Birdsong "Performance" Studies: A Sad Commentary

Donald Kroodsma

Department of Biology, University of Massachusetts Amherst, MA. U.S.A.

Birdsong performance studies have embraced the significance of a scatterplot between trill rate and frequency bandwidth, using the distance that a given sound plots from an upper bound (i.e., "vocal deviation") as a measure of how difficult it is to perform that sound; the relative difficulty of the performance then becomes a means by which to assess the quality of the song and the singer (review in Kroodsma, 2017). Those who promote this literature defend it by its wide acceptance: ". . . vocal deviation has been used widely as a composite index of vocal performance . . ." (Podos, Moseley, Goodwin, McClure, et al., 2016:204) and “Vocal deviation . . . has indeed been adopted widely in tests of song function” (Goodwin & Podos, 2015:1). "Studies of trilled vocalizations," with their inherent "tradeoff between syllable repetition rate and frequency bandwidth," are hailed as "a premiere illustration of how performance constraints shape the evolution of mating displays . . . [with] sexual selection favoring high performance trills" (Wilson, Bitton, Podos, & Mennill, 2014:214).

In reality, however, no credible scientific data support these ideas. The considerable literature that has developed on this topic is instead a premiere illustration of how highly flawed methods have been used by prolific, influential authors to promote an intuitively appealing but false story, all of which has been accepted uncritically by others. What is especially disturbing about this literature is that those who author, promote, and defend it are among the perceived

leaders of the Animal Behavior Society: three presidents, two recipients of the Exemplar Award, and one recipient of the Young Investigator Award.

The three responses to my Forum (Cardoso, 2017; Podos, 2017; Vehrencamp, de Kort, & Illes, 2017) do not refute the big-picture and methodological problems that I raise. Podos distracts from the big issues, resorting to defense of the indefensible largely by addressing minor, secondary issues or nonissues altogether. Vehrencamp et al. (all in the first paragraph) declare that my "criticism is flawed" because of my ". . . weaknesses . . . mistakes . . . weaknesses . . . outright errors . . . misrepresentations . . . highly selective citation of the literature . . . convoluted logic . . . weaknesses . . . and a misunderstanding . . ."; I stand by all of my original critique. Cardoso doesn't really address the main problems I discuss, but instead offers advice on how to disentangle ideas and facts in performance-related research.

Below, I offer a few final thoughts on this topic.

Overuse of the word "performance" obfuscates—two examples

As I described in my Forum, using the word "performance" to describe measured vocal deviation serves to turn an assumption (that a sound with low vocal deviation is difficult to produce) into the conclusion that birds with low vocal deviation perform better and are therefore higher quality birds. I offer two examples of this kind of obfuscation from publications that were submitted after a draft of my Forum became widely available in December 2014.

To try the exercise I suggest in my Forum, read Podos et al. (2016) without the word "performance," which occurs 139 times, substituting instead neutral words of your choice. Very quickly the paper has an entirely different feel; no longer is it on the cutting edge of sexual selection science, but instead it becomes a rather prosaic description of syllable complexity

among songs, with no information on the relative difficulty of producing those syllables or whether the birds care. There's nothing wrong with that. A good description will last forever and would contribute more to our understanding of the natural world than all of the performance experiments I have critiqued.

Also try reading Cardoso and Atwell (2016), again removing "performance" and substituting neutral words. The title states that "Shared songs are of lower performance in the dark-eyed junco" (*Junco hyemalis*), but without the word "performance" to interfere, I am freer to ask questions about the nature of the shared songs. And I start to fret, for several reasons. Classifying the songs of more than 150 different males into more than 250 song types is a huge task, and a highly subjective one. Also, the word "shared" implies that different renditions of the same song type by different males are part of the same cultural tradition, i.e., males "share" these songs because they have learned them from each other. But if the amount of sharing *within* a junco population drops to near zero at about one kilometer (Figure 3 in Newman, Yeh, & Price, 2008), how can any songs *between* two sites 80 km distant be truly "shared," i.e., part of the same cultural tradition? It's more likely that songs at these two distant locations are simply different dialects, which can have different trill rates and frequency bandwidths, i.e., different "performances" (as in swamp sparrows, my Figures 6 and 8 in Kroodsma, 2017). Any similarity between songs of the two locations would be by chance alone.

I then predicted a bias in Cardoso and Atwell's analyses, in that simpler syllables with fast trill rates would be more likely to be classified as shared, simply because less detail in the sonagrams would be available for the human eye to classify them as "unshared." I predicted correctly, finding that the syllable rates of shared songs between these two locations were significantly faster than unshared songs (one-tailed t-test; $n_1 = 21$ shared songs, median = 10.6

syllables/sec; n_2 = 241 unshared songs, median = 13.0 syllables/sec; t = 2.46, p = 0.007, or t = 2.64, p = 0.004, depending on whether the 21 shared syllable rates come from the "ML" or "UCSD" data). By cutting through the performance verbiage (the word "performance" occurs almost once in each sentence), I come to doubt that the classification of songs as shared or unshared has any relevance to the birds themselves across these two locations, much less to sexual selection or any index of song performance.

