

# Birdsong Performance Studies: A Contrary View

Donald Kroodsma

52 School Street, Hatfield MA 01038

Email: [DonaldKroodsma@gmail.com](mailto:DonaldKroodsma@gmail.com)

Phone: 413-247-3367

Word count: 10,480

## ABSTRACT

Birdsong biologists interested in sexual selection and honest signaling have repeatedly confirmed, over more than a decade, the significance of a scatter plot between trill rate and frequency bandwidth. This “motor constraints hypothesis” of Podos (1997) proposes that the closer a song plots to an upper bound on this graph, the more difficult the song is to sing, and the more difficult the song the higher quality the singer, so that song quality honestly reveals male quality. In reviewing the confirming literature, however, I can find no support for this “performance hypothesis.”

The scatter in the graph for songbirds is explained not by male quality but by social factors and song learning. When songbirds learn their songs from each other, multiple males in a neighborhood will sing the same song type. The need to conform to the local dialect of song types guides males to learn a typical example of each song type for that population, not to take a memorized song and diminish or exaggerate it in trill rate or frequency bandwidth to honestly demonstrate his relative prowess. When data in this scatter plot are coded both by song type and by male, it is the song type and the need to conform that explains the variability, not the quality of different males. There is no reliable information in the song performance measures that can be used to evaluate a singing male.

## KEY WORDS

Birdsong, performance, confirmation bias, advocacy, chipping sparrow, swamp sparrow, banded wren

## INTRODUCTION

Something in how a male songbird delivers his songs must convey something about his relative quality to those who listen, especially females, but identifying those somethings has proven challenging. In the study of birdsong repertoires and female choice, for example, it has been widely accepted that "Females of many songbird species show a preference for mating with males that have larger song repertoires" (Nowicki, Hasselquist, Bensch, et al., 2000), but in spite of a host of studies confirming that relationship, there is no strong evidence that males or females attend to the number of different songs a male can sing (Byers and Kroodsma, 2009).

Another idea that has over the last decade gained much traction is the motor constraints hypothesis, or the performance hypothesis, of Podos (1997). Scatter plots of trill rates and frequency bandwidths show an inverse relationship, the more rapid the trill, the narrower the bandwidth (see Figures 4, 5, and 10 for examples). Blank areas with no data beyond an upper

40 bound suggest a motor constraint, i.e., the birds can't produce those combinations of trill rates  
 41 and bandwidths (but see Figure 4). The interesting hypothesis is that how close a song plots to  
 42 the upper bound might reveal the difficulty of producing that song, so that songs near the upper  
 43 bound honestly reveal a high quality singer; both prospective mates and competing males might  
 44 then use those high-performance songs to detect high quality singers.

45 This hypothesis has “been adopted widely in tests of song function” (Goodwin and  
 46 Podos, 2015), and has been repeatedly confirmed over the past decade. My careful scrutiny of  
 47 those studies here, however, reveals that the hypothesis has become largely an assumption, never  
 48 truly tested, and simply biologically implausible, if not impossible. Here I review the confirming  
 49 studies, beginning with a most recent paper, on chipping sparrows (*Spizella passerina*), because  
 50 it reveals especially clearly the methods used to confirm the hypothesis; I then proceed to the  
 51 studies of swamp sparrows (*Melospiza georgiana*), before briefly reviewing other species.

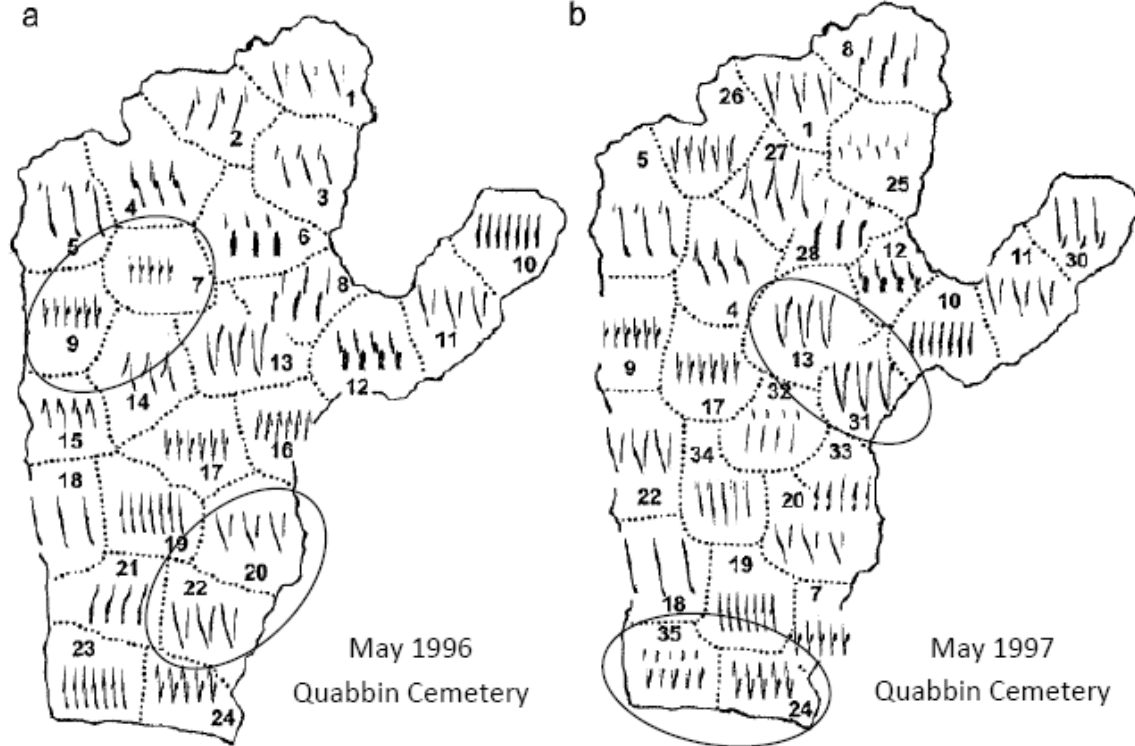
52 First, however, I offer apologies to those who feel my approach too frank, or too blunt, or  
 53 overkill. My goals, however, are not only to present sufficient evidence to make the case that the  
 54 performance hypothesis is not supported in these studies, but also to demonstrate how  
 55 confirmation bias taints the literature on birdsong and sexual selection, in hopes that it will be  
 56 avoided in the future.

## 57 **CHIPPING SPARROW**

58 I begin by illustrating how a chipping sparrow acquires his song, because the roots of  
 59 implausibility for the performance hypothesis lie in the basic biological basis for song learning.

### 60 *The biology of song learning by chipping sparrow*

61 A young chipping sparrow acquires his song by copying the song of an adult next to  
 62 whom he settles, as illustrated by Liu and Kroodsma (2006; Figure 1). The adult's song is copied  
 63 whether the trill is delivered slowly (males 13 & 31, 20 & 22) or more rapidly (7 & 9, 24 & 35),  
 64 based on the social bond between the adult tutor and the youngster who is establishing his first  
 65 territory. This conclusion is based on unequivocal field evidence by color-banding 324 young  
 66 chipping sparrows and following them during dispersal.



67

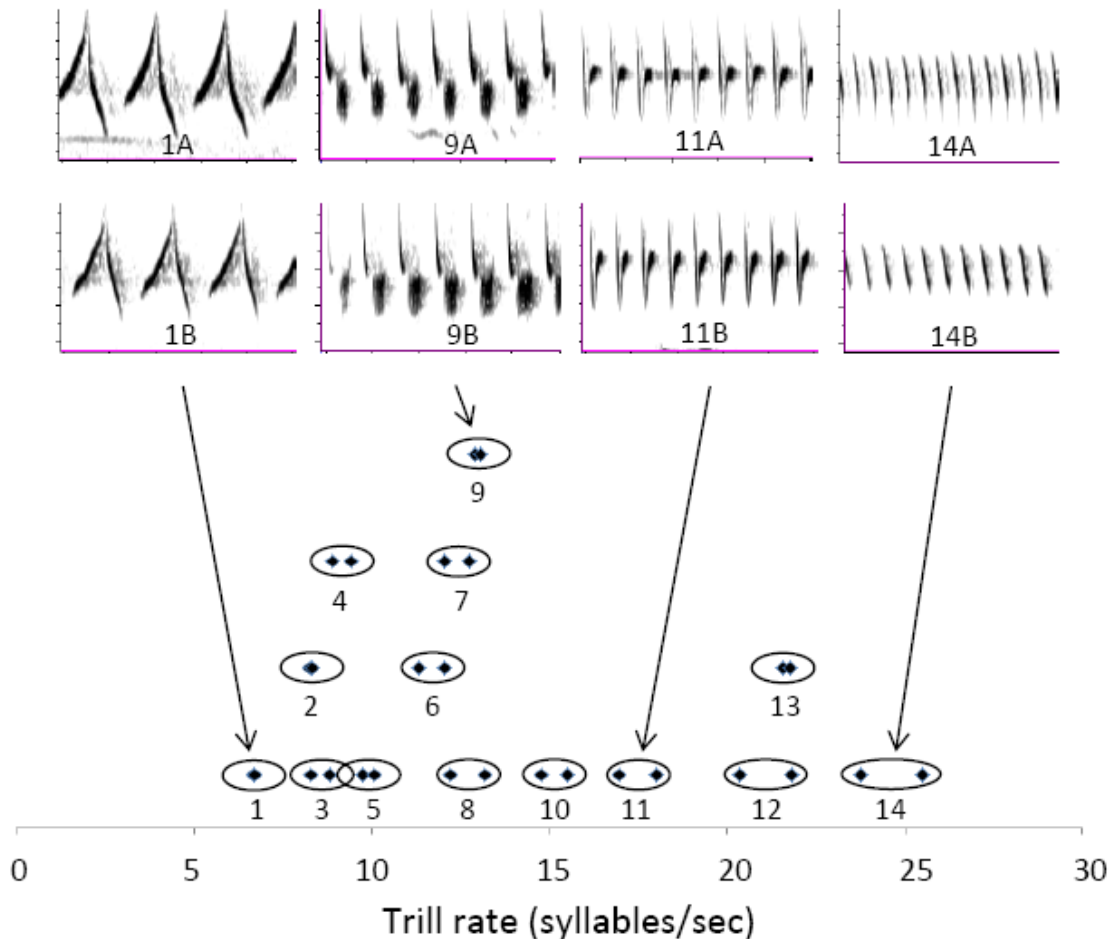
68 Figure 1. From Liu and Kroodtsma (2006). “Yearling Chipping Sparrows imitate songs of an immediate  
 69 neighbor, but the instability of territories results in only short-term song sharing among neighbors. (a)  
 70 In May of 1996, 24 territorial males (numbered 1-24) were found in the Quabbin Cemetery, and a  
 71 portion of each male’s song type (0.35 sec) is illustrated. Males 7 and 9 share similar song types, as do  
 72 males 20 and 22. (b) In May of 1997, 26 territorial males were found in the cemetery, 16 returning  
 73 adults (at least two years old) from the previous year and 10 birds breeding there for the first time.  
 74 Males 7 and 9 both returned but are no longer neighbors, and male 22 did not return. Male 35 did not  
 75 learn the song of his father (male 5) or his father’s close neighbors in 1996 or 1997, but instead  
 76 appeared to learn from his immediate neighbor in 1997 (male 24). Male 31 also appeared to acquire his  
 77 song from an immediate neighbor (male 13) in 1997, not from his father (male 10). Both 1997 yearlings  
 78 (31 and 35) hatched late in the 1996 breeding season, and each most likely acquired his song during  
 79 1997, as a yearling.”

