

1                   No Support for Honest Signaling in Birdsong and the  
2                                   Motor Constraints Hypothesis

3                                   Donald Kroodsma

4                                   52 School Street, Hatfield MA 01038

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

6                                   Phone: 413-247-3367

7                                   Word count: 15,488

8  
9    **ABSTRACT**

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

17           The scatter in the graph for songbirds is explained not by male quality but largely, if not  
18 entirely, by social factors and song learning. When songbirds learn their songs from each other,  
19 multiple males in a neighborhood will sing the same song type. The need to conform to the local  
20 dialect of song types guides males to learn a typical example of each song type for that  
21 population, not to take a memorized song and diminish or exaggerate it in trill rate or frequency

22 bandwidth to honestly demonstrate his relative prowess. When data in this scatter plot are coded  
23 both by song type and by male, it is the song type and the need to conform that explains the  
24 variability, not the quality of different males. There is no reliable information in the song  
25 performance measures that can be used to evaluate a singing male.

26           Understanding how and why birds sing will require an open and honest attempt to falsify  
27 multiple working hypotheses, rather than attempts to confirm a favored one.

## 28 **KEY WORDS**

29 Birdsong, performance, confirmation, advocacy, honesty, integrity, chipping sparrow, swamp  
30 sparrow

## 31 **INTRODUCTION**

32           Something in how a male songbird delivers his songs must convey something about his  
33 relative quality to those who listen, especially females, but identifying those somethings has  
34 proven challenging. It has been argued that those who study sexual selection seem to be  
35 particularly challenged when it comes to testing hypotheses and considering possible alternatives  
36 (e.g., Prum 2010, 2012). In the study of birdsong repertoires and female choice, for example, it  
37 has been repeatedly confirmed that larger repertoires are better, but in spite of a host of studies  
38 confirming the relationship, there is no credible scientific evidence that males or females attend  
39 to the number of different songs a male can sing (Byers and Kroodsma 2009).

40           Another idea that has over the last decade gained much traction is the motor constraints  
41 hypothesis, or the performance hypothesis, of Podos (1997). Scatter plots of trill rates and  
42 frequency bandwidths show an inverse relationship, the more rapid the trill, the narrower the

43 bandwidth (see Figures 4 and 5 for examples). Blank areas with no data beyond an upper bound  
 44 suggest a motor constraint, i.e., the birds can't produce those combinations of trill rates and  
 45 bandwidths (but see Figure 4). The interesting hypothesis is that how close a song plots to the  
 46 upper bound might reveal the difficulty of producing that song, so that songs near the upper  
 47 bound honestly reveal a high quality singer; both prospective mates and competing males might  
 48 then use those high-performance songs to detect high quality singers.

49 This hypothesis has “been adopted widely in tests of song function [Podos et al. 2009]”  
 50 (Goodwin and Podos 2015), and has been repeatedly confirmed over the past decade. My careful  
 51 scrutiny of those studies here, however, reveals that the hypothesis has become largely an  
 52 assumption, never truly tested, and simply implausible. Here I review the confirming studies,  
 53 beginning with the most recent paper, on chipping sparrows, because it reveals especially clearly  
 54 the methods used to confirm the hypothesis; I then proceed to the many studies of swamp  
 55 sparrows, before reviewing other species.

## 56 **CHIPPING SPARROW**

57 **Goodwin, S. E., and J. Podos. 2014. Team of rivals: alliance formation in territorial**  
 58 **songbirds is predicted by vocal signal structure**

59 The claims made in this paper are substantial, and novel, beginning with the title (above),  
 60 and continuing in the Abstract:

61 Our results provide the first evidence that animals like chipping sparrows rely on  
 62 precise assessments of mating signal features, as well as relative comparisons of  
 63 signal properties among multiple animals in communication networks, when

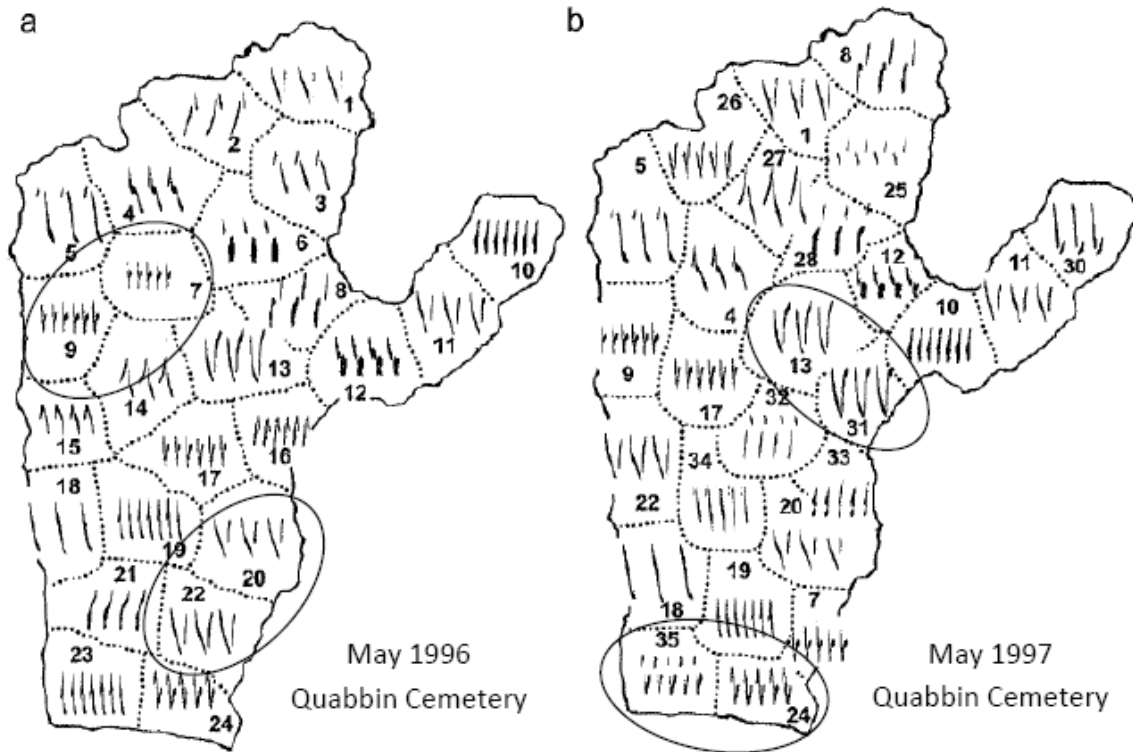
64           deciding when and with whom to form temporary alliances against a backdrop of  
65           competition and rivalry.

66           According to Goodwin and Podos, a chipping sparrow male precisely assesses the  
67           relative prowess of another male based on the trill rate of his song, the faster the trill the higher  
68           the quality of the male. Based on trill rate, a male then decides with whom to form a coalition  
69           when defending a territory.

70           These claims are vigorously defended in Goodwin and Podos (2015; see also Akçay and  
71           Beecher 2015). Before refuting those claims, I first summarize what was already known about  
72           chipping sparrow song development and singing behavior when Goodwin and Podos (2014) was  
73           published.

74           *How a chipping sparrow acquires his song*

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



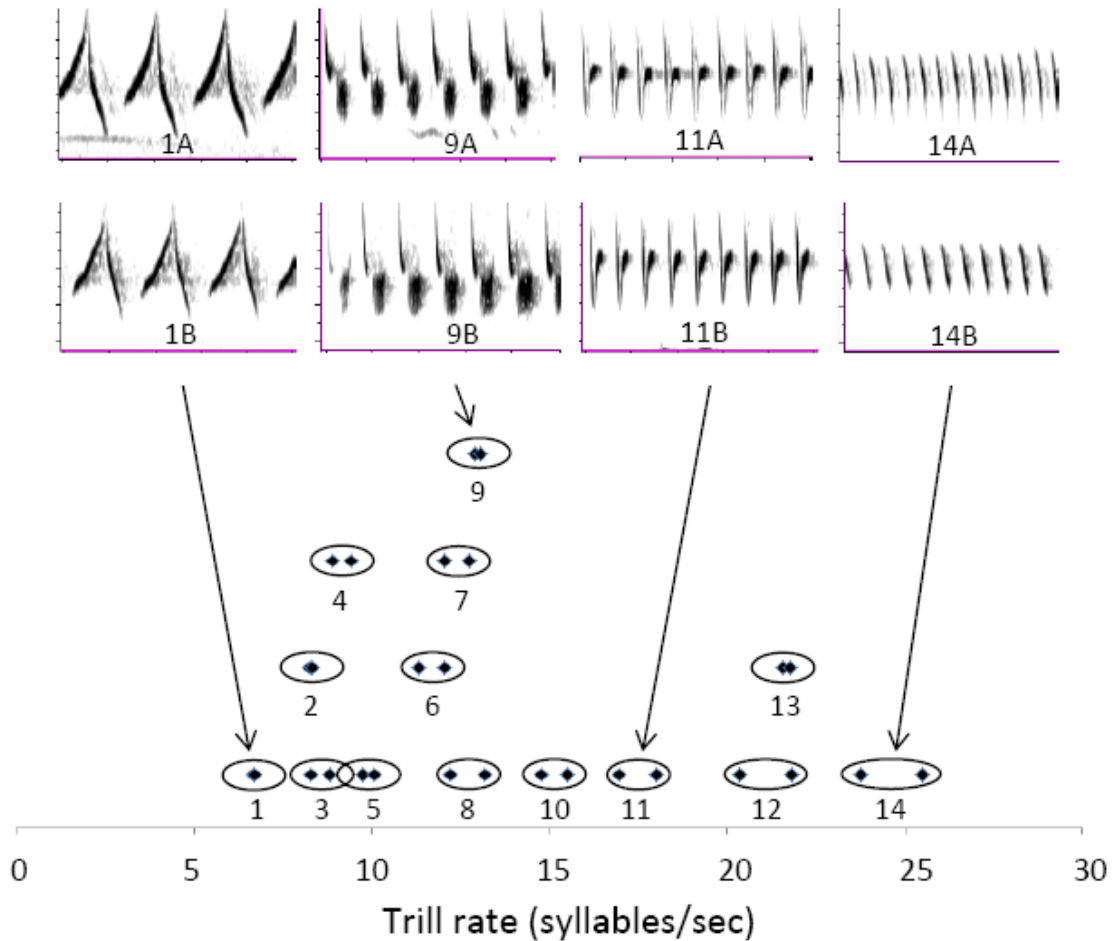
81

82 Figure 1. From Liu and Kroodsma (2006). "Yearling Chipping Sparrows imitate songs of an immediate  
 83 neighbor, but the instability of territories results in only short-term song sharing among neighbors. (a)  
 84 In May of 1996, 24 territorial males (numbered 1-24) were found in the Quabbin Cemetery, and a  
 85 portion of each male's song type (0.35 sec) is illustrated. Males 7 and 9 share similar song types, as do  
 86 males 20 and 22. (b) In May of 1997, 26 territorial males were found in the cemetery, 16 returning  
 87 adults (at least two years old) from the previous year and 10 birds breeding there for the first time.  
 88 Males 7 and 9 both returned but are no longer neighbors, and male 22 did not return. Male 35 did not  
 89 learn the song of his father (male 5) or his father's close neighbors in 1996 or 1997, but instead  
 90 appeared to learn from his immediate neighbor in 1997 (male 24). Male 31 also appeared to acquire his  
 91 song from an immediate neighbor (male 13) in 1997, not from his father (male 10). Both 1997 yearlings  
 92 (31 and 35) hatched late in the 1996 breeding season, and each most likely acquired his song during  
 93 1997, as a yearling."

94

95           To further illustrate how a young male chipping sparrow learns rather precisely the song  
96 of his adult tutor, and especially the tutor's trill rate, I recorded chipping sparrows during early  
97 May (2015) when they first returned from migration, before post-learning dispersal might occur.  
98 In two populations, one on a golf course in Lewiston, Michigan, the other in a city park in  
99 Northampton, Massachusetts, I recorded 67 different males. Equipment used were a Sound  
100 Devices 722 digital recorder and a stereo Telinga microphone. Birds were not banded, but most  
101 of the birds were recorded in rapid succession by moving directly from one singer to the next, so  
102 that the previous and next singer could be heard while recording a given male. If songs of  
103 suspected neighbors were identical, and I could not distinguish their songs, I conservatively  
104 assumed they were the same male and discarded one of the recordings from the data set. Using  
105 Raven pro software, I measured trill rates and frequency bandwidths for three high quality songs  
106 for each male, and used the median value in analyses ("spectrogram window size" in Raven: 110  
107 for temporal measures, 2050 for frequency; lower and upper frequencies measured as -24 dB  
108 down from max power).

109           Among these 67 males, I found 14 pairs of adjacent males with essentially identical songs  
110 (see Figure 2), as one would expect based on how chipping sparrows learn their songs. As is  
111 clear in Figure 2, song types and trill rates are determined by where and from a male learns his  
112 song and cannot reflect any measure of his quality, in the sense of Podos (1997).



113

114 Figure 2. A few dozen different songs can occur within a chipping sparrow population (only four  
 115 illustrated here: 1, 9, 11, 14), but neighboring males (A and B) often have nearly identical songs, the  
 116 result of a young male copying the song of a nearby adult singer (Liu and Kroodsma 1999, 2006); all  
 117 features of a male's song, including his trill rate as illustrated here (14 examples), are determined by  
 118 that adult tutor. In the lower graph, each oval encircles the two data points (pairs 1 and 2 are identical)  
 119 for trill rates from two neighboring males with the same song (data are distributed vertically for easier  
 120 visibility). Each data point is the median of three measurements for a given male.

