. Pure-tone birdsong by resonance filtering of harmonic overtones. Proceedings of the National Academy of Sciences of the USA 100:7372–7376.

Beebee, M. D. 2004. Variation in vocal performance in the songs of a wood-warbler: evidence for the function of distinct singing modes. Ethology 110:531–542.

Blackburn, T. M., J. H. Lawton, and J. N. Perry. 1992. A method of

estimating the slope of upper bounds of plots of body size and abundance in natural animal assemblages. Oikos 65:107–112.

Byers, J., E. Hebets, and J. Podos. 2010. Female mate choice based upon male motor performance. Animal Behaviour 79:771–778.

Cade, B. S., J. W. Terrell, and R. L. Schroeder. 1999. Estimating effects of limiting factors with regression quantiles. Ecology 80:311–323.

Cardoso, G. C., J. W. Atwell, Y. Hu, E. D. Ketterson, and T. D. Price. 2012. No correlation between three selected trade-offs in birdsong performance and male quality for a species with song repertoires. Ethology 118:584–593.

Cardoso, G. C., J. W. Atwell, E. D. Ketterson, and T. D. Price. 2007. Inferring performance in the songs of dark-eyed juncos (*Junco hyemalis*). Behavioral Ecology 18:1051–1057.

Cardoso, G. C., and Y. Hu. 2011. Birdsong performance and the evolution of simple (rather than elaborate) sexual signals. American Naturalist 178:679–686.

Caro, S. P., K. B. Sewall, K. G. Salvante, and K. W. Sockman. 2010. Female Lincoln's sparrows modulate their behavior in response to variation in male song quality. Behavioral Ecology 21:562–569.

Christensen, R., S. Kleindorfer, and J. Robertson. 2006. Song is a reliable signal of bill morphology in Darwin's small tree finch *Camarhynchus parvulus*, and vocal performance predicts male pairing success. Journal of Avian Biology 37:617–624.

Cramer, E. R. A. 2013. Vocal deviation and trill consistency do not affect male response to playback in house wrens. Behavioral Ecology 24:412–420.

Cramer, E. R. A., and J. J. Price. 2007. Red-winged blackbirds *Ageliaus phoeniceus* respond differently to song types with different performance levels. Journal of Avian Biology 38:122–127.

Darwin, C. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. J. Murray, London.

de Kort, S. R., E. R. B. Eldermire, S. Valderrama, C. A. Botero, and S. L. Vehrencamp. 2009. Trill consistency is an age-related assessment signal in banded wrens. Proceedings of the Royal Society B: Biological Sciences 276:2315–2321.

Derryberry, E. P. 2009. Ecology shapes birdsong evolution: variation in morphology and habitat explains variation in white-crowned sparrow song. American Naturalist 174:24–33.

Derryberry, E. P., N. Seddon, S. Claramunt, J. A. Tobias, A. Baker, A. Aleixo, and R. T. Brumfield. 2012. Correlated evolution of beak morphology and song in the Neotropical woodcreeper radiation. Evolution 66:2784–2797.

Draganoiu, T. I., L. Nagle, and M. Kreutzer. 2002. Directional female preference for an exaggerated male trait in canary (*Serinus canaria*) song. Proceedings of the Royal Society B: Biological Sciences 269:2525–2531.

DuBois, A. L., S. Nowicki, and W. A. Searcy. 2009. Swamp sparrows, *Melospiza georgiana*, modulate vocal performance in an aggressive context. Biology Letters 5:163–165.

———. 2011. Discrimination of vocal performance by male swamp sparrows. Behavioral Ecology and Sociobiology 65:717–726.

Futuyma, D. J. 2009. Evolution. 2nd ed. Sinauer, Sunderland, MA.

Gibbs, H. L., and P. R. Grant. 1987. Oscillating selection on Darwin's finches. Nature 327:511–513.

Grant, B. R., and P. R. Grant. 2002. Lack of premating isolation at the base of a phylogenetic tree. American Naturalist 160:1–19.

Gutenbrunner, C., and J. Jureckova. 1992. Regression rank scores and regression quantiles. Annals of Statistics 20:305–330.

Hartley, H. O., and H. A. David. 1954. Universal bounds for mean range and extreme observation. Annals of Mathematical Statistics 25:85–99.

Hartley, R. S., and R. A. Suthers. 1989. Airflow and pressure during canary song: direct evidence for mini-breaths. Journal of Comparative Physiology A 165:15–26.