80

81 To further illustrate how a young male chipping sparrow learns rather precisely the song  
 82 of his adult tutor, and especially the tutor’s trill rate, I recorded chipping sparrows during early  
 83 May (2015) when they first returned from migration, before post-learning dispersal might occur.  
 84 In two populations, one on a golf course in Lewiston, Michigan, the other in a city park in  
 85 Northampton, Massachusetts, I recorded 67 different males. Equipment used were a Sound  
 86 Devices 722 digital recorder and a stereo Telinga microphone. Birds were not banded, but most  
 87 of the birds were recorded in rapid succession by moving directly from one singer to the next, so  
 88 that the previous and next singer could be heard while recording a given male. If songs of  
 89 suspected neighbors were identical, and I could not distinguish their songs in sonagrams, I

90 conservatively assumed they were the same male and discarded one of the recordings from the  
 91 data set. Using Raven pro software, I measured trill rates and frequency bandwidths for three  
 92 high quality songs for each male, and used the median value in analyses (“spectrogram window  
 93 size” in Raven: 110 for temporal measures, 2050 for frequency; lower and upper frequencies  
 94 measured as -24 dB down from max power; I believe my methods match those routinely used in  
 95 performance studies).

96 Among these 67 males, I found 14 pairs of adjacent males with essentially identical songs (see  
 97 Figure 2), as one would expect based on how chipping sparrows learn their songs. As is clear in  
 98 Figure 2, song types and trill rates are determined by where and from whom a male learns his  
 99 song and cannot reflect any measure of his quality, in the sense of Podos (1997). A male with a  
 100 trill rate of 25 is no “better” than a male with trill rate of seven (and there is no evidence for song  
 101 learning in any songbird species or especially in chipping sparrows (Liu and Kroodsma, 1999,  
 102 2006) that a male is in any way limited in what naturally occurring trill rate he can learn).



103

104 Figure 2. A few dozen different songs can occur within a chipping sparrow population (only four  
 105 illustrated here: 1, 9, 11, 14), but neighboring males (A and B) often have nearly identical songs, the  
 106 result of a young male copying the song of a nearby adult singer (Liu and Kroodsma 1999, 2006); all

107 features of a male's song, including his trill rate as illustrated here (14 examples), are determined by  
 108 that adult tutor. In the lower graph, each oval encircles the two data points (pairs 1 and 2 are identical)  
 109 for trill rates from two neighboring males with the same song (data are distributed vertically for easier  
 110 visibility). Each data point is the median of three measurements for a given male.

111

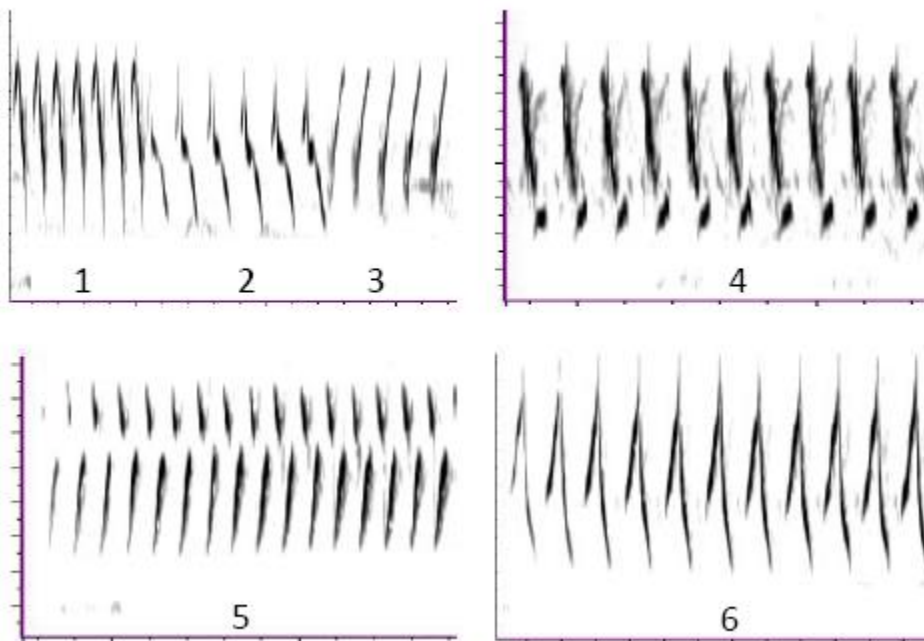
112 ***The biology of song use in chipping sparrows***

113 Well before sunrise, during the dawn chorus, male chipping sparrows range widely over  
 114 space, especially into neighboring territories, but they can also display with other males in arenas  
 115 far removed from their daytime centers of nesting activity. For example,

116 If territories are widely dispersed, it seems that the males still convene at a traditional  
 117 location, sparring there even if some of the males don't own territories that border that  
 118 place (Kroodsma, 2005:319; see also Liu, 2004).

119 One example suffices, from a location in eastern Missouri (see Figure 3). In that example,  
 120 four males displayed simultaneously in a lek-like arena during the dawn chorus, all singing near  
 121 one another on a paved road, but before sunrise they all dispersed to their daytime centers of  
 122 activity. Replacing those four males after sunrise were two other males, each now on his daytime  
 123 center of activity, each of which was presumably displaying elsewhere during the dawn chorus.  
 124 Male chipping sparrows thus routinely intrude on the daytime activity centers (i.e., "territories")  
 125 of other males and display there competitively with lek-like behavior.

126



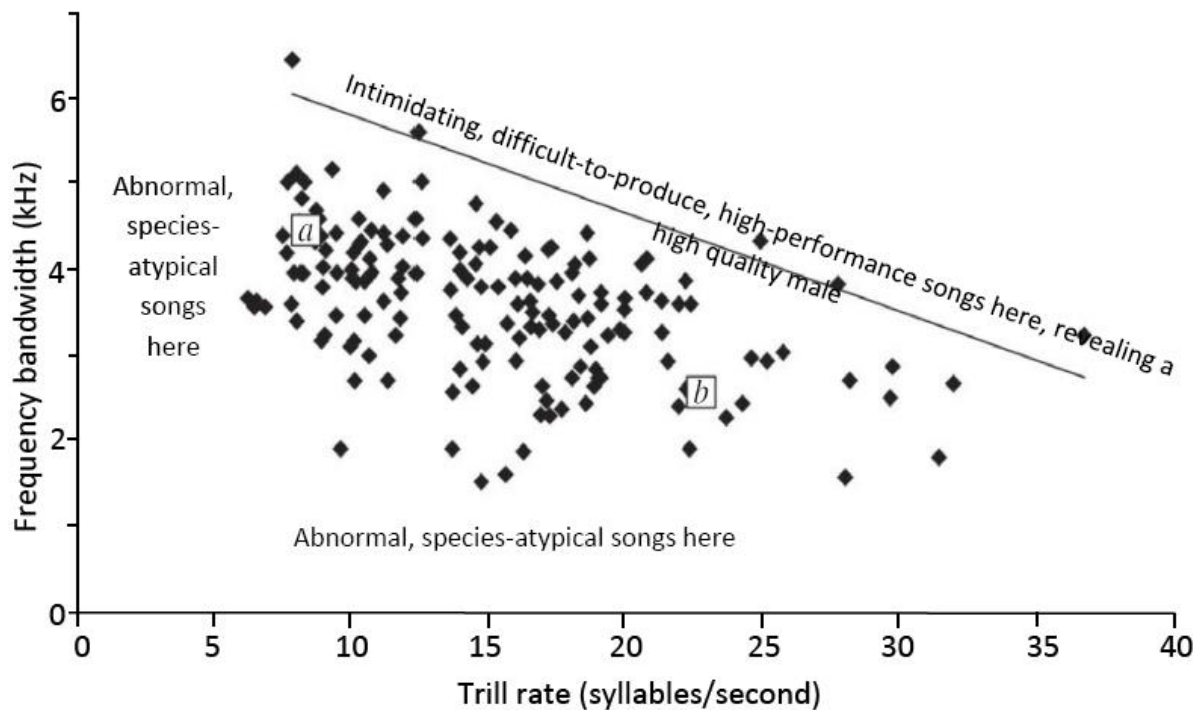
127

128 Figure 3. During the dawn chorus, chipping sparrow males can gather in lek-like, competitive singing  
 129 arenas well away from their daytime territories. In this example, during the dawn chorus, four males (1-  
 130 4) gathered in a lek-like arena on a paved road, displaying in the dark, sight unseen; individuality in their  
 131 songs allowed each to be identified. After the dawn singers had dispersed to their daytime territories,  
 132 two other males (5, 6), who presumably were elsewhere during the dawn chorus, sang at this location  
 133 on their daytime territories.

134

### 135 *The trill rate/frequency bandwidth graph*

136 The standard graph provided in studies of performance is the scatterplot of frequency  
 137 bandwidth versus trill rate (Figure 4). The distance from a given plotted point to the upper bound  
 138 regression line (i.e., the deviation from the line) is then interpreted as a measure of a male's  
 139 performance or proficiency on that particular song. A small deviation is a high-performance  
 140 song, a large deviation a low-performance song. Because basic ornithological data on song type  
 141 and individual males are not encoded in the data, however, the biological significance of the data  
 142 is obscured.



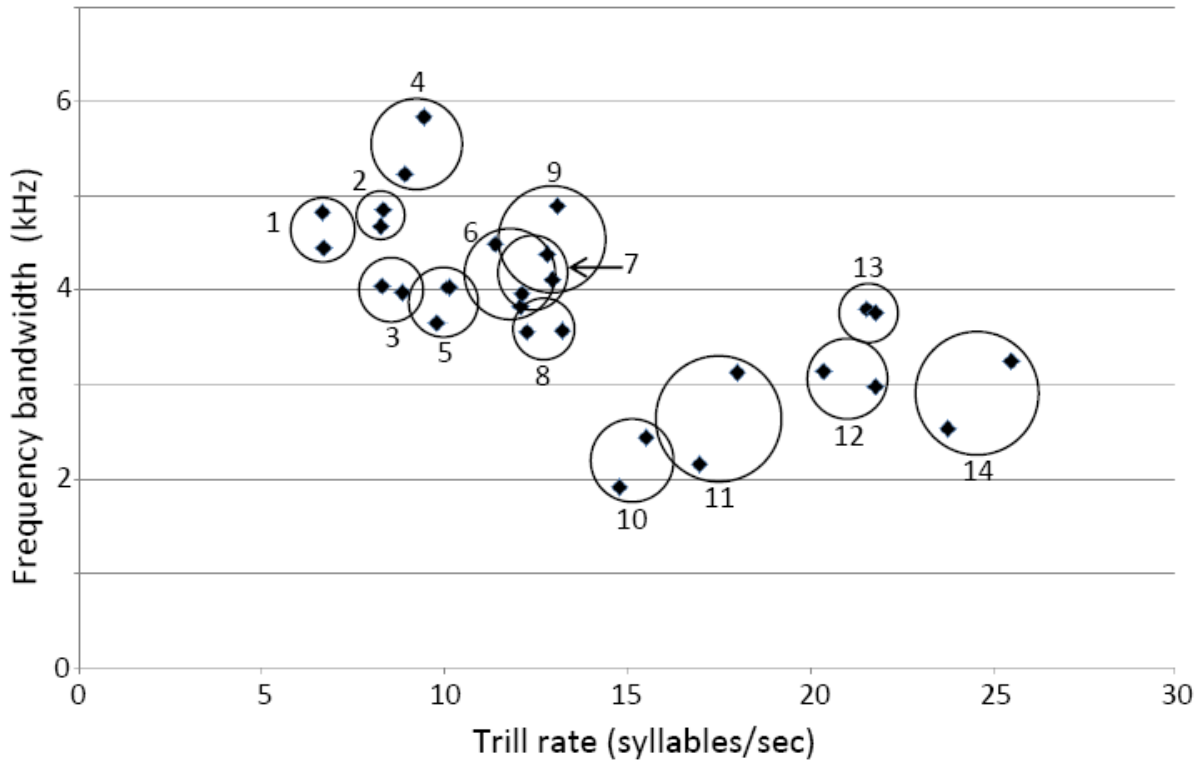
143

144 Figure 4. Data from Goodwin and Podos (2014): "Chipping sparrow songs show evidence of a vocal  
 145 constraint . . . Biplot of trill rate and frequency bandwidth (n = 160 males) reveals a performance trade-  
 146 off in vocal production . . ." (letters 'a' and 'b' refer to a portion of the original figure not illustrated  
 147 here). Data are replotted on expanded axes to show the open space below and to the left of the data  
 148 points.