121

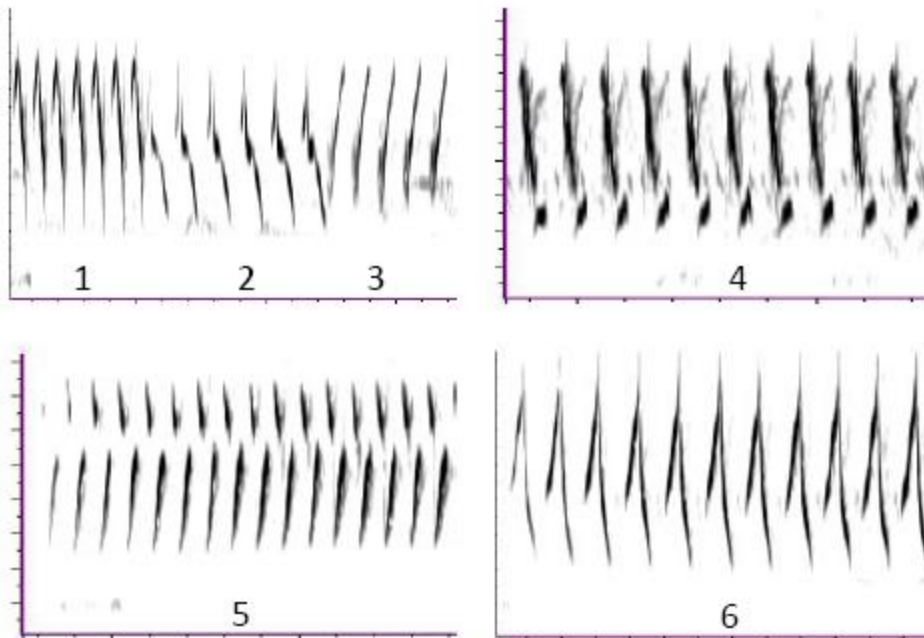
122            *How and where a chipping sparrow uses his song*

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

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

129            One example suffices, from a location in eastern Missouri (see Figure 3). In that example,  
130 four males displayed in a lek-like arena during the dawn chorus, all singing near one another on a  
131 paved road, but before sunrise they all dispersed to their daytime centers of activity. Replacing  
132 those four males after sunrise were two other males, each now on his daytime center of activity,  
133 each of which was presumably displaying elsewhere during the dawn chorus.





134

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

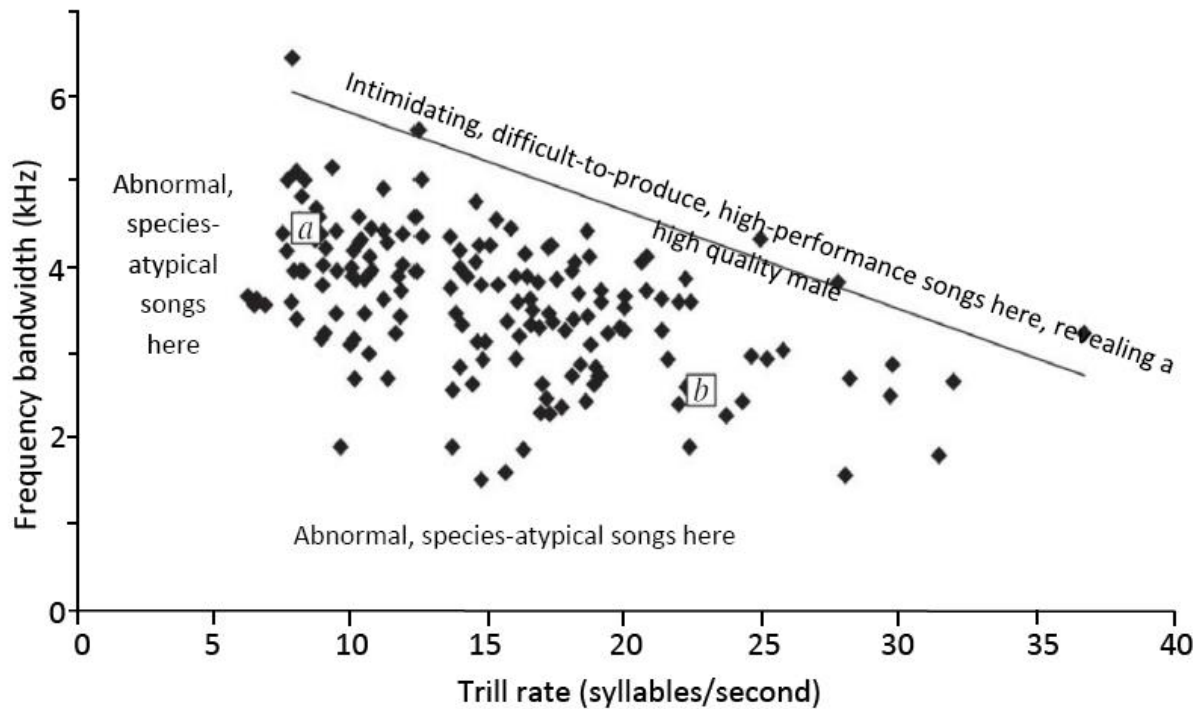
141

142 Male chipping sparrows thus routinely intrude on the daytime activity centers (i.e.,  
 143 “territories”) of other males and display there competitively with lek-like behavior.

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

145 The standard graph provided in studies of performance is the scatterplot of frequency  
 146 bandwidth versus trill rate (Figure 4). The distance from a given plotted point to the upper bound

147 regression line (i.e., the deviation from the line) is then interpreted as (or often assumed to be) a  
 148 measure of a male's performance or proficiency on that particular song. A small deviation is a  
 149 high-performance song, a large deviation a low-performance song. Because basic ornithological  
 150 data on song type and individual males are not encoded in the data, however, the biological  
 151 significance of the data is obscured.

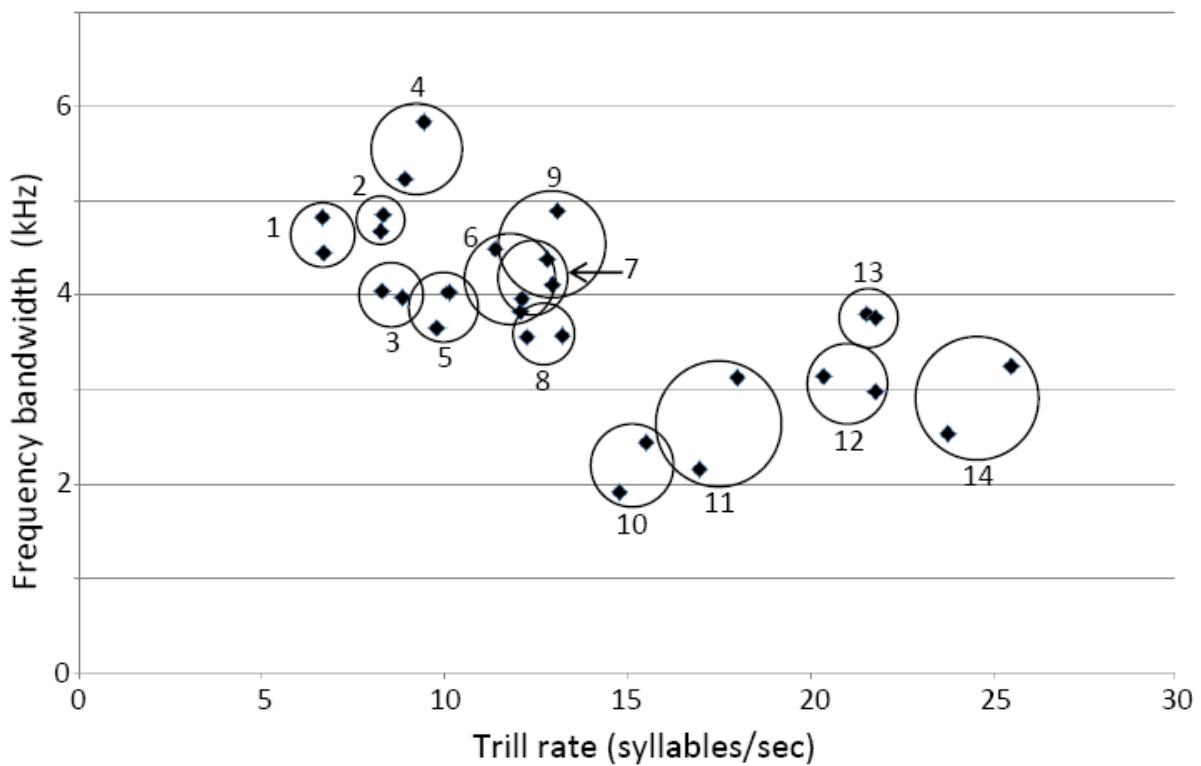


152

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

158

159 Consider, then, a graph of this sort that includes the information necessary to interpret it  
 160 in a biological context (Figure 5). Given how a chipping sparrow learns his song from a  
 161 neighboring adult, it is clear from this figure that social factors and song learning explain not  
 162 only 1) the variability in trill rates within a population (as in Figure 2), but also 2) the variability  
 163 in frequency bandwidth (Figure 5), and therefore also 3) the scatter in the plot from Goodwin  
 164 and Podos (Figure 4). Scatter in the graph is explained not by trill rates or frequency bandwidths  
 165 that reflect male quality, but instead by song types, reflecting the social bonds that play a crucial  
 166 role in the process by which a young male learns his song.

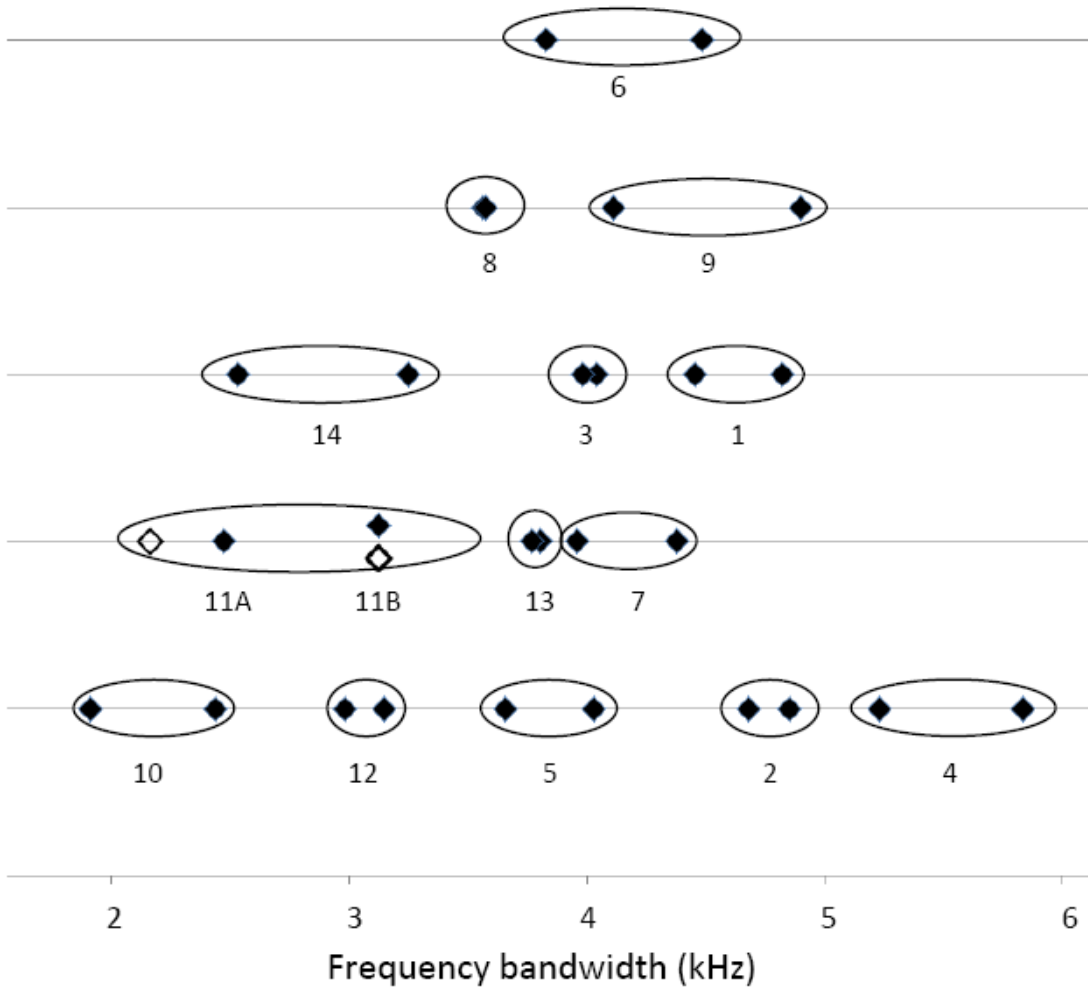


167  
 168 Figure 5. All aspects of a male chipping sparrow's song, including the trill rate and frequency bandwidth,  
 169 are determined largely by the song that he imitates from an adult male, so that song types dictate the  
 170 scatter in the plot, not relative male quality. Fourteen pairs of similar songs among neighboring males  
 171 (see Figure 2) are plotted, with the two data points for each of the 14 pairs of songs encircled.

172

173           Proponents of the performance hypothesis might cling to the hope that individual males  
174 reveal their prowess by expanding or diminishing their frequency bandwidths, but regraphing  
175 these data to show only frequency bandwidths reveals little support for that idea either (Figure  
176 6). The median percentage difference among the 14 pairs of birds, with the larger bandwidth  
177 divided by the smaller, is 10%, just a little larger than the median 6% difference between the  
178 largest and smallest of three bandwidth measurements made for each of the 28 males. (Less  
179 variability is found among trill rates: Among the 14 pairs, trill rates differ by a median of 6%  
180 (Figure 2), whereas the median difference in trill rates among the three measures for each of the  
181 28 birds is only 1%.)