Herrel, A., J. Podos, S. K. Huber, and A. P. Hendry. 2005. Bite performance and morphology in a population of Darwin's finches: implications for the evolution of beak shape. Functional Ecology 19:43–48.

Herrel, A., J. Podos, B. Vanhooydonck, and A. P. Hendry. 2009. Force-velocity trade-off in Darwin's finch jaw function: a biomechanical basis for ecological speciation? Functional Ecology 23:119–125.

Hoese, W. J., J. Podos, N. C. Boetticher, and S. Nowicki. 2000. Vocal tract function in birdsong production: experimental manipulation of beak movements. Journal of Experimental Biology 203:1845–1855.

Huber, S. H., and J. Podos. 2006. Beak morphology and song features covary in a population of Darwin's finches (*Geospiza fortis*). Biological Journal of the Linnean Society 88:489–498.

Illes, A. E., M. L. Hall, and S. L. Vehrencamp. 2006. Vocal performance influences male receiver response in the banded wren. Proceedings of the Royal Society B: Biological Sciences 273:1907–1912.

Janicke, T., S. Hahn, M. S. Ritz, and H.-U. Peter. 2008. Vocal performance reflects individual quality in a non-passerine. Animal Behaviour 75:91–98.

Juola, F. A., and W. A. Searcy. 2011. Vocalizations reveal body condition and are associated with visual traits in great frigatebirds (*Fregata minor*). Behavioral Ecology and Sociobiology 65:2297–2303.

Kagawa, H., and M. Soma. 2013. Song performance and elaboration as potential indicators of male quality in Java sparrows. Behavioural Processes 99:138–144.

Koenker, R. 2012. Quantreg: quantile regression. R package version 4.91. Accessed December 6, 2012. http://CRAN.R-project.org/package=quantreg.

Koenker, R., and G. Bassett Jr. 1978. Regression quantiles. Econometrica 46:33–50.

Koenker, R., and K. Hallock. 2001. Quantile regression: an introduction. Journal of Economic Perspectives 15:143–156.

Lahti, D. C., D. L. Moseley, and J. Podos. 2011. A trade-off between accuracy and performance in bird song learning. Ethology 117:802–811.

Liu, I. A., B. Lohr, B. Olsen, and R. Greenberg. 2008. Macrogeographic vocal variation in subspecies of swamp sparrow. Condor 110:102–109.

Logue, D. M., C. Chalmers, and A. H. Gowland. 2008. The behavioural mechanisms underlying temporal coordination in black-bellied wren duets. Animal Behaviour 75:1803–1808.

MacLeod, C. D., and R. C. MacLeod. 2009. The relationship between body mass and relative investment in testes mass in amniotes and other vertebrates. Oikos 118:903–916.

Maddison, C. J., R. C. Anderson, N. H. Prior, M. D. Taves, and K. K. Soma. 2012. Soft song during aggressive interactions: seasonal changes and endocrine correlates in song sparrows. Hormones and Behavior 62:455–463.

Moseley, D. L., D. C. Lahti, and J. Podos. Forthcoming. Responses to song playback vary with the vocal performance of both signal senders and receivers. Proceedings of the Royal Society B: Biological Sciences 280:20131401. http://dx.doi.org/10.1098/rspb.2013.1401.

Naguib, M., R. Schmidt, P. Sprau, T. Roth, C. Flörcke, and V. Amrhein. 2008. The ecology of vocal signaling: male spacing and communication distance of different song traits in nightingales. Behavioral Ecology 19:1034–1040.

Narango, D. L. 2012. Causes and consequences of urban-associated song variation: a study of vocal behavior in the northern cardinal (*Cardinalis cardinalis*). MSc thesis. Ohio State University, Columbus.

Nowicki, S., M. Westneat, and W. Hoese. 1992. Birdsong: motor function and the evolution of communication. Seminars in Neuroscience 4:385–390.

Pasch, B., A. S. George, P. Campbell, and S. M. Phelps. 2011. Androgen-dependent male vocal performance influences female preference in Neotropical singing mice. Animal Behaviour 82:177–183.

Podos, J. 1996. Motor constraints on vocal development in a songbird. Animal Behaviour 51:1061–1070.

———. 1997. A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae). Evolution 51:537–551.

———. 2001. Correlated evolution of morphology and vocal signal structure in Darwin's finches. Nature 409:185–188.

———. 2010. Acoustic discrimination of sympatric morphs in Darwin's finches: a behavioral mechanism for assortative mating? Philosophical Transactions of the Royal Society B: Biological Sciences 365:1031–1039.