149

150 Consider, then, a graph of this sort that includes the information necessary to interpret it  
 151 in a biological context (Figure 5). Given how a chipping sparrow learns his song from a

152 neighboring adult, it is clear from this figure that social factors and song learning explain not  
 153 only 1) the variability in trill rates within a population (as in Figure 2), but also 2) the variability  
 154 in frequency bandwidth (Figure 5), and therefore also 3) the scatter in the plot from Goodwin  
 155 and Podos (Figure 4). Scatter in the graph is explained not by trill rates or frequency bandwidths  
 156 that reflect male quality, but instead by song types, reflecting the social bonds that play a crucial  
 157 role in the process by which a young male learns his song.



158  
 159 Figure 5. All aspects of a male chipping sparrow's song, including the trill rate and frequency bandwidth,  
 160 are determined largely by the song that he imitates from an adult male, so that song types dictate the  
 161 scatter in the plot, not relative male quality. Plotted data are from Figure 2, and each of the 14 circles  
 162 encompasses the songs of two neighboring males.

163  
 164 ***A focused critique of Goodwin and Podos (2014)***

165 The claims made by Goodwin and Podos are substantial, and novel (quotes from the title  
 166 and abstract, with my edits italicized in brackets):

167 Team of rivals: alliance formation [*a cooperative fighting team*] in territorial  
 168 songbirds is predicted by vocal signal structure [*trill rate*] . . . Our results provide  
 169 the first evidence that animals like chipping sparrows rely on precise assessments  
 170 of mating signal features [*trill rates*], as well as relative comparisons of signal  
 171 properties [*trill rates*] among multiple animals in communication networks, when  
 172 deciding when and with whom to form temporary alliances [*cooperative fighting*  
 173 *teams*] against a backdrop of competition and rivalry.

174           These claims are made, however, by omitting reference to two basic ornithological facts  
 175 about the subject animal that were published on the same population of chipping sparrows. 1)  
 176 Trill rate reflects song learning from neighbors, not male quality (see above), and males cannot  
 177 therefore assess one another based on trill rate, let alone precisely; and 2) published information  
 178 on male behaviors would lead one to believe that these gatherings of singing males in small  
 179 singing arenas are competitive, not cooperative, as simply assumed by Goodwin and Podos  
 180 (2014).

181           A third major problem with Goodwin and Podos (2014) lies in their statistics. First, as  
 182 pointed out by Akçay and Beecher (2015), the three tests supporting coalitions are simply done  
 183 wrong; even when “corrected” by Goodwin and Podos (2015), the one remaining significant test  
 184 ( $p = 0.03$ ) remains problematical (see [https://caglarakcay.wordpress.com/2015/07/02/new-](https://caglarakcay.wordpress.com/2015/07/02/new-critique-of-chipping-sparrow-coalition-paper/)  
 185 [critique-of-chipping-sparrow-coalition-paper/](https://caglarakcay.wordpress.com/2015/07/02/new-critique-of-chipping-sparrow-coalition-paper/) ). Furthermore, the reported statistically significant  
 186 tests are gleaned from a much larger, unreported series of nonsignificant tests. The authors  
 187 analyzed data on 1) frequency bandwidths alone, 2) trill rates alone, and 3) a combination of  
 188 frequency bandwidths and trill rates. Even though only the bandwidth/rate combination makes  
 189 any sense for the performance literature, the authors report only data on trill rates, because the  
 190 only statistically significant tests were found there. When undisclosed “multiple comparisons”  
 191 have been made, however, building a biologically significant story only on reported statistical  
 192 tests that just barely reach  $\alpha = 0.05$  is problematic, because any correction for multiple  
 193 comparisons (e.g., Bonferroni) would render the reported tests nonsignificant; for example, the  
 194 only remaining, “corrected” significant test that Goodwin and Podos (2015) can claim, at  $p =$   
 195  $0.03$ , is nonsignificant when it is corrected ( $\alpha = 0.05/3 = 0.017$ ) for the three admitted tests that  
 196 were done. Include the undisclosed tests in the correction and there remains nothing on which to  
 197 base their story. As Simmons, Nelson, and Simonsohn (2011) succinctly write, “Our job as  
 198 scientists is to discover truths about the world . . . [but] . . . Undisclosed flexibility in data  
 199 collection and analysis allows presenting anything as significant.”

200           A fourth serious problem with Goodwin and Podos (2014) is revealed in this seemingly  
 201 innocuous statement: “We created stimuli by increasing or decreasing trill rate while ensuring  
 202 the song was within the observed population range.” This oft-used method, with a single-minded  
 203 focus on trill rate alone, dates back to Podos (1996) and simply creates highly abnormal,  
 204 experimental song stimuli. Consider a song with trill rate of 28, for example (see Figure 4, or  
 205 song 14 in Figure 2); if three of four syllables are replaced with silence, thus reducing the trill  
 206 rate to seven, the experimental trill rate is still “within the observed population range” of trill  
 207 rates and is therefore still considered normal, even though other dimensions of the song are  
 208 highly abnormal and unlike anything a chipping sparrow would ever hear or sing (see also Figure  
 209 9 for an illustration of this problem with swamp sparrows). If trill rates of 21, 14, and 7 are  
 210 created from a wild-type trill rate of 28, those songs become increasingly abnormal, yet only the  
 211 declining “performance value” of the song is considered relevant in performance studies.  
 212 Without acknowledging how abnormal the song stimuli are (see below for pointed critique of  
 213 Moseley, Lahti, and Podos, 2013), reactions to these increasingly abnormal songs are used as  
 214 “evidence that trill rates are a salient vocal feature in the assessment of territorial rivals”  
 215 (Goodwin and Podos, 2014): “. . . aggressive behaviours . . . (approached the speaker more  
 216 closely . . . spent more time within 2 m of the speaker . . . attacked the mount more often) . . .  
 217 were significantly greater in response to *fast trill rates*,” i.e., to *increasingly normal stimuli*  
 218 (emphases mine). Responses to these experimental stimuli reflect only how abnormal they are



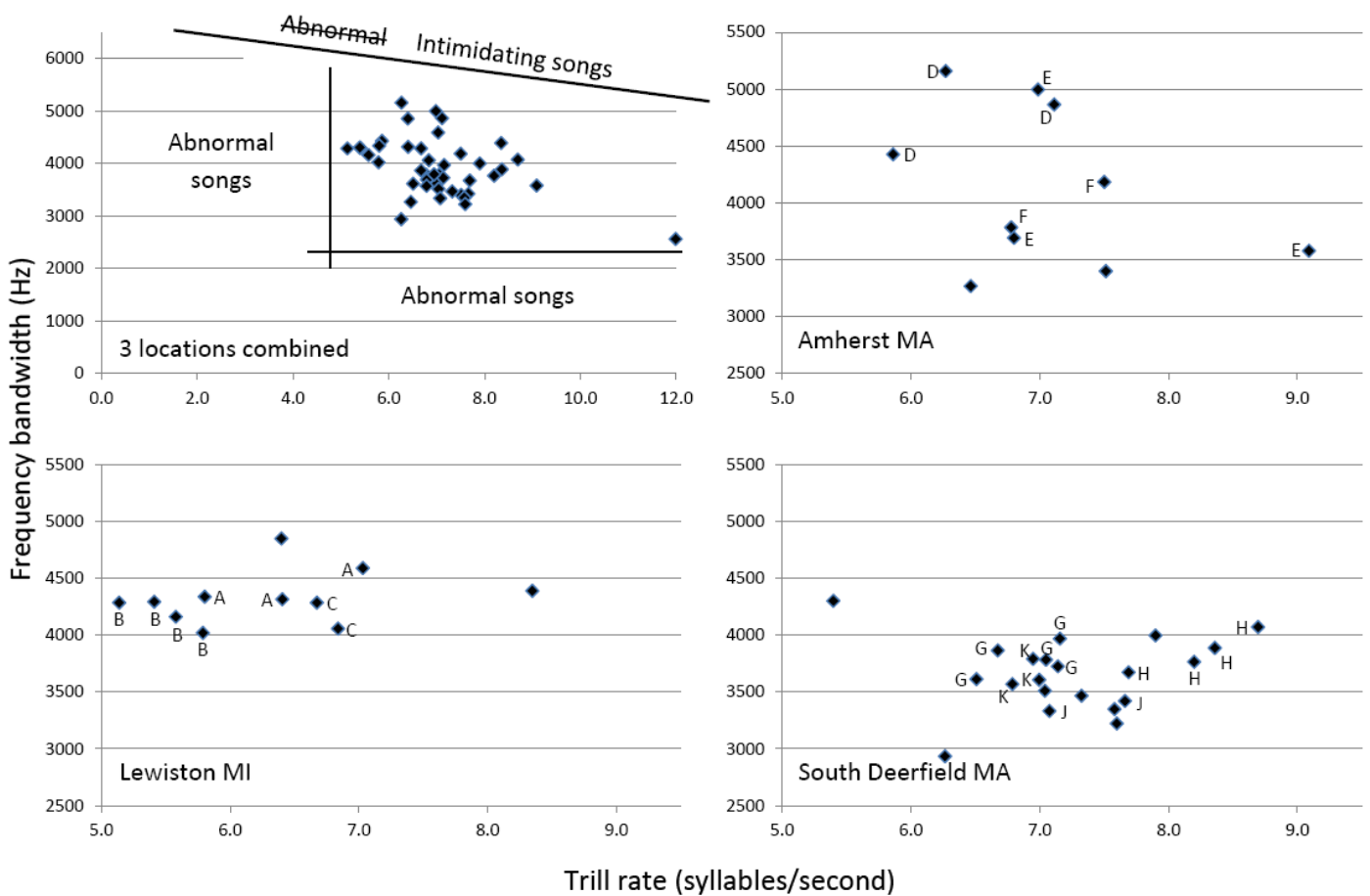
219 (or how much song stimulus is delivered; see also below, for Moseley et al., 2013), not how far  
 220 these strange, experimental songs plot from an upper bound on the performance scale.