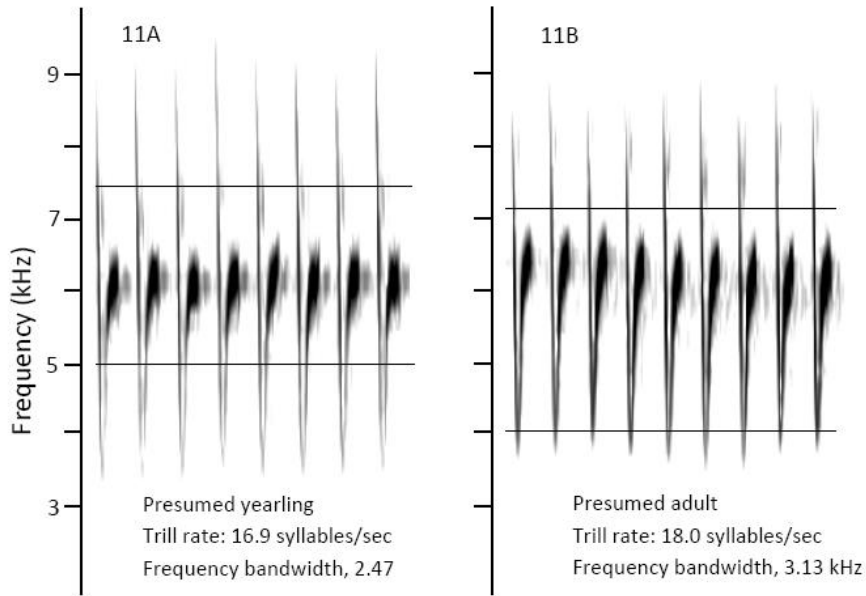
182           Some of the variability in frequency bandwidth is undoubtedly related to early season  
183 variability in songs of yearling males. Male 11A, for example (Figure 6), behaved much like a  
184 yearling just settling on his territory, as he repeatedly attacked his neighbor 11B and sang  
185 relatively plastic songs. His song appeared to be a work in progress, but already matched well the  
186 song of his neighbor 11B (Figure 7). The narrower frequency bandwidth may be typical of all  
187 yearling males, as song development routinely proceeds from the earliest, narrow-frequency  
188 scratches and often pure tones to more well-developed songs, spanning a broader frequency  
189 range when fully developed. (This does not mean, of course, that “performance” of younger  
190 birds is worse than that of older birds, in the sense of Podos (1997); it simply means that their  
191 song development has not yet been completed.)



192

193 Figure 6. Neighboring males with highly similar songs (same 14 pairs as in Figure 2) also have highly  
 194 similar frequency bandwidths. As in Figure 2, each oval encircles the two solid data points (bandwidths  
 195 for the two birds in pair 8 are identical) that represent the median bandwidth as measured from three  
 196 different songs for each male (data are distributed vertically for easier visibility). For bird 11A, a  
 197 presumed yearling learning the song of male 11B, frequency bandwidths were highly variable, as  
 198 indicated by the open two symbols, one of which matches exactly the frequency bandwidth of the  
 199 presumed adult tutor 11B.

200



201

202 Figure 7. A half second section of sonagrams for birds 11A (a presumed yearling) and 11B (presumed  
 203 adult) reveal highly similar songs. In trill rate, they differ by about six percent (16.9 vs 18.0  
 204 syllables/second). The median low and high frequency (measured as -24dB down from max power; n =  
 205 3 for each bird) are indicated by horizontal lines through the sonagrams, revealing a frequency  
 206 bandwidth that is 27 percent larger in the adult, though shifted to a lower frequency range. These two  
 207 birds were recorded just a few days after they had returned from migration, when the song of the  
 208 presumed yearling was still somewhat variable (see Figure 6), but apparently converging on the song of  
 209 the nearby adult with whom he was aggressively interacting.

210

### *A focused critique of Goodwin and Podos (2014)*

211 Contrary to the hidden assumption of Goodwin and Podos, a young male chipping  
 212 sparrow does not alter the trill rate of his tutor song to maximize his supposed performance  
 213 ability or vocal proficiency (sensu Podos 1997). Nor is there evidence for song learning in any  
 214 songbird species or especially in chipping sparrows (Liu and Kroodsma 1999, 2006) that males  
 215 are in any way limited in what naturally occurring trill rates they can learn, such that a young

216 male would choose a territory adjacent to an adult singer whose song the young bird knew he  
217 would be able to master. Goodwin and Podos (2014) would have to make the far-fetched claim,  
218 for example, that male 1A or 1B (Figure 2) settled where he did, next to a male singing an  
219 especially slow trill rate of about seven syllables/second, because each knew as a young bird that  
220 he would not be able to master the faster trill rates of eight to 25 found among other males in  
221 these populations.

222 (The one justification Goodwin and Podos (2014) use for focusing on trill rate alone is  
223 that “in songbirds accumulating evidence suggests that . . . fast trills, are especially threatening”  
224 (Moseley et al. 2013). But, as revealed in my review below, the paper by Moseley et al. is also  
225 seriously flawed and contains no evidence that fast trills are threatening.)

226 The second unwarranted assumption by Goodwin and Podos (2014) is that males have  
227 restricted areas (i.e., “territories”) in which they carry out all of their activities, and the authors  
228 therefore “mapped the territories of singing males” during the daytime when the birds could be  
229 seen. By omitting all reference to how chipping sparrows range widely over space, and by  
230 omitting all reference to how chipping sparrows acquire their songs, Goodwin and Podos  
231 generate a story with no biological basis.

232 The third major problem with Goodwin and Podos (2014) is that the reported statistically  
233 significant results are gleaned from a much larger, unreported series of nonsignificant tests on  
234 frequency bandwidths alone and a combination of frequency bandwidths and trill rates. As  
235 Simmons et al. (2011) write, “Our job as scientists is to discover truths about the world . . . [but]  
236 . . . Undisclosed flexibility in data collection and analysis allows presenting anything as  
237 significant.” Furthermore, the statistical tests of Goodwin and Podos (2014) are done improperly

238 (Akçay and Beecher 2015), even when revised in their 2015 rebuttal (see  
 239 <https://caglarakcay.wordpress.com/>).

240 Other issues were addressed by Akçay and Beecher (2015). Those issues, together with  
 241 the three major problems addressed above, serve as an effective primer on how the performance  
 242 hypothesis of Podos (1997) has been repeatedly confirmed in the literature since its inception.

243 ***Summary statements***

244 1) Goodwin and Podos (2014, 2015) omit all reference to how chipping sparrows actually  
 245 acquire their songs (Liu 2001; Liu and Kroodsma 1999, 2006) and instead falsely assume that a  
 246 chipping sparrow acquires a song with a trill rate that honestly conveys his performance ability  
 247 and his overall quality (in line with Podos 1997). The trill rate of a male is determined by the  
 248 song of his adult tutor, not by his relative prowess or “performance ability.”

249 2) There is no evidence for song learning in any songbird species or especially in  
 250 chipping sparrows (Liu and Kroodsma 1999, 2006) that a male is in any way limited in what  
 251 naturally occurring trill rate he can learn. The trill rates of ~7 syllables/second for birds 1A and  
 252 1B (Figure 2) and trill rates of ~25 for birds 14A and 14B were determined by where the males  
 253 settled on their first territory and do not reflect a measure of male quality.

254 3) The scatterplots of Figures 4 and 5 have no relevance to anything about 1) male  
 255 quality, 2) honesty in signaling, 3) song proficiency, or 4) performance.

256 4) Goodwin and Podos (2014, 2015) also omit all reference to how male chipping  
 257 sparrows routinely display competitively in lek-like arenas well off their own territories (Liu  
 258 2004, Kroodsma 2007), and instead invent cooperative alliances and coalitions to explain any  
 259 gathering of singing males. (These omitted facts are not obscure, I should point out, as they were



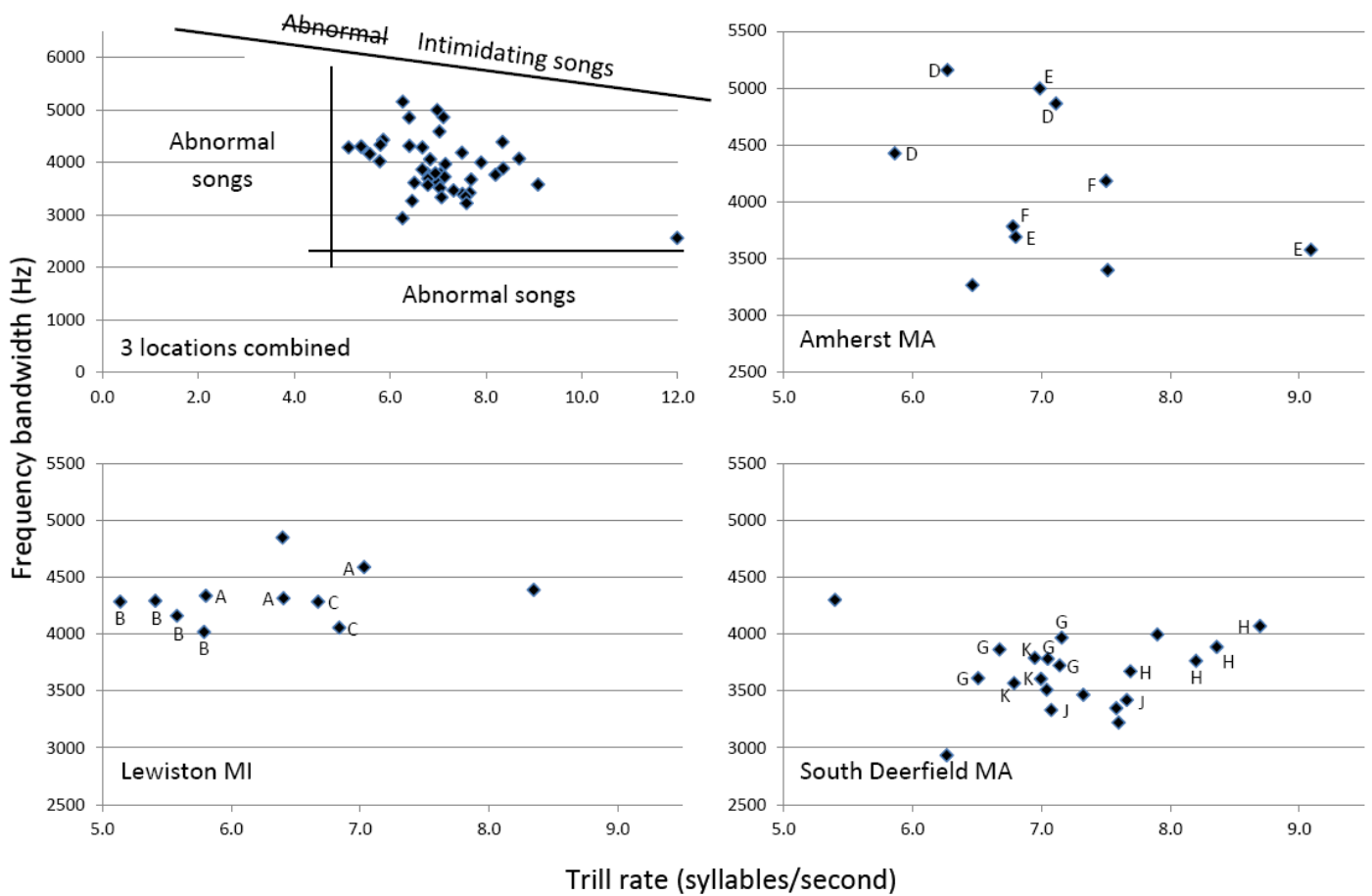
260 published in mainstream journals by a graduate student in the same department as Goodwin and  
261 Podos, working on the same population of chipping sparrows (see Liu 2001.)

262 5) Conclusion: There are no precise assessments of mating signal features, no teams of  
263 rivals, no alliance formations, and no coalitions, as reported in Goodwin and Podos (2014, 2015).

## 264 **SWAMP SPARROWS**

265 Since 2004, for both male and female swamp sparrows, numerous studies have confirmed  
266 the significance of song performance (sensu Podos 1997). I address each of those studies in turn,  
267 but first illustrate in several graphs why I cannot believe the claims of these swamp sparrow  
268 studies either.

269 My swamp sparrow data were collected during 2015 from three locations (Figure 8). At  
270 each site, I used a stereo Telinga parabolic microphone, and either a Sound Devices 722 or  
271 Marantz PMD661 digital recorder. Birds were unbanded, but each male sang repeatedly over a  
272 few hours from the same predictable locations, and attributing each recording to a particular  
273 male was not difficult; if any doubts existed as to the origin of a song, it was discarded from the  
274 analyses. Songs were then analyzed on Raven Pro 1.4 software (settings the same as for chipping  
275 sparrows), and the median of three examples of each song type from each bird was used in the  
276 analyses.



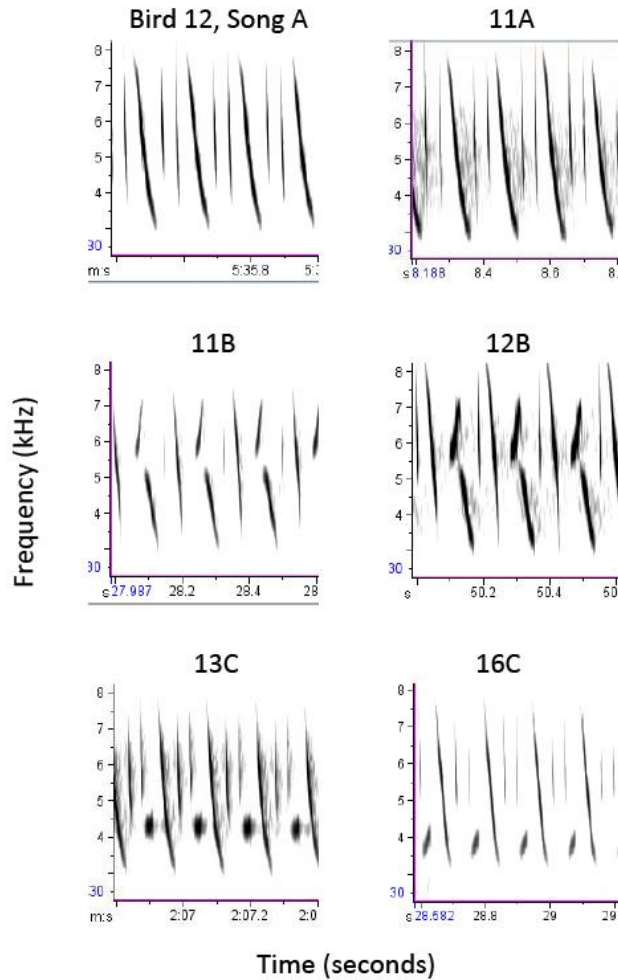
278 Figure 8. Scatterplots of trill rate and frequency bandwidth for swamp sparrows from three locations,  
 279 revealing that scatter is largely due to song types (letters A-K; song types deemed unique to an individual  
 280 are not labeled), leaving little if any information available about the quality of individual singers. In the  
 281 upper left, data for all three locations are combined, and the axes meet at the origin (0, 0); the graphs for  
 282 the three separate locations are drawn to a different scale, expanded to better show the variation within  
 283 locations (note that the outlying data point at 12 syllables/second is omitted from the lower right graph;  
 284 excluding it has no bearing on the conclusions I wish to draw from this figure). The upper bound line in  
 285 the upper left subfigure is from Ballentine et al. (2004); the bounds below and to the left of the data  
 286 points are placed arbitrarily.