Future research

I had an initial say in my Forum (Kroodsma, 2017), though I had at first reached more broadly into the birdsong and sexual selection literature. The "nutritional stress hypothesis" (Nowicki, Peters, & Podos, 1998), for example, also needs a full treatment. Someone wishing to tackle this topic could begin with Nowicki, Hasselquist, Bensch, & Peters (2000), one of the supporting pillars of this literature, in which readers are led to believe that females can assess the quality of a male based on repertoire size, when only about one percent of the variation in repertoire size is statistically explained by a correlation with a proxy of physical condition. That notion was rebutted by Kroodsma (2004) but perpetuated in Peters, Searcy, and Nowicki (2014), and promoted over a hundred times in the literature as of January 2017 (Web of Science). I contend that the same kinds of faulty methods used in the vocal deviation literature have been exploited in the song repertoire literature (see review by Byers & Kroodsma, 2009) and nutritional stress literature.

I agree with Cardoso and Podos in that there is much to learn in trying to understand what birds hear and how they might assess one another based on song. But which research program can pursue these topics with sufficient objectivity, independence, and credibility? One of the best clues is offered by the citations that authors choose to support their own work. With my Forum

article in hand, Podos et al. (2016) continued to cite at face value almost all of the papers that I had found severely flawed, either implicitly or explicitly defending these works (see also Goodwin & Podos, 2015; Podos, 2017). In contrast, Byers, Akresh, and King (2016), when studying "song and male quality in prairie warblers" (*Setophaga discolor*; quote from their title), had ample opportunity to do so but chose to cite not a single one. Authors choose the framework for their publications but lose their own credibility when they cite and promote work that is not credible.

Science is the search for truths about the natural world, but when the chosen methods cannot reveal truths, science suffers, as does the credibility of all scientists in all disciplines, including climate science. Producing good science is a responsibility shared by all of us, by authors, referees, editors, among others. How this performance literature (and specifically the "vocal deviation literature") has flourished for so long is disconcerting, and does not speak well for the gate-keepers of science.

The primary problem is described well by Gitzen (2007:748), who discusses how authors ". . . stretch available data, gloss over uncertainties in their evidence, and ignore contrary results . . . [largely because] . . . the personal rewards of . . . [such behaviors] . . . far outweigh risks . . ." Therein lies the key: Only when the costs of producing faulty science outweigh the benefits will sound science prevail. In the performance literature I have reviewed, the benefits have until now far outpaced the costs.

Acknowledgments

For steeling my resolve to address these issues, I thank the target authors who tried numerous ways to stymie my message. I also thank a small group of supporters who said, for birdsong science, and for science in general, "you must persevere"—you know who you are. And the anonymous referees have been thorough, knowledgeable, and extraordinarily helpful—*thank you.* For a full history of my Forum article, beginning May 2014, see http://donaldkroodsma.com/?page_id=1596.

References

Byers, B. E., Akresh, M., & King, D. I. (2016). Song and male quality in prairie warblers. *Ethology*, 122, 660-670.

Byers, B. E., & Kroodsma, D. E. (2009). Female mate choice and songbird song repertoires. *Animal Behaviour*, 77, 13-22.

Cardoso, G. C. (2017). Advancing the inference of performance in birdsong. *Animal Behaviour*, 125, E29-E32.

Cardoso, G. C., & Atwell, J. W. (2016). Shared songs are of lower performance in the dark-eyed junco. *Royal Society Open Science*, 3, 160341.

Gitzen, R. A. (2007). The dangers of advocacy in science. Science, 317, 748.

Goodwin, S. E., & Podos, J. (2015). Reply to Akçay & Beecher: Yes, team of rivals in chipping sparrows. *Biology Letters*, 11(7). http://dx.doi.org/10.1098/rsbl.2015.0319.

Kroodsma, D. E. (2004). The diversity and plasticity of song development. In P. Marler, & H. Slabbekoorn (Eds.), *Nature's Music. The Science of Birdsong* (pp. 108-131). Amsterdam, Netherlands: Elsevier Academic Press.

Kroodsma, D. (2017). Birdsong performance studies: a contrary view. *Animal Behaviour*, 125, E1-E16.

Newman, M. M., Yeh, P. J., & Price, T. D. (2008). Song Variation in a recently founded population of the dark-eyed junco (*Junco hyemalis*). *Ethology*, 114, 164–173.

Nowicki, S., Peters, S., & Podos, J. (1998). Song learning, early nutrition and sexual selection in songbirds. *American Zoologist*, 38, 179-190.

Nowicki, S., D. Hasselquist, D., Bensch, S., & Peters, S. (2000). Nestling growth and song repertoire size in great reed warblers: evidence for song learning as an indicator mechanism in mate choice. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 267, 2419-2424.

Peters, S., Searcy, W. A., & Nowicki, S. (2014). Developmental stress, song-learning, and cognition. *Integrative and Comparative Biology*, 54, 555-567.

Podos, J. (2017). Birdsong performance studies: reports of their death have been greatly exaggerated. *Animal Behaviour*, 125, E17-E24.

Podos, J., Moseley, D. L., Goodwin, S. E., McClure, J., Taft, B. N., Strauss, A. V. H., et al. (2016). A fine-scale, broadly applicable index of vocal performance: Frequency excursion. *Animal Behaviour*, 116, 203e212.

Vehrencamp, S. L., de Kort, S. R., and Illes, A. E. (2017). Response to Kroodsma's critique of banded wren song performance research. *Animal Behaviour*, 125, E25-E28.

Wilson, D. R., Bitton, P. P., Podos, J., & Mennill, D. J. (2014). Uneven sampling and the analysis of vocal performance constraints. *American Naturalist*, 183, 214-228.