221 Still other issues have been addressed by Akçay and Beecher (2015). Those issues,  
 222 together with the four major problems addressed above, illustrate, in part, how the performance  
 223 hypothesis of Podos (1997) has been repeatedly confirmed in the literature since its inception.

224 **SWAMP SPARROWS**

225 *The biology of song learning by swamp sparrows*

226 To illustrate the implications of song learning for swamp sparrows, I recorded birds at  
 227 three locations during 2015 (Figure 6). At each site, I used a stereo Telinga parabolic  
 228 microphone, and either a Sound Devices 722 or Marantz PMD661 digital recorder. Birds were  
 229 unbanded, but each male sang repeatedly over a few hours from the same predictable locations,  
 230 and attributing each recording to a particular male was not difficult; if any doubts existed as to  
 231 the origin of a song, it was discarded from the analyses. Songs were then analyzed on Raven Pro  
 232 1.4 software (settings the same as for chipping sparrows), and the median of three examples of  
 233 each song type from each bird was used in the analyses.



235 Figure 6. Scatterplots of trill rate and frequency bandwidth for swamp sparrows from three locations,  
 236 revealing that scatter is largely due to song types (letters A-K; song types deemed unique to an individual

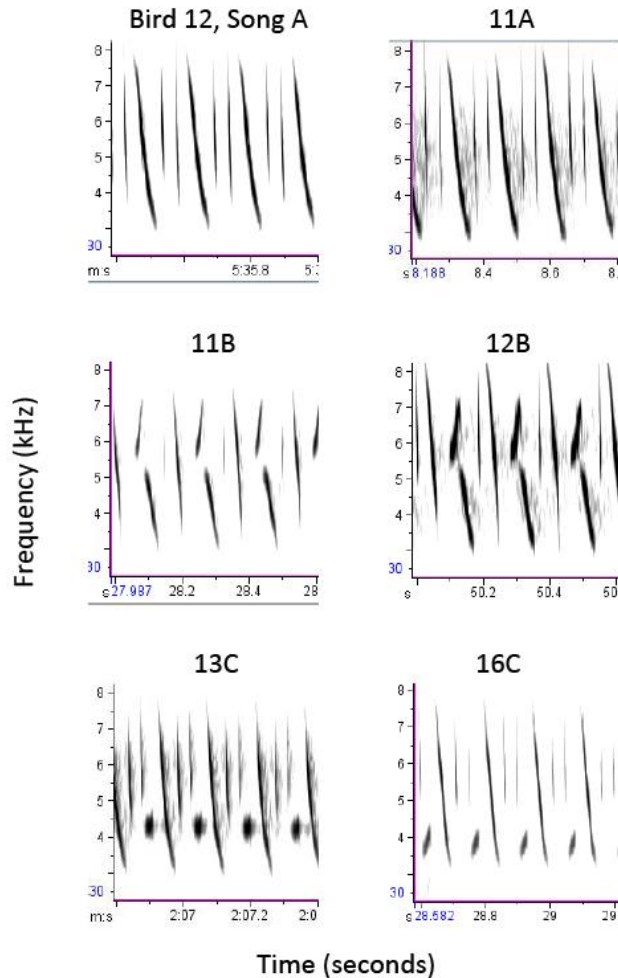
237 are not labeled), leaving little if any information available about the quality of individual singers. In the  
 238 upper left, data for all three locations are combined, and the axes meet at the origin (0, 0); the graphs for  
 239 the three separate locations are drawn to a different scale, expanded to better show the variation within  
 240 locations (note that the outlying data point at 12 syllables/second is omitted from the lower right graph;  
 241 excluding it has no bearing on the conclusions to be drawn from this figure). The upper bound line in the  
 242 upper left subfigure is from Ballentine, Hyman, and Nowicki (2004); the bounds below and to the left of  
 243 the data points are placed arbitrarily. In the three subplots, letters label different renditions of a given  
 244 song type from different males; because of local dialects, song types were not shared across locations.

245 Several important points are revealed in these data (Figure 6):

246 1) Normal, wild-type swamp sparrow songs are restricted to a relatively limited set of all  
 247 possible trill rates and frequency bandwidths (upper left subfigure in Figure 6). Outside of this  
 248 restricted area, all songs are, by definition, abnormal. When bounded lines are added to the  
 249 graph, they draw attention to the limits. Abnormal songs that approach the bound above the sea  
 250 of data are considered supernormal and especially high performance (Podos, 1997), so  
 251 intimidating and threatening that listening males might well flee them (e.g., Illes, Hall, and  
 252 Vehrencamp, 2006); songs to the left and below the normal songs are just abnormal.

253 2) Trill rates and frequency bandwidths can vary significantly by location, depending on  
 254 the local dialect. Frequency bandwidths from Lewiston MI, for example, are mostly above 4  
 255 kHz, those from South Deerfield MA below 4 kHz; trill rates are correspondingly slower at  
 256 Lewiston. Geographic differences occur, no doubt, because males learn the local dialect, which  
 257 varies from place to place.

258 3) Much of the scatter in the data is explained by song types (see also Figure 7): Birds  
 259 learn their songs (including trill rates and corresponding bandwidths) from one another, and as a  
 260 result, many songs are shared within the population, so that songs of the same song type from  
 261 different birds tend to plot near one another (especially clear for Lewiston and South Deerfield).



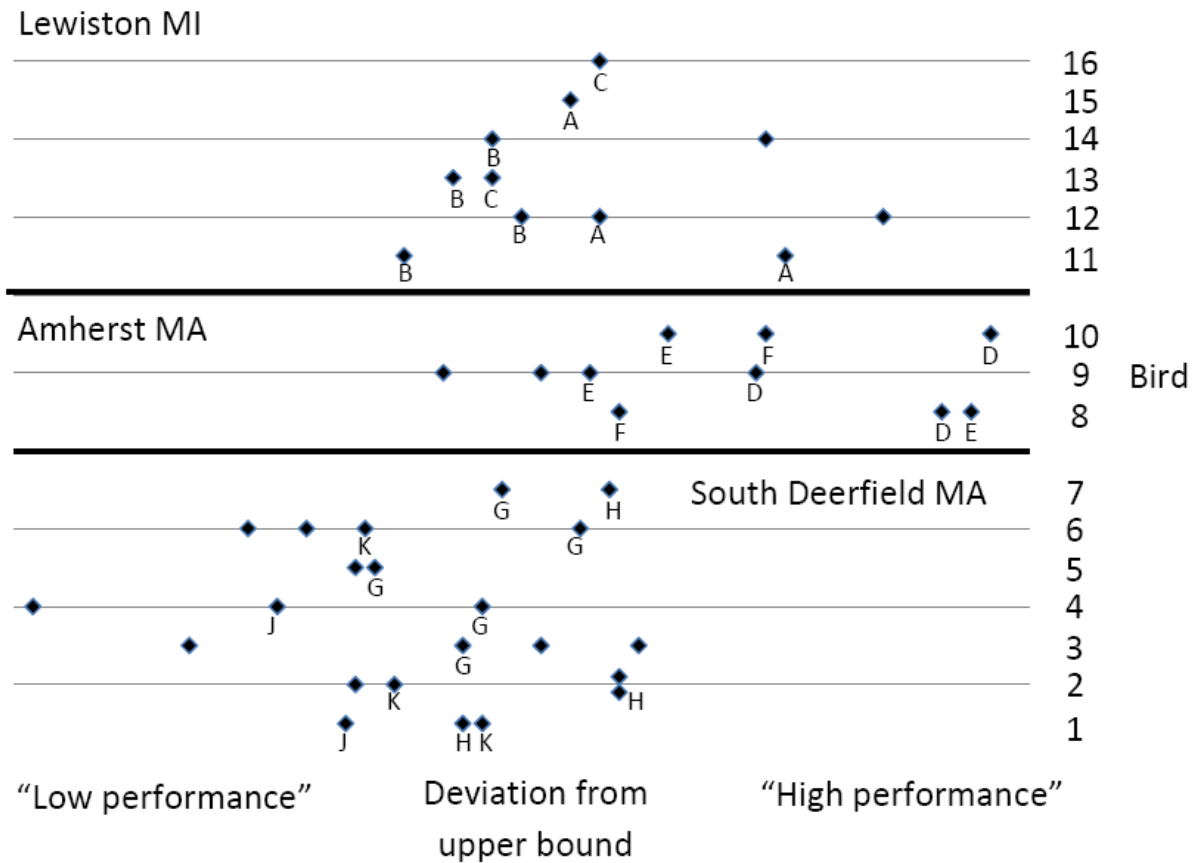
262

263 Figure 7. Swamp sparrow males within a marsh learn their songs from one another and, as a  
 264 result, many songs are shared among birds in a population. From Lewiston, Michigan, two  
 265 examples are illustrated for three different song types (A, B, C). Songs of higher performance  
 266 (lower deviation from the upper bound, as shown in upper left of Figure 6) are in the second  
 267 column. Birds 11 and 12 (first two rows) were immediate neighbors; note that bird 11 had the  
 268 “better” A song, bird 12 the “better” B song.

269 A critical but untested feature of the Podos (1997) performance hypothesis is that songs  
 270 actually provide reliable, honest signals of male quality. If these scatterplots with the upper  
 271 bound are at all relevant to how male and female swamp sparrows might assess a singer, then the  
 272 performance measures must provide consistently reliable information about the singer. If no  
 273 reliable information is provided, the relative performance of different males cannot be used as an  
 274 honest signal of his relative quality.

275 As revealed in Figure 8, performance measures provide no reliable information about  
 276 male quality. Because measures for different males are broadly overlapping, and a given male  
 277 might have both the “best” and the “worst” song in his repertoire, performance measures cannot  
 278 be used by either other swamp sparrows or by humans to assess the relative quality of a singer.

279 The data provide no support for the feasibility of the performance hypothesis of Podos (1997),  
 280 and essentially render the hypothesis false for these swamp sparrows.



281  
 282 Figure 8. Song performance measures (sensu Podos, 1997) can provide no reliable information about  
 283 inherent male quality in swamp sparrows. Data plotted here have been extracted from Figure 6 by  
 284 measuring the shortest distance to the upper bound for each song (all measures are relative,  
 285 depending on a number of factors, so no absolute scale is provided for performance). Performance  
 286 measures for different males are broadly overlapping, such that a male might have the highest  
 287 performance on one of his song types but the lowest on another (e.g., compare bird 3 from South  
 288 Deerfield with birds 1, 2, 5, 6, and 7; song types are lettered as in Figure 6). Note that Amherst swamp  
 289 sparrows significantly “outperform” the South Deerfield birds only 20 kilometers distant (the two  
 290 populations were recorded within a few days of each other during late June 2015; I can think of no  
 291 methodological issues that would have produced such results).

292 *A focused critique of swamp sparrows and performance studies*

293 My conclusion is at odds with all of the published studies on swamp sparrows. In an  
 294 attempt to understand why, I next examine each of those studies in chronological order.