287 Several important points are revealed in these data (Figure 8):

288           1) Normal, wild-type swamp sparrow songs are restricted to a relatively limited set of all  
289 possible trill rates and frequency bandwidths (upper left subfigure in Figure 8). Outside of this  
290 restricted area, all songs are, by definition, abnormal. When bounded lines are added to the  
291 graph, they draw attention to the limits. Abnormal songs that approach the bound above the sea  
292 of data are considered supernormal and especially high performance (Podos 1997), so  
293 intimidating and threatening that listening males might well flee them (e.g., Illes et al. 2006);  
294 songs to the left and below the normal songs are just abnormal.

295           2) Trill rates and frequency bandwidths can vary significantly by location. Frequency  
296 bandwidths from Lewiston MI, for example, are mostly above 4 kHz, those from South Deerfield  
297 MA below 4 kHz; trill rates are correspondingly slower at Lewiston.

298           3) Much of the scatter in the data is explained by song types (see also Figure 9): Birds  
299 learn their songs (including trill rates and corresponding bandwidths) from one another, and as a  
300 result, many songs are shared within the population, so that songs of the same song type from  
301 different birds tend to plot near one another (especially clear for Lewiston and South Deerfield).



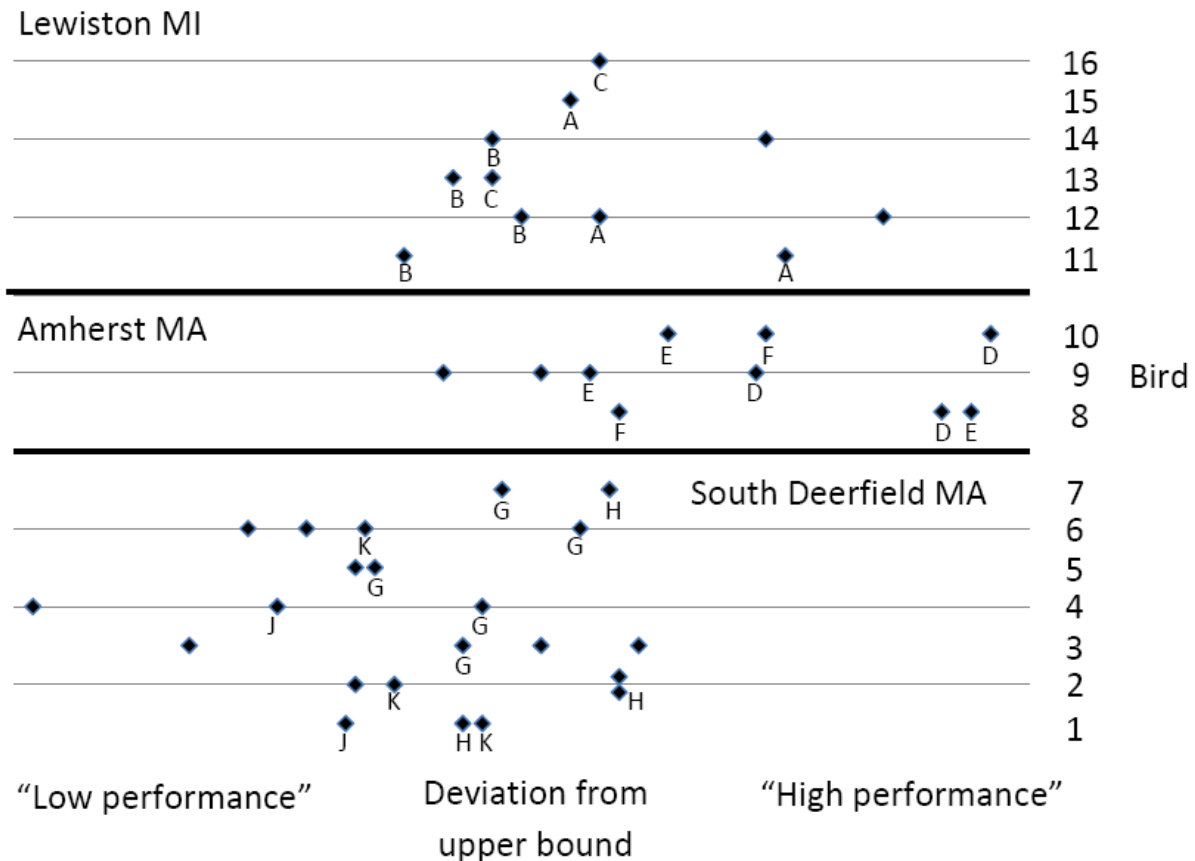
302

303 Figure 9. Swamp sparrow males within a marsh learn their songs from one another and, as a  
 304 result, many songs are shared among birds in a population. From Lewiston, Michigan, two  
 305 examples are illustrated for three different song types (A, B, C). Songs of higher performance  
 306 (lower deviation from the upper bound, as shown in upper left of Figure 11) are in the second  
 307 column. Birds 11 and 12 (first two rows) were immediate neighbors; note that bird 11 had the  
 308 “better” A song, bird 12 the “better” B song.

309 A critical but untested feature of the Podos (1997) performance hypothesis is that songs  
 310 actually provide reliable, honest signals of male quality. If these scatterplots with the upper

311 bound are at all relevant to how male and female swamp sparrows might assess a singer, then the  
 312 performance measures must provide consistently reliable information about the singer. If no  
 313 reliable information is provided, the relative performance of different males cannot be used as an  
 314 honest signal of his relative quality.

315 As revealed in Figure 10, performance measures provide no reliable information about  
 316 male quality. Because measures for different males are broadly overlapping, and a given male  
 317 might have both the “best” and the “worst” song in his repertoire, performance measures cannot  
 318 be used by either other swamp sparrows or by humans to assess the relative quality of a singer.  
 319 The data provide no support for the feasibility of the performance hypothesis of Podos (1997),  
 320 and essentially render the hypothesis false for these swamp sparrows.



322 Figure 10. Song performance measures (sensu Podos 1997) can provide no reliable information about  
 323 inherent male quality in swamp sparrows. Data plotted here have been extracted from Figure 11 by  
 324 measuring the shortest distance to the upper bound for each song (all measures are relative,  
 325 depending on a number of factors, so no absolute scale is provided for performance). Performance  
 326 measures for different males are broadly overlapping, such that a male might have the highest  
 327 performance on one of his song types but the lowest on another (e.g., compare bird 3 from South  
 328 Deerfield with birds 1, 2, 5, 6, and 7; song types are lettered as in Figure 8). Note that Amherst swamp  
 329 sparrows significantly “outperform” the South Deerfield birds only 20 kilometers distant (the two  
 330 populations were recorded within a few days of each other during late June 2015; I can think of no  
 331 methodological issues that would have produced such results).

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

334 **Podos, J., S. Peters, and S. Nowicki. 2004. Calibration of song learning targets**  
 335 **during vocal ontogeny in swamp sparrows, *Melospiza georgiana*.**

336 The basic claim here is that, when a young swamp sparrow learns a given song type, he  
 337 adjusts the trill rate or frequency bandwidth to match his own proficiency at producing that song,  
 338 so as to acquire as high a performance song as he possibly can. A dozen times throughout the  
 339 paper the authors declare how their data are consistent with or support this “calibration  
 340 hypothesis”:

- 341 1) The calibration hypothesis was supported by two lines of evidence . . .
- 342 2) . . . in directions consistent with expectations [of the calibration hypothesis]
- 343 3) Consistent with the calibration hypothesis, we found . . .

- 344 4) The direction of this transition is consistent with the calibration hypothesis
- 345 5) The observation is consistent with our expectation . . .
- 346 6) Consistent with this result, we also found that differences in trill rates . . .
- 347 7) The main support for the calibration hypotheses comes from analysis of . . .
- 348 8) . . . songs more consistent with expected levels of proficiency
- 349 9) Another line of evidence that provides support for the calibration hypothesis . . .
- 350 10) This bias is consistent with the calibration hypothesis . . .
- 351 11) A third line of evidence . . . offers additional . . . support for the calibration
- 352 hypothesis
- 353 12) These results were consistent with the calibration hypothesis

354 I provide excerpts from all examples, because all together they illustrate how selective

355 authors can be in reporting evidence to support the performance hypothesis. The impression

356 conveyed is that, given how much is consistent with the calibration hypothesis, it must therefore

357 be true. But the words “consistent with” are also red flags for readers to ask what other

358 hypotheses the data might be consistent with.

359 The figures I provide on the biology of swamp sparrow song soundly reject the authors’

360 conclusions about calibration. When learning a song, a swamp sparrow conforms to the

361 particular song type of the local dialect; he does not adjust features of what he learns in any way

362 consistent with an attempt to calibrate a song to his own abilities. As a result, a male may have

363 the worst “performance” on one song type, the best on another.

364 Not once do the authors disclose that all of their data are also consistent with a very  
 365 simple alternative explanation, that no matter what recognizable features of a song a swamp  
 366 sparrow hears, he tries to develop as normal a song as possible, making a fine-tuned effort to  
 367 take whatever he hears and produce a normal song (the only logical conclusion also for Lahti et  
 368 al. 2013—see review below). Contrary to the title and all of the statements in this paper, there is  
 369 no credible evidence that an individual male “calibrates” songs to his particular proficiency.

370 Science requires a balanced interpretation, not careful selection of data or details that  
 371 might support only one’s favored explanation. “Details that could throw doubt on your  
 372 interpretation must be given” (Feynman 1985; see fuller quote in Discussion).

373 It was in 2004, the very year this paper was published, that I strongly objected to another  
 374 paper by the first author, writing to him the following:

375 Science is the search for truth, regardless of how good the story is; “marketing and  
 376 advertising” are the search for a good story, regardless of the truth (October 2004)

377 **Ballentine, B., J. Hyman, and S. Nowicki. 2004. Vocal performance influences**  
 378 **female response to male bird song: an experimental test.**

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

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



386 Those who cite this paper are equally convinced, such as the following quotes reveal:

387 . . . it is known that females prefer trilled songs closer to the production limit (Illes et al.  
388 2006)

389 . . . females have been shown to prefer males that are singing higher performance trills  
390 (Schmidt et al. 2008)

391 . . . females are able to assess a male's quality as a potential mate using vocal  
392 performance (Dubois et al. 2011)

393 . . . females . . . are known to discern fine features of song in the functional contexts of  
394 mate choice (Lachlan et al. 2014)

395 In fact, "As of May/June 2014, this 'highly cited paper' received enough citations to  
396 place it in the top 1% of the academic field of Plant & Animal Science based on a highly cited  
397 threshold for the field and publication year" (Web of Science). In a survey entitled "25 Years of  
398 Behavioral Ecology," a review article for the journal cites the importance of this paper; Simmons  
399 (2014) reports that this paper is

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

404 I do not believe the results of this study primarily because they make no biological sense.  
405 Realize, for example, that if experimental songs were selected from Figure 9, the authors claim  
406 that females would display more to the songs in the right column than to those in the left column,

407 even though males 11 and 12 provide mixed messages with song types A and B. Or study  
 408 Figures 8 and 12, seeing how different renditions of the same song type are clustered in the plots;  
 409 the authors claim that the females would display more to any rendition that lies to the upper right  
 410 of the others, i.e., closer to the boundary line drawn across the top of the plot.

411 As with chipping sparrows, the hidden premise is that a swamp sparrow is free to develop  
 412 as high a quality song as it can, so as to impress males or females. Yet all qualities of a song that  
 413 a male learns are constrained by the song type that he is learning, and the supposed performance  
 414 of that song is necessarily limited to the relatively small range of performance values available  
 415 for that song type. In other words, a male “willingly” sacrifices his “performance status” to sing  
 416 a particular low-performance song type, such as male 12 from Lewiston learning low-  
 417 performance song B even though on another song type he has the best in show (Figure 10). It is  
 418 as if only the song type mattered to him and his overall “song proficiency or performance” did  
 419 not.

420 How does one account for the results of this study? Briefly, I offer three possibilities.  
 421 First, “Believing is seeing,” it might be said, the results stemming from nonblind observers with  
 422 strong expectations for the results (i.e., observer bias). From the outset, the concept of  
 423 “performance” is already a given, not a hypothesis (“our knowledge of song production  
 424 mechanisms allows us to identify a priori which songs are produced with greater vocal  
 425 proficiency”—Introduction), and variations of the word “perform” with its attending  
 426 assumptions are used 87 times throughout this paper (for bias generated by mere use of the word  
 427 “performance,” see Discussion).

428 Second, the equipment used might have been an issue: two different cassette tape  
 429 recorders (Sony TCM-5000 EV and Sony TCD-5M) and, more importantly, two different

430 parabolic reflectors, one with 13” diameter (Sony PBR-330) and one with 18” diameter (Saul  
431 Mineroff SME PR-1000). For a study that wishes to compare variability of songs within and  
432 among males, and across sessions, and ask females to assess subtle features of these songs, it  
433 would be essential to insure that all of the recording gear gave identical results. The 13”  
434 parabolic reflector, for example, is rather nondirectional, and would be expected to record songs  
435 with more reverberation than the 18” reflector.

436         Third, I find it distressing that alternative explanations for results are never considered in  
437 this study or similar studies. Consider only the reverberation that can vary from one recording to  
438 another, depending on which parabola was used, or how far the bird was away from the  
439 microphone, or the amount of intervening vegetation between the singer and the microphone.  
440 There is a large parallel literature, none of it cited in any of these studies of performance, that  
441 shows how birds vary their response to playback songs depending on how much reverberation is  
442 in the recording (e.g., Naguib 1996), prompting Morton et al. (1986) to write the following:

443             Sufficient evidence now exists to suggest that sound degradation, and the ability  
444             of birds to use it, should be taken into account in studies using responses to  
445             playback of bird song.

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

450             **Dubois, A. L., S. Nowicki, and W. A. Searcy. 2009. Swamp sparrows modulate vocal**  
451             **performance in an aggressive context.**

452 The authors' main conclusion (p. 163, from the Abstract):

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

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

460 1) The performance of particular song types that males used in aggressive contexts were  
461 no different from song types chosen for control contexts. When it matters most, then, when a  
462 male is challenged on his territory, he chooses a random song from his repertoire, not a song that  
463 best conveys his overall quality. This important result is highly inconsistent with the  
464 performance hypothesis (though not mentioned in the abstract), yet the authors puzzlingly  
465 conclude “. . . we do not think this means that males are not trying to maximize their vocal  
466 performance during aggressive signaling . . .”