Podos, J., S. K. Huber, and B. Taft. 2004a. Bird song: the interface of evolution and mechanism. Annual Review of Ecology, Evolution, and Systematics 35:55–87.

Podos, J., D. C. Lahti, D. L. Moseley. 2009. Vocal performance and sensorimotor learning in songbirds. Advances in the Study of Behavior 40:159–195.

Podos, J., and S. Nowicki. 2004. Beaks, adaptation, and vocal evolution in Darwin's finches. BioScience 54:501–510.

Podos, J., J. A. Southall, and M. R. Rossi-Santos. 2004b. Vocal mechanics in Darwin's finches: correlation of beak gape and song frequency. Journal of Experimental Biology 207:607–619.

Price, J. J., and L. M. Whalen. 2009. Plumage evolution in the oropendolas and caciques: different divergence rates in polygynous and monogamous taxa. Evolution 63:2985–2998.

R Development Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org/.

Reichert, M. S., and H. C. Gerhardt. 2012. Trade-offs and upper limits to signal performance during close-range vocal competition in gray tree frogs *Hyla versicolor*. American Naturalist 180:425–437.

Riede, T., R. A. Suthers, N. H. Fletcher, and W. E. Blevins. 2006. Songbirds tune their vocal tract to the fundamental frequency of their song. Proceedings of the National Academy of Sciences of the USA 103:5543–5548.

Roberts, S. G. B., R. I. M. Dunbar, T. V. Pollet, and T. Kuppens. 2009. Exploring variation in active network size: constraints and ego characteristics. Social Networks 31:138–146.

Roff, D. A., and D. J. Fairbairn. 2007. The evolution of trade-offs: where are we? Journal of Evolutionary Biology 20:433–447.

Scharf, F. S., F. Juanes, and M. Sutherland. 1998. Inferring ecological relationships from the edges of scatter diagrams: comparison of regression techniques. Ecology 79:448–460.

Schluter, D., and P. R. Grant. 1984. Determinants of morphological patterns in communities of Darwin's finches. American Naturalist 123:175–196.

Schmidt, R., H. P. Kunc, V. Amrhein, and M. Naguib. 2008. Aggressive responses to broadband trills are related to subsequent pairing success in nightingales. Behavioral Ecology 19:635–641.

Searcy, W. A., and S. Nowicki. 2005. The evolution of animal communication: reliability and deception in signaling systems. Princeton University Press, Princeton, NJ.

Seddon, N. 2005. Ecological adaptation and species recognition drives vocal evolution in Neotropical suboscine birds. Evolution 59:200–215.

Slabbekoorn, H., and T. B. Smith. 2000. Does bill size polymorphism affect courtship song characteristics in the African finch *Pyrenestes ostrinus*? Biological Journal of the Linnean Society 71:737–753.

Sockman, K. H. 2009. Annual variation in vocal performance and its relationship with bill morphology in Lincoln's sparrows, *Melospiza lincolnii*. Animal Behaviour 77:663–671.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry: the principles and practice of statistics in biological research. 3rd ed. W. H. Freeman, New York.

Stearns, S. C. 1992. The evolution of life histories. Oxford University Press, Oxford.

Suthers, R. A., E. Vallet, and M. Kreutzer. 2012. Bilateral coordination and the motor basis of female preference for sexual signals in canary song. Journal of Experimental Biology 215:2950–2959.

van Damme, R., R. S. Wilson, B. Vanhooydonck, and P. Aerts. 2002. Performance constraints in decathletes. Nature 415:755–756.

Vehrencamp, S. L., J. Yantachka, M. L. Hall, and S. R. de Kort. 2013. Trill performance components vary with age, season, and motivation in the banded wren. Behavioral Ecology and Sociobiology 67:409–419.

Westneat, M. W., J. H. Long Jr., W. Hoese, and S. Nowicki. 1993. Kinematics of birdsong: functional correlation of cranial movements and acoustic features in sparrows. Journal of Experimental Biology 182:147–171.

Wilson, D. R., P. Bitton, J. Podos, and D. J. Mennill. 2013. Data from: Uneven sampling and the analysis of vocal performance constraints. American Naturalist, Dryad Digital Repository http://dx.doi.org/10.5061/dryad.36847.

Zollinger, S. A., J. Podos, E. Nemeth, F. Goller, and H. Brumm. 2012. On the relationship between, and measurement of, amplitude and frequency in birdsong. Animal Behaviour 84:E1–E9.