295 **Podos, Peters, and Nowicki. 2004. Calibration of song learning targets during vocal**  
 296 **ontogeny in swamp sparrows, *Melospiza georgiana*.**

297 The basic claim here is that, when a young swamp sparrow learns a given song type, he  
 298 adjusts the trill rate or frequency bandwidth to match his own proficiency at producing that song,  
 299 so as to acquire as high a performance song as he possibly can (i.e., closest to the upper bound on  
 300 the graph). The authors strongly claim that their data are "consistent with" or "support" this  
 301 "calibration hypothesis," a dozen times making statements such as "This bias is consistent with  
 302 the calibration hypothesis . . .," or, summing it all up, "These results were consistent with the  
 303 calibration hypothesis."

304 The impression conveyed by these strong claims is that, given how everything is  
 305 consistent with the calibration hypothesis, it must therefore be true. But the words "consistent  
 306 with" are also red flags for readers to ask what other hypotheses the data might be consistent  
 307 with, or what data are not consistent with the hypothesis.

308 The figures I provide on the biology of swamp sparrow song learning reject the authors'  
 309 conclusions about calibration. When learning a song, a swamp sparrow conforms to the  
 310 particular song type of the local dialect; he does not adjust features of what he learns in any way  
 311 consistent with an attempt to calibrate a song to his own abilities. As a result, a male may have  
 312 what would have to be labeled the worst "performance" on one song type, the best on another.

313 Not once do the authors disclose that all of their data are also consistent with a very  
 314 simple alternative explanation, that no matter what recognizable features of a song a swamp  
 315 sparrow hears, he tries to develop as normal a song as possible, making a fine-tuned effort to  
 316 take whatever he hears and produce a normal, wild-type song (the only logical conclusion also  
 317 for Lahti, Moseley, and Podos, 2013—see review below). Contrary to the title and all of the  
 318 statements in this paper, there is no credible evidence that an individual male "calibrates" songs  
 319 to his particular proficiency.

320 Science requires a balanced interpretation, not careful selection of data or details that  
 321 might support only one's favored explanation. "Details that could throw doubt on your  
 322 interpretation must be given" (Feynman 1985).

323 **Ballentine, Hyman, and Nowicki. 2004. Vocal performance influences female**  
 324 **response to male bird song: an experimental test.**

325 Female swamp sparrows are shown to display more to high-performance than to low-  
 326 performance songs of the same song type recorded from different males. The authors conclude  
 327 the following:

328 . . . we can conclude with certainty that females are attending to subtle differences  
 329 in song reflecting male motor capabilities. Thus, our results provide a crucial  
 330 piece of evidence in support of the general hypothesis that female birds assess  
 331 male quality on the basis of vocal performance.

332 Those who cite this paper in support of their own work are equally convinced: ". . . it is  
 333 known that females prefer trilled songs closer to the production limit" (Illes, Hall, and  
 334 Vehrencamp, 2006; reviewed below); ". . . females are able to assess a male's quality as a  
 335 potential mate using vocal performance" (Dubois, Nowicki, and Searcy, 2011; reviewed below);  
 336 ". . . females . . . are known to discern fine features of song in the functional contexts of mate  
 337 choice" (Lachlan, Anderson, Peters et al., 2014).

338 In fact, “As of May/June 2014, this ‘highly cited paper’ received enough citations to  
 339 place it in the top 1% of the academic field of Plant & Animal Science based on a highly cited  
 340 threshold for the field and publication year” (Web of Science). In a survey entitled “25 Years of  
 341 Behavioral Ecology,” a review article for the journal cites the importance of this paper; Simmons  
 342 (2014) reports that this paper is

343 [one of] . . . the 10 articles from Behavioral Ecology which have received the  
 344 most number of citations . . . females are more likely to solicit copulations from  
 345 males capable of vocalizing at the upper boundaries of the performance limit;  
 346 female swamp sparrows prefer males with the most elaborate sexual displays

347 Now consider this study in the biological context that I have presented above. Realize, for  
 348 example, that if experimental songs were selected from Figure 7, the authors claim that females  
 349 would display more to the songs in the right column than to those in the left column, even though  
 350 males 11 and 12 provide conflicting information with song types A and B. Males are simply too  
 351 inconsistent in their performance measures (*sensu* Podos 1997; see Figure 8), and these measures  
 352 therefore cannot serve as a reliable index of male quality.

353 As with chipping sparrows, the hidden premise is that a swamp sparrow is free to develop  
 354 as high a quality song as it can, so as to impress males or females. Yet all qualities of a song that  
 355 a male learns are constrained by the song type that he is learning, and the supposed performance  
 356 of that song is necessarily limited to the relatively small range of performance values available  
 357 for that song type. In other words, a male will “willingly” sacrifice his “performance status” to  
 358 sing a particular “low-performance song type,” such as male 12 from Lewiston learning low-  
 359 performance song B even though on another song type he has the best in show (Figure 8). It is  
 360 as if only the song type mattered to him and his overall “song proficiency or performance” did  
 361 not.

362 How does one account for the results of this study by Ballentine et al. (2004)? Briefly, I  
 363 offer three possibilities. First, “Believing is seeing,” it might be said, the results stemming from  
 364 nonblind observers with strong expectations for the results (i.e., observer bias). From the outset,  
 365 the concept of “performance” is already a given, not a hypothesis (“our knowledge of song  
 366 production mechanisms allows us to identify a priori which songs are produced with greater  
 367 vocal proficiency”—Introduction), and variations of the word “perform” with its attending  
 368 assumptions are used nearly 100 times throughout this paper (in all “performance” studies, there  
 369 is an inherent assumption and built-in bias that is generated by mere use of the nonneutral word  
 370 “performance”).

371 Second, the equipment used might have been an issue: two different cassette tape  
 372 recorders (Sony TCM-5000 EV and Sony TCD-5M) and, more importantly, two different  
 373 parabolic reflectors, one with 13” diameter (Sony PBR-330) and one with 18” diameter (Saul  
 374 Mineroff SME PR-1000). For a study that wishes to compare variability of songs within and  
 375 among males, and across sessions, and ask females to assess subtle features of these songs, it  
 376 would be essential to insure that all of the recording gear gave identical results. The 13”  
 377 parabolic reflector, for example, is rather nondirectional, and would be expected to record songs  
 378 with considerably more reverberation than the 18” reflector.

379 Third, alternative explanations for results are never considered in this study (and rarely, if  
 380 ever, in other performance studies). Consider only the reverberation that can vary from one  
 381 recording to another, depending on which parabola was used, for example, or how far the bird

382 was away from the microphone, or the amount of intervening vegetation between the singer and  
 383 the microphone. There is a large parallel literature, none of it cited in any of these studies of  
 384 performance, that shows how birds vary their response to playback songs depending on how  
 385 much reverberation is in the recording (e.g., Naguib 1996), prompting Morton, Gish, and Van  
 386 Der Voort (1986) to write the following:

387       Sufficient evidence now exists to suggest that sound degradation . . . should be  
 388       taken into account in studies using responses to playback of bird song.

389 My measures of frequency bandwidths show a decrease of up to 13 percent in reverberated  
 390 recordings. Songs with reverberation could thus be rated “low-performance” songs, in which  
 391 case both females and males would be expected to respond less to them based on reverberation  
 392 alone.

393       **Dubois, Nowicki, and Searcy. 2009. Swamp sparrows modulate vocal performance**  
 394       **in an aggressive context.**

395       The authors’ main conclusion (p. 163, from the Abstract):

396       . . . we show that male swamp sparrows . . . increase the vocal performance of  
 397       individual song types in aggressive contexts by increasing both the trill rate and  
 398       frequency bandwidth.

399       Male swamp sparrows were played either a control song (that of a white-crowned  
 400       sparrow, *Zonotrichia leucophrys*) or an aggressive song (that of a conspecific), and the authors  
 401       then measured the trill rates and frequency bandwidths of the songs delivered in these two  
 402       contexts. Two results stand out:

403       1) The particular song types used in aggressive and control contests did not differ. When  
 404       it matters most, then, when a male is challenged on his territory, he chooses a random song from  
 405       his repertoire, not a song that best conveys his overall quality. This important result is highly  
 406       inconsistent with the performance hypothesis (though not mentioned in the abstract), yet the  
 407       authors puzzlingly conclude “. . . we do not think this means that males are not trying to  
 408       maximize their vocal performance during aggressive signaling . . .”

409       2) The following results are stated in the title of the paper and the abstract: “. . . males  
 410       increased both the trill rate . . . and the frequency bandwidth . . . during the aggressive trial. This  
 411       results in significantly higher vocal performance . . . during the aggressive trial . . .” (results are  
 412       based on an overall average among  $n = 23$  males, with increases of from 6.94 to 7.10  
 413       syllables/sec, 4870.4 to 4960.9 Hz). These authors would later declare that male swamp sparrows  
 414       “actively increase” and “exaggerate” their vocal performance in aggressive situations (Dubois et  
 415       al. 2011).

416       Yet, one must ask, how could it possibly be biologically meaningful to increase the trill  
 417       rate by 2.3 percent or frequency bandwidth by 1.8 percent? In Figure 6, for example, consider a  
 418       song with trill rate of 6.0 syllables/sec and frequency bandwidth of 4000 Hz that is  
 419       “exaggerated” to 6.1 syllables/second and a frequency bandwidth of 4072 Hz. The exaggerated  
 420       data point on the scatterplot is moved a miniscule distance. If a male really wanted to increase  
 421       his performance during aggressive contexts, he could switch to a more impressive song in his  
 422       repertoire, but he doesn’t do that, as if performance did not matter. Moreover, two years later, the  
 423       authors would accept that these “exaggerations” are biologically meaningless (Dubois et al.  
 424       2011; see below).

425 Furthermore, the title of the paper may be true, but it is highly misleading, because  
 426 swamp sparrows also modulate their songs in nonaggressive contexts. Using two lengthy  
 427 recordings from my collection, for example, I measure that trill rates vary from one to three  
 428 percent within a neutral session, spanning the two percent change the authors measured from  
 429 neutral to aggressive contexts. Frequency bandwidth is also modulated within neutral sessions,  
 430 varying by a median of 1.1 percent among the three measures taken from all swamp sparrows  
 431 that I analyzed for this study.

432 It should also be noted that DuBois et al. (2009) measured frequency at a resolution of  
 433 172 Hz, yet the frequency difference between neutral and aggressive contexts was reported as 91  
 434 Hz, about half the magnitude of the measurement error, thus rendering their frequency  
 435 measurements inadequate. Also, frequency measures may have been influenced by how close the  
 436 singer was to the microphone: likely very close during playback in aggressive contexts, more  
 437 distant in neutral contexts.

438

439 **Ballentine. 2009. The ability to perform physically challenging songs predicts age**  
 440 **and size in male swamp sparrows, *Melospiza georgiana*.**

441 The author "used the highest performance song in a male's repertoire to determine  
 442 each male's vocal performance," but that rationale is questionable. As discussed above, a  
 443 necessary condition for honesty and reliability is that males consistently use songs within  
 444 a relatively narrow range of performance abilities. If the performance values of males  
 445 broadly overlap (see my Figure 8), so that a male can rank highest on one of his songs  
 446 and lowest on another (as also revealed in Kagawa and Soma, 2013), and a male doesn't  
 447 even use his "best" songs in aggressive encounters when it matters most (see Dubois et  
 448 al., 2009), it makes little sense to rate a male only by the one song of highest performance  
 449 ability.