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

473            Yet, one must ask, how could it possibly be biologically meaningful to increase the trill  
474 rate by 2.3 percent or frequency bandwidth by 1.8 percent? In Figure 8, for example, consider a  
475 song with trill rate of 6.0 syllables/sec and frequency bandwidth of 4000 Hz that is  
476 “exaggerated” to 6.1 syllables/second and a frequency bandwidth of 4072 Hz. The data point on  
477 the scatterplot is moved a miniscule distance. If a male really wanted to increase his performance  
478 during aggressive contexts, he could switch to a more impressive song in his repertoire, but he  
479 doesn’t do that, as if performance did not matter. Two years later, the authors would accept that  
480 these “exaggerations” are biologically meaningless (Dubois et al. 2011; see below).

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

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

494            I do not believe that this study by Dubois et al. (2009) demonstrates anything relevant to  
495 performance and honesty in signaling.

496

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

499 I stumble over one key sentence in the methods:

500 I used the highest performance song in a male's repertoire to determine each  
501 male's vocal performance.

502 That makes no biological sense to me. As discussed with other studies, a necessary  
503 condition for honesty and reliability is that males consistently use songs within a relatively  
504 narrow range of performance abilities. If the performance values of males broadly overlap (see  
505 my Figure 10), so that a male can rank highest on one of his songs and lowest on another (as also  
506 revealed in Kagawa and Soma 2013), and a male doesn't even use his "best" songs in aggressive  
507 encounters when it matters most (see Dubois et al. 2009), it makes no sense to rate a male only  
508 by the one song of highest performance ability.

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

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

523           Given that all song types were recorded from all males in this study, the author missed an  
524 opportunity to show, as I have, that song performance cannot be a reliable measure of male  
525 quality (my Figures 8 and 10). I do not understand how so important an analysis, crucial for the  
526 performance hypothesis to be true, can be so shunned.

527

528           **Podos, J., D. C. Lahti, and D. L. Moseley. 2009. Vocal performance and**  
529           **sensorimotor learning in songbirds.**

530           In this review one finds unflagging support for the performance hypothesis. After  
531 reviewing the literature in a much different light than I am here, the authors declare the  
532 following:

533           Emerging descriptive and experimental evidence thus suggests that vocal  
534 performance varies among individuals, and suggests that singers who maximize  
535 vocal performance gain advantages in song function and ultimately in  
536 reproductive success (p. 170).

537           There is no credible scientific evidence cited to support that conclusion, nor will there be  
538 any in the years to follow.

539

540 **Dubois, A. L., S. Nowicki, and W. A. Searcy. 2011. Discrimination of vocal**  
541 **performance by male swamp sparrows.**

542 In experiment 1, males are asked to discriminate between high- and low-performance  
543 songs of the same song type as sung by different males.

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

548 For three reasons, I have trouble accepting these results on face value:

549 1. Based on my Figures 8 and 10, the results do not make biological sense. Song types  
550 plot in difference spaces on the scatterplots, because males conform to the features of that song  
551 type when learning it. All indications are that the conforming is important, not any exaggeration  
552 of trill rate or frequency bandwidth to reveal one’s prowess on a particular song type. I see no  
553 reason how a little variation about the population average for trill rate or bandwidth for a given  
554 song type could possibly be meaningful.

555 2. The songs used in playbacks are the same songs that were used by Ballentine et al.  
556 (2004) and Ballentine (2009), and I am concerned that the two parabolic microphones (one 13”,  
557 the other 18”) could have produced different results, with the less directional 13” reflector  
558 recording songs with greater reverberation and therefore lower performance measures. If songs  
559 are low-performance because they consistently have higher reverberation, the playback results  
560 cannot be attributed to the chosen explanation of performance ability.



561           3. Among authors who appear so committed to the performance hypothesis, credibility of  
562 results would be enhanced if blind observers were used to make the numerous judgments about  
563 the responses of birds to playback stimuli: Was that a loud song (i.e., a “broadcast” song) or a  
564 “soft song”? And just how far from the speaker is the bird? Flags at 2, 4, and 8 meters “in either  
565 direction from the speaker” (in two directions?) may help in estimating distance but are no  
566 guarantee of objectivity, especially when the bird is not directly between flagging, and when  
567 distances of <16 or >16 meters from the speaker must be estimated. When authors appear so  
568 committed to confirming a favored idea, data collected nonblindly will always be suspect.

569           Experiments 2 and 3 are similar to each other, each of them asking if males respond  
570 differently to the kind of within-male differences in vocal performance observed in DuBois et al.  
571 (2009), where trill rates and frequency bandwidths increased on average about 2 % from neutral  
572 to aggressive performances. No significant differences in response were found, i.e., males  
573 responded no differently to the “extremes” of high- and low-performance versions of a particular  
574 song type that a given male might sing.

575           But no matter what results are obtained in experiments 2 and 3, the results can still be  
576 interpreted to support honesty in signaling and the motor constraints hypothesis. If, during  
577 contexts when it matters most (e.g., aggressive encounters), males do not increase their  
578 performance enough to be detected, then the songs are a good “example of an index signal, since  
579 it cannot be effectively cheated” (p. 725); males can thus “garner information about an opponent  
580 based on his vocal performance, whether or not any information is conveyed through modulation  
581 of this characteristic” (p. 726). And if the differences in modulation had been sufficient to be  
582 detected, then listeners could still “garner information about an opponent based on his vocal

583 performance.” It would appear that there is no way to falsify the motor constraints hypothesis of  
 584 Podos (1997), no way that performance cannot honestly convey a male’s quality.

585 **Lahti, D. C., D. L. Moseley, and J. Podos. 2011. A tradeoff between performance**  
 586 **and accuracy in bird song learning.**

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

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

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

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

601 I struggle with these statements. What is certainly true is that the young swamp sparrows  
 602 removed silent intervals from odd, slowed tutor songs to produce more normal, wild-type songs.  
 603 That result, however, based on abnormal, experimental songs, does not warrant any conclusion  
 604 about a young swamp sparrow either in nature or in the laboratory taking a natural tutor song that

605 it hears, foregoing “imitative accuracy,” and adjusting that song in trill rate or frequency  
606 bandwidth to match his own capabilities, all so that he can honestly broadcast his individual  
607 quality. There are no data in this paper or elsewhere demonstrating that a young swamp sparrow  
608 adjusts a normal or abnormal song to match his own individual proficiency, only data showing  
609 how young birds strive to produce normal, species-typical songs. There are no data that  
610 demonstrate any honesty in the signaling of trill rate or frequency bandwidth.

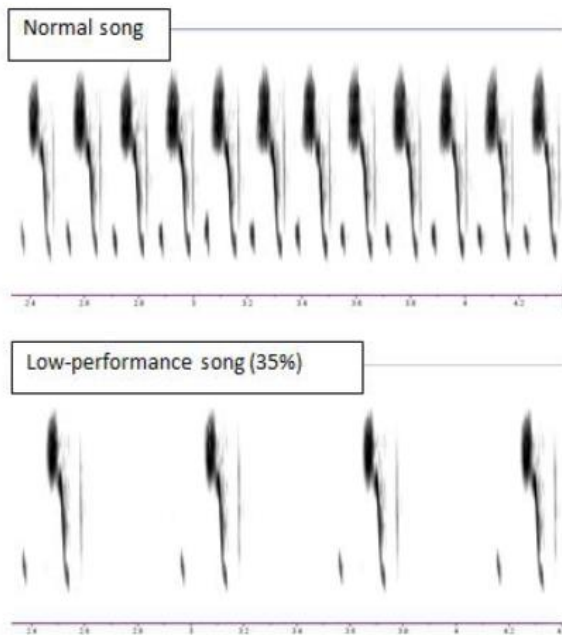
611 One hint that the authors perceive the conflict between their data and their interpretation,  
612 however briefly, is provided in the following quote (*italics mine*):

613 . . . our results reveal that vocal ontogeny can be shaped . . . by a premium on high  
614 performance. Again, performance in this case refers to the trill rate of songs, all  
615 other features being equal, and *high performance being that of typical songs*  
616 *recorded from the field* as compared with our experimentally slowed versions.

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

624 **Moseley, D. L., D. C. Lahti, and J. Podos. 2013. Responses to song playback vary**  
 625 **with the vocal performance of both signal senders and receivers.**

626 The authors use time-honored methods (e.g., Podos et al. 2004, Lahti et al. 2011) to  
 627 produce highly abnormal test stimuli: Two-second songs for playback to swamp sparrows are  
 628 prepared from normal songs by either inserting or deleting silent spaces between the song  
 629 elements; the manipulated songs then contain anywhere from 35% (a “low performance” song)  
 630 to 155% (a “high performance” song) of the elements in control songs, with trill rates for those  
 631 particular songs thus ranging from 35% to 155% of normal. As is evident in their figure 1 and  
 632 my Figure 11, three obvious features of the songs have changed from the original song: 1) the  
 633 trill rate is slower or faster, 2) the quantity of stimulus is correspondingly less or more, and 3) the  
 634 more silence edited into or out of the song, the more abnormal it is, unlike anything a swamp  
 635 sparrow has ever sung or would hear.



636

637 Figure 11. To produce playback stimuli, Moseley et al. (2013) altered a normal song (top) by adding or  
638 deleting silent spaces between the song elements. In this illustration (bottom), silent intervals are  
639 increased 8.3 times, producing a highly abnormal song with four syllables delivered at 35% the rate of  
640 normal.

641

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

647 Second, the stimulus songs are simply highly abnormal. To a swamp sparrow, a song  
648 slowed to 35% of normal must sound odd indeed, heard as a staccato, halting sequence of  
649 perhaps familiar song elements all out of sync, as these songs fall far outside the range of what  
650 any swamp sparrow would ever sing in nature. The authors in the Discussion do eventually admit  
651 that about “one-third of all stimuli . . . [were] . . . outside the natural range of swamp sparrows’  
652 trill rates,” but they discount the abnormality of those songs by reference to previous (flawed)  
653 studies, though not to their own study two years before. In that previous study, Lahti et al. (2011)  
654 concluded from their results that songs below 55% or above 115% were “so unlike typical  
655 swamp sparrow songs that males do not consider them as targets for learning,” i.e., they are so  
656 abnormal as to not be biologically meaningful; nevertheless, Moseley et al. use songs well  
657 outside that range.

658 The confounding variables are of no concern to the authors when they conclude the  
659 following:

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

663 When that conclusion is rewritten to focus on one of the confounding variables, it  
664 becomes rather less interesting and probably unpublishable:

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

668 It's curious that the authors found a relationship between the vocal performance of the  
669 responding male and how aggressively he responded to the playback stimuli. I don't understand  
670 how that relationship could possibly be true (see Figures 8-10), given that 1) males have several  
671 song types in their repertoire, 2) those song types vary widely in vocal performance, 3) such that  
672 song performance offers no reliable indication of male quality (Figure 10), 4) the particular song  
673 type a male chooses to use in aggressive contexts is random with respect to the purported vocal  
674 performance capabilities of that male (DuBois et al. 2009), 5) the song he does use is not  
675 exaggerated in performance in any detectable way, and 6) and the authors measured the vocal  
676 performance of the responding male only by that one randomly chosen song he used during the  
677 playback responses. Nevertheless, all previous studies notwithstanding, and in contradiction to  
678 the basic biology of swamp sparrow singing behavior, the authors "predicted that subjects'  
679 tendencies to engage simulated intruders would vary positively with their own vocal  
680 performance," and then proved it.

681 I find the logic troubling throughout this paper. Two examples suffice:

682           1) “. . . we predicted that stimuli with performance levels increased slightly would be  
683 responded to aggressively, whereas stimuli increased to the highest performance levels would be  
684 avoided, because of the higher perceived risk” of a supernormal stimulus (from Introduction).  
685 The highest performance songs, by the authors’ definition, are also the most abnormal, yet the  
686 authors argue that these highly abnormal songs would cause subject males to flee. I do not know  
687 how anyone can distinguish between failing to respond to a highly abnormal stimulus and fleeing  
688 a strong stimulus, yet the authors do so.

689           2) The Discussion is a tangled series of ad hoc explanations for why males

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

692           b) might respond strongly to high performance (abnormal) songs (high threat for loss of  
693 paternity to intruding superior male), or

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

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

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

702

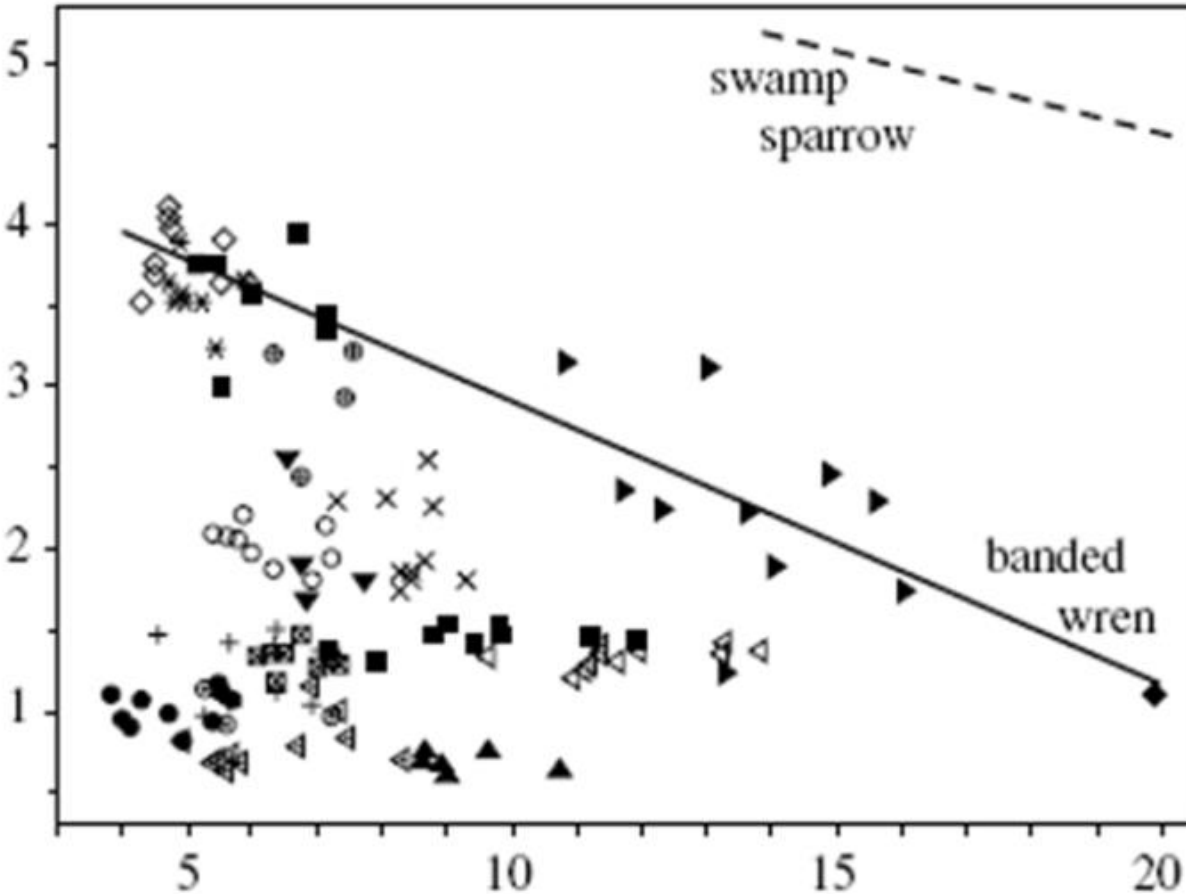
703 **OTHER SPECIES**

704 Nor in the following papers can I find any credible scientific support for the performance  
705 hypothesis (Podos 1997). The following reviews are in chronological order:

706 **Illes, A. E., M. L. Hall, and S. L. Vehrencamp. 2006. Vocal performance influences**  
707 **male receiver response in the banded wren.**

708 Given that each male banded wren (*Thryothorus pleurostictus*) has about 20 different  
709 songs, each learned from other males in the local dialect, the scatterplot of frequency bandwidth  
710 and till rate contains a wealth of information (Figure 12). Foremost, to me at least, it reveals  
711 great variation in “performance” among different song types. Some song types are low  
712 performance, some high performance, so that like swamp sparrows, the scatter in the plot seems  
713 dictated by song type, having little if anything to do with individual differences in performance.





714

715 Figure 12. From Illes et al. (2006). "Graph of trill rate versus frequency bandwidth for 695 trills recorded  
 716 from 13 individuals and including 16 trill types. Each symbol type represents a different trill type. The  
 717 banded wren upper-bound limit is shown with a solid line and the swamp sparrow limit (Ballentine et al.  
 718 2004) with a dashed line."

719         Given the revealing data I provide for chipping sparrows and swamp sparrows, I cannot  
 720 understand how the authors could 1) invest an enormous amount of energy in testing for a  
 721 possible difference in how birds might respond to two slightly different versions of a given song  
 722 type (if performance were really important, why not compare a low-performance song type with  
 723 a high-performance song type? and why are so many low-performance song types maintained in  
 724 the population?), 2) report significant differences in responses, 3) interpret an apparent reduced

725 response to a song as the strongest possible response (songs closer to the upper bound “posed a  
726 threat so extreme that they effectively repelled rivals, even territory owners”; p. 1911), and then  
727 4) expect readers to believe that result.

728 I cannot believe, for all of the reasons laid out for the swamp sparrow papers, though I  
729 recognize that I may stand alone, given how this paper has been cited since 2006 (74 times as of  
730 October 2015).

731 **Cramer, E. R. A., and J. J. Price. 2007. Red-winged blackbirds *Agelaius phoeniceus***  
732 **respond differently to song types with different performance levels.**

733 Congratulations are due these authors for collecting their playback data blindly, but I find  
734 no support for Podos’ “motor constraints hypothesis” in the data.

735 As reported by the authors, there is no trade-off between trill rate and frequency  
736 bandwidth; for the vast majority of songs, no matter what the trill rate, the frequency bandwidth  
737 falls within roughly the same range. Any measures of performance, then, would seem irrelevant  
738 to the blackbirds. That appears to be the case: When males responded aggressively to playback  
739 intrusions, the songs used during or after the event were random with respect to the performance  
740 level as determined by bandwidth and trill rate.

741 Those data do not support Podos’ performance hypothesis, but it seems as if the authors  
742 feel compelled to come to a different conclusion in their playback experiment. Songs of high and  
743 low performance were played to the blackbirds, and the birds responded more strongly to the low  
744 performance songs. The performance hypothesis would again seem to be contradicted, unless  
745 one is convinced that the hypothesis must be true. If true, then males only appeared to respond

746 more strongly to the low performance songs, because they were in fact fleeing the more  
747 intimidating, high-performance songs. Such is the conclusion of the authors:

748 Our results show that male red-winged blackbirds can discriminate between  
749 different song performance levels, suggesting that vocal proficiency plays a role  
750 in male-male interactions.

751 **De Kort, S. R., E. R. B. Eldermire, E. R. A. Cramer, and S. L. Vehrencamp. 2009.**  
752 **The deterrent effect of bird song in territory defense.**

753 The authors use a software program called “pitch bender” to increase or decrease by 1  
754 kHz the bandwidth of a normal banded wren song. These altered songs with an abnormal, low  
755 frequency bandwidth (i.e., low performance) and abnormal, high frequency bandwidth (high  
756 performance) are then played back to the wrens. The results of a multitude of statistical tests are  
757 presented, and all results are interpreted in the context of performance, ignoring any other  
758 possible interpretations.

759 For me, one conclusion provides the flavor of the paper: The apparent lack of response to  
760 the abnormal songs with high frequency bandwidth is in fact the highest response possible, the  
761 authors conclude, because birds were intimidated and therefore avoided that song.

762 The entire paper invites an explanation in terms of how abnormal the stimuli were, of  
763 course, and what the birds actually heard in them, and which statistical tests were done and  
764 which reported, and how one decides an apparent lack of response is in fact the strongest  
765 response possible, and all of the logic I have used to rebut the swamp sparrow papers and Illes et  
766 al. (2006). Instead, I will simply let the authors speak for themselves, to illustrate the logic

767 necessary to confirm the performance hypothesis (my italics show other references reviewed in  
768 this document):

769 Banded wrens responded differently to song stimuli that were manipulated to vary  
770 in frequency bandwidth and therefore performance level (p. 201)

771 With the High-performance stimuli, we attempted to simulate intruders that were  
772 superior to all individuals in the population and therefore should have been  
773 intimidating to all subjects (p. 205)

774 . . . subjects responded differentially to the 3 stimulus types that were presumed to  
775 differ in their threat level. These results are in agreement with an earlier study that  
776 showed that banded wrens responded differently to playback songs that differed  
777 in performance level (*Illes et al. 2006*) (p. 204)

778 These results are in accordance with theoretical models of aggressive encounters  
779 in which low-performance opponents are challenged without further assessment.  
780 Median- and high-performance opponents, however, may require further  
781 assessment, and the latter may be perceived as too intimidating for approach (p.  
782 200, Abstract).

783 Consistent with earlier studies (. . . *Ballentine et al. 2004; Illes et al. 2006;*  
784 *Cramer and Price 2007*), this study provides additional data to support the  
785 hypothesis that trill performance level, a structural component of song (*Podos*  
786 *1997*), affects response behavior by receivers. (p. 205)

787 This study therefore completes the body of work showing unambiguously that it is  
788 the bounded relationship between trill rate and frequency bandwidth that provides  
789 receivers with tools to assess the competitiveness of the singer (p. 205)

790 A possible playback control or two would have been appropriate. If the authors had  
791 considered the alternative explanation that their manipulated songs were at all abnormal, they  
792 could have easily compared unmanipulated songs of low, or medium, or high performance, from  
793 the same male or different males; two songs of equal performance value, regardless of song type,  
794 should evoke equal responses from territorial males. Or why not simply test males with normal  
795 (unaltered) high- and narrow-frequency bandwidth songs at the same trill rate, regardless of song  
796 type, as it is largely assumed that only performance matters, not song types. With no controls for  
797 how abnormal the song stimuli were, and for a variety of other reasons, I can have no confidence  
798 in the stated results of this paper.

799 **Juola, F. A., and W. A. Searcy. 2011. Vocalizations reveal body condition and are**  
800 **associated with visual traits in great frigatebirds (*Fregata minor*).**

801 Eventually, the authors admit that vocal deviation, i.e., performance, has no predictive  
802 value for listeners:

803 . . . results suggest that vocalizations provide honest information about male body  
804 condition and gular pouch size in great frigatebirds but do not influence male  
805 success in pairing with females. (Abstract)

806 We found that vocal deviation was significantly associated with body condition, in  
807 the direction that would be expected: lower vocal deviation, which equates with  
808 higher vocal performance, was associated with better body condition. Again,

809           however, the association was relatively weak, so that a female would not actually  
810           learn that much about a male's body condition by judging his vocal deviation.  
811           Moreover, the association would not be considered statistically significant if a  
812           correction were made for multiple comparisons. (Discussion)

813           **Cramer, E. R. A., M. L. Hall, S. R. De Kort, I. J. Lovette, and S. L.**  
814           **Vehrencamp. 2011. Infrequent extra-pair paternity in the banded wren, a**  
815           **synchronously breeding tropical passerine**

816           Overall, a fascinating paper. From the Results:

817           [Extra-pair] “. . . males sang trill notes with more consistent structure . . . and . . .  
818           had higher performance levels . . . than the . . . males they cuckolded . . . [and] . . .  
819           tended to have larger song repertoires and to be older and larger . . .”

820           From the Discussion:

821           Although our sample size for comparing EP and WP [extra-pair and cuckolded]  
822           males was prohibitively small, we did find differences . . . EP males sang with  
823           higher trill performance and higher levels of trill consistency than did the WP  
824           males they cuckolded; playback studies have identified these measures as  
825           important signals in this species (*Illes et al. 2006, deKort et al. 2009 . . .*).

826           The existence of a statistical correlation, of course, does not mean that the birds  
827           themselves use the correlated features to assess each other. It is entirely reasonable, however,  
828           based on the biology of song learning, that yearling males could be distinguished from older  
829           birds based solely on the consistency with which songs are produced; in just a few songs, all a

830 listener has to do is listen for imperfections, for the hints of the plasticity that would identify him  
 831 as a youngish bird still perfecting his songs.

832 In contrast, I find it inconceivable that banded wrens would assess adult singers based on  
 833 the performance in their songs (again, sensu Podos 1997). The scatterplot for banded wrens  
 834 (Figure 12) looks much like that for swamp sparrows (Figure 8), in that multiple songs types are  
 835 distributed in restricted spaces throughout the plot. I have no doubt that a figure for banded  
 836 wrens revealing the deviation of all song types for all males would look much like that figure for  
 837 swamp sparrows (Figure 10), revealing that no reliable information about male quality could  
 838 possibly be gleaned from those performance measures. Even if there were a statistical trend for  
 839 some adults (not yearlings) to be “better” singers than other adults, the sampling effort required  
 840 to assess performance among the males on all ~20 of their song types would be prohibitive.

841 **Vehrencamp, S. L., J. Yantachka, M. L. Hall, and S. R. De Kort. 2013. Trill**  
 842 **performance components vary with age, season, and motivation in the banded wren.**

843 From the opening sentence of the Abstract (emphases mine), I am predictably troubled:

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

847 This sentence squarely places the context and rationale for this study in the realm of  
 848 performance and honesty in signaling, with “difficult-to-execute” sounds revealing male quality.  
 849 Everything will be interpreted in this context, yet I continue to see no scientific justification for  
 850 doing so. According to the scatterplot of trill rate and bandwidth for banded wrens (Figure 12),  
 851 for example, relatively few songs are difficult to execute as defined in this performance context,

852 because most songs fall far from the upper bound on the graph. Furthermore, every male  
853 “willingly” learns those “low-performance,” easy-to-execute songs in order to have particular  
854 song types in his repertoire, *as if performance did not matter*. There is no obvious selection for  
855 high-performance, difficult-to-execute songs that honestly reflect male condition or quality, only  
856 selection to conform to the local dialect.

857 In this study, it seems that always, just beneath the surface, there is the assumption that a  
858 male is better if his song plots closer to the upper bound drawn on the scatter plot of trill rate and  
859 frequency bandwidth, and that birds use this performance criterion to assess one another. With a  
860 correlation in hand, the authors, without any credible evidence that the females are actually  
861 listening to these features of male song, state that

862 Females also appear to attend to trill fine structure, since males that obtained  
863 extra-pair fertilizations had both lower vocal deviation and greater note  
864 consistency than the males they cuckolded (*Cramer et al. 2011*). [p. 210; see my  
865 review of this topic under *Ballentine 2009*.]

866 It is assumed that, if a measure shows some statistical difference, the birds must use that  
867 feature to evaluate each other:

868 More aggressive birds tended to show a stronger seasonal improvement in  
869 bandwidth and vocal deviation . . . (p. 415)

870 The assumption becomes the conclusion when it is declared that

871 . . . overall trill performance . . . [is] . . . a good indicator of intrinsic quality (p.

872 417)



873           This paper has some laudable descriptive data showing how songs might change over  
 874 various time scales, both within and among males, but no evidence that the birds actually pay  
 875 any attention to the features described. Furthermore, as stated, “The changes are often small in  
 876 magnitude yet highly significant” (p. 406)—statistically significant, I would qualify, with no  
 877 information on the possible biological significance of the features described, especially related to  
 878 honesty in signaling, performance measures, and assessment of male quality.