450 Also, suppose a female is to use performance, as measured in this paper, to distinguish  
 451 first-year from older birds. For each male that she'd want to assess, she would have to 1) listen to  
 452 his entire repertoire over an extended period, 2) rate and remember each of his song types on the  
 453 performance scale, 3) eventually dismiss as irrelevant all the song types of lowest performance  
 454 value (but why?), and 4) focus only on the one song type that plots closest to the upper bound on  
 455 the scatter plot, because that is the song type to be used to predict this male's age and quality.  
 456 And she would have to accomplish this task for a number of males before making decisions  
 457 about relative male quality.

458 Identifying a first year bird does not require that much effort. Songs of first year birds are  
 459 typically more plastic and less repeatable than those of older birds, and this plasticity alone could  
 460 readily identify a young bird in just a few songs. Merely writing repeatedly that the data  
 461 "support" the hypothesis that birds attend to performance ability, and not mentioning (less  
 462 exciting) alternative explanations, does not make the hypothesis true (see also my above critique  
 463 of Podos et al., 2004).

464 Given that all song types were recorded from all males in this study, the author missed an  
 465 opportunity to show, as I have, that song performance cannot be a reliable measure of male  
 466 quality (my Figures 6 and 8). This important analysis, crucial for the performance hypothesis to  
 467 be true, has been so shunned in all studies over the years.



468

469 **Podos, Lahti, and Moseley. 2009. Vocal performance and sensorimotor learning in**  
 470 **songbirds.**

471 In this review one finds unflagging support for the performance hypothesis:

472 Emerging descriptive and experimental evidence thus suggests that vocal  
 473 performance varies among individuals, and suggests that singers who maximize  
 474 vocal performance gain advantages in song function and ultimately in  
 475 reproductive success (p. 170).

476 I can find no credible scientific evidence to support that conclusion, either in the  
 477 literature up to 2009 or the years to follow.

478

479 **Dubois, Nowicki, and Searcy. 2011. Discrimination of vocal performance by male**  
 480 **swamp sparrows.**

481 Three experiments are performed. In experiment 1, males are asked to discriminate  
 482 between high- and low-performance songs of the same song type as sung by different males.

483 Responses were greater toward high-performance song on all five univariate  
 484 measures, and the differences were significant for three of these . . . This result  
 485 supports our . . . hypothesis that males assess individual differences in vocal  
 486 performance” (p. 722).

487 Three issues can be raised about these conclusions:

488 1. As can be seen in Figures 6 and 8, song types plot in different spaces on the  
 489 scatterplots, because males conform to the features of that song type when learning it. All  
 490 indications are that the conforming is important, not any exaggeration of trill rate or frequency  
 491 bandwidth to reveal one’s prowess on a particular song type.

492 2. The songs used in playbacks are the same songs that were used by Ballentine et al.  
 493 (2004) and Ballentine (2009), and the two parabolic microphones (one 13”, the other 18”) could  
 494 have easily produced different results, with the less directional 13” reflector recording songs with  
 495 greater reverberation and therefore lower performance measures. If songs are low-performance  
 496 because they consistently have higher reverberation, the playback results cannot be attributed to  
 497 the chosen explanation of performance ability.

498 3. When investigators are deeply committed to an hypothesis, it is essential that  
 499 observations be done blindly (but they were not): Was that a loud song (i.e., a “broadcast” song)  
 500 or a “soft song”? And just how far from the speaker is the bird? Flags at 2, 4, and 8 meters “in  
 501 either direction from the speaker” (in two directions?) may help in estimating distance but are no  
 502 guarantee of objectivity, especially when the bird is not directly between flagging, and when  
 503 distances of <16 or >16 meters from the speaker must be estimated.

504 Experiments 2 and 3 are similar to each other, each of them asking if males respond  
 505 differently to the kind of within-male differences in vocal performance observed in DuBois et al.  
 506 (2009), where trill rates and frequency bandwidths increased on average about 2 % from neutral  
 507 to aggressive performances. No significant differences in response were found, i.e., males

508 responded no differently to the “extremes” of high- and low-performance versions of a particular  
509 song type that a given male might sing.

510 It would seem that no matter what results are obtained in experiments 2 and 3, the results  
511 can still be interpreted to support honesty in signaling and the motor constraints hypothesis. If,  
512 during contexts when it matters most (e.g., aggressive encounters), males do not increase their  
513 performance enough to be detected, then the songs are a good “example of an index signal, since  
514 it cannot be effectively cheated” (p. 725); males can thus “garner information about an opponent  
515 based on his vocal performance, whether or not any information is conveyed through modulation  
516 of this characteristic” (p. 726). But if any differences in modulation had been sufficient to be  
517 detected, then listeners could still “garner information about an opponent based on his vocal  
518 performance.” It would seem that there is no way to falsify the motor constraints hypothesis of  
519 Podos (1997), no way that performance cannot honestly convey a male’s quality.

520 **Lahti, Moseley, and Podos. 2011. A tradeoff between performance and accuracy in**  
521 **bird song learning.**

522 Experimental songs are produced by adding or deleting silent intervals between song  
523 elements, yielding songs that swamp sparrows would never by themselves have produced or  
524 heard in nature. Young swamp sparrows are then tutored with these odd songs.

525 Our main finding is that birds elevated the trill rates of low-performance models,  
526 but at the expense of imitative accuracy.

527 The elevation of trill rates of slowed models supports the hypothesis that birds  
528 calibrate learned vocal output to match their individual performance capabilities  
529 (Podos et al. 2004, 2009) . . . Prior work in swamp sparrows showed calibration . .  
530 . (Podos et al. 2004).

531 . . . our data imply that selection has favored birds that . . . [produce] . . . trill rates  
532 that maximize birds’ vocal capabilities . . . A bias toward increasing the  
533 performance level of songs would enable birds to indicate their performance  
534 capacities; otherwise, the quality of a tutor’s song would set a ceiling on the  
535 performance level a learner could attain

536 These interpretations are troublesome. What is certainly true is that the young swamp  
537 sparrows removed silent intervals from odd, slowed tutor songs to produce more normal, wild-  
538 type songs. That result, however, based on abnormal, experimental songs, does not warrant any  
539 conclusion about a young swamp sparrow either in nature or in the laboratory taking a natural  
540 tutor song that it hears, foregoing “imitative accuracy,” and adjusting that song in trill rate or  
541 frequency bandwidth to match his own capabilities, all so that he can honestly broadcast his  
542 individual quality. There are no data in this paper or elsewhere demonstrating that a young  
543 swamp sparrow adjusts a normal or abnormal song to match his own individual proficiency, only  
544 data showing how young birds strive to produce normal, species-typical songs. There are no data  
545 that demonstrate any honesty in the signaling of trill rate or frequency bandwidth.

546 One hint that the authors perceive the conflict between their data and their interpretation,  
547 however briefly, is provided in the following quote (*italics mine*):

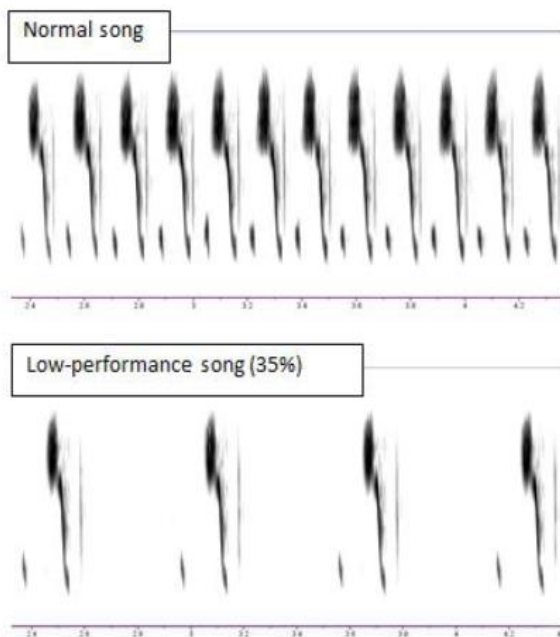
548 . . . our results reveal that vocal ontogeny can be shaped . . . by a premium on high  
549 performance. Again, performance in this case refers to the trill rate of songs, all

550 other features being equal, and *high performance being that of typical songs*  
 551 *recorded from the field* as compared with our experimentally slowed versions.

552 Except for these 11 words, throughout the paper “performance” refers to vocal  
 553 proficiencies of individuals, to an individual adjusting a tutor’s song to the “best” song he  
 554 possibly can produce (i.e., relatively fast trill and broad frequency bandwidth), thus revealing his  
 555 individual proficiency and quality, as in the first sentence of this quote. Puzzlingly, it is as if  
 556 another author speaks, suggesting that high performance should be defined simply as a normal,  
 557 wild-type song, that calibration is to some species-typical trait, not to some individual measure of  
 558 proficiency—but that voice then disappears.

559 **Moseley, Lahti, and Podos. 2013. Responses to song playback vary with the vocal**  
 560 **performance of both signal senders and receivers.**

561 The authors use time-honored methods (e.g., Podos et al. 2004, Lahti et al. 2011) to  
 562 produce highly abnormal test stimuli: Two-second songs for playback to swamp sparrows are  
 563 prepared from normal songs by either inserting or deleting silent spaces between the song  
 564 elements; the manipulated songs then contain anywhere from 35% (a “low performance” song)  
 565 to 155% (a “high performance” song) of the elements in control songs, with trill rates for those  
 566 particular songs thus ranging from 35% to 155% of normal. As is evident in their figure 1 and  
 567 my Figure 9, three obvious features of the songs have changed from the original song: 1) the trill  
 568 rate is slower or faster, 2) the quantity of stimulus is correspondingly less or more, and 3) the  
 569 more silence edited into or out of the song, the more abnormal it is, unlike anything a swamp  
 570 sparrow has ever sung or would hear.



571  
 572 Figure 9. To produce playback stimuli, Moseley et al. (2013) altered a normal song (top) by adding or  
 573 deleting silent spaces between the song elements. In this illustration (bottom), silent intervals are  
 574 increased 8.3 times, producing a highly abnormal song with four syllables delivered at 35% the rate of  
 575 normal.

576

577           The two confounding variables are a serious problem. First, it is entirely reasonable to  
 578 expect that a “normal” song with three times as much stimulus as the lowest performance song  
 579 might elicit a stronger response, based on stimulus quantity alone. But the authors offer no  
 580 control for such an alternative explanation for their results. Without somehow controlling for this  
 581 confounding factor of stimulus quantity, one cannot attribute response strength to trill rate alone.

582           Second, the stimulus songs are simply highly abnormal. To a swamp sparrow, a song  
 583 slowed to 35% of normal must sound odd indeed, heard as a staccato, halting sequence of  
 584 perhaps familiar song elements all out of sync, as these songs fall far outside the range of what  
 585 any swamp sparrow would ever sing in nature. Two years before, in fact, these same authors  
 586 (Lahti et al., 2011) had concluded that songs with trill rates below 55% or above 115% of normal  
 587 were “so unlike typical swamp sparrow songs that males do not consider them as targets for  
 588 learning,” i.e., they are so abnormal as to not be biologically meaningful; nevertheless, Moseley  
 589 et al. use songs well outside that range.

590           The confounding variables of stimulus quantity and abnormality are of no concern to the  
 591 authors when they conclude the following:

592           territorial male swamp sparrows responded significantly less strongly to low-performance  
 593 than to control-performance playback stimuli, consistent with our hypothesis that  
 594 receivers should attribute limited threat to low-performance songs . . .