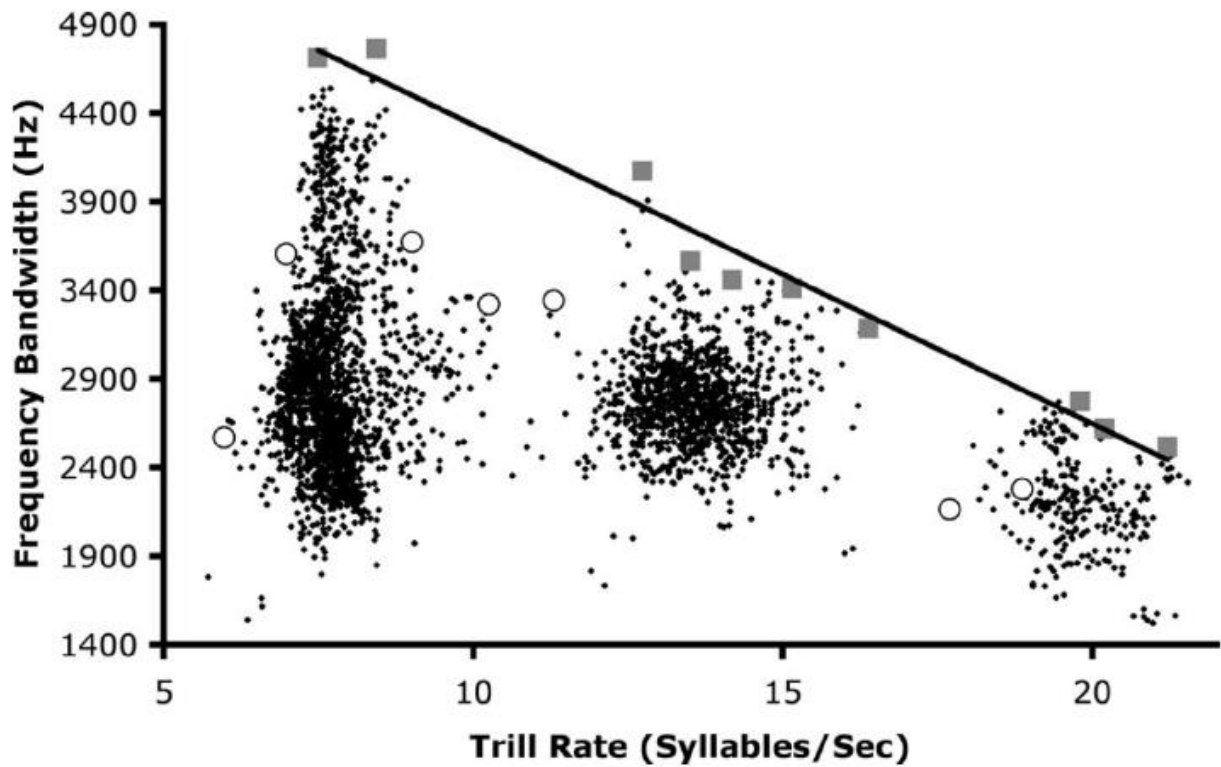
879           **Cramer, E. R. A. 2013. Vocal deviation and trill consistency do not affect male**  
 880 **response to playback in house wrens.**

881           The house wren (*Troglodytes aedon*), more than other species, varies a given syllable  
 882 type over a broad frequency range, depending on where the trill is in the song. The house wren is  
 883 therefore an unlikely species in which males or females would assess other individuals based  
 884 either on how consistently a song is delivered or on how far a song plots from the upper bound  
 885 on the scatterplot. And that is what the author found:

886           To the best of my knowledge, no published study . . . has failed to find an effect  
 887 of vocal deviation and trill consistency on receiver responses . . . [but I did]

888           The standard graph has a unique appearance (Figure 13). With the upper bound line on  
 889 the graph, and given the developing literature about that line, I can understand why an  
 890 investigator might (feel compelled to) study the implications of that line. The line draws  
 891 attention to it and seems to dictate one’s thoughts, but take the line away, perhaps redraw the  
 892 graph so that the origin is at (0,0), and one begins to ask entirely different questions. What are  
 893 the three “blobs” of data points hanging there in space? Why the vacant areas between the blobs  
 894 and all around them? Is this a local dialect, or would songs of all house wren populations plot

895 like this?



896

897 Figure 13. From Cramer (2013). "Frequency bandwidth as a function of trill rate in house wrens ( $n =$   
 898 3299 trills measured in 51 males in 2009; 2010 songs not shown). Large gray squares indicate the points  
 899 used to calculate the upper-bound regression line estimating the performance limit on trill rate and  
 900 frequency bandwidth (line); white circles are the data points from bins that lacked high-performance  
 901 trill exemplars, which I excluded from analysis (see main text for details). Black points are songs not  
 902 used to estimate the performance limit."

903

904 **Sprau, P., T. Roth, V. Amrhein, and M. Naguib. 2013. The predictive value of trill**  
 905 **performance in a large repertoire songbird, the nightingale *Luscinia megarhynchos*.**

906 Nightingales are recorded when they first return in spring, when no females have yet arrived.  
907 The main comparison is then between first-year males who are singing during their first week on  
908 a territory and older males who have had at least one more singing season on territory.

909 Several quotes give the flavor of the paper:

910 Because individuals are limited in the performance of trills, singing close to the  
911 performance limit, i.e. singing trills with a high repetition rate and a high  
912 frequency bandwidth, is considered a trait that honestly reflects the quality of a  
913 signaler.

914 . . . older males sang trills that were closer to the upper bound than did males that  
915 were in their second calendar year [i.e., yearlings]. . . vocal performance of trills .  
916 . . may function as an indicator of male quality . . . trill performance is likely to be  
917 a useful indicator of age . . .

918 [Our study shows that] . . . a songbird species with large song repertoire can  
919 encode valuable information such as signaler age in the performance of physically  
920 challenging trills. These characteristics make trills a suitable song trait that may  
921 enable listeners to readily assess male quality.

922 I would argue that performance measures are a highly unlikely feature to assess male age.  
923 If nightingales are like so many other songbirds, a first-year male on his territory who has just  
924 returned from migration can give abundant hints in his singing that he is still perfecting his  
925 repertoire of about 200 learned songs. Some plasticity typically remains, with successive  
926 syllables in a trill not yet identical, so that an astute listener (either bird or human) can readily  
927 identify the younger birds in a population. (See similar discussion under Ballentine 2009, above).

928 To distinguish young birds from old, there is no need for a listener to integrate measures of trill  
929 rate and frequency bandwidth in coming to some assessment.

930 Consider, in fact, the task that a female nightingale faces in using performance measures  
931 to assess a male. First, the vast majority of songs plot well away from the upper bound,  
932 indicating that most are not at all challenging to produce, so in the vast majority of songs she  
933 hears there will be no predictive value of male quality. If she hears a male sing a particularly  
934 challenging song type, she needs to know how he compares to other males, so she'd have to hear  
935 another male singing the same song type, and given that a nightingale learns 200 or so different  
936 songs from other males, she faces an arduous task. And then she needs to do the same with more  
937 song types, because the predictive value of any one song type is very low.

938 I find it inconceivable that female nightingales would use performance measures, as  
939 defined by Podos (1997) and promoted throughout the subsequent literature, especially in a  
940 large-repertoire bird like a nightingale. Showing a statistically significant difference in a  
941 complex measure between older and younger birds is far from showing that the birds actually use  
942 that measure, or that the measure has any biological relevance to the birds, especially when the  
943 relationship between the two measures has essentially no predictive value.

944 **Kagawa, H., and M. Soma. 2013. Song performance and elaboration as potential**  
945 **indicators of male quality in Java sparrows [*Lonchura oryzibora*].**

946 In the authors' own words:

947 We found a great individual variation in trill proportion, trill performance, and  
948 song complexity. A positive association between trill performance and body size  
949 suggested that trills can serve as an indicator of male quality (p. 138, Abstract).

950 While the effect of body size on average VD [vocal deviation] was not  
 951 statistically significant ( $P = 0.10$ ), minimum VD was significantly negatively  
 952 associated with body size ( $P = 0.004 . . .$ ) (p. 141, Results)

953 Data in the standard scatterplot are coded by male, so that one can see how each of the 24  
 954 males' 1-5 trills are scattered about the graph. As with swamp sparrows, there is no reliable  
 955 information about male quality in a Java sparrow's songs, as a male can have the worst and  
 956 (almost) the best songs in different song types.

957 A statistically astute listener might increase its chances of assessing the body weight (and  
 958 therefore quality) of the singer if he (or she) 1) assayed all of a male's songs, eventually 2)  
 959 identified his "best" song (shortest distance from upper bound on the graph), 3) discarded  
 960 information in all of his other songs (but why?), then 4) compared that best song to the best  
 961 songs of other males in the population. That the birds would go through this process seems rather  
 962 unlikely, especially discarding as irrelevant all of a male's songs except his "best."

963

#### 964 **DARK-EYED JUNCOS**

965 **Cardoso, G. C., J. W. Atwell, E. D. Ketterson, and T. D. Price. 2007. Inferring**  
 966 **performance in the songs of dark-eyed juncos (*Junco hyemalis*).**

967 We conclude that bird receivers may best evaluate how well a song is performed  
 968 if they integrate multiple cues and discuss how researchers may similarly devise  
 969 measures of song performance.

970 Excellent! It is the consideration of multiple song traits, of multiple explanations, of  
 971 multiple working hypotheses (Chamberlain 1964) that will enable progress in understanding  
 972 birdsong, an approach that avoids all of the pitfalls of trying to confirm or prove a favorite idea.

973 **Cardoso, G. C., J. W. Atwell, E. D. Ketterson, and T. D. Price. 2009. Song types,**  
974 **song performance, and the use of repertoires in dark-eyed juncos (*Junco hyemalis*).**

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

982 Here is the analysis for which I have been yearning, and the conclusion is much the same  
983 that one inevitably comes to when looking at the figures in Illes et al. (2006), Liu and Kroodsma  
984 (2006), and Kagawa and Soma (2013), and in my analyses of chipping sparrows and swamp  
985 sparrows (Figures 1-9). What matters most to these singing males is to have a song type like  
986 other birds in the population, and the relative performance abilities in singing that particular song  
987 type are almost certainly irrelevant.

988 Curiously, for the 14 citations listed in Web of Science by authors other than Cardoso  
989 himself (October 2015), not a single author mentions that Cardoso et al. have offered a  
990 fundamentally different interpretation of the scatter plot that was introduced by Podos (1997) and  
991 confirmed by multiple authors since then.

992 **Cardoso, G. C., J. W. Atwell, Y. Hu, E. D. Ketterson, and T. D. Price. 2012. No**  
993 **correlation between three selected trade-offs in birdsong performance and male**  
994 **quality for a species with song repertoires.**

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

997 These results complement a previous study on this population showing that most  
998 variation in performance is found among song types rather than among males.

999 Collectively, the lack of association between trade-off-based aspects of song  
1000 performance and male age or condition, plus variation among song types that  
1001 interferes with rapid assessment of a male's best performance, indicate that these  
1002 aspects of song performance do not allow a good assessment of male quality in  
1003 juncos, and perhaps more generally in species with song repertoires [p. 584; and I  
1004 would add any species without repertoires as well]

1005 As of October 2015, there has been little time to incorporate this Cardoso et al. (2012)  
1006 paper into literature citations, but the overall work of Cardoso et al. has been critically belittled  
1007 by Podos and colleagues (Zollinger et al. 2012) because of how frequency bandwidths were  
1008 measured. In a wide-ranging critique, Cardoso et al. are instructed on 1) proper measurements  
1009 and methodology, 2) interpretation of data, 3) validity of results, 4) experimental rigor, 5)  
1010 alternative explanations and hypotheses for data, 6) the ability to reject hypotheses, 7)  
1011 appropriate use of skepticism, 8) problems in published papers that “undermine the validity of  
1012 the results reported and the conclusions reached,” and 9) “basic principles” of science. Podos and  
1013 his coauthors are concerned, more broadly, with 10) how papers failing on these measures will  
1014 “have a profound adverse effect on the way the research field is viewed by the rest of the  
1015 scientific community.” In the context of all that I have reviewed here, this document by  
1016 Zollinger, Podos, Nemeth et al. (2012) is extraordinary.

1017

1018 **DISCUSSION**

1019 Others have spoken eloquently on how science is done, and best not done, and I will rely  
 1020 largely on quotes from published work here.

1021 *What is . . . Science? Advocacy? Marketing?*

1022 In his 1974 commencement address to CalTech students, Nobel-laureate Richard  
 1023 Feynman (1985:341) grappled with the difficulty of defining science, but superbly described  
 1024 how one recognizes good science (emphasis his):

1025 It's a kind of scientific integrity, a principle of scientific thought that corresponds  
 1026 to a kind of utter honesty—a kind of leaning over backwards. For example, if  
 1027 you're doing an experiment, you should report everything that you think might  
 1028 make it invalid—not only what you think is right about it: other causes that could  
 1029 possibly explain your results . . . Details that could throw doubt on your  
 1030 interpretation must be given . . . You must do the best you can—if you know  
 1031 anything at all wrong, or possibly wrong—to explain it. If you make a theory, for  
 1032 example, and advertise it . . . then you must also put down all the facts that  
 1033 disagree with it, as well as those that agree with it . . . In summary, the idea is to  
 1034 give *all* of the information to help others to judge the value of your contribution;  
 1035 not just the information that leads to judgment in one particular direction or  
 1036 another.

1037 This scientific integrity described by Feynman is in contrast to what Gitzen (2007) calls  
 1038 “advocacy.” Advocates



1039 . . . stretch available data, gloss over uncertainties in their evidence, and ignore  
1040 contrary results . . . Once bold claims about a poorly tested method or weak result  
1041 are published, their sins are forgiven and they can be worked into future  
1042 introductions and discussions at will . . . Often, this form of advocacy is obvious  
1043 only to the small percentage of any journal's readers that have scientific expertise  
1044 in a specialized area—a small pool of appropriate reviewers . . . the personal  
1045 rewards of . . . [advocacy] . . . far outweigh risks . . .

1046 The costs of advocacy are high, damaging the credibility of all scientists in every field of  
1047 endeavor (consider climate science alone), but especially within the community of scholars who  
1048 should be policing their own. Again, Feynman (1985):

1049 . . . although you may gain some temporary fame and excitement, you will not  
1050 gain a good reputation as a scientist if you haven't tried to be very careful in this  
1051 kind of work. And it's this type of integrity, this kind of care not to fool yourself,  
1052 that is missing to a large extent in much of the research in cargo cult science.

1053 Studies of sexual selection tend to be plagued by advocacy, as eloquently addressed by  
1054 Richard Prum:

1055 . . . the study of sexual selection has become a weak science that largely seeks to  
1056 confirm the adaptive hypotheses it assumes—i.e. that natural selection on mating  
1057 preferences is the determining force in intersexual selection. In this intellectual  
1058 environment, failure to confirm an honest indication or adaptive signaling  
1059 hypothesis merely means that the researchers have failed to work hard enough to  
1060 do so . . . the possibility that traits are not indicating anything is rarely even  
1061 entertained. Sexual selection has become a field in which the role of natural

1062 selection on mating preferences is usually assumed, rarely discussed, largely  
 1063 beyond testing and even redefined into the definition of sexual selection itself.  
 1064 (Prum 2012:2253)

1065 . . . the goal of much empirical work in intersexual selection is to confirm the  
 1066 origin of the signal honesty and sensory efficiency rather than to test its existence.  
 1067 . . . In confirmationist research, negative results are interpreted as failure to have  
 1068 yet looked hard enough to find the evidence of additional selection on preferences  
 1069 . . . Much of intersexual selection research is an extant remnant of the  
 1070 “adaptationist programme” (Gould and Lewontin 1979) in which the deterministic  
 1071 power of natural selection is assumed and alternative explanations are defined out  
 1072 of existence or treated as irrelevant (Prum 2010:3086)

1073 ***Word choice***

1074 One root of the problem is the prosaic one of the very words we use to describe the world  
 1075 around us. Words like “performance” and “proficiency” and “constraints” (and all their cousins),  
 1076 as used throughout the literature that I review here, are all nonneutral, loaded terms with the  
 1077 implicit assumption that where a song plots on a graph tells “how well” a male sings, or how  
 1078 “proficiently” he sings, and therefore how “good” a male he is. Repeated use of the term  
 1079 “performance” leads to reification, with the concept no longer a hypothesis to be tested but  
 1080 instead a proven fact, or an assumption so hidden that it is accepted as fact. Functional,  
 1081 nonneutral terms like “performance” inevitably shape the mind and unconsciously block  
 1082 alternative views from being entertained, as they implicitly define the universe of discourse. As a  
 1083 result, “Our job as scientists . . . to discover truths about the world” (Simmons et al. 2011: 1359)  
 1084 is severely hampered.

1085 As Marler and Hamilton (1966:716) wrote (emphasis mine),  
 1086 The process of description is intimately involved with naming, and here too a  
 1087 degree of discipline is called for. Studies of communicatory behavior in animals  
 1088 have often included in their primary descriptions such terms as domination and  
 1089 subordination behavior, inferiority and superiority postures, intimidation,  
 1090 distraction, threat, and appeasement displays [and “performance,” I might add].  
 1091 These terms are liable to prejudge the function of behavior . . . *clear separation of*  
 1092 *description from function is desirable ... There should be a maximum reliance on*  
 1093 *intrinsic properties of the behavior and a minimum of interpretation*

1094 I have felt bound to use the same terminology in this review that is used in the papers  
 1095 reviewed, though I flinch every time I write the word “performance,” because the very attempt to  
 1096 address the problems is already half-defeated by the very use of such a loaded word. Substituting  
 1097 neutral descriptive terms for the loaded terms can be a mind-expanding experience. Consider, for  
 1098 example, the terms “low-performance” (relatively slow trills and narrow frequency bandwidth)  
 1099 and “high-performance” (faster trills, broader bandwidth). The intellectual landscape is released  
 1100 from single-minded explanations by merely labeling these songs “slow-narrow” trills and “fast-  
 1101 broad” trills (or some such descriptive terms); with the descriptions no longer rooted in terms  
 1102 that focus on only one functional interpretation, one can more comfortably acknowledge a null  
 1103 hypothesis and alternative hypotheses, and do one’s best to falsify them in turn.

1104 ***Going forward***

1105 Only when multiple working hypotheses are considered (Chamberlain 1964), and only  
 1106 when the “scientific integrity” and “utter honesty” of Feynman (1985) are adopted, will progress

1107 be made in understanding how and why birds sing, and what information listeners extract from  
1108 the singers.

#### 1109 **ACKNOWLEDGMENTS**

1110 I thank Pavel Linhart, Gonçalo Cardoso, David Lahti, and especially Becky Cramer for  
1111 responding to an early draft of my document. Others who have provided helpful advice were  
1112 Caglar Akçay, Bruce Byers, Sylvia Halkin, Ted Miller, Gene Morton, and Walter Berry.

1113

#### 1114 **REFERENCES**

1115 Akçay, C., and M. D. Beecher. 2015. Team of rivals in chipping sparrows? A comment on  
1116 Goodwin & Podos. *Biology Letters*. 11:20141043.

1117 Ballentine, B. 2009. The ability to perform physically challenging songs predicts age and size in  
1118 male swamp sparrows, *Melospiza georgiana*. *Animal Behaviour* 77:973-978.

1119 Ballentine, B., J. Hyman, and S. Nowicki. 2004. Vocal performance influences female response  
1120 to male bird song: an experimental test. *Behavioral Ecology* 15:163-168.

1121 Byers, B. E., and D. E. Kroodsma. 2009. Female mate choice and songbird song repertoires.  
1122 *Animal Behaviour* 77:13-22.

1123 Cardoso, G. C., J. W. Atwell, E. D. Ketterson, and T. D. Price. 2007. Inferring performance in  
1124 the songs of dark-eyed juncos (*Junco hyemalis*). *Behavioral Ecology* 18:1051-1057.

1125 Cardoso, G. C., J. W. Atwell, E. D. Ketterson, and T. D. Price. 2009. Song types, song  
1126 performance, and the use of repertoires in dark-eyed juncos (*Junco hyemalis*). *Behavioral*  
1127 *Ecology* 20:901-907.

- 1128 Cardoso, G. C., J. W. Atwell, Y. Hu, E. D. Ketterson, and T. D. Price. 2012. No correlation  
1129 between three selected trade-offs in birdsong performance and male quality for a species with  
1130 song repertoires. *Ethology* 118:584-593.
- 1131 Chamberlin, T. C. 1965. The method of multiple working hypotheses; with this method the  
1132 dangers of parental affection for a favorite theory can be circumvented. *Science* 148:754-759.
- 1133 Cramer, E. R. A. 2013. Physically challenging song traits, male quality, and reproductive success  
1134 in house wrens. *PLoS ONE* 8(3): e59208. doi:10.1371/ journal.pone.0059208
- 1135 Cramer, E. R. A., and J. J. Price. 2007. Red-winged blackbirds *Agelaius phoeniceus* respond  
1136 differently to song types with different performance levels. *Journal of Avian Biology* 38:122-  
1137 127.
- 1138 Cramer, E. R. A., M. L. Hall, S. R. De Kort, I. J. Lovette, and S. L. Vehrencamp. 2011.  
1139 Infrequent extra-pair paternity in the banded wren, a synchronously breeding tropical passerine.  
1140 *Condor* 113:637-645.
- 1141 De Kort, S. R., E. R. B. Eldermire, E. R. A. Cramer, and S. L. Vehrencamp. 2009. The deterrent  
1142 effect of bird song in territory defense. *Behavioral Ecology* 20:200-206.
- 1143 Dubois, A. L., S. Nowicki, and W. A. Searcy. 2009. Swamp sparrows modulate vocal  
1144 performance in an aggressive context. *Biology Letters* 5:163-165.
- 1145 Dubois, A. L., S. Nowicki, and W. A. Searcy. 2011. Discrimination of vocal performance by  
1146 male swamp sparrows. *Behavioral Ecology and Sociobiology* 65:717-726.
- 1147 Feynman, R. P. 1985. *Surely You're Joking, Mr. Feynman!* W. W. Norton & Company, New  
1148 York City.
- 1149 Gitzen, R. A. 2007. The dangers of advocacy in science. *Science* 317:748-748.

- 1150 Goodwin, S. E., and J. Podos. 2014. Team of rivals: alliance formation in territorial songbirds is  
1151 predicted by vocal signal structure. *Biology Letters* 10:Article Number: 20131083.
- 1152 Goodwin, S. E., and J. Podos. 2015. Reply to Akçay & Beecher: yes, team of rivals in chipping  
1153 sparrows. *Biology Letters*. 11:20150319.
- 1154 Illes, A. E., M. L. Hall, and S. L. Vehrencamp. (2006). Vocal performance influences male  
1155 receiver response in the banded wren. *Proceedings of the Royal Society B-Biological Sciences*  
1156 273(1596): 1907-1912.
- 1157 Juola, F. A., and W. A. Searcy. 2011. Vocalizations reveal body condition and are associated  
1158 with visual traits in great frigatebirds (*Fregata minor*). *Behavioral Ecology and Sociobiology*  
1159 65:2297–2303.
- 1160 Kagawa, H., and M. Soma. 2013. Song performance and elaboration as potential indicators of  
1161 male quality in Java sparrows. *Behavioural Processes* 99:138-144.
- 1162 Kroodsma, D. E. 2005. *The Singing Life of Birds. The Art and Science of Listening to Birdsong.*  
1163 Houghton-Mifflin Co., Boston, Massachusetts.
- 1164 Lachlan, R. F., R. C. Anderson, S. Peters, W. A. Searcy, and S. Nowicki. 2014. Typical versions  
1165 of learned swamp sparrow song types are more effective signals than are less typical versions.  
1166 *Proceedings of the Royal Society B-Biological Sciences* 281:Article Number: 20140252.
- 1167 Lahti, D. C., D. L. Moseley, and J. Podos. 2011. A tradeoff between performance and accuracy  
1168 in bird song learning. *Ethology* 117:802-811.
- 1169 Liu, W.-c. 2001. Song development and singing behavior of the chipping sparrow (*Spizella*  
1170 *passerina*) in western Massachusetts. Ph. D. Thesis, Department of Biology. University of  
1171 Massachusetts, Amherst.

- 1172 Liu, W.-c. 2004. The effect of neighbours and females on dawn and daytime singing behaviours  
1173 by male chipping sparrows. *Animal Behaviour* 68:39-44.
- 1174 Liu, W.-c., and D. E. Kroodsma. 1999. Song development by field sparrows (*Spizella pusilla*)  
1175 and chipping sparrows (*Spizella passerina*). *Animal Behaviour* 57:1275-1286.
- 1176 Liu, W.-c., and D. E. Kroodsma. 2006. Song learning by chipping sparrows: when, where, and  
1177 from whom. *The Condor* 108:509-517.
- 1178 Marler, P. and W. J. Hamilton III. 1966. *Mechanisms of Animal Behavior*. John  
1179 Wiley & Sons, New York
- 1180 Morton, E. S., S. L. Gish, and M. Van Der Voort. 1986. On the learning of degraded and  
1181 undegraded songs in the Carolina wren. *Animal Behaviour* 34:815-820.
- 1182 Moseley, D. L., D. C. Lahti, and J. Podos. 2013. Responses to song playback vary with the vocal  
1183 performance of both signal senders and receivers. *Proceedings of the Royal Society B-Biological*  
1184 *Sciences* 280. Issue: 1768 Article Number: 20131401
- 1185 Naguib, M. 1996. Auditory distance estimation in song birds: Implications, methodologies and  
1186 perspectives. *Behavioural Processes* 38:163-168.
- 1187 Nowicki, S., S. Peters, and J. Podos. 1998. Song learning, early nutrition and sexual selection in  
1188 songbirds. *American Zoologist* 38:179-190.
- 1189 Podos, J. 1996. Motor constraints on vocal development in a songbird. *Animal Behaviour*  
1190 51:1061-1070.
- 1191 Podos, J. 1997. A performance constraint of the evolution of trilled vocalizations in a songbird  
1192 family (Passeriformes: Emberizidae) *Evolution* 51:537-551.

- 1193 Podos, J., D. C. Lahti, and D. L. Moseley. 2009. Vocal performance and sensorimotor learning in  
1194 songbirds, p. 159-195. *In: Advances in the Study of Behavior*, Vol. 40. M. Naguib, K.  
1195 Zuberbuhler, N. Clayton, and V. Janik (eds.). Elsevier Academic Press, Inc., San Diego,  
1196 California.
- 1197 Podos, J., S. Peters, and S. Nowicki. 2004. Calibration of song learning targets during vocal  
1198 ontogeny in swamp sparrows, *Melospiza georgiana*. *Animal Behaviour* 68:929-940.
- 1199 Prum, R. O. 2010. The Lande-Kirkpatrick Mechanism is the null model of evolution by  
1200 intersexual selection: implications for meaning, honesty, and design in intersexual signals.  
1201 *Evolution* 64:3085-3100.
- 1202 Prum, R. O. 2012. Aesthetic evolution by mate choice: Darwin's really dangerous idea.  
1203 *Philosophical Transactions of the Royal Society B-Biological Sciences* 367:2253-2265.
- 1204 Schmidt, R., H. P. Kunc, V. Amrhein, and M. Naguib. 2008. Aggressive responses to broadband  
1205 trills are related to subsequent pairing success in nightingales. *Behavioral Ecology*. 19:635-641.
- 1206 Simmons, J. P., L. D. Nelson, and U. Simonsohn. 2011. False-positive psychology: undisclosed  
1207 flexibility in data collection and analysis allows presenting anything as significant. *Psychological*  
1208 *Science*. 22:1359-1366.
- 1209 Simmons, L. W. 2014. 25 years of Behavioral Ecology. *Behavioral Ecology*. 25:1-3.
- 1210 Snijders, L., E. P. van Rooij, J. M. Burt, C. A. Hinde, K. van Oers, and M. Naguib. 2014. Social  
1211 networking in territorial great tits: slow explorers have the least central social network positions.  
1212 *Animal Behaviour*. 98:95-102.



- 1213 Sprau, P., T. Roth, V. Amrhein, and M. Naguib. 2013. The predictive value of trill performance  
1214 in a large repertoire songbird, the nightingale *Luscinia megarhynchos*. *Journal of Avian Biology*  
1215 44:576-574.
- 1216 Vehrencamp, S. L., J. Yantachka, M. L. Hall, and S. R. De Kort. 2013. Trill performance  
1217 components vary with age, season, and motivation in the banded wren. *Behavioral Ecology and*  
1218 *Sociobiology* 67:409-419.
- 1219 Wilson, D. R., P. P. Bitton, J. Podos, and D. J. Mennill. 2014. Uneven sampling and the analysis  
1220 of vocal performance constraints. *American Naturalist* 183:214-228.
- 1221 Zollinger, S. A., J. Podos, E. Nemeth, F. Goller, and H. Brumm. 2012. On the relationship  
1222 between, and measurement of, amplitude and frequency in birdsong. *Animal Behaviour* 84:E1-  
1223 E9.