595           When that conclusion is rewritten to focus on one of the confounding variables, it  
 596 becomes uninteresting and almost certainly unpublishable:

597           territorial male swamp sparrows responded significantly less strongly to abnormal than to  
 598 normal playback stimuli, consistent with our hypothesis that receivers should attribute  
 599 limited threat to abnormal songs . . .

600           The authors found a relationship between the vocal performance of the responding male  
 601 and how aggressively he responded to the playback stimuli. It is difficult to understand how that  
 602 relationship could be true (see Figures 6-8), given that 1) males have several song types in their  
 603 repertoire, 2) those song types vary widely in vocal performance, 3) such that song performance  
 604 offers no reliable indication of male quality (Figure 8), 4) the particular song type a male chooses  
 605 to use in aggressive contexts is random with respect to the purported vocal performance  
 606 capabilities of that male (DuBois et al., 2009), 5) the song he does use is not exaggerated in  
 607 performance in any detectable way, and 6) and the authors measured the vocal performance of  
 608 the responding male only by that one randomly chosen song he used during the playback  
 609 responses. Nevertheless, all previous studies notwithstanding, and in contradiction to the basic  
 610 biology of swamp sparrow singing behavior, the authors “predicted that subjects’ tendencies to  
 611 engage simulated intruders would vary positively with their own vocal performance,” and then  
 612 proved it.

613           The logic is troubling throughout this paper. Two examples suffice:

614           1) “. . . we predicted that stimuli with performance levels increased slightly would be  
 615 responded to aggressively, whereas stimuli increased to the highest performance levels would be  
 616 avoided, because of the higher perceived risk” of a supernormal stimulus (from Introduction).  
 617 The highest performance songs, by the authors’ definition, can also be the most abnormal, yet the  
 618 authors argue that these highly abnormal songs cause subject males to flee. The authors do not

619 explain, however, how they can distinguish between failing to respond to a highly abnormal  
620 stimulus and fleeing a strong stimulus.

621         2) The Discussion is a string of ad hoc explanations for why males

622             a) might not respond strongly to low-performance (abnormal) songs (e.g., low threat from  
623 low quality intruding male who is no threat in extra-pair matings for the resident male),

624             b) might respond strongly to high performance (perhaps relatively normal) songs (high  
625 threat for loss of paternity to intruding superior male), or

626             c) might not respond strongly to even higher performance (perhaps most abnormal)  
627 songs, because then the responding male should flee, though now the apparent lack of response  
628 to the stimulus is because the test stimulus is high threat, not low threat as before.

629         In spite of all these issues, the authors conclude the following:

630         Taken together, our results provide a novel line of support for the hypothesis that  
631         vocal performance provides a reliable signal of aggressive threat . . . Most  
632         broadly, our data contribute to a general understanding of how animals respond to  
633         signals or signalers that are threatening.

634

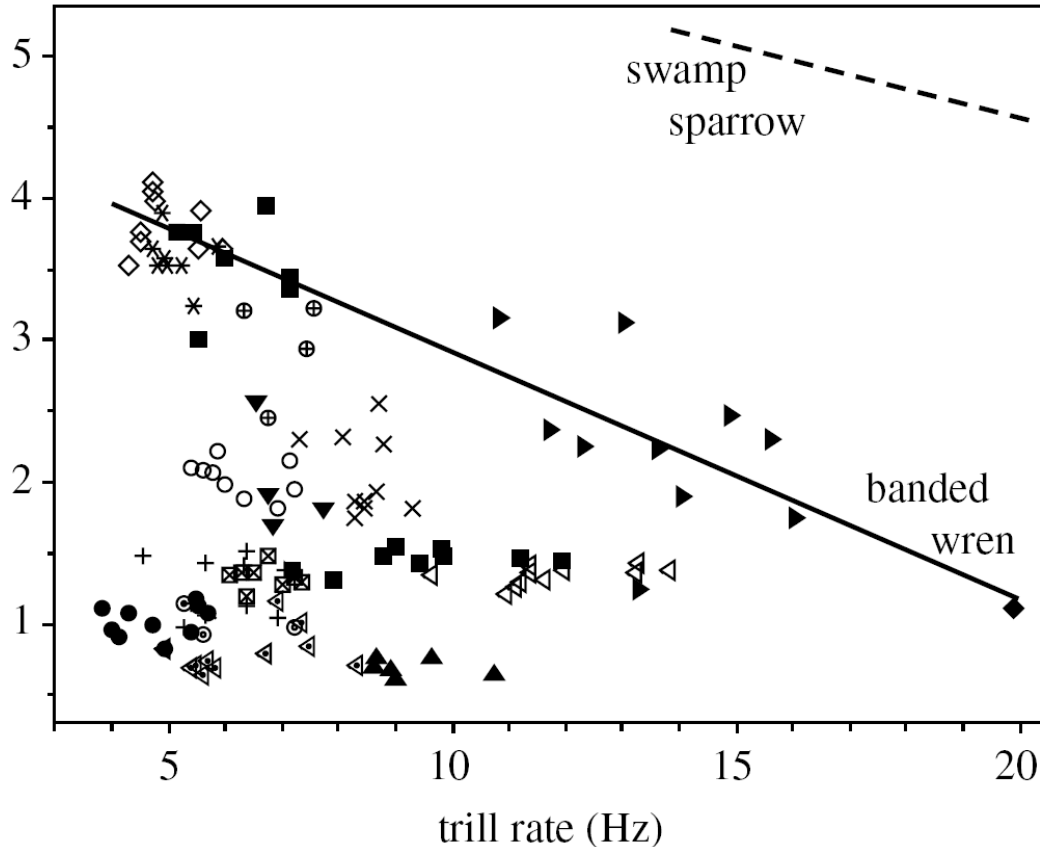
## 635 **OTHER SPECIES**

636 I offer comments on just a few more papers, although I have reviewed many others searching for  
637 evidence supporting the performance hypothesis (e.g., Cramer and Price 2007, DeKort et al.  
638 2009, Juola and Searcy 2011, Cramer et al. 2011, Cramer 2013, Sprau et al. 2013, Kagawa and  
639 Soma 2013).

640         **Illes, Hall, and Vehrencamp. 2006. Vocal performance influences male receiver**  
641         **response in the banded wren.**

642         Given that each male banded wren (*Thryothorus pleurostictus*) has about 20 different  
643 songs, each learned from other males in the local dialect, the scatterplot of frequency bandwidth  
644 and till rate contains a wealth of information (Figure 10). Foremost, to me at least, it reveals  
645 great variation in “performance” among different song types. Some song types are low  
646 performance, some high performance, so that like swamp sparrows, the scatter in the plot seems  
647 dictated by song type, having little if anything to do with individual differences in performance.

648



649

650 Figure 10. From Illes et al. (2006). “Graph of trill rate versus frequency bandwidth for 695 trills recorded  
 651 from 13 individuals and including 16 trill types. Each symbol type represents a different trill type. The  
 652 banded wren upper-bound limit is shown with a solid line and the swamp sparrow limit (Ballentine et al.  
 653 2004) with a dashed line.”

654 Given the biological basis for this kind of graph for chipping sparrows and swamp  
 655 sparrows (see above), it is difficult to understand how the authors could 1) invest an enormous  
 656 amount of energy in testing for a possible difference in how birds might respond to two slightly  
 657 different versions of a given song type (if performance were really important, for example, why  
 658 not compare a low-performance song type with a high-performance song type? and why are so  
 659 many low-performance song types maintained in the population?); 2) report statistically  
 660 significant differences in responses; 3) interpret an apparent reduced response to a song as the  
 661 strongest possible response (songs closer to the upper bound “posed a threat so extreme that they  
 662 effectively repelled rivals, even territory owners”; p. 1911); and then 4) expect discriminating  
 663 readers to believe that result.

664 Also puzzling is the narrow scope of the one statistical test that is presented in support of  
 665 the performance hypothesis. Given the enormous complexity of the playback design and all of  
 666 the ways in which a male’s response could have been evaluated, the authors present only one  
 667 statistical test: “The 16 males that entered the 10 m fast circle at some point during the trial spent  
 668 less time there the higher the performance score of their stimulus trill . . .  $p = 0.020$  . . .” To be  
 669 comfortable that result, one would need to know how many other undisclosed statistical tests had  
 670 been done that were not significant, to determine if one needed to adjust for multiple

671 comparisons. Given how undisclosed methods “allows presenting anything as significant”  
 672 (Simmons et al., 2011), authors will enhance their credibility only if they explicitly state all they  
 673 have done to study their results.

674 For all of the reasons laid out for the swamp sparrow papers, the results and conclusions  
 675 of this paper are questionable. Its results are not questioned, however, by the community of  
 676 biologists who cite it so frequently in the literature, 79 times as of February 2016.

677 **Vehrencamp, Yantachka, Hall, and De Kort. 2013. Trill performance components**  
 678 **vary with age, season, and motivation in the banded wren.**

679 From the opening sentence of the Abstract (emphases mine):

680 Acoustic displays with *difficult-to-execute* sounds are often subject to *strong*  
 681 *sexual selection* because *performance levels* are related to the sender’s *condition*  
 682 or genetic *quality* (p. 409).

683 This sentence squarely places the context and rationale for this study in the realm of  
 684 performance and honesty in signaling, with “difficult-to-execute” sounds revealing male quality.  
 685 Everything will be interpreted in this context, yet there is no obvious scientific justification for  
 686 doing so. According to the scatterplot of trill rate and bandwidth for banded wrens (Figure 10),  
 687 for example, relatively few songs are difficult to execute as defined in this performance context,  
 688 because most songs fall far from the upper bound on the graph. Furthermore, every male  
 689 “willingly” learns those “low-performance,” easy-to-execute songs in order to have particular  
 690 song types in his repertoire, *as if performance did not matter*. There is no obvious selection for  
 691 high-performance, difficult-to-execute songs that honestly reflect male condition or quality, only  
 692 selection to conform to the local dialect of particular song types.

693 **Cardoso, Atwell, Ketterson, and Price. 2009. Song types, song performance, and the**  
 694 **use of repertoires in dark-eyed juncos (*Junco hyemalis*).**

695 We found low but significant correlations of performance measures among the  
 696 song types of individual males. This contrasts with highly consistent differences  
 697 in performance among song types, regardless of which males sing them (p. 901) .  
 698 . . The main conclusion from our results is that, because most of the variation in  
 699 performance depends on the song type, a receiver that compares a few song types  
 700 from different males is likely to obtain little information about performance  
 701 differences between males . . . (p. 905)

702 Here is the analysis for which I have been yearning, and the conclusion is much the same  
 703 that one inevitably comes to when looking at my analyses of chipping sparrows and swamp  
 704 sparrows (Figures 1-8), and the figures in Illes et al. (2006), Liu and Kroodsma (2006), and  
 705 Kagawa and Soma (2013). What matters most to these singing males is to have a song type like  
 706 other birds in the population, and the relative performance abilities in singing that particular song  
 707 type are almost certainly irrelevant.

708 **Cardoso, Atwell, Hu, Ketterson, and Price. 2012. No correlation between three**  
 709 **selected trade-offs in birdsong performance and male quality for a species with song**  
 710 **repertoires.**

711 Here is the same message, that performance of songs as plotted on the graph of trill rate  
 712 and frequency bandwidth (Podos 1997) has little predictive value:

713 These results complement a previous study on this population showing that most  
 714 variation in performance is found among song types rather than among males.  
 715 Collectively, the lack of association between trade-off-based aspects of song  
 716 performance and male age or condition, plus variation among song types that  
 717 interferes with rapid assessment of a male's best performance, indicate that these  
 718 aspects of song performance do not allow a good assessment of male quality in  
 719 juncos, and perhaps more generally in species with song repertoires [p. 584; and I  
 720 would add any species without repertoires as well]

721 The overall work of Cardoso et al. has been critically belittled (Zollinger, Podos, Nemeth  
 722 et al., 2012) because of how frequency bandwidths were measured (manually from sonagrams).  
 723 In a wide-ranging critique, Cardoso et al. are instructed on 1) proper measurements and  
 724 methodology, 2) interpretation of data, 3) validity of results, 4) experimental rigor, 5) alternative  
 725 explanations and hypotheses for data, 6) the ability to reject hypotheses, 7) appropriate use of  
 726 skepticism, 8) problems in published papers that “undermine the validity of the results reported  
 727 and the conclusions reached,” and 9) “basic principles” of science. These authors are concerned,  
 728 more broadly, with 10) how papers failing on these measures will “have a profound adverse  
 729 effect on the way the research field is viewed by the rest of the scientific community.” I  
 730 profoundly agree with these prescriptions for good science and the consequences of bad science,  
 731 and the scientific community should in fact have a profound adverse reaction to the kinds of  
 732 birdsong performance studies that I review here.

733 What I find surprising is that Cardoso et al. (2009, 2012) have, in my opinion, offered (or  
 734 at least hinted at) a new and correct interpretation of the trill rate/bandwidth scatterplot, but that  
 735 contribution to science has gone unrecognized. Instead, these authors have been beaten down by  
 736 a technicality, on how frequency bandwidth was measured. For the 14 citations of Cardoso et al.  
 737 (2009) listed in Web of Science by authors other than Cardoso himself (October 2015), for  
 738 example, no one mentions that Cardoso et al. have offered a fundamentally different  
 739 interpretation for the significance of the scatter plot that was introduced by Podos (1997) and  
 740 confirmed repeatedly since then.

741

## 742 **SUMMARY**

743 The scatter in the trill rate/bandwidth plot of Podos (1997) is readily explained by social  
 744 factors and song learning: Males strive to acquire a species-typical song, often imitating rather  
 745 precisely typical songs that are representative of the local dialect. This biological basis for song  
 746 learning renders the performance hypothesis implausible, if not impossible.

747 No compelling evidence suggests that either males or females attend to trill rate,  
 748 frequency bandwidth, or a combination of the two in assessing the quality of the singer, i.e., no  
 749 compelling evidence supports the performance hypothesis.

750

## 751 **DISCUSSION**

752 In the spirit of a Forum article, I will let others begin the Discussion. I have only two  
 753 requests of those who might respond.



754 First, please first read the following two items:

755 1) pp. 338-346 in Feynman (1985; also available here,  
756 <http://www.lhup.edu/~DSIMANEK/cargocul.htm>, and other places)

757 2) Gitzen (2007)

758 Second, if anyone chooses to defend the performance hypothesis, please identify the  
759 strongest support for this hypothesis that has been published.

760

## 761 **ACKNOWLEDGMENTS**

762 I thank Pavel Linhart, Gonçalo Cardoso, David Lahti, and especially Becky Cramer for  
763 responding to an early draft of my document. Others who have provided helpful advice were  
764 Caglar Akçay, Mike Beecher, Bruce Byers, Sylvia Halkin, Ted Miller, Gene Morton, and Walter  
765 Berry.

766

## 767 **REFERENCES**

768 Akçay, C., and M. D. Beecher. 2015. Team of rivals in chipping sparrows? A comment on  
769 Goodwin & Podos. *Biology Letters*. 11:20141043.

770 Ballentine, B. 2009. The ability to perform physically challenging songs predicts age and size in  
771 male swamp sparrows, *Melospiza georgiana*. *Animal Behaviour* 77:973-978.

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

774 Byers, B. E., and D. E. Kroodsma. 2009. Female mate choice and songbird song repertoires.  
775 *Animal Behaviour* 77:13-22.

776 Cardoso, G. C., J. W. Atwell, E. D. Ketterson, and T. D. Price. 2009. Song types, song  
777 performance, and the use of repertoires in dark-eyed juncos (*Junco hyemalis*). *Behavioral*  
778 *Ecology* 20:901-907.

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

782 Cramer, E. R. A. 2013. Physically challenging song traits, male quality, and reproductive success  
783 in house wrens. *PLoS ONE* 8(3): e59208. doi:10.1371/ journal.pone.0059208

784 Cramer, E. R. A., and J. J. Price. 2007. Red-winged blackbirds *Agelaius phoeniceus* respond  
785 differently to song types with different performance levels. *Journal of Avian Biology* 38:122-  
786 127.

787 Cramer, E. R. A., M. L. Hall, S. R. De Kort, I. J. Lovette, and S. L. Vehrencamp. 2011.  
788 Infrequent extra-pair paternity in the banded wren, a synchronously breeding tropical passerine.  
789 *Condor* 113:637-645.

790 De Kort, S. R., E. R. B. Eldermire, E. R. A. Cramer, and S. L. Vehrencamp. 2009. The deterrent  
791 effect of bird song in territory defense. *Behavioral Ecology* 20:200-206.

- 792 Dubois, A. L., S. Nowicki, and W. A. Searcy. 2009. Swamp sparrows modulate vocal  
793 performance in an aggressive context. *Biology Letters* 5:163-165.
- 794 Dubois, A. L., S. Nowicki, and W. A. Searcy. 2011. Discrimination of vocal performance by  
795 male swamp sparrows. *Behavioral Ecology and Sociobiology* 65:717-726.
- 796 Feynman, R. P. 1985. *Surely You're Joking, Mr. Feynman!* W. W. Norton & Company, New  
797 York City.
- 798 Gitzen, R. A. 2007. The dangers of advocacy in science. *Science* 317:748-748.
- 799 Goodwin, S. E., and J. Podos. 2014. Team of rivals: alliance formation in territorial songbirds is  
800 predicted by vocal signal structure. *Biology Letters* 10:Article Number: 20131083.
- 801 Goodwin, S. E., and J. Podos. 2015. Reply to Akçay & Beecher: yes, team of rivals in chipping  
802 sparrows. *Biology Letters*. 11:20150319.
- 803 Illes, A. E., M. L. Hall, and S. L. Vehrencamp. (2006). Vocal performance influences male  
804 receiver response in the banded wren. *Proceedings of the Royal Society B-Biological Sciences*  
805 273(1596): 1907-1912.
- 806 Juola, F. A., and W. A. Searcy. 2011. Vocalizations reveal body condition and are associated  
807 with visual traits in great frigatebirds (*Fregata minor*). *Behavioral Ecology and Sociobiology*  
808 65:2297–2303.
- 809 Kagawa, H., and M. Soma. 2013. Song performance and elaboration as potential indicators of  
810 male quality in Java sparrows. *Behavioural Processes* 99:138-144.
- 811 Kroodsma, D. E. 2005. *The Singing Life of Birds. The Art and Science of Listening to Birdsong.*  
812 Houghton-Mifflin Co., Boston, Massachusetts.
- 813 Lachlan, R. F., R. C. Anderson, S. Peters, W. A. Searcy, and S. Nowicki. 2014. Typical versions  
814 of learned swamp sparrow song types are more effective signals than are less typical versions.  
815 *Proceedings of the Royal Society B-Biological Sciences* 281:Article Number: 20140252.
- 816 Lahti, D. C., D. L. Moseley, and J. Podos. 2011. A tradeoff between performance and accuracy  
817 in bird song learning. *Ethology* 117:802-811.
- 818 Liu, W.-c. 2004. The effect of neighbours and females on dawn and daytime singing behaviours  
819 by male chipping sparrows. *Animal Behaviour* 68:39-44.
- 820 Liu, W.-c., and D. E. Kroodsma. 1999. Song development by field sparrows (*Spizella pusilla*)  
821 and chipping sparrows (*Spizella passerina*). *Animal Behaviour* 57:1275-1286.
- 822 Liu, W.-c., and D. E. Kroodsma. 2006. Song learning by chipping sparrows: when, where, and  
823 from whom. *The Condor* 108:509-517.
- 824 Morton, E. S., S. L. Gish, and M. Van Der Voort. 1986. On the learning of degraded and  
825 undegraded songs in the Carolina wren. *Animal Behaviour* 34:815-820.
- 826 Moseley, D. L., D. C. Lahti, and J. Podos. 2013. Responses to song playback vary with the vocal  
827 performance of both signal senders and receivers. *Proceedings of the Royal Society B-Biological*  
828 *Sciences* 280. Issue: 1768 Article Number: 20131401
- 829 Naguib, M. 1996. Auditory distance estimation in song birds: Implications, methodologies and  
830 perspectives. *Behavioural Processes* 38:163-168.

- 831 Nowicki, S., D. Hasselquist, S. Bensch, and S. Peters. 2000. Nestling growth and song repertoire  
 832 sire in great reed warblers: evidence for song learning as an indicator mechanism in mate choice.  
 833 Proceedings of the Royal Society of London Series B-Biological Sciences. 267:2419-2424.
- 834 Podos, J. 1997. A performance constraint of the evolution of trilled vocalizations in a songbird  
 835 family (Passeriformes: Emberizidae) Evolution 51:537-551.
- 836 Podos, J., D. C. Lahti, and D. L. Moseley. 2009. Vocal performance and sensorimotor learning in  
 837 songbirds, p. 159-195. *In: Advances in the Study of Behavior*, Vol. 40. M. Naguib, K.  
 838 Zuberbuhler, N. Clayton, and V. Janik (eds.). Elsevier Academic Press, Inc., San Diego,  
 839 California.
- 840 Podos, J., S. Peters, and S. Nowicki. 2004. Calibration of song learning targets during vocal  
 841 ontogeny in swamp sparrows, *Melospiza georgiana*. Animal Behaviour 68:929-940.
- 842 Schmidt, R., H. P. Kunc, V. Amrhein, and M. Naguib. 2008. Aggressive responses to broadband  
 843 trills are related to subsequent pairing success in nightingales. Behavioral Ecology. 19:635-641.
- 844 Simmons, J. P., L. D. Nelson, and U. Simonsohn. 2011. False-positive psychology: undisclosed  
 845 flexibility in data collection and analysis allows presenting anything as significant. Psychological  
 846 Science. 22:1359-1366.
- 847 Simmons, L. W. 2014. 25 years of Behavioral Ecology. Behavioral Ecology. 25:1-3.
- 848 Sprau, P., T. Roth, V. Amrhein, and M. Naguib. 2013. The predictive value of trill performance  
 849 in a large repertoire songbird, the nightingale *Luscinia megarhynchos*. Journal of Avian Biology  
 850 44:576-574.
- 851 Vehrencamp, S. L., J. Yantachka, M. L. Hall, and S. R. De Kort. 2013. Trill performance  
 852 components vary with age, season, and motivation in the banded wren. Behavioral Ecology and  
 853 Sociobiology 67:409-419.
- 854 Zollinger, S. A., J. Podos, E. Nemeth, F. Goller, and H. Brumm. 2012. On the relationship  
 855 between, and measurement of, amplitude and frequency in birdsong. Animal Behaviour 84:E1-  
 856 